Universidade Federal de Campina Grande

Centro de Engenharia Elétrica e Informática

Coordenação de Pós-graduação em Ciência da Computação

Detecção de Pornografia Infanto-juvenil Baseada em Pornografia Adulta e Estimativa de Idade Facial

Danilo Coura Moreira

Campina Grande, PB, Brasil

2021

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Detecção de Pornografia Infanto-juvenil Baseada em Pornografia Adulta e Estimativa de Idade Facial

Danilo Coura Moreira

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Visão Computacional

Eanes Torres Pereira (Orientador) - Marco Alvarez (Coorientador Estrangeiro)

Campina Grande, Paraíba, Brasil

M838d Moreira, Danilo Coura.
Detecção de pornografia infanto-juvenil baseada em pornografia adulta e estimativa de idade facial / Danilo Coura Moreira. - Campina Grande, 2021.
173 f. : il. Color

Tese (Doutorado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2021.
"Orientação: Prof. Dr. Eanes Torres Pereira, Prof. Dr. Marco Alvarez".
Referências.

1. Computação Forense. 2. Visão Computacional. 3. Classificação de Imagens. 4. Pornografia Infanto-juvenil. 5. Aprendizado Profundo. 6. Redes Neurais Convolucionais. I. Pereira, Eanes Torres. II. Alvarez, Marco. III. Título.

CDU 004:393.48(043)

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECÁRIA ITAPUANA SOARES DIAS CRB-15/93

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO CIENCIAS DA COMPUTACAO
Rua Aprígio Veloso, 882, - Bairro Universitário, Campina Grande/PB, CEP 58429-900

FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

DANILO COURA MOREIRA

DETECÇÃO DE PORNOGRAFIA INFANTO-JUVENIL BASEADA EM PORNOGRAFIA ADULTA E ESTIMATIVA DE IDADE FACIAL

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Doutor em Ciência da Computação.

Aprovada em: 22/02/2021

Prof. Dr. EANES TORRES PEREIRA, UFCG, Orientador

Prof. Dr. HERMAN MARTINS GOMES, UFCG, Examinador Interno

Prof. Drª. JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, UFCG, Examinadora Interna

Prof. Dr. CLÁUDIO ROSITO JUNG, UFRS, Examinador Externo

Prof. Dr. FRANCISCO MADEIRO BERNARDINO JUNIOR, UNICAP, Examinador Externo

Prof. Dr. MARCO ANTONIO ALVAREZ VEGA, UNIVERSITY OF RHODE ISLAND, Examinador Externo

---

Documento assinado eletronicamente por **JOSEANA MACEDO FECHINE**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 22/02/2021, às 17:30, conforme horário oficial de Brasília, com fundamento no art. 8º,

caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **EANES TORRES PEREIRA**, **PROFESSOR 3 GRAU**, em 22/02/2021, às 17:30, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **HERMAN MARTINS GOMES**, **PROFESSOR 3 GRAU**, em 22/02/2021, às 17:31, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **FRANCISCO MADEIRO BERNARDINO JUNIOR**, **Usuário Externo**, em 22/02/2021, às 20:12, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

A autenticidade deste documento pode ser conferida no site https://sei.ufcg.edu.br/autenticidade, informando o código verificador **1286118** e o código CRC **2D0C28DC**.

Referência: Processo nº 23096.007336/2021-10 SEI nº 1286118

Resumo

A evolução tecnológica certamente trouxe e vem trazendo grandes avanços sociais e econômicos para a geração atual, entretanto, esse desenvolvimento também tem sido utilizado por alguns indivíduos para a prática de novos crimes ou, até mesmo, auxiliar na prática de antigos delitos. É o que acontece com a violência sexual contra crianças e adolescentes, um crime realizado ao longo dos anos sem auxílio da tecnologia que, nas últimas décadas, tem sido impulsionado pela posse e compartilhamento de arquivos digitais possuindo conteúdo pornográfico infanto-juvenil. O aumento do cometimento de crimes dessa natureza acaba influenciando a demanda da realização dos exames periciais em busca de conteúdo pornográfico infanto-juvenil. Esses exames, na maioria das vezes, são realizados de maneira não automatizada nos institutos de polícia científica do Brasil. Baseado nessa necessidade, foi desenvolvida uma técnica que, sem a utilização de imagens de conteúdo pornográfico infanto-juvenil, possibilita detectar conteúdo pornográfico e inferir, por meio de faces detectadas, a probabilidade de esse tipo de imagem retratar crianças e/ou adolescentes. Para a detecção de pornografia, foi proposta uma nova base de dados pornográfica (*Pornographic and Explicit Dataset 376K*) aliada ao uso de uma estratégia gulosa para a seleção e ajuste fino de uma rede neural convolucional. Para estimar a menoridade penal dos envolvidos, foi proposta uma nova técnica baseada também em aprendizado profundo para a estimação de idade real, a partir de faces humanas, que utiliza dados de idade aparente para aprimorar seu resultado final. Por fim, ainda foi proposta uma técnica baseada em aprendizado de máquina tradicional que aprimorou os resultados referentes à probabilidade de uma face humana pertencer a um individuo menor de idade. Essa abordagem, composta por módulos inovadores na detecção de pornografia e na estimação de idade real facial, que superaram pesquisas inseridas no estado da arte em suas respectivas áreas, também atingiu resultados compatíveis com o estado da arte da detecção de pornografia infanto-juvenil.

Palavras-chave: Computação Forense, Visão Computacional, Classificação de Imagens, Pornografia Infanto-juvenil, Aprendizado Profundo, Redes Neurais Convolucionais.

Abstract

Surely the technological evolution has been bringing a huge social and economic advancement to our generation, however, this development has also been used by some individuals to commit new kinds of crimes or even support the old ones. This is what happens with sexual child and teen exploitation, a kind of crime that was committed through the years without technological support, but in the last decades has been boosted by the possession and sharing of digital files with child and teen pornographic content. The increase in these crimes influences directly the demand for digital exams that look for child pornography. These exams almost always are made non-automatically in police scientific departments of Brazil. Because of this need, it was developed a technique, without using any illicit pornographic content, which allows the users to detect pornographic content and infer, through detected human faces, the likelihood of this kind of image to portray child or teens. In order to detect pornography, it was proposed a novel dataset (Pornographic and Explicit Dataset 376K) used with a greed strategy to, respectively, choose and fine-tune a convolutional neural network and its hyperparameters. In order to estimate if the related people are underage, it was presented a novel technique also based on deep learning to estimate the real age from human faces using data from apparent age to improve its results. In the end, it was still proposed a machine learning technique that improved the results related to the likelihood of a human face to belong to an underage person. This approach, composed of innovative modules in pornography detection and real age estimation, which outperforms state-of-the-art researches in those respective fields, also achieved compatible results with the state-of-the-art of child and adolescent pornography detection.

Keywords: Digital Forensics, Computer Vision, Image Classification, Child and Adolescent Pornography, Deep Learning, Convolutional Neural Networks.

Dos Termos Legais

Salienta-se que:

1. Todo o processo de construção do conjunto de dados contendo pornografia infanto-juvenil se deu em momento de atividade pericial do Perito Oficial Criminal Danilo Coura Moreira, acontecendo sempre nas dependências e utilizando o aparato tecnológico do Setor de Computação Forense do Núcleo de Criminalística de Campina Grande - PB. O Perito, agindo no estrito cumprimento de dever legal, de acordo com o inciso III do Art. 23 do Código Penal, é excluído da ilicitude na realização de um fato típico relativo à manipulação das referidas imagens, por força do desempenho de uma obrigação imposta por lei.

2. Construída a base de dados, toda a análise experimental (também realizada pelo Perito Oficial Criminal Danilo Coura Moreira nas dependências e utilizando o aparato tecnológico do Setor de Computação Forense do Núcleo de Criminalística de Campina Grande - PB) utilizou apenas os tensores referentes a cada imagem, não sendo necessária e nem realizada a visualização das imagens em questão.

3. O conjunto de dados contendo pornografia infanto-juvenil encontra-se armazenado de maneira segura no Setor de Computação Forense do Núcleo de Criminalística de Campina Grande - PB, para eventual reprodutibilidade dos experimentos e não será distribuída, senão por expressa decisão judicial que vise a colaboração para o desenvolvimento de ferramentas que possam alavancar técnicas para a melhoria da detecção de pornografia infanto-juvenil.

Agradecimentos

Agradeço, primeiramente, a Deus, que me proporcionou saúde, determinação e uma família maravilhosa, que me estruturou para que eu pudesse ter chegado até aqui.

A meu pai, Antonimário (*in memorian*), por todo o amor, carinho e dedicação. Pelo exemplo de hombridade, responsabilidade, honestidade e por sempre ter priorizado nossa educação (minha e de meus irmãos).

A minha mãe, Conceição, por todo amor, carinho e dedicação. Pelas preocupações, cuidados constantes e por ser exemplo de determinação, mostrando-me que por mais difícil que fosse o caminho, eu não deveria desistir.

Aos meus irmãos, Toninho e Herlinha, pelo carinho e suporte contínuos, fazendo-me sentir apoiado em todos os momentos.

A minha esposa, Luciane, por todo o amor, carinho, renúncia, companheirismo e cumplicidade, sempre apoiando-me e erguendo-me nos momentos mais difíceis.

A meu filho, Henrique, que me fez conhecer o amor incondicional de pai, dando-me ainda mais gana de ser sua referência.

A minha sogra, Miriam, pela dedicação, suporte e cuidados prestados a nosso filho Henrique. A conclusão desse trabalho teria sido muito mais difícil sem a sua ajuda.

Aos mestres, de toda minha trajetória escolar e acadêmica, pelo altruísmo por dividir o conhecimento conquistado. Em especial, à Profa. Dra. Joseana Fechine, pelo direcionamento e pela orientação dedicada nos primeiros dois anos e meio do doutorado, ao Prof. Dr. Eanes Torres Pereira, por se propor a orientar-me a partir de então, contribuindo incisivamente no desenvolvimento da nossa pesquisa e ao Prof. Dr. Marco Alvarez, pelo convite e todo apoio acadêmico/pessoal para a realização do Programa de Doutorado Sanduíche no Exterior (PSDE).

Ao colega e amigo, Prof. Dr. Dimas Cassimiro, pelo suporte e orientações acadêmicas prestadas ao longo da minha jornada na pós-graduação.

Ao Governo do Estado da Paraíba, pela licença concedida para a realização do Programa de Doutorado Sanduíche no Exterior (PSDE).

Aos colegas/amigos do Instituto de Polícia Científica da Paraíba (IPC-PB), pelo apoio e incentivo. Em especial, ao Diretor Geral do IPC-PB, Marcelo Lopes Burity, pela anuência da

liberação dos serviços laborais durante a realização do Programa de Doutorado Sanduíche no Exterior (PSDE) e ao então Chefe do Núcleo de Criminalística de Campina Grande, Elton Ferreira Frazão, por ser um dos maiores incentivadores para a realização do referido programa.

Ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande (COPIN/UFCG), pela oportunidade que me foi dada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro para a realização do Programa de Doutorado Sanduíche no Exterior (PSDE), que aconteceu entre os meses de agosto de 2019 e maio de 2020.

Conteúdo

Lista de Símbolos

ASGD - *Asynchronous Stochastic Gradient Descent*

AVP - *Average Precision*

BOVW - *Bag of Visual Words*

CC - *Camada de Convolução*

CP - *Camada de Pooling*

CTC - *Camadas Totalmente Conectadas*

CCP - *Código Penal Brasileiro*

CPU - *Central Processing Unit*

CVPR - *Conference on Computer Vision and Pattern Recognition*

DMCNet - *Deep Multicontext Network*

EER - *Equal Error Rate*

ECA - *Estatuto da Criança e do Adolescente*

EMA - *Erro Médio Absoluto*

FN - *Falso Negativo*

FP - *Falso Positivo*

GPU - *Graphics Processing Unit*

HSV - *Hue-Saturation-Value*

ICCV - *International Conference of Computer Vision*

ILSVRC - *ImageNet Large Scale Visual Recognition Challenge*

LUT - *Look-up Table*

MAE - *Mean Absolute Error*

MIL - *Multiple Instance Learning*

MLP - *Multilayer Perceptron*

MTCNN - *Multitask Cascaded Convolutional Networks*

NSFW - *Not-safe-for-work*

ORB - *Oriented FAST and Rotated BRIEF*

PCA - *Principal Component Analysis*

PEDA 376K - *Pornographic and Explicit Database 376K*

RGB - *Red-Green-Blue*

RNC - *Rede Neural Convolucional*

ROC - *Receiver Operating Characteristic*

SDH/PR - *Secretaria de Direitos Humanos da Presidência República*

SGD - *Stochastic Gradient Descent*

SVM - *Suport Vector Machine*

SURF - *Speeded Up Robust Features*

t-SNE - *t-Distributed Stochastic Neighbor Embedding*

TA - *Taxa de Aprendizado*

TVN - *Taxa de Verdadeiro Negativo*

TVP - *Taxa de Verdadeiro Positivo*

URL - *Uniform Resource Locator*

VN - *Verdadeiro Negativo*

VP - *Verdadeiro Positivo*

Lista de Figuras

Lista de Tabelas

Capítulo 1

Considerações Iniciais

1.1 Introdução

A inexorável popularização dos microcomputadores e *smartphones*, aliada ao amplo acesso à rede mundial de computadores, possibilitaram a agilidade na comunicação e na troca de informações entre os usuários. É inquestionável o grande avanço que o desenvolvimento dessas tecnologias trouxe e vem trazendo em aspectos sociais e econômicos. Contudo essa evolução também vem possibilitando que infratores realizem novas práticas ilegais (Crime Cibernético Próprio[1]) ou até mesmo aperfeiçoem crimes já tipificados (Crime Cibernético Impróprio[2]) (ELEUTÉRIO; MACHADO, 2011).

Anteriormente a essa popularização, atos preparatórios, de execução ou de consumação de um crime muitas vezes sequer eram registrados. Quando da ocorrência desse registro, se dava por meios analógicos de difícil propagação e alcance limitado (CAPPELLARI; VERONEZI, 2005). Atualmente, em poucos segundos, esses dispositivos computacionais são facilmente capazes de capturar, armazenar e compartilhar imagens (GANGWAR et al., 2017).

Os crimes relacionados à violência sexual contra crianças e adolescentes não fogem desse padrão. Esse tipo de crime, especificamente o compartilhamento e posse de arquivos pornográficos, também vem sofrendo os reflexos do avanço tecnológico. Os referidos crimes acabam recebendo um impulsionamento em seu cometimento devido: (i) à facilidade de

[1] São os crimes exclusivamente cibernéticos. Dependem da existência de ambiente computacional para a sua existência (VECCHIA, 2014)

[2] São os crimes onde o ambiente computacional serve apenas como ferramenta para a sua realização (VECCHIA, 2014)

captura das imagens por meio dos novos dispositivos que apresentavam essa funcionalidade (e.g. *smartphones*); (ii) à acessibilidade a dispositivos capazes de compartilhar esse tipo de material (e.g. *smartphones*, computadores pessoais); (iii) à popularização do acesso à internet, possibilitando que os arquivos ilícitos sejam facilmente compartilhados; e (iv) à falsa sensação de anonimato e segurança sentida pelo agressor (VELHO et al., 2016).

Com o objetivo de se adequar a essa nova realidade em que a sociedade encontra-se inserida, em de 25 de novembro de 2008 a posse de arquivos possuindo conteúdo pornográfico infanto-juvenil passou a ser caracterizado como crime, por meio da vigência da Lei N° 11.829/2008 (BRASIL, 2008). Anteriormente, em 1990, era apenas tipificado o compartilhamento de arquivos dessa natureza, de acordo com o artigo 241 do Estatuto da Criança e do Adolescente (BRASIL, 1990), tendo sido aprimorado somente em 2003, por meio da Lei N° 10.764, de 12 de novembro de 2003 (BRASIL, 2003).

Dessa forma, cada vez mais foi e vem sendo essencial a realização de exames periciais capazes de detectar essa espécie de material ilícito. Não se trata da realização de um exame de maneira discricionária, mas sim obrigatória, pois de acordo com o artigo 158 do Código de Processo Penal (CPP), "Quando a infração deixar vestígios, será indispensável o exame de corpo de delito, direto ou indireto, não podendo supri-lo a confissão do acusado" (BRASIL, 1941).

Embora os vestígios estejam presentes nos dispositivos de armazenamento, esses aparelhos vêm conseguindo, paulatinamente, armazenar mais arquivos de mídia, o que dificulta a realização do trabalho do perito criminal. A pesquisa de Polastro e Eleutério (2010) mostrou a existência de mais de 300.000 arquivos de imagens em um dispositivo dessa natureza. Por fim, os autores afirmam que, após a realização do exame pericial, foram encontrados pouco mais de uma centena de arquivos que evidenciaram algum tipo de crime.

Ademais, Platzer, Stuetz e Lindorfer (2014) afirmam que a realização de uma inspeção humana em busca de imagens ilícitas, por um longo período de tempo, resulta na diminuição momentânea da capacidade cognitiva e de concentração humana. Esses efeitos se dão devido à lentidão e monotonia da referida tarefa. Dessa forma, essa atividade acaba estando sujeita ao erro humano, fazendo com que imagens ilícitas passem despercebidas. Dessa forma, se faz necessário que esta busca seja automatizada, diminuindo os erros cometidos por uma análise manual, assim como o tempo demandado para a realização dessa tarefa.

Além do mais, a distinção entre imagens pornográficas de não pornográficas muitas vezes torna-se difícil de ser julgada pelo próprio ser humano, muito devido à existência de conteúdo subjetivo em algumas destas mídias. Dessa forma, diferentes pessoas poderão possuir classificações distintas para uma mesma imagem (PUTRO; ADJI; WINDURATNA, 2015). Portanto, não se trata de uma tarefa trivial, tornando-se mais árdua quando tem-se que inferir se determinado ator já atingiu ou não a maioridade penal (dezoito anos no Brasil) (ELEUTÉRIO; MACHADO, 2011).

Por fim, a junção da: (i) facilidade de disseminação de arquivos de imagens, (ii) alta capacidade de armazenamento dos dispositivos computacionais; (iii) lentidão e falibilidade de uma inspeção totalmente humana; e, principalmente, (iv) tipificação criminal da posse destes arquivos faz com que seja cada vez mais necessária uma ferramenta que auxilie as atividades do Perito Criminal da área da Computação Forense a identificar esse tipo de arquivo em dispositivos computacionais de maneira eficaz.

1.2 Problematização

Baseado na publicação dos dados de um serviço mantido pela Secretaria de Direitos Humanos da Presidência da República (SDH/PR), o Disque 100, foram registrados em 2017 mais de 84 mil denúncias relativas à violência sexual contra crianças e adolescentes. Entretanto trata-se apenas de uma parcela da realidade, visto que, na maioria dos casos, a denúncia é inexistente ou sequer chega a ser descoberta a realização do crime.

É importante esclarecer que a violência sexual contra a criança e o adolescente é caracterizada de duas maneiras: a (i) exploração sexual infanto-juvenil e o (ii) abuso sexual. Na primeira forma de violência sexual o grande interesse é o aspecto financeiro, em que crianças e/ou adolescentes são tratados como mercadorias, por meio da realização do turismo sexual, do tráfico de pessoas e da **produção e comercialização de pornografia**. No segundo tipo, a motivação maior é a satisfação sexual própria de um adulto, existindo ou não o consentimento do menor de idade.

Esses abusos tanto podem acontecer por meio de relações sexuais, realização de carícias, exposição a situações em que encontram-se vulneráveis (parcial ou totalmente desnudos em conotação sexual) e **produção de material pornográfico** (VELHO et al., 2016). Dessa

forma, fica bem claro que a produção, disseminação e até mesmo a posse de pornografia infanto-juvenil são caracterizadas como violência sexual contra crianças e adolescentes, ora como exploração, ora como abuso.

De acordo com Velho et al. (2016), esse tipo de violência contra a criança e o adolescente sempre existiu ao longo da história da humanidade. Entretanto os avanços tecnológicos possibilitaram a produção de material pornográfico e, principalmente nas últimas décadas, alavancaram a disseminação desse material, por meio das facilidades expostas pela Internet e pelo uso de computadores pessoais e *smartphones*, tornando o crime de posse e compartilhamento de pornografia infanto-juvenil cada vez mais comum.

1.3 Motivação e Justificativa da Pesquisa

Atualmente, a maior parcela dos institutos de polícia científica do Brasil não possui ferramentas para a detecção de material pornográfico, quiçá voltados especificamente para o âmbito infanto-juvenil. Dessa forma, esse tipo de exame pericial acaba tendo que ser realizado de maneira não automática.

A realização dessa análise nos referidos moldes sujeita-se a falhas humanas, por se tratar de uma tarefa repetitiva e exaustiva. Essas falhas acontecem principalmente devido à grande quantidade de imagens que os atuais dispositivos objetos dos exames periciais conseguem armazenar.

Ademais, a demanda de exames desse tipo vem aumentando inexoravelmente, devido ao crescimento do número de casos de crimes relacionados à posse e ao compartilhamento de material pornográfico infanto-juvenil ao longo das últimas décadas. Esse aumento torna-se um fator agravante no desenvolvimento das atividades periciais dos institutos de polícia científica.

Portanto, percebeu-se a necessidade da criação de uma ferramenta, voltada para os institutos de polícia científica do Brasil, que fosse capaz de auxiliar o perito criminal na realização dos exames periciais dessa natureza, minimizando seu tempo de realização e, ao mesmo tempo, aumentando a confiabilidade final do exame, por meio do uso de técnicas inovadoras.

1.4 Objetivos

Na sequência, serão explicitados o objetivo geral e os objetivos específicos referentes à pesquisa em questão.

1.4.1 Objetivo Geral

Propor, modelar e validar uma nova metodologia, no nível do estado da arte, que seja capaz de identificar conteúdo pornográfico infanto-juvenil em imagens.

1.4.2 Objetivos Específicos

Partindo-se do objetivo geral, são destacados os seguintes objetivos específicos:

1. Especificar uma arquitetura que, por meio de técnicas de Visão Computacional, seja capaz de identificar imagens pornográficas de qualquer natureza (i.e. pornografia adulta ou pornografia infanto-juvenil);

2. Propor uma metodologia que, dada uma imagem pornográfica, aponte a probabilidade de os indivíduos envolvidos serem menores de idade, por meio do uso de estimativa de idade facial;

3. Desenvolver o modelo proposto sem a utilização de imagens de conteúdo pornográfico infanto-juvenil (utilizadas apenas para validação dos resultados finais), dada sua escassez, dificuldade de aquisição e restrições de manipulação e;

4. Contribuir para a evolução de Perícia Criminal na área da Computação Forense, por meio do uso de uma metodologia inovadora.

1.5 Questões de Pesquisa

Baseado nos Objetivos Geral e Específicos, foram elaboradas as seguintes Questões de Pesquisa (QP) que guiaram este estudo:

1. QP1 - É possível superar o estado da arte na detecção de pornografia infanto-juvenil em imagens sem utilizar imagens dessa natureza para a construção do modelo?

2. QP2 - Como desenvolver uma técnica competitiva de reconhecimento de pornografia quando comparada a serviços de moderação de imagens inseridos no estado da arte?

3. QP3 - É viável superar o estado da arte na estimativa de idade real em faces e aplicar essa técnica para determinação de menoridade penal de um indivíduo?

4. QP4 - Como determinar, com associação de uma probabilidade, o resultado do reconhecimento de menores de idade nas imagens pornográficas?

As referidas questões foram respondidas por meio da realização das atividades a seguir:

1. Uso de uma arquitetura em série que, em um primeiro momento, seja capaz de classificar uma imagem como pornográfica ou não, utilizando um modelo treinado apenas com imagens contendo pornografia adulta e imagens não ofensivas. Em um segundo momento, caso a imagem seja classificada como pornográfica, estimar a idade facial dos indivíduos envolvidos por meio de um modelo treinado com faces de imagens não ofensivas (referente à **QP1**).

2. Desenvolvimento de uma base de dados pornográfica possuindo uma grande quantidade e diversidade de dados, com critérios objetivos sobre definição de pornografia que resultaram em rótulos confiáveis referentes às classes: (i) pornografia e (ii) não pornografia (referente à **QP2**).

3. Proposição de uma arquitetura baseada em aprendizado profundo, escolhida por meio de uma estratégia gulosa que definiu a melhor rede neural convolucional a ser utilizada para a detecção de pornografia, assim como seus principais hiperparâmetros (referente à **QP2**).

4. Padronização, por meio de aprendizado de máquina tradicional, dos resultados oriundos dos serviços de moderação de imagens, para que seus desempenhos pudessem ser comparadas entre si e com a arquitetura proposta para detecção de pornografia (referente à **QP2**).

5. Apresentação de uma nova técnica de estimação de idade real em faces, por meio do uso de um modelo baseado em aprendizado profundo. A referida técnica utiliza como

rótulo de treinamento uma distribuição de probabilidade discreta de idades gerada a partir de informações de idade real e aparente de faces (referente à **QP3**).

6. Apresentação de uma análise comparativa entre técnicas utilizadas para estabelecer a probabilidade de uma face pertencer a um indivíduo com menos de dezoito anos (referente à **QP4**).

1.6 Contribuições

As principais contribuições deste trabalho de doutorado são apresentadas a seguir:

- Construção e uso de uma nova base de dados, a *Pornographic and Explicit Database 376K (PEDA 376K)*. Essa base de dados contém mais de 376 mil imagens, categorizadas manualmente em duas classes: (i) pornográficas e não pornográficas. Foi utilizada uma definição objetiva de pornografia por meio de regras explícitas, baseadas na pesquisa de Wang, Jin e Tan (2016), que resultaram em rótulos confiáveis para cada umas das classes.

- Aplicação de uma estratégia gulosa, aliada ao uso da base de dados PEDA 376K, para ajustar uma arquitetura baseada em aprendizado profundo capaz de diferenciar imagens pornográfica de não pornográficas, superando serviços de moderação de imagens no estado da arte.

- Proposição de uma técnica de estimação de idade real em faces, também baseada em aprendizado profundo, capaz de superar as pesquisas do estado da arte que também utilizaram como parâmetro a base de dados APPA-REAL **??**. Essa abordagem se diferencia das demais por também utilizar dados de idade aparente em faces para construir distribuições de probabilidade discreta das idades reais para treinamento do modelo.

- Apresentação de uma abordagem, utilizando os módulos propostos, capaz de detectar imagens pornográficas e, dada(s) a(s) face(s) do(s) indivíduo(s), determinar a(s) probabilidade(s) desse(s) ser(em) menor(es) de idade, sem o uso de imagens pornográficas de crianças e/ou adolescentes na etapa de construção do modelo.

- Aplicação do Classificador de Menoridade Penal, por meio do uso de uma rede neural multicamada perceptron, com o intuito de minimizar o erro médio das probabilidades estimadas de indivíduos possuírem menos de dezoito anos, baseado nos metadados e na idade predita da cada uma de suas face.

1.7 Estrutura do Documento

No Capítulo 2, é realizada uma contextualização da Computação Forense na Criminalística e sua relação com a pornografia infanto-juvenil, além do tratamento da legislação brasileira no tocante a crimes dessa natureza.

No Capítulo 3, é apresentada uma fundamentação teórica necessária para o desenvolvimento da pesquisa, considerando-se os conceitos de Visão Computacional, aprendizado de máquina tradicional, redes neurais convolucionais e detecção de pele humana.

No Capítulo 4 são discriminados os estudos contidos na literatura que apresentaram relação com a pesquisa, visando obter maior conhecimento sobre a propositura levantada.

No Capítulo 5, é detalhada a arquitetura adotada para a realização da tarefa alvo da pesquisa, a detecção de pornografia infanto-juvenil. Ademais, contextualiza-se as abordagens propostas nos Capítulos 6, 7 e 8 na arquitetura em questão, apontando seus papeis e onde encontram-se inseridas na arquitetura em questão.

No Capítulo 6, é demonstrado todo o processo de criação de uma grande base de dados de imagens pornográficas, a *Pornographic and Explicit Database 376K* (PEDA 376K). Também é descrito o porquê da sua necessidade, além de um breve levantamento bibliográfico das bases de dados com esse tipo de imagem disponíveis na literatura. Ademais, é proposta uma arquitetura, com base em aprendizado profundo, treinada com os dados da PEDA 376K. Por fim, é apresentada uma metodologia comparativa entre a arquitetura proposta e os serviços de moderação de imagens inseridos no estado da arte, por meio da padronização dos resultados baseada em aprendizado de máquina, que viabiliza a realização dessa análise comparativa.

No Capítulo 7, é apresentado um estudo comparativo entre técnicas consolidadas de estimação de idade real facial e técnicas propostas capazes de aprimorar a estimativa de idade real em faces, utilizando dados de idade aparente facial em alguns casos.

No Capítulo 8, é apresentado o Classificador de Menoridade Penal, que visa direcionar

o Perito Criminal na decisão de determinar se uma imagem pornográfica contem crianças e adolescentes ou não, por meio do aprimoramento da probabilidade de uma determinada face contida na imagem pertencer a um indivíduo com menos de 18 anos.

No Capítulo 9, são expostos os resultados obtidos na análise experimental, que contempla as proposições dos Capítulos 6, 7 e 8, assim como da arquitetura como um todo, apresentada no Capítulo 5. Por fim, tempos de processamento da abordagem proposta são detalhados por meio da realização de uma análise de desempenho computacional.

No Capítulo 10, são expostas as considerações finais do trabalho e as respostas das questões de pesquisa referentes ao estudo em questão.

Por fim, no Apêndice A, são expostas proposições iniciais de detecção de pornografia fundamentada em detecção de pele humana, detecção de faces e aprendizado de máquina tradicional.

Capítulo 2

A Computação Forense e a Pornografia Infanto-juvenil

Este capítulo tem como objetivo esclarecer os conceitos básicos inerentes à Computação Forense no geral, seu papel na Criminalística, assim como a relação existente com a pornografia infanto-juvenil.

2.1 Computação Forense

Em 1965, Gordon Moore afirmou que, a cada 18 meses, a quantidade de transistores impressos em uma pastilha seria duplicada sem afetar seu custo de fabricação. Por muitos anos essa tendência fez com que os computadores se tornassem cada vez mais acessíveis e, consequentemente, mais presentes no cotidiano dos indivíduos em geral (ELEUTÉRIO; MACHADO, 2011). Mesmo a "Lei de Moore" não perdurando até os dias atuais, a ainda rápida evolução computacional continua potencializando paulatinamente essa popularização.

Entretanto essa propagação abriu margem para o uso desse meio para o cometimento de novos crimes, ou até mesmo antigos, mas de uma maneira mais atual e sofisticada. Acompanhando essa popularização, a Computação Forense vem progressivamente atuando na área da Criminalística com o intuito de desvendar esses crimes cibernéticos.

2.1.1 A Computação Forense na Criminalística

A utilização do conjunto de conhecimento de várias ciências, com o objetivo de obter informações de vestígios em locais de crime, que auxiliem na materialização e/ou autoria do fato ocorrido é denominada Criminalística (ESPÍNDULA et al., 2007). Sendo assim, por meio dos conhecimentos oriundos da Ciência da Computação, a Computação Forense enquadra-se como uma ramificação da Criminalística, tratando especificamente do levantamento de vestígios cibernéticos em locais de infrações penais (VELHO et al., 2016).

De acordo com Edmond Locard[1], o princípio básico da ciência forense contempla que, em locais de crime, sempre existe uma troca em que o criminoso leva consigo algo do local e deixa alguma coisa para trás quando parte. Mesmo tendo sido um princípio postulado no início do século XX, em que os dispositivos computacionais sequer existiam, é amplamente aplicável à Computação Forense (POLLITT, 2008).

Tendo em vista a obrigatoriedade imposta no Código Penal Brasileiro (CPP) (BRASIL, 1941), em seu artigo 158, no qual afirma ser indispensável a realização de exame pericial em locais de crime contendo vestígios, torna-se imprescindível o exame pericial em locais de crimes cibernéticos[2] (ELEUTÉRIO; MACHADO, 2011).

Portanto, o objetivo principal da Computação Forense é indicar a: (i) dinâmica, (ii) autoria e (iii) materialidade dos crimes relacionados à área de Informática, por meio do levantamento de evidências digitais, transformando-as em provas materiais de crime, utilizando métodos técnico-científicos que conferem-lhes validade probatória em juízo (ELEUTÉRIO; MACHADO, 2011).

2.1.2 Crimes Cibernéticos

Em decorrência dos avanços tecnológicos no final do século XX, surgiram novos crimes que até então não eram sequer tipificados, visto que o Código Penal Brasileiro (BRASIL, 1940) não previa a existência desse novo ambiente que possibilitaria essa prática. Dessa forma, até a criação da Lei N° 12.737/2012 (“Lei Carolina Dieckmann”), os operadores do direito não possuíam lei específica para tipificar esse tipo de conduta, utilizando de analogias para

[1] Edmond Locard (1877-1966) foi um pioneiro da ciência forense, conhecido também como o Sherlock Holmes da França.

[2] Trata-se de um local de crime convencional contendo dispositivos computacionais que podem apresentar relação com a investigação.

caracterizar os crimes dessa natureza (VECCHIA, 2014).

Os crimes cibernéticos não são necessariamente novas modalidades de crimes, pois crimes previamente tipificados podem ser realizados com o auxílio de um ambiente computacional. Essas duas modalidades de crimes cibernéticos são denominadas: (i) Crime Cibernético Próprio, que exige e depende do ambiente computacional para sua realização e (ii) Crime Cibernético Impróprio, no qual se utiliza o ambiente computacional apenas como ferramenta para a prática do crime (ELEUTÉRIO; MACHADO, 2011)

2.2 Pornografia Infanto-juvenil

Baseado no Protocolo Facultativo para a Convenção sobre os Direitos da Criança sobre a venda de crianças, prostituição e pornografia infantis[3] e no Estatuto da Criança e do Adolescente (BRASIL, 1990), a pornografia infanto-juvenil é caracterizada quando da situação que envolva criança ou adolescente em atividade sexual explícita, real ou até mesmo simulada (inclusive montagem) ou exibição, com fins sexuais, de pelo menos um de seus órgãos genitais.

2.2.1 Pornografia Infanto-juvenil ao Longo dos Anos

A pornografia infanto-juvenil é um dos crimes mais comuns que envolvem a utilização de dispositivos computacionais. Muito se deu pela modernização dos meios de comunicação e facilidade de dispositivos em capturar e disseminar fotos e vídeos. Além do mais, a falsa sensação de anonimato faz com que o agente delituoso perca o receio em praticar esse tipo de crime, por meio da posse ou até mesmo o compartilhamento desse tipo de material ilícito (VELHO et al., 2016).

A violência sexual contra crianças e adolescentes não é uma prática recente. Contudo ganhou e vem ganhando impulso, em decorrência do avanço da tecnologia nas últimas décadas. Mesmo tendo sido alavancado pelo avanço tecnológico, esse tipo de infração é caracterizada como crime cibernético impróprio, visto que sua prática não é estritamente dependente do ambiente computacional, pois a posse e a disseminação desse material ilícito pode ser reali-

[3]Protocolo adotado pela Assembleia Geral das Nações Unidas, em 25 de maio de 2000 e promulgado pelo Decreto N° 5.007, de 8 de março de 2004

zado sem dispositivos computacionais, como por meio de fotografias analógicas ou impressões em papel (VECCHIA, 2014).

2.2.2 A Pornografia Infanto-juvenil e a Legislação Brasileira

A partir de 1990, a legislação brasileira começou a contemplar crimes que possuíssem relação com material pornográfico infanto-juvenil. Originalmente, os artigos 240 e 241 do Estatuto da Criança e do Adolescente (ECA) (BRASIL, 1990) apresentavam tipificação basicamente para produção teatral, televisiva ou cinematográfica de material pornográfico infanto-juvenil, assim como a sua publicação.

Em 2003, a Lei N° 10.764, de 12 de novembro de 2003 (BRASIL, 2003) passou a vigorar, dando uma nova redação para os artigos 240 e 241 do ECA. Essa reformulação passou a contemplar também atividades fotográficas e outros meios visuais, publicação na internet, além de punir também a venda desse material.

Até o ano de 2008 somente a posse de material pornográfico infantil não era considerada crime, o que dificultava bastante a caracterização do crime, pois o exame pericial tinha que comprovar a criação ou a distribuição desse material. Entretanto a vigência da Lei N° 11.829, de 25 de novembro de 2008 (BRASIL, 2008) deu uma segunda redação para o artigo 240 e desmembrou o artigo 241 do ECA em cinco novos artigos (241-A, 241-B, 241-C, 241-D e 241-E).

Essa reformulação, dentre outras mudanças, passou a considerar crime a simples posse de material pornográfico que, até então, não era caracterizado como tal. A partir de então, a constatação de posse desse conteúdo tornou-se operação constante e comum no âmbito da Computação Forense, gerando uma demanda antes inexistente (ELEUTÉRIO; MACHADO, 2011).

2.2.3 Determinação e Estimativa de Idade de Crianças e Adolescentes

Até então, não existe uma maneira exata de determinar a idade de um ser humano, a não ser por meio de prova documental. Existem técnicas presenciais capazes de lançar estimativas da idade de pessoas. Contudo não coincidem com a idade cronológica. Algumas dessas técnicas são: desenvolvimento ósseo e dental, desenvolvimento dos caracteres sexuais secundários e

presença de doenças sistêmicas. No que diz respeito à determinação da idade em imagens, cuja interação com os personagens não é possível, acaba sendo uma tarefa ainda mais árdua. Alguns estudos utilizam a mensuração de elementos macroscópicos como proporcionalidade entre membros e órgãos (VELHO et al., 2016).

Com relação à apreciação das características dos órgão genitais, muito se utiliza a "Escala de Tanner" ou "Estagiamento de Tanner" para determinação do estágio púbere de um indivíduo e pode ser utilizado para estimar idades nas vítimas de pornografia infanto-juvenil (GREENBERGER et al., 1975). Essa técnica é dividida em cinco níveis, levando-se em consideração a distribuição dos pelos pubianos, do tamanho do órgão sexual masculino e das mamas femininas.

Dessa forma, ainda é desafiador estimar se determinado indivíduo possui menos de 18 anos, principalmente quando isso envolve adolescentes, pois muitos atingem maturidade corporal de maneira antecipada devido a diversas causas, tais como fatores climáticos, sociais econômicos, genéticos e raciais. No caso de bebês e crianças, afirmar que são menores de 18 anos é uma tarefa bem mais simples, visto as singularidades inerentes a essas duas classes (VELHO et al., 2016).

Além das dificuldades relatadas, muitas imagens fotográficas apresentam baixa qualidade, ângulos desfavoráveis ou mostram apenas partes específicas do corpo. Esses fatores acabam dificultando ainda mais a estimação da idade dos indivíduos envolvidos (VELHO et al., 2016).

2.3 Considerações Finais

Neste capítulo, foi exposta a relação existente entre a Computação Forense e a pornografia infanto-juvenil. Inicialmente, mostrou-se onde se encontra inserida a Computação Forense no universo da Criminalística, assim como a definição de crimes cibernéticos e seus tipos.

Em seguida, definiu-se categoricamente a pornografia infanto-juvenil, além da prática dessa modalidade de crime ao longo dos anos, tendo sido traçado o histórico evolutivo da legislação brasileira em relação às tipificações da prática desse tipo de crime.

Por fim, foram citadas técnicas de determinação e estimativa de idade em crianças e adolescentes. Entretanto ratifica-se que essas técnicas não são exatas, traçando apenas uma

aproximação da idade. Destacou-se também a dificuldade ainda maior da aplicação dessas técnicas em imagens que apresentam baixa qualidade, ângulo desfavorável, dentre outros fatores dificultadores.

Capítulo 3

Fundamentação Teórica

Este capítulo tem como objetivo explicitar os conceitos básicos e terminologias relativas ao entendimento do presente trabalho. Os principais conceitos apresentados são: (i) Visão Computacional; (ii) Aprendizado de Máquina; (iii) Redes Neurais Convolucionais e (iv) Detecção de Pele Humana.

3.1 Visão Computacional

Trata-se de um campo interdisciplinar que, por meio da junção de software e hardware, automatiza tarefas realizadas pelo sistema visual humano, sendo muitas vezes capaz de interpretar imagens e vídeos digitais de maneira mais aprimorada (SONKA; HLAVAC; BOYLE, 2014).

A Visão Computacional compreende diversas áreas: (i) análise de movimento: trata-se do processamento de diversas imagens em sequência, tendo como objetivo estimar a velocidade de determinado objeto; (ii) reconstrução de cena: área que visa computar um modelo em três dimensões, dadas duas ou mais imagens de uma determinada cena; (iii) restauração de imagens: tarefa que tem como fim primordial a remoção de ruídos em imagens e (iv) reconhecimento: o domínio principal da Visão Computacional, cujo objetivo principal é determinar se uma imagem possui algum objeto específico, característica ou atividade (SZELISKI, 2010).

3.1.1 Classificação de Imagem

A Classificação de Imagem, inserida na área de reconhecimento da Visão Computacional, se resume a, dada uma imagem qualquer de entrada, realizar a inferência para estimar a probabilidade de essa pertencer a uma determinada classe (LU; WENG, 2007). Na visão computacional, esse procedimento é realizado por meio de técnicas capazes de extrair características relevantes das imagens, em que padrões de cada categoria são reconhecidos e, dessa forma, podem ser categorizados como tal (DESHPANDE, 2016).

Para os seres humanos, trata-se de uma tarefa natural e muitas vezes inconsciente, apoiando-se na detecção de características para a tomada dessa decisão. Por exemplo, para identificar uma determinada espécie de animal, observa-se a existência de algumas singularidades em cada ser (i.e. asas, patas, tamanho, cor, mamas).

3.1.2 Detecção de Faces

A detecção de faces, subárea da detecção de objetos, é um tópico clássico e um dos mais estudados na Visão Computacional (MATHIAS et al., 2014). Seu papel é determinar a presença e a localização de uma face em uma imagem, distinguindo-a de todos os outros padrões contidos no ambiente (GUPTA; SHARMA, 2014).

De acordo com Comaschi (2016), os detectores de faces podem ser categorizados em classes: (i) modelos rígidos, (ii) modelos de partes deformáveis e (iii) redes neurais convolucionais. Os modelos rígidos têm como linha principal de pesquisa o trabalho de Viola e Jones (2004), tendo servido de base para muitos dos melhores detectores de face historicamente descritos na literatura. Por muitos anos os modelos de partes deformáveis figuraram no estado da arte da detecção de faces. Originalmente foram desenvolvidos para detecção de objetos (FELZENSZWALB; HUTTENLOCHER, 2005) e, posteriormente, aperfeiçoados para a detecção específica de faces (FELZENSZWALB et al., 2010).

As redes neurais convolucionais têm sido amplamente utilizadas na Visão Computacional para a classificação de imagens, isso deve-se muito ao avanço do poder de processamento dos computadores e ao acesso a um grande número de dados rotulados. Dessa forma, por se tratar de um subproblema da área, diversas pesquisas têm utilizado dessa técnica para o reconhecimento de faces, atingindo resultados do estado da arte (VITORINO et al., 2018).

3.2 Aprendizado de Máquina Tradicional

Nesta seção, é apresentado o conceito de aprendizado de máquina, assim como quatro algoritmos desse campo de estudo que foram utilizados nesta pesquisa. Apesar da sua popularização ter sido impulsionada nos últimos anos, o campo de estudo de aprendizado de máquina não é tão recente. Em 1959, um dos pioneiros nessa área, Arthur Samuel[1], já o definia como a habilidade dos computadores aprenderem sem serem explicitamente programados (SAMUEL, 1959).

Modelos baseados em aprendizado de máquina são capazes de aprender e incrementar seus desempenhos por meio da entrada de dados e, por muitas vezes, generalizando-os para obter melhores resultados. A aquisição desse conhecimento se dá por meio da aplicação de algoritmos matemáticos e estatísticos nos dados passados, fazendo com que seja possível realizar uma predição (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007).

De maneira geral, os algoritmos de aprendizado de máquina podem ser subdivididos em aprendizado não supervisionado, supervisionado e por reforço. Os algoritmos supervisionados recebem os dados de treinamento de maneira rotulada, ou seja, de maneira prévia é inferido o valor ou a classe daquele determinado dado e o algoritmo é orientado a aprender baseado na relação entre dado propriamente dito e seu rótulo (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007). Nos algoritmos não supervisionados não existe um rótulo relativo a cada dado disponibilizado, dessa forma, o algoritmo agrupa os dados de acordo com a similaridade entre eles, distribuindo-os em subgrupos (GHAHRAMANI, 2003). No aprendizado por reforço, o modelo aprende a melhor decisão a ser tomada por meio de tentativas, resultando em acertos ou erros que são, respectivamente, recompensados ou punidos (SUTTON; BARTO, 2018).

Um dos problemas clássicos no qual o aprendizado de máquina supervisionado pode ser aplicado é o de classificação. Para que o classificador desempenhe sua função é necessário que haja uma etapa inicial de treinamento. Nessa etapa, o modelo recebe dados rotulados no domínio das classes que serão classificados. De maneira concomitante, é realizado o ajuste fino dos principais hiperparâmetros do modelo para que o classificador possa generalizar os dados, visando à etapa de teste, em que serão utilizados dados diversos do conjunto de

[1] Arthur Lee Samuel (1901-1990) foi um pioneiro no campo de jogos de computadores e inteligência artificial, tendo o mesmo criado o termo “Aprendizado de Máquina”, em 1959.

treinamento (NASCIMENTO FILHO, 2017).

O problema em questão pode ser formalizado matematicamente na notação vetorial, em que o domínio é dado por $X = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$, tal que x_i representa um vetor de atributos do i-ésimo exemplo e y_i representa seu respectivo rótulo de classe que pertence ao conjunto $Y = \{y_1, y_2, ..., y_k\}$ que possui k classes.

Sendo assim, um classificador é uma função dada por $g : X \rightarrow Y$, utilizada para classificar instâncias de um conjunto de teste E que não estão contidas na base de treinamento inicial T, ou seja, $T \cap E = \emptyset$, dado que $T \subset X \times Y$ e $E \subset X \times Y$, X representa o espaço de atributos e Y é um conjunto finito e discreto que representa os rótulos das classes.

3.2.1 Regressão Logística Binária

Desde o início do século XX, a regressão logística vem sendo utilizada em aplicações científicas, com o intuito de classificar variáveis categóricas por meio de uma relação linear entre os valores de entrada e os valores de saída (SWAMINATHAN, 2018).

Entretanto esse método não é capaz de distinguir dados que não são separados linearmente, senão por meio do uso de engenharia de características (e.g. multiplicação e potenciação dos valores). Em sua disposição original, essa abordagem geralmente não é recomendada para lidar com informações complexas (MOREIRA; FECHINE, 2018a).

Representação do Modelo

O modelo corresponde a uma função de dimensão n, que representa a quantidade de parâmetros utilizados (JAMES et al., 2013), como pode ser observado na Equação (3.1).

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_n x_n \tag{3.1}$$

Considerando θ como uma matriz $1 \times (n + 1)$ contendo os valores $\theta_0, \theta_1, \theta_2, ..., \theta_n$ e X como uma matriz $1 \times (n + 1)$ contendo os valores $1, x_1, x_2, ..., x_n$, a função que define o modelo de regressão linear pode ser reescrita conforme a Equação (3.2).

$$h_\theta(x) = \theta^T X \tag{3.2}$$

Como os únicos resultados desejados para a função $h(x)$ são 0 e 1, a condição $0 \leq$

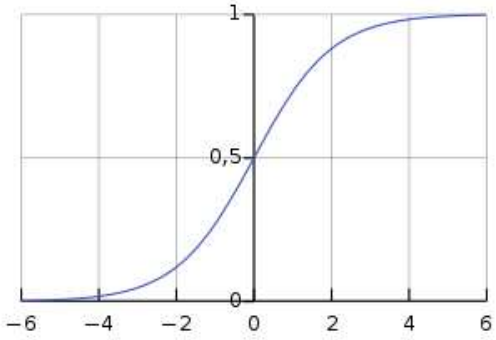
$h(x) \leq 1, h(x) \in \mathbb{R}$ deve ser satisfeita. Para tal, é aplicada a função logística, também conhecida por função sigmóide, de acordo com as Equações (3.3), (3.4) e a Figura 3.1, resultando em uma nova função $h_\theta(x)$, conforme explicitado na Equação (3.5).

$$z = \theta^T X \tag{3.3}$$

$$g(z) = \frac{1}{1 + e^{-z}} \tag{3.4}$$

$$h_\theta(x) = g(\theta^T X) \tag{3.5}$$

Figura 3.1: Função Sigmóide.



Fonte: Autor.

Por fim, para que haja a categorização, o valor retornado pela função $h(x)$ deve pertencer ao universo dos números naturais, $h(x) \in \mathbb{N}$. Essa transformação é dada na Equação (3.6).

$$\begin{aligned} h_\theta(x) \geq 0,5 \rightarrow y = 1 \\ h_\theta(x) < 0,5 \rightarrow y = 0 \end{aligned} \tag{3.6}$$

Função de Custo

De acordo com Géron (2019), o ponto de partida da criação do modelo é que os valores $\theta = \theta_0, \theta_1, ..., \theta_n$ sejam inicializados (geralmente com valor zero) para que possa ser calculado o seu custo ($J(\theta)$). A perda para cada valor de entrada $(x^{(p)}, y^{(p)})$ é calculada, sendo $p = 1, 2, ..., m$. A função de custo J é dada pela média do somatório das perdas L, conforme

pode ser visto na Equação (3.7), assim como nas Figuras 3.1 e 3.2.

$$\begin{aligned} J(\theta) &= \frac{1}{m}\sum_{i=1}^{m} \mathrm{L}(h_\theta(x^{(i)}), y^{(i)}) \\ \mathrm{L}(h_\theta(x), y) &= \begin{cases} -\log(h_\theta(x)), & \text{se y = 1} \\ -\log(1 - h_\theta(x)), & \text{se y = 0} \end{cases} \end{aligned} \tag{3.7}$$

A função de custo $J(\theta)$ pode ser reescrita de maneira simplificada, removendo-se seu formato condicional, conforme explicitado na Equação (3.8).

$$\mathrm{L}(h_\theta(x), y) = -y\ log(h_\theta(x)) - (1 - y)log(1 - h_\theta(x)) \tag{3.8}$$

Por fim, o custo total $J(\theta)$ pode ser calculado da seguinte maneira, conforme explicitado na Equação (3.9).

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} -y^{(i)}\ log(h_\theta(x^{(i)})) - (1 - y^{(i)})log(1 - h_\theta(x^{(i)})) \tag{3.9}$$

Gradiente Descendente

A partir do cálculo da função de custo $J(\theta)$ do modelo, é possível realizar sua minimização, otimizando-se os parâmetros θ por meio da técnica de gradiente descendente. Salienta-se a importância do uso da função logarítmica na função de custo $J(\theta)$, pois isso altera seu comportamento para convexo, impedindo que o algoritmo do gradiente descendente estagne em um mínimo local e não atinja o mínimo global (GÉRON, 2019).

A otimização de cada parâmetro é dada por seu valor subtraído da derivada da função de custo, sempre multiplicada pela taxa de aprendizado α, conforme se vê na Equação (3.10). Os valores de θ são atualizados até que seja atingido o valor mínimo, conforme critérios de parada adotados.

$$\begin{aligned} \theta_{j'} &= \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta) \\ \theta_{j'} &= \theta_j - \frac{\alpha}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \end{aligned} \tag{3.10}$$

Regularização

Por muitas vezes, um modelo se encontra muito ajustado aos dados da etapa de treinamento (*overfitting*) e acaba não generalizando bem para a predição de novos dados. Esse comportamento se dá devido à complexidade e ao grande número de parâmetros utilizados no modelo (GRUS, 2019).

A regularização tem a função de penalizar os parâmetros θ (geralmente aqueles de maior complexidade) no momento de sua otimização, simplificando o modelo, de forma que possa apresentar maior generalização no momento da predição dos dados. A regularização se dá por intermédio da adição da parcela $\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$ no final do cálculo do Custo $J(\theta)$, conforme pode ser visualizado na Equação (3.11). O parâmetro λ é o fator de regularização e tem como função controlar quão regularizado o modelo será.

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\ log(h_\theta(x^{(i)})) - (1-y^{(i)})log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2 \quad (3.11)$$

A alteração no cálculo da função de custo $J(\theta)$, consequentemente, altera o algoritmo do gradiente descendente. Salienta-se que o limiar (θ_0) possui uma equação diferenciada por não estar inserido na regularização. Na Equação (3.12), mostra-se a nova formulação do gradiente descendente com a inclusão da regularização.

$$\begin{aligned} \theta_{0'} &= \theta_0 - \frac{\alpha}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)} \\ \theta_{j'} &= \theta_j - \frac{\alpha}{m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} + \lambda\theta_j\right], \text{para j = \{1, 2, ..., n\}} \end{aligned} \quad (3.12)$$

3.2.2 Perceptron Multicamadas

As redes neurais artificiais têm como principal inspiração o cérebro humano, simulando as conexões reais dos neurônios por meio do uso de interconexões de neurônios artificiais, na tentativa de reproduzir o alto poder de processamento da mente humana (AL-MOHAIR; SALEH; SUANDI, 2015a).

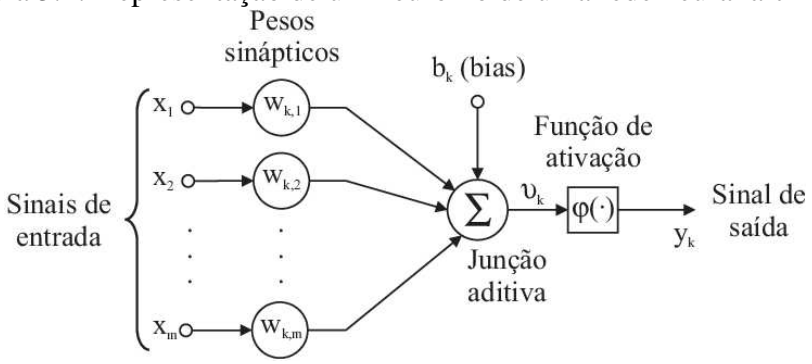
Perceptron multicamadas é uma rede neural artificial composta por três ou mais camadas

de neurônios conectados por pesos entre cada camada adjacente. Com exceção dos nodos de entrada, cada neurônio passa seus dados para a próxima camada por meio de uma função de ativação não linear (e.g. Sigmoide, Tanh, ReLu). Na etapa de treinamento, é utilizado um algoritmo de retropropagação do erro, denominado *backpropagation*, para a minimização da função de custo por meio da otimização dos parâmetros. As múltiplas camadas e as funções de ativação não lineares são capazes de classificar dados de maior complexidade (BISHOP, 2006).

Representação do Modelo

A Figura 3.2 representa o modelo artificial de um neurônio. Essa modelagem apresenta um sinal de limiar (b) que regula a ativação do neurônio. As conexões são representadas pelos pesos (W) que amplificam cada um dos sinais recebidos da camada anterior (ou de entrada) (X) e a função de ativação (φ) define a saída da camada da rede neural, modelando a forma como o neurônio responde ao nível de excitação, conforme as Equações (3.13), (3.14) e (3.15) (BISHOP, 2006).

Figura 3.2: Representação de um neurônio de uma rede neural artificial.



Fonte: Adaptada de Zanetti et al. (2008).

$$a(x) = \varphi(W^T X) \tag{3.13}$$

$$z = W^T X \tag{3.14}$$

$$\varphi(z) = \frac{1}{1 + e^{-z}} \tag{3.15}$$

Dessa forma, a arquitetura da rede neural artificial perceptron multicamadas é composta por uma camada de entrada $X = \{x_1, x_2, ..., x_m\}$ contendo m vetores de características de tamanho n, k camadas escondidas ou unidades de ativação $a^{(k)}(x)$ e uma camada final de saída.

Essa arquitetura necessita que a matriz de pesos da camada intermediária tenha dimensão $k \times n$, em que k é a quantidade de nodos de ativação e n é o número de pesos da camada anterior mais um referente ao limiar (b). Dessa forma, poderá ser realizada a multiplicação entre as matrizes $z = W^T X$ e, em seguida, submetida à função de ativação $\varphi(z) = \dfrac{1}{1+e^{-z}}$, concluindo a camada de ativação $a(x) = \varphi(W^T X)$. Seu resultado irá alimentar a camada subsequente ou retornará o resultado final da rede neural artificial.

Função de Custo e Regularização

Conforme Bishop (2006), a função de custo para redes neurais $J(\Theta)$ é uma generalização da função de custo $J(\theta)$ com regularização usada na regressão logística (vide Equação (3.11)). Dado L como sendo o total de camadas na rede neural, s_l o número de unidades na primeira camada e K o número de classes na camada de saída, tem-se a seguinte Função de Custo $J(\Theta)$, conforme a Equação (3.16).

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\log((h_\Theta(x^{(i)}))_k) + (1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right] + \\ + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{j,i}^{(l)})^2 \tag{3.16}$$

Algoritmo de Retropropagação (*Backpropagation*)

O algoritmo de retropropagação (*backpropagation*) tem como objetivo minimizar a função de custo, de modo análogo ao gradiente descendente da regressão logística. Essa minimização é feita por meio da sua derivação parcial $\frac{\partial}{\partial \Theta_{i,j}^{(l)}} J(\Theta)$, que deverá ser armazenada em uma matriz de três dimensões $\Delta_{i,j}^{(l)}$ (BISHOP, 2006).

Em seguida, para cada um dos dados do conjunto de treinamento $\{(x^{(i)}, y^{(i)}), ..., (x^{(m)}, y^{(m)})\}$ deve ser atribuído $a(1) := x^{(t)}$ e em seguida realizada a propagação das ativações nas demais camadas $a(l)$, tal que $l = \{2, 3, ..., L\}$. A derivada

de cada camada de ativação deve ser calculada, iniciando pela camada final por meio da Equação (3.17).

$$\delta^{(L)} = a^{(L)} - y^{(t)} \tag{3.17}$$

Para as demais camadas, faz-se o uso da regra da cadeia para facilitar o cálculo da derivada, realizando-se a multiplicação elementar (produto de Hadamard) entre os vetores envolvidos conforme as Equações (3.18), (3.19) e (3.20).

$$\delta^{(l)} = ((\Theta^{(l)})^T \delta^{(l+1)}) \circ g'(z^{(l)}) \tag{3.18}$$

$$g'(z^{(l)}) = a^{(l)} \circ (1 - a^{(l)}) \tag{3.19}$$

$$\delta^{(l)} = ((\Theta^{(l)})^T \delta^{(l+1)}) \circ a^{(l)} \circ (1 - a^{(l)}) \tag{3.20}$$

De posse de todas as derivadas das camadas de ativação, os valores da matriz $\Delta_{i,j}^{(l)}$ devem ser atualizados por meio da adição do produto entre os valores da camada de ativação atual e a derivada da camada de ativação subsequente, conforme a Equação (3.21).

$$\Delta_{i',j'}^{(l)} = \Delta_{i,j}^{(l)} + a_j^{(l)} \delta^{(l+1)} \tag{3.21}$$

Por fim, após iterar sob todos os dados, é calculada a matriz $D_{i,j}^{(l)}$ que é equivalente ao $\frac{\partial}{\partial \Theta_{i,j}^{(l)}} J(\Theta)$, conforme a Equação (3.22). Salienta-se para o não uso de regularização para os termos de limiar (j=0).

$$\begin{aligned} D_{i',j'}^{(l)} &= \frac{1}{m}\left(\Delta_{i,j}^{(l)} + \lambda \Theta_{i,j}^{(l)}\right), \text{para j } \neq 0 \\ D_{i',j'}^{(l)} &= \frac{1}{m}\left(\Delta_{i,j}^{(l)}\right), \text{para j = 0} \end{aligned} \tag{3.22}$$

3.2.3 Árvore de Decisão

O classificador baseado em árvore de decisão é um modelo que possui uma regra de classificação em cada ligação dos nodos não folha que tem como objetivo agrupar as amostras com o mesmo rótulo por meio de divisões recursivas. Ao atingir o nodo folha, a classificação terá sido dada por completo (GÉRON, 2019).

Representação do Modelo

Uma árvore de decisão é construída por meio de indução, ou seja, realizando divisões dos dados em subconjuntos. Esses subconjuntos são escolhidos de acordo com o maior ganho de informação obtido nessa divisão, baseado na impureza dos dados (SANJEEVI, 2017).

Essa impureza, que é a quantidade de incerteza de um determinado conjunto de dados, é mensurada por meio da entropia, que tem seu cálculo baseado nas proporções $p(c)$ de número de elementos de cada classe $C = \{c_1, c_2, ..., c_m\}$ em relação ao total de dados, conforme a Equação (3.23).

$$H(C) = \sum_{c \in C} -p(c) log_2(p(c)) \tag{3.23}$$

O ganho de informação $G(A, C)$ tem a função de calcular a diferença entre as entropias nos momentos anterior e posterior ao particionamento do conjunto C pelo atributo A, conforme a Equação (3.24).

$$G(A, C) = H(C) - \sum_{t \in T} p(t) H(t) \tag{3.24}$$

em que T é o subconjunto criado pelo particionamento do conjunto C pelo atributo A, $p(t)$ é a proporção do número de elementos em t em relação ao número de elementos em C e $H(t)$ é a entropia do subconjunto t.

Dessa forma, o modelo é otimizado pela escolha das regras que maximizem o ganho de informação de maneira recursiva em todos os subconjuntos de dados.

Regularização

Assim como os demais modelos, a árvores de decisão também estão sujeitas ao *overfitting*, se ajustando de maneira incisiva aos dados na etapa de treinamento e não sendo generalistas a ponto de não predizer bem dados desconhecidos. Dessa forma, existem basicamente duas maneiras de simplificar o modelo: (i) a parada antecipada e (ii) a poda (BISHOP, 2006).

A parada antecipada tem como objetivo finalizar a árvore de decisão, antes que esta se torne muito complexa e acabe se ajustando demasiadamente aos dados de treinamento. Essa parada se dá por meio de algumas condições. Os valores dessas condições sempre são

determinados pelo desempenho das métricas de avaliação nos dados de validação.

A profundidade é uma dessas condições. Por meio do monitoramento das métricas de avaliação nos dados de validação, é possível determinar a profundidade que apresenta melhor desempenho nesses dados distintos aos de treinamento. De maneira análoga, pode-se utilizar como condição a quantidade mínima de dados em um nodo para que o mesmo seja particionado.

Também se pode utilizar como condições para parada antecipada a percepção de não melhoria dos erros de classificação de cada nodo. A ausência de uma evolução determina a não continuidade da árvore naquela ramificação, simplificando-a. Contudo essa condição pode apresentar falhas por ser uma abordagem de cima para baixo (*top-down*). Dessa forma, recomenda-se o uso da técnica de poda, por utilizar uma abordagem de baixo para cima (*bottom-up*)

A técnica da poda utiliza uma função de custo J baseada na soma de uma medida de ajuste e de complexidade, ou seja, quão bem ajustado o modelo está somado a sua complexidade. Esse ajuste é dado pelo erro de classificação $L(T)$ e a complexidade se dá pelo número $F(T)$ de folhas da árvore, além do fator de regularização λ, conforme a Equação (3.25).

$$J(T) = E(T) + \lambda F(T) \tag{3.25}$$

Sendo assim, para cada nodo de decisão é aplicada a poda e verificada se houve uma minimização dos custos. Em caso afirmativo, parte-se para o próximo nodo, até que não haja minimização do custo total da árvore.

3.2.4 Floresta Aleatória

Para realizar a classificação, a floresta aleatória utiliza simultaneamente um conjunto de árvores de decisão, por meio da técnica de *bootstrap aggregation*. Geralmente, essa técnica apresenta melhores resultados, quando comparada à arvore de decisão. Entretanto não é tão rápida quanto uma única árvore de decisão, tanto na etapa de treinamento, quanto na etapa de testes (GRUS, 2019).

Dessa forma, a partir do conjunto $F(X) = \{f_1(X), f_2(X), ..., f_N(X)\}$ de tamanho N de árvores de decisão construídas com dados aleatórios da base de dados de treinamento, o

algoritmo realiza uma votação entre os classificadores, sendo a moda o resultado da predição $\hat{y}$, conforme a Equação (3.26).

$$\hat{y} = moda(F(X)) \tag{3.26}$$

3.3 Redes Neurais Convolucionais

Com o aumento paulatino da disponibilidade de mídia, assim como da capacidade de processamento dos computadores, o aprendizado profundo vem ganhando espaço na resolução de problemas resolvidos anteriormente por meio de outras técnicas. Na área da Visão Computacional, as redes neurais convolucionais (RNC) vêm se destacando ao longo dos anos, principalmente nas subáreas de reconhecimento de imagem e de objetos em imagem (GOODFELLOW et al., 2016).

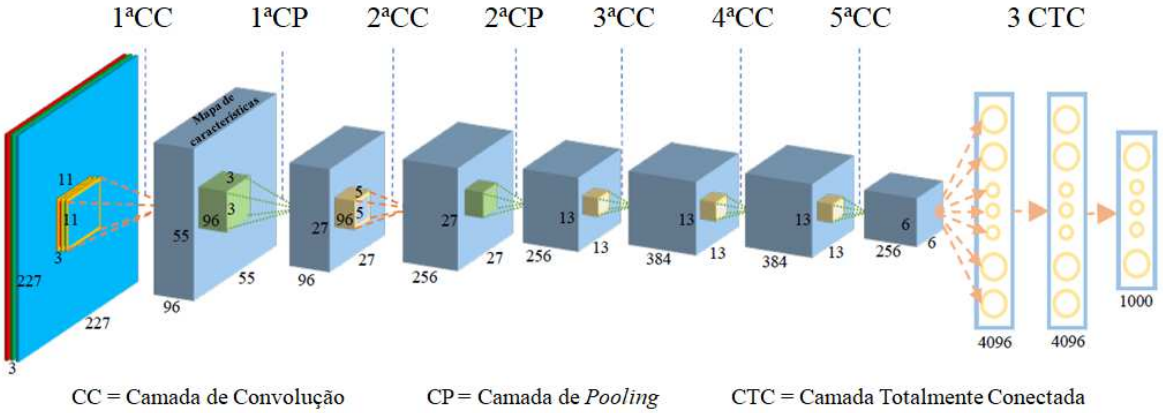
As redes neurais convolucionais (RNC) apresentam uma estrutura em camadas inspirada na organização hierárquica do córtex visual humano, em que cada uma dessas camadas possui neurônios capazes de extrair e aprender características dos dados de entrada da rede neural. Essa hierarquia faz com que cada camada represente características em diferentes patamares. Os neurônios contidos nas camadas iniciais são responsáveis por representar características de mais baixo nível, enquanto que os neurônios localizados nas camadas finais representam características de mais alto nível. Essa estratégia faz com que se evite o levantamento de características pelo projetista, deixando a cargo da própria rede neural essa função. Em contrapartida esse tipo de rede neural geralmente requer uma grande quantidade de dados em sua etapa de treinamento, por possuir uma grande quantidade de parâmetros a serem aprendidos (QIN et al., 2018).

3.3.1 Estrutura de Uma Rede Neural Convolucional

Uma rede neural convolucional possui diversos componentes que, utilizados em conjunto, são capazes de realizar o treinamento do modelo e, posteriormente, a predição dada a existência de um modelo já treinado. Essa estrutura pode ser visualizada na Figura 3.3, que ilustra a rede neural convolucional *CaffeNet* (JIA et al., 2014), uma replicação da *AlexNet* (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) contendo cinco camadas de convo-

lução (CC), duas camadas de *pooling* (CP) seguidas de três camadas totalmente conectadas (CTC).

Figura 3.3: Estrutura da Rede Neural Convolucional *CaffeNet*.



Fonte: Adaptada de Qin et al. (2018).

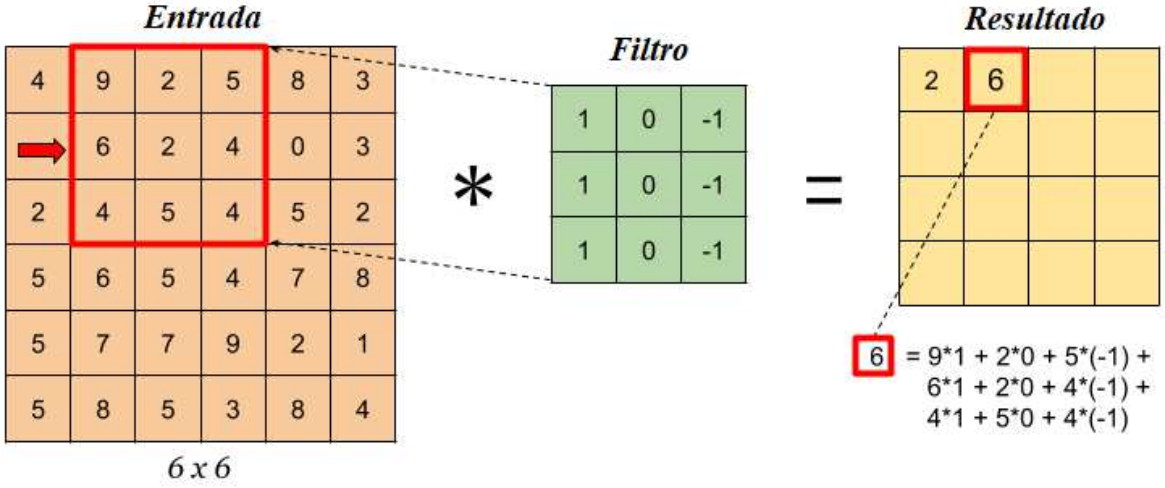
Convolução

A primeira e mais importante estrutura é a convolução. Esse componente é semelhante ao de um detector de contornos, que possui um filtro específico para cada tipo de contorno a ser detectado. Esse filtro é formado por uma matriz de valores fixos que percorre toda a imagem aplicando a convolução em cada etapa dessa varredura, como mostra a Figura 3.4. Por fim, uma imagem apresentando os contornos é gerada, sendo representada por uma matriz contendo a convolução resultante de cada etapa da referida varredura. A convolução aplicada pelo filtro corresponde ao somatório da multiplicação de cada um de seus elementos pelo seu correspondente na imagem (GOODFELLOW et al., 2016).

A convolução apresenta a mesma dinâmica do detector de contornos, contudo, é capaz de representar diversas características visuais além de simples contornos. Isso se dá pelo uso de valores variáveis em seus filtros, esses sendo atualizados até a última iteração da fase de treinamento do modelo por meio do *backpropagation*. Dessa forma, o modelo buscará por características visuais que individualizem cada categoria de imagem.

O filtro convolucional é um tensor que, obrigatoriamente, deve apresentar sua terceira dimensão idêntica a da imagem (ou mapa de características, no caso de não ser a convolução inicial) na qual será aplicada a convolução. Geralmente, as duas primeiras dimensões dos

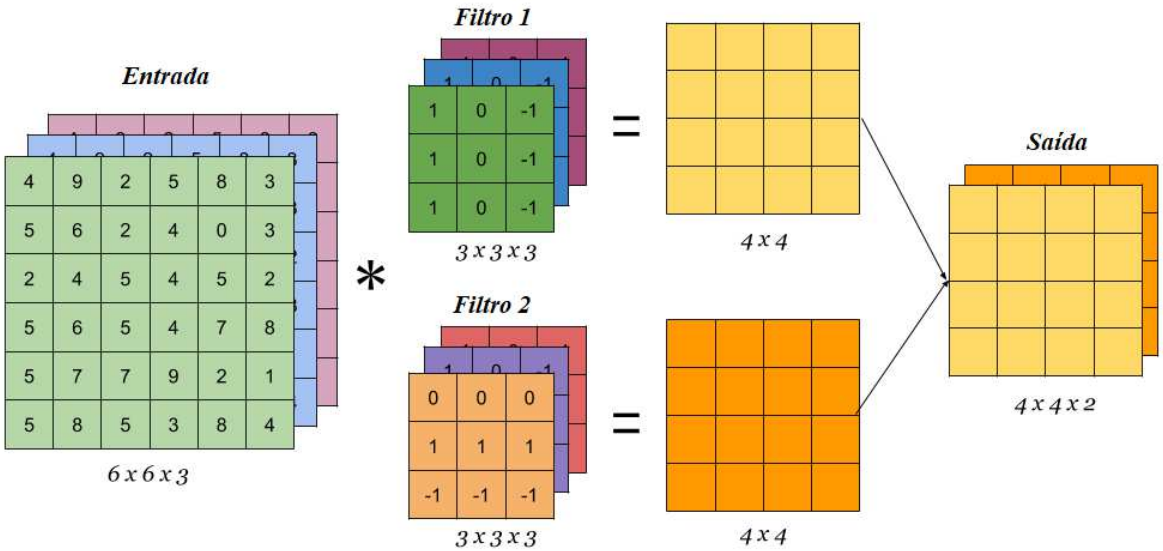
Figura 3.4: Filtro Prewitt Vertical sendo aplicado sobre imagem para detecção de contornos.



Fonte: Adaptada de Prijono (2018).

filtros apresentam uma forma quadrada, sendo geralmente mais adotados nos formatos 3x3, 5x5, 7x7 e 9x9. A quantidade de filtros utilizados é o que ditará o tamanho da terceira dimensão do tensor resultante dessa etapa, conforme representado na Figura 3.5.

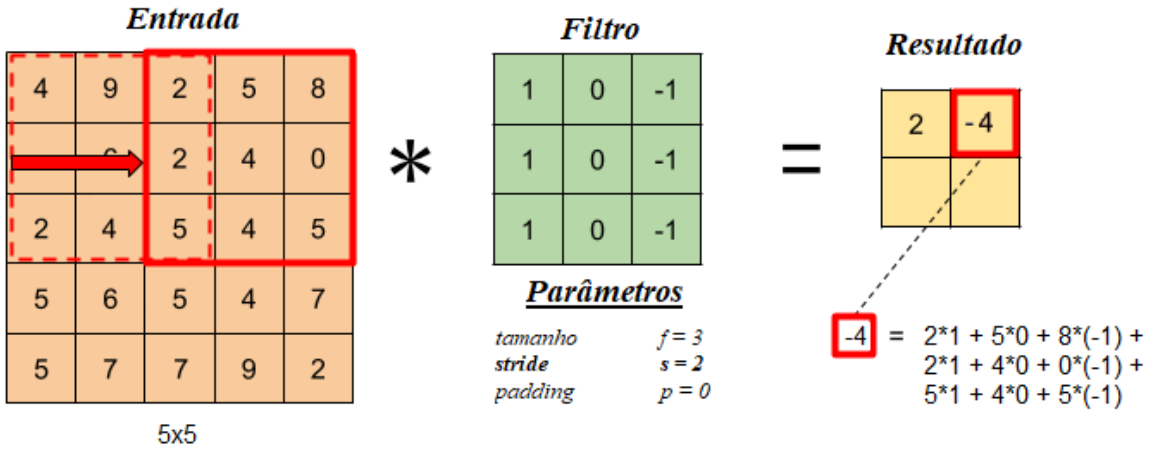
Figura 3.5: Convolução em imagem de tamanho 6x6x3 utilizando dois filtros de tamanho 3x3x3, resultando em uma matriz de tamanho 4x4x2.



Fonte: Adaptada de Prijono (2018).

A maneira como a varredura do filtro é aplicada na imagem também pode ser escolhida, decidindo a quantidade de movimentação do filtro horizontal e verticalmente. Esse parâmetro é chamado de *stride* e reflete diretamente no tamanho da matriz de saída, conforme mostrado na Figura 3.6.

Figura 3.6: Convolução em imagem de tamanho 5x5 utilizando um filtro de tamanho 3x3, aplicando stride = 2 e padding = 1, resultando em uma matriz de tamanho 2x2.



Fonte: Adaptada de Prijono (2018).

O modo como a varredura vem sendo aplicada faz com que os píxeis das bordas tenham menos ênfase comparados aos píxeis localizados no interior da imagem. Visando-se evitar esse problema, faz-se o uso de uma borda de valores iguais a zero e de tamanho variável, conforme ilustrado na Figura 3.7. Além do mais, o uso desse artifício denominado *padding* permite, caso seja desejado, que a matriz resultante mantenha seu tamanho original, denominado como *same padding*. A terminologia *valid padding* implica que nenhuma borda de *padding* foi utilizada.

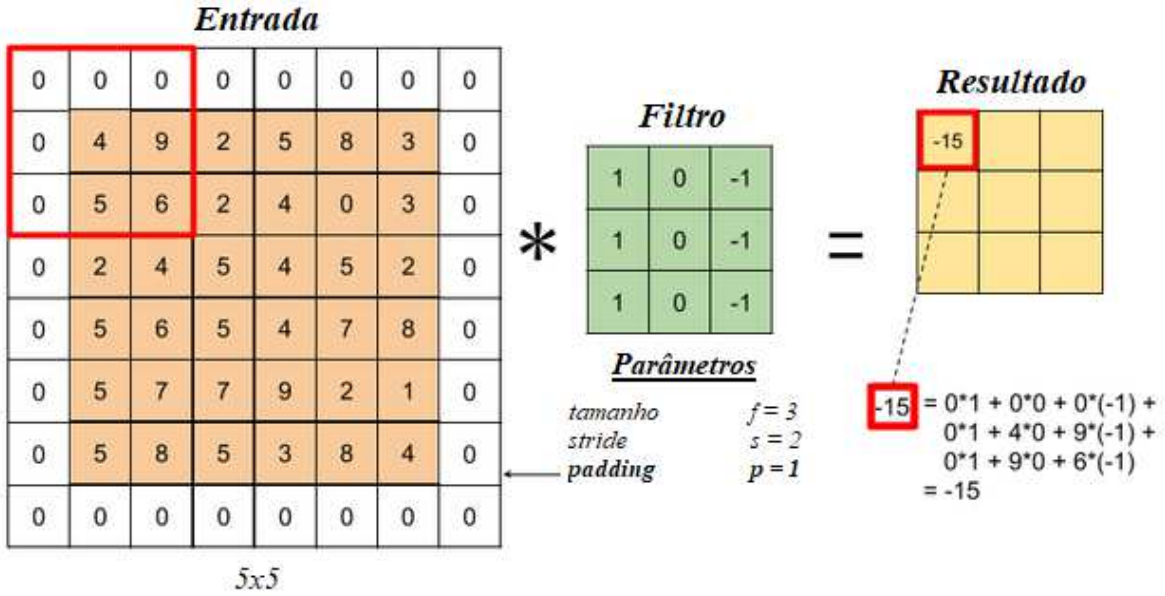
Por fim, sendo l a camada atual, por meio da Equação (3.27) é possível calcular o tamanho da matriz resultante $n^{[l]}$, com base no tamanho da matriz de entrada $n^{[l-1]}$, no tamanho do filtro $f^{[l]}$, no *stride* $s^{[l]}$ e no *padding* $p^{[l-1]}$ adotados.

$$n^{[l]} = \lfloor \frac{n^{[l-1]} + 2p^{[l-1]} - f^{[l]}}{s^{[l]}} + 1 \rfloor \tag{3.27}$$

Função de Ativação

A função de ativação tem um papel fundamental não apenas nas redes neurais convolucionais, mas em todos os tipos de redes neurais. Seu propósito é adicionar não linearidade ao modelo, visto que na etapa de convolução são aplicadas apenas operações lineares (i.e. multiplicação entre elementos e somatório). A ausência dessa função limitaria o modelo para lidar apenas com discriminação de dados lineares, o que não condiz com a complexidade de

Figura 3.7: Convolução em imagem de tamanho 5x5 utilizando um filtro de tamanho 3x3 e aplicando stride = 2, resultando em uma matriz de tamanho 3x3.



Fonte: Adaptada de Prijono (2018)

classificar imagens em diversas categorias.

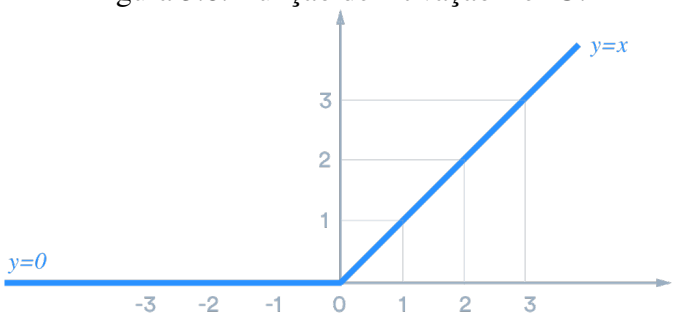
Inicialmente as funções Sigmóide e Tanh eram amplamente utilizadas para esse fim. Contudo, atualmente, a função ReLU (NAIR; HINTON, 2010) vem sendo comumente aplicada por possuir diversas vantagens em detrimento das demais, mesmo tratando-se de uma função de implementação extremamente trivial. Se comporta como uma função identidade para os valores positivos, mantendo-os, e altera todos os valores negativos para 0, conforme mostram a Equação (3.28) e Figura 3.8.

$$f(x) = max(0, x) \tag{3.28}$$

Essa simplicidade faz com que apresente melhor desempenho comparada às funções de ativação Sigmóide e Tanh, pois o cálculo da sua derivada é muito mais simples que as demais (resultando apenas 0 ou 1), além de também eliminar o esvaecimento dos gradientes (efeito "*vanishing*"), que acontece quando os valores são muito pequenos ou muito grandes nas funções Sigmoide e Tanh, implicando valores muito pequenos de suas derivadas.

Outro ponto positivo da função de ativação ReLU é a criação de uma rede neural esparsa, pois nem todos os neurônios são ativados. Quando o somatório dos pesos multiplicados pelo retorno da camada de ativação anterior (ou entrada, no caso da primeira camada) é negativo,

Figura 3.8: Função de Ativação ReLU.



Fonte: Extraída de Peixeiro (2019).

a função retornará zero e, consequentemente, seu gradiente também será. Dessa forma a rede se tornará menos densa, resultando em uma rede menos custosa e mais rápida.

Contudo, o excesso de não ativações dos neurônios pode prejudicar a rede. Esse problema é conhecido como "*Dying ReLu Problem*". Visando mitigar esse problema são utilizadas variações da função de ativação ReLu (e.g. LeakyReLu, ELU), que não retornam zero para os valores negativos, mas valores negativos próximos a zero.

Camada de Convolução

A união do componente de convolução seguido da função de ativação F formam a camada de convolução, como pode ser visualizado na Figura 3.9. O resultado de uma camada de convolução l é denominado mapa de características. Esses mapas atuam como entrada para a próxima camada de convolução (QIN et al., 2018). A Equação (3.29) descreve matematicamente como se comporta um determinado filtro i dessa camada:

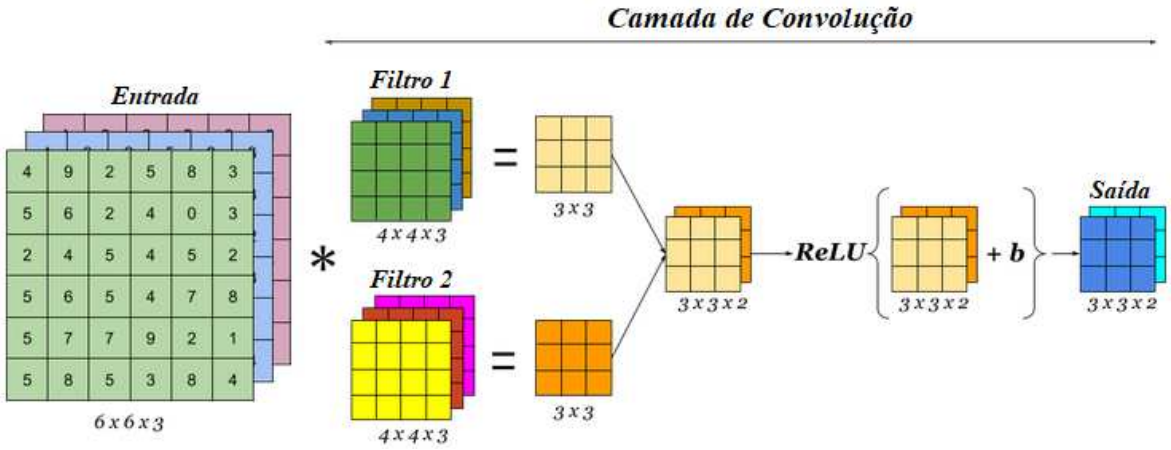
$$a_{i,l+1} = F\left(\sum w_{i,l} a_{i,l} + b_{i,l}\right), \tag{3.29}$$

em que w e b representam, de maneira respectiva, os pesos e o limiar do filtro em questão.

Camada de *Pooling*

Geralmente, após a camada de convolução, é aplicada uma camada de *pooling*. Seu uso se dá basicamente por dois motivos: (i) minimizar o custo computacional, visto que seu uso faz com que haja uma diminuição do tamanho do tensor (altura e largura, mantendo a profundidade inalterada) e (ii) generalizar o modelo, pois se trata de um regularizador,

Figura 3.9: Representação de uma camada de convolução de uma rede neural convolucional.



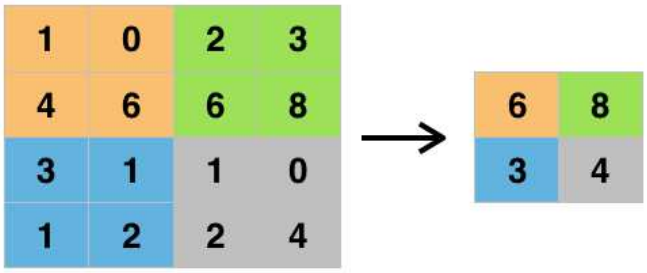
Fonte: Adaptada de Prijono (2018).

devido à redução do número de parâmetros (GOODFELLOW et al., 2016).

A camada de *pooling* é aplicada independentemente em cada canal e seus hiperparâmetros referentes ao tamanho e ao *stride* são variáveis. Geralmente, utiliza-se o filtro e *stride* de tamanho 2, que retorna uma matriz de tamanho $\frac{n}{2} \times \frac{n}{2}$, ou seja, correspondendo a 25% do tamanho original.

Normalmente, utiliza-se o *pooling* retornando valor máximo encontrado (*maxpooling*), conforme ilustrado na Figura 3.10, mas também pode ser utilizada a média (*average pooling*) ou a norma L2 dos valores (*L2-norm pooling*). Salienta-se também que, após escolhidos, seus hiperparâmetros são fixos, não sendo modificados pelo algoritmo de retropropagação (*backpropagation*).

Figura 3.10: *Maxpooling* utilizando filtro e *stride* de tamanho 2.



Fonte: Adaptada de Prijono (2018).

No caso do classificador de imagens, assim como em outros problemas relacionados à Visão Computacional, o *pooling* é utilizado sem nenhum prejuízo, pois o objetivo principal da rede neural convolucional é detectar a presença de uma determinada característica (alto

valor de ativação) na imagem e não sua localização exata na imagem.

Camadas Totalmente Conectadas

Esse componente geralmente é utilizado para categorizar as características extraídas pelas camadas de convolução. A saída da última camada convolucional é linearizada e utilizada como entrada para as camadas totalmente conectadas, cujo resultado é submetido à função *SoftMax*, definida matematicamente na Equação (3.30). Essa função tem como objetivo obter um vetor de probabilidades $P = \{P_1, P_2, ..., P_n\}$ de N dimensões, em que cada dimensão está relacionada a uma classe.

$$P_i = \frac{e^{a_i}}{\sum_{n=1}^{N} e^{a_n}}, \tag{3.30}$$

em que a_i é o i-ésimo neurônio da última camada das camadas totalmente conectadas (QIN et al., 2018).

3.3.2 Algoritmo de Retropropagação (*Backpropagation*)

Assim como nas redes neurais *perceptron* multicamadas, as redes neurais convolucionais também utilizam o algoritmo de retropropagação (*Backpropagation*) para atualizar seus parâmetros w e b (i.e. pesos e limiar, respectivamente) com o objetivo de minimizar sua função de custo J, vide Equação (3.31). Isso se dá por meio da diminuição da função de perda P, ou seja, reduzindo o erro entre os valores reais e os valores preditos das amostras de treinamento $X = \{x_1, x_2, ..., x_n\}$ e seus respectivos rótulos $Y = \{y_1, y_2, ..., y_n\}$, vide Equação (3.32).

$$J(w, b) = \frac{1}{n} \sum_{i=1}^{n} P(w, b, x_i, y_i) \tag{3.31}$$

$$P(w, b, x_i, y_i) = (y_i - f(w, b, x_i))^2, \tag{3.32}$$

em que $F(.)$ representa os valores preditos pela rede neural convolucional, detalhada na Equação (3.33)

$$f(w,b,x) = F(\sum w_{i,l} F(\sum w_{i,l-1} F(...F(w_{i,1}x_{i,0} + b_{i,0})...) + b_{i,l-1}) + b_{i,l}). \quad (3.33)$$

Por fim, para minimizar a função de custo $J(w,b)$, é calculada a derivada parcial com respeito a cada peso w e o limiar b, conforme a Equação (3.34), aplicando a retropropagação em todas as camadas da Rede Neural Convolucional.

$$\frac{\partial J}{\partial (w,b)} \quad (3.34)$$

Para que os parâmetros da rede sejam atualizados, se faz necessário o uso de um otimizador (e.g. gradiente descendente). O procedimento de atualização utilizando esse método acontece de forma iterativa, em que o peso da iteração $j+1$ é dado pelo valor do peso da iteração j menos o produto entre a taxa de aprendizado e derivada parcial na Equação (3.34), como pode ser observado na Equação (3.35) (QIN et al., 2018).

$$w_{j+1} = w_j - \eta \cdot \frac{\partial J(w; x, y)}{\partial w} \quad (3.35)$$

3.3.3 Transferência de Aprendizado (*Transfer Learning*)

A dependência de dados é um dos principais problemas ao lidar com modelos baseados em aprendizado profundo. Para que esse tipo de modelo possa aprender os padrões dos dados, em que não existe um levantamento manual das características, se faz necessário uma grande quantidade de dados na etapa de treinamento. Todavia é extremamente difícil construir um conjunto de dados: (i) rotulado, (ii) em larga escala e (iii) com alta qualidade (TAN et al., 2018).

A transferência de aprendizado atenua o problema da ausência de uma grande quantidade de dados no aprendizado profundo. Especificamente na área de classificação de imagens, essa técnica utiliza um modelo previamente treinado com imagens inseridas em um domínio distinto do domínio alvo. Essa transferência de conhecimento faz com que os neurônios das camadas iniciais da rede neural aprendam características de baixo nível que, mesmo oriundas de outros tipos de imagens, são aproveitadas para a classificação das imagens do domínio

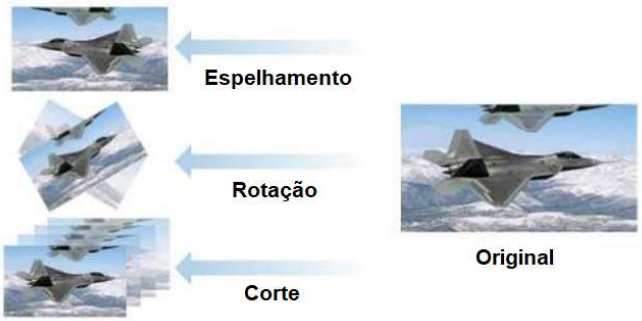
alvo. Ademais, esse procedimento diminui o tempo de treinamento, pois os parâmetros já estão previamente treinados, além de também, em boa parte dos casos, não requerer o retreinamento de toda a rede neural (GOODFELLOW et al., 2016).

3.3.4 Aumento de Dados (*Data Augmentation*)

Assim como a técnica de transferência de aprendizado, o aumento de dados também tem como objetivo principal minimizar o problema da ausência de uma grande quantidade de dados para treinamento dos modelos baseados em aprendizado profundo. Esse aumento de dados é dado pela geração de imagens artificiais a partir das imagens já contidas na base de dados (SHORTEN; KHOSHGOFTAAR, 2019).

Essas imagens artificiais podem ser criadas por meio de técnicas básicas, como remoção de parte da imagem, mudanças em seus espaços de cores ou até mesmo mistura de imagens. Transformações geométricas também são amplamente utilizadas, conforme ilustrado na Figura 3.11. As imagens são rotacionadas nos mais diversos ângulos, espelhadas verticalmente e/ou horizontalmente, cortadas de maneira aleatória ou distorcidas (SHORTEN; KHOSHGOFTAAR, 2019).

Figura 3.11: Criação de diversas imagens a partir de uma única por meio de espelhamento, rotações e cortes aleatórios.



Fonte: Adaptada de Taylor e Nitschke (2017).

Técnicas mais elaboradas, baseadas em aprendizado profundo, também são capazes de intensificar a quantidade de dados do conjunto de treinamento. Redes Adversárias Generativas são capazes de criar imagens artificiais baseadas em um conjunto de dados, retendo características semelhantes ao conjunto original. Transferências de estilo também são utilizadas, em que o estilo de uma imagem é mesclado com o conteúdo de uma segunda ima-

gem (SHORTEN; KHOSHGOFTAAR, 2019).

3.4 Detecção de Pele Humana

A detecção de pele pode ser considerada um problema de classificação binário, em que seu classificador infere se um determinado píxel é considerado pele ou não. Geralmente utilizam-se os espaços de cores como características básicas para essa detecção. Esses espaços de cores são formados por elementos de crominância e luminância, que se referem ao valor das cores e de luz, respectivamente. Esses espaços de cores podem ser divididos em quatro tipos (XIONG; LI, 2012):

1. Básicos (e.g. RGB, sRGB, CIE-XYZ);
2. Ortogonais (e.g. YCbCr, YDbDr, YPbPr, YUV, YIQ);
3. Perceptuais (e.g. HSV, HSL, HSI) e;
4. Perceptualmente Uniformes (e.g. CIE-Lab, CIE-Luv, CIE-LCH, CIE-CAT02 LMS).

Sendo assim, escolhe-se um ou mais espaços de cores e se extrai um ou mais de seus componentes, utilizando-os em sua forma bruta ou modificados, por meio de operações aritméticas com outros componentes de cores. Também podem ser utilizadas características da textura da pele, pois objetos ou planos de fundo contidos nas imagens podem possuir a mesma tonalidade da pele humana (MAHMOODI; SAYEDI, 2016).

3.4.1 Classificador Fundamentado em Regras

O classificador fundamentado em regras, dentre todos os classificadores, é o menos complexo, pois demanda menos processamento de máquina por possuir cálculos mais simples em relação aos demais, além de não ser necessária a fase de treinamento (PLATZER; STUETZ; LINDORFER, 2014; MUHAMMAD; ABU-BAKAR, 2015).

Funciona por meio do uso de expressões lógicas, envolvendo um ou mais componentes de cores, que determinam se um píxel é considerado de pele ou não. Essa técnica é ideal para ambientes embarcados e aplicações em tempo real, devido a seu tempo de resposta

reduzido. Contudo geralmente apresenta piores resultados quando comparado aos demais classificadores (MUSTAFAH; AZMAN, 2012)

3.4.2 Classificador Baseado em Histograma de Frequência

Trata-se de um modelo baseado no uso de histograma de frequência de píxeis, sem o uso de nenhuma função de densidade de probabilidade. Realiza-se um mapeamento quantificando os píxeis de pele em toda a distribuição de cores do espaço de cores utilizado. Dessa forma, é possível atribuir um valor de probabilidade de que cada píxel no espaço de cor seja de pele (MAHMOODI; SAYEDI, 2016).

Essas probabilidades são obtidas por meio da construção de um histograma, denominado de *Look-up Table* (LUT), de dimensão igual à quantidade de canais de cores do espaço de cor (geralmente 3) e cada dimensão possui o tamanho da faixa de píxeis abrangida pelo respectivo canal. Por exemplo, uma LUT do espaço de cor RGB terá três dimensões, possuindo o tamanho $i \times i \times i$, em que i é igual a 256. Entretanto, pode-se reduzir o tamanho da tabela (e.g. $128 \times 128 \times 128$, $64 \times 64 \times 64$, $32 \times 32 \times 32$) e computar as incidências dos píxeis por subfaixas em vez dos próprios.

O processo de aprendizado é simples, contudo, requer uma grande base de dados para obtenção de bons resultados. A probabilidade final de cada píxel ser de pele é dada pela Equação (3.36).

$$P(R_iG_iB_i|pele) = \frac{N(R_iG_iB_i)}{T}, \tag{3.36}$$

em que $N(R_iG_iB_i)$ é o número de ocorrências de um determinado píxel e T é total de píxeis de pele computados.

Também se faz uso de uma técnica com dois histogramas, como pode ser visto na pesquisa de Jones e Rehg (2002), esses seguindo o mesmo formato do LUT. Contudo o segundo histograma, de maneira análoga ao primeiro, tem a função de contabilizar os píxeis de não pele. Como é sabida a existência da sobreposição de píxeis de pele e não pele, utiliza-se um classificador bayesiano que considera as probabilidade de um mesmo píxel ser de pele e não pele, conforme a Equação (3.37).

$$P(pele|R_iG_iB_i) = \frac{P(R_iG_iB_i|pele)P(pele)}{P(R_iG_iB_i|pele)P(pele) + P(R_iG_iB_i|não\ pele)P(não\ pele)} \quad (3.37)$$

Dessa forma, por meio dos histogramas, é possível calcular as probabilidades de um determinado píxel ocorrer dado que ele seja pele ($P(R_iG_iB_i|pele)$) ou não pele ($P(R_iG_iB_i|não\ pele)$). As probabilidades de qualquer píxel ser de pele ($P(pele)$) ou não pele ($P(não\ pele)$) podem ser levantadas por meio de um conjunto de imagens rotuladas ou pela simples inferência que possuem a mesma probabilidade.

De acordo com Poudel et al. (2013), é possível simplificar a razão das probabilidades de um píxel ser de pele ($P(R_iG_iB_i|pele)$) ou não pele ($P(R_iG_iB_i|não\ pele)$), como pode ser observado nas Equações (3.38), (3.39) e (3.40). Essa simplificação tem como objetivo utilizar um limiar que determina a razão para definir se um píxel é de pele ou não. Esse limiar θ é composto por uma variável K que multiplica a razão das probabilidades de um píxel ser de pele $P(pele)$ e não pele $P(não\ pele)$, como pode ser observado na Equação (3.41).

$$P(pele|R_iG_iB_i) = \frac{P(R_iG_iB_i|pele)P(pele)}{P(R_iG_iB_i)} \quad (3.38)$$

$$P(não\ pele|R_iG_iB_i) = \frac{P(R_iG_iB_i|não\ pele)P(não\ pele)}{P(R_iG_iB_i)} \quad (3.39)$$

$$\frac{P(pele|R_iG_iB_i)}{P(não\ pele|R_iG_iB_i)} = \frac{P(R_iG_iB_i|pele)P(pele)}{P(R_iG_iB_i|não\ pele)P(não\ pele)} \quad (3.40)$$

$$\frac{P(R_iG_iB_i|pele)}{P(R_iG_iB_i|não\ pele)} > \theta, \theta = K\left(\frac{P(não\ pele)}{P(pele)}\right) \quad (3.41)$$

3.4.3 Classificador Baseado em Aprendizado de Máquina

Classificadores baseados em aprendizado de máquina também são capazes de diferenciar entre píxeis de pele e não pele. Esses classificadores apresentam maior complexidade quando comparados aos baseados em regras e em histogramas, pois necessitam de uma fase de treinamento mais densa para aprendizado dos padrões, além de demandar maior processamento

de máquina por possuírem cálculos mais complexos, por isso apresentam maior tempo nas fases de treinamento e teste. Entretanto atingem melhores resultados quando comparados aos demais classificadores (BRANCATI et al., 2017).

Diversas são as máquinas de aprendizado existentes que podem ser aplicadas no reconhecimento de pele humana. As paramétricas são baseadas em modelos gaussianos (e.g. *Single Gaussian Models*, *Gaussian Mixture Models*, *Cluster of Gaussian Models*) e elípticos (e.g. *Elliptic Boundary*). Esses modelos simulam uma distribuição real, como as dos classificadores baseados em histograma, por meio do uso de funções de densidade de probabilidade, sendo possível assim utilizar menos dados em sua fase de treinamento sem comprometer o resultado final, como adotado nos trabalhos de Dong et al. (2012), Mustafa, Elbashir e Babikir (2015) e Du et al. (2012).

Também são bastante utilizadas máquinas baseadas nos conceitos de Árvores (e.g. Árvores de Decisão, Árvores de Regressão, Florestas Aleatórias), Redes Neurais (e.g. Perceptron Multicamadas), máquina de vetores de suporte (*support vector machine - SVM*) entre outras (KHAN et al., 2012; MA et al., 2014; MUSTAFA; ELBASHIR; BABIKIR, 2015).

Entretanto, o reconhecimento de pele humana não é uma tarefa trivial. É possível enumerar alguns fatores que alteram a padronização das cores relacionadas à pele humana, tais como: iluminação, plano de fundo, equipamento de captura e características pessoais (MAHMOODI; SAYEDI, 2016).

O fator mais depreciador do desempenho da detecção de pele está atrelado à constância de cor, estando estritamente relacionada à iluminação desarmônica do ambiente. As técnicas de correção de cor e cancelamento de iluminação propõem amenizar essa heterogeneização e, consequentemente, melhorar o desempenho da detecção. Os planos de fundo que possuem características de cores semelhantes à pele humana (e.g. paredes, madeiras e tijolos), assim como objetos no ambiente, também são ditos como aspectos desafiadores para essa tarefa, impactando consideravelmente o desempenho caso o método escolhido não transponha esse empecilho. A não padronização das características das câmeras (e.g. sensor de resposta, lentes, configurações) faz com que a distribuição das cores da pele humana varie de dispositivo para dispositivo, degradando o desempenho de detectores baseados em um espaço de cor específico. Dentre os vários subfatores existentes nas características pessoais relativas à detecção de pele, o mais desafiador sem dúvida é a diversidade étnica, que faz com que a

cor da pele varie entre as cores preta, amarela e branca. Gênero, idade e condições de saúde são fatores menos preocupantes na variação da coloração da pele humana, porém também devem ser considerados (MAHMOODI; SAYEDI, 2016).

3.5 Considerações Finais

Neste capítulo, foi apresentado um entendimento geral acerca dos principais tópicos pertencentes ao domínio da pesquisa em questão. Primordialmente, foram introduzidos os conceitos de Visão Computacional e de Classificação de Imagem. Na sequência foi apresentado o conceito de Aprendizado de Máquina. Foram discriminados quatro dos algoritmos mais utilizados (i.e. regressão logística, perceptron multicamadas, árvore de decisão e floresta aleatória), expondo a representação de cada modelo, assim como métodos de regularização.

Em seguida, ainda no âmbito do aprendizado de máquina, abordou-se sobre redes neurais convolucionais, enfatizando sua importância atual na área da Visão Computacional, principalmente no reconhecimento de imagens. Ademais, foram esclarecidas suas principais etapas, como convolução, função de ativação e camada de *pooling*. Também foram expostas duas importantes técnicas para minimização dos efeitos de uma base de dados de tamanho limitado: a i) Transferência de Aprendizado (*Transfer Learning*) e o ii) Aumento de Dados (*Data Augumentation*).

Por fim, foi apresentado o conceito de detecção de pele humana, técnica precursora para a detecção de imagens pornográficas. Foram enumeradas as técnicas fundamentadas em regras, histogramas de frequência e aprendizado de máquina.

Capítulo 4

Trabalhos Relacionados

Este capítulo contempla a revisão bibliográfica referente aos principais trabalhos relacionados que foram encontrados na literatura, assim como suas respectivas propostas. Os trabalhos encontram-se alocados em três grupos: trabalhos que investigam a detecção de conteúdo pornográfico adulto, a detecção de pornografia infanto-juvenil e a estimação de idade por meio de reconhecimento facial.

4.1 Reconhecimento de Conteúdo Pornográfico Adulto

Com o crescimento do compartilhamento de imagens e vídeos na Internet, o reconhecimento de imagem pornográfica vem ganhando importância nos últimos anos. Seu principal objetivo é que esse conteúdo não seja exposto a públicos inadequados, como crianças e adolescentes (LI et al., 2016).

Anteriormente a essa ampla disseminação, era possível lidar com esse problema por meio de técnicas elementares. A utilização de uma lista negra bloqueando sites com conteúdo pornográfico era uma delas. Também utilizava-se a identificação de sites e/ou arquivos por meio de palavras-chave que inferissem que esse tipo de conteúdo era presente. No entanto, o número de sites dessa natureza cresce sem precedentes todos os dias, tornando-se impraticável incluir manualmente todos os sites pornográficos em um lista negra, assim como identificar por meio textual sites e/ou arquivos que já prevêem essa técnica falha (NIAN et al., 2016).

Portanto, tornou-se necessária a utilização de técnicas que pudessem detectar automaticamente imagens desse tipo por meio do seu conteúdo. Sendo assim, iniciou-se uma busca

por técnicas oriundas da Visão Computacional para que essas imagens pudessem ser classificadas. A partir de então, esse tipo de reconhecimento vem sendo amplamente utilizado em muitas aplicações na Internet, tais como mecanismos de pesquisa de imagens, redes sociais de compartilhamento de fotos, provedores de serviços de hospedagem de arquivos e sites de transmissão de vídeo (LI et al., 2016).

Sendo assim, o reconhecimento de imagens possuindo conteúdo adulto vem sendo estudado por diversos pesquisadores. Ao longo dos anos, o método mais utilizado foi o baseado em detecção de pele, fazendo uso de técnicas adicionais como textura de pele, morfologia humana e características locais. Atualmente, as redes neurais profundas vêm mostrando melhores resultados na detecção desse tipo de imagem (LI et al., 2016; OU et al., 2017).

4.1.1 Baseado em Detecção de Pele Humana

O reconhecimento de conteúdo pornográfico baseado nessa técnica tem como primeira etapa um detector de píxeis de pele. Esse detector define quais os píxeis da imagem questionada são pele e não pele. Passada essa etapa, define-se quais características baseadas nos dados de pele são relevantes. A partir de então, essas características são utilizadas para o classificador definir se determinada imagem é pornográfica ou não.

O trabalho proposto por Ap-Apid (2005) foi um dos precursores no reconhecimento de imagens pornográficas. O autor utilizava regras estáticas tanto para determinar se um píxel era de pele, quanto para inferir se a imagem era pornográfica. Os píxeis eram considerados de pele se obedecessem às regras nos espaços de cores nRGB e HSV, de acordo com as expressões lógicas a seguir:

$$\begin{aligned}(R > 220 \ \& \ G > 210 \ \& \ B > 170 \ \& \ |R-G| > 15 \ \& \ R > B \ \& \ G > B) \ | \\ (R > 95 \ \& \ G > 40 \ \& \ B > 20 \ \& \ max(R,G,B) - min(R,G,B) > 15 \\ \& \ |R-G| > 15 \ \& \ R > G \ \& \ R > B)\end{aligned}$$

$$(0 \leq H \leq 50 \ | \ 340 \leq H \leq 360) \ \& \ 0,2 < S \ \& \ 0,35 < V$$

Depois de detectados os píxeis de pele, Ap-Apid (2005) segmentava todas as regiões

de pele e as ordenava de maneira decrescente por tamanho, além de criar um polígono que abarcasse as três maiores regões de pele. A classificação das imagens como pornográficas ou não era realizada por meio de um conjunto de regras estáticas que consideravam uma série de critérios (e.g. proporção de píxeis de pele, maiores regiões de pele, número de regiões de pele).

As pesquisas de Polastro e Eleutério (2010) e Medina e Palladino (2017) utilizaram abordagens similares à pesquisa de Ap-Apid (2005), no que diz respeito à classificação dos píxeis de pele e do tipo de imagem. O trabalho de Polastro e Eleutério (2010) diferenciou-se por combinar outras técnicas, tais como: (i) a análise dos nomes dos arquivos e (ii) o uso de uma lista negra contendo o código *hash* de arquivos previamente catalogados como ilegais. O estudo de Medina e Palladino (2017) se diferenciava principalmente em dois aspectos: (i) no formato do delimitador (retangular) para segmentar as três maiores regiões de pele contidas na imagem e (ii) nos espaços de cores adotados (RGB e YCbCr) para a detecção de pele.

No estudo de Platzer, Stuetz e Lindorfer (2014) foi utilizado como classificador das imagens uma máquina de vetores de suporte, um modelo baseado em aprendizado de máquina tradicional. Na etapa de extração de características, foi utilizada uma grande variedade de atributos (e.g. porcentagem de pele, compacidade, elipticidade, retangularidade, excentricidade, orientação, *Hmean*) considerando as cinco maiores regiões, gerando um vetor de tamanho 43. Por fim, foram adicionadas etapas de pré e pós processamento para o detector de pele, além de implementar várias regras estáticas visando a redução de falsos positivos.

Na pesquisa de Ma et al. (2014) foi proposta a utilização de particularidades da morfologia humana aliada às características extraídas por meio da detecção de pele humana. Dessa forma, era possível definir se píxeis atribuídos como pele humana estavam contidos em uma região com formato similar a partes do corpo humano, eliminando falsos positivos em função da similaridade das cores da pele com planos de fundo e objetos do cenário. No entanto, devido a uma estrutura complicada, é difícil considerar todas as possíveis posições relativas das partes do corpo. Além disso, esses métodos apresentam uma alta complexidade computacional, tornando-os inadequados para uso generalizado (OU et al., 2017).

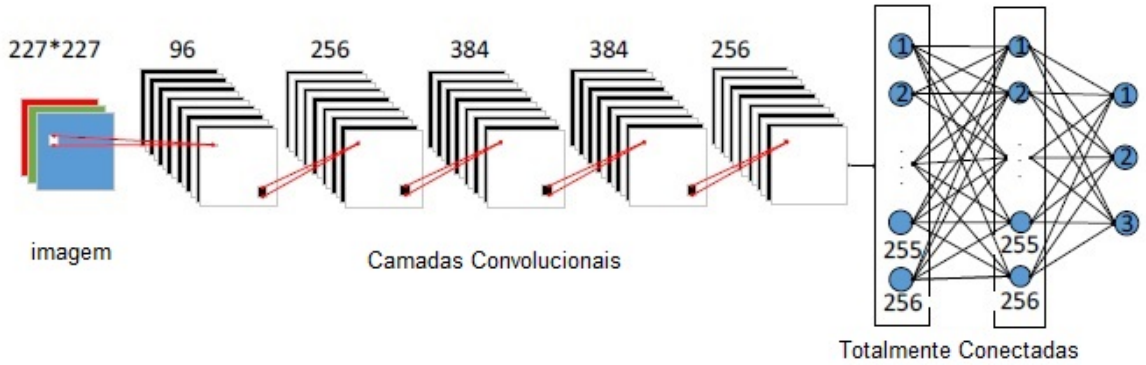
Apesar de ter sido amplamente estudada e ainda aplicada atualmente, os métodos baseados nesta categoria acabam sendo falhos por inferirem que a existência de uma grande quantidade de píxeis de pele está relacionada à pornografia, assim como o oposto, o que não

é uma verdade absoluta. É possível perceber a existência de atos sexuais com pessoas vestidas, assim como imagens com alta exposição de pele de pessoas em trajes de banho ou em práticas desportivas. Essa generalização acaba gerando um acréscimo nos falsos negativos e falsos positivos, respectivamente (GANGWAR et al., 2017).

4.1.2 Baseado em Redes Neurais Convolucionais

O trabalho Huang e Kong (2016) tem como objetivo identificar uma nova categoria de imagem pornográfica, denominada de *upskirt*. Trata-se de imagens fotográficas capturadas sob saias de indivíduos do sexo feminino. Os autores propõem identificar, além desta categoria, também imagens pornográficas convencionais e não pornográficas, somando ao todo três categorias em que uma imagem pode ser classificada. Foi utilizada uma variação da rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), alterando o tamanho de alguns filtros e das unidades escondidas da rede neural totalmente conectada no final, além de três categorias de saída: pornográfica, *upskirt* e não pornográfica, como pode ser visto na Figura 4.1. Em seguida, foram treinadas sete redes idênticas com imagens distintas, tendo todas essas demonstrado melhor acurácia, na detecção de pornografia, quando comparadas a métodos tradicionais baseados em detecção de pele e de formato humano. Por fim, os autores também criaram diversos conjuntos (*ensembles*) contendo entre duas e sete dessas redes neurais convolucionais, tendo apresentado melhores resultados que as redes neurais convolucionais isoladas.

Figura 4.1: Variação da rede neural convolucional AlexNet.



Fonte: Adaptada de Huang e Kong (2016).

Os estudos de Zhou et al. (2016) e Surendran e Stephen (2017), apesar de utilizarem

a mesma rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) como base, propõem uma abordagem distinta com relação à detecção de imagens pornográficas, fazendo uso de duas etapas sequenciais: (i) detecção grosseira e (ii) detecção refinada. A detecção grosseira é baseada em regras que levam em consideração a quantidade de píxeis de pele e do tamanho de faces nas imagens, fazendo com que a grande maioria das imagens possam ser classificadas corretamente como normais. Já a detecção refinada é realizada por meio do uso da referida rede neural convolucional, restando apenas as imagens mais difíceis de serem categorizadas. Os autores fazem uso da técnica de transferência de aprendizado, mantendo os pesos das três primeiras camadas de convolução da rede neural convolucional e retreinando apenas as duas últimas, juntamente com a rede neural totalmente conectada e seu classificador. Comparou-se o método proposto com dois métodos baseados em características manualmente extraídas das imagens, SURF-HSV (GENG et al., 2015) e ORB+HSV (ZHUO et al., 2016), obtendo melhores resultados e desempenho computacional comparado a ambos.

O estudo de Wang, Jin e Tan (2016) propõe o uso de um modelo que utiliza uma rede neural profunda baseada na arquitetura GoogLeNet (SZEGEDY et al., 2015) aliada à técnica de MIL (*Multiple Instance Learning*), que em vez de receber imagens rotuladas, recebe um conjunto de subimagens individualmente rotuladas. Dessa forma, a imagem será classificada como não pornográfica quando todas as subimagens assim forem classificadas, sendo necessário que haja apenas uma subimagem rotulada como pornográfica para que a imagem seja rotulada dessa maneira. Os autores fazem uso de uma base de dados que contém 117.000 imagens pornográficas e 117.000 imagens não pornográficas. Por fim, é possível concluir que o modelo adotado apresenta melhor taxa de detecção que a rede neural sem o uso da técnica de MIL, além das técnicas tradicionais baseadas em recuperação (SHIH; LEE; YANG, 2007) e em saco de palavras visuais (*bag-of-visual-words - BOVW*) (LOPES et al., 2009).

Usando uma base de dados privada com aproximadamente 650.000 imagens, o estudo de Li et al. (2016) mostra que a rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) apresentou resultados melhores que dois outros classificadores, no que diz respeito à detecção de imagens pornográficas, esses baseados em detecção de pele (ZUO; HU; WU, 2010) e em saco de palavras visuais (*bag-of-visual-words - BOVW*) (AVILA et al., 2013). Ademais, os autores propuseram um modelo híbrido que alia

a rede neural convolucional a um saco de palavras visuais (*bag-of-visual-words - BOVW*), superando o uso isolado da rede neural convolucional em três das cinco configurações apresentadas.

O estudo de Ou et al. (2017) tem como objetivo classificar imagens entre três categorias: não pornográfica, pornográfica e imprópria para crianças utilizando uma arquitetura denominada DMCNet (*Deep Multicontext Network*), na qual a sua maior parte consiste em uma rede neural convolucional profunda. Foram utilizadas três tipos dessas redes: a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), a VGG-16 (SIMONYAN; ZISSERMAN, 2014) e a GoogLeNet (SZEGEDY et al., 2015). A rede neural convolucional VGG-16 (SIMONYAN; ZISSERMAN, 2014) apresentou melhores resultados, no entanto, a rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) mostrou-se mais rápida que as demais.

Surinta e Khamket (2019) compararam diferentes modelos baseados em aprendizado de máquina utilizando a base de dados pornográfica TI-UNRAM Pornographic Image Dataset (WIJAYA et al., 2015). Tais modelos incluem redes neurais convolucionais, saco de palavras visuais (*bag-of-visual-words - BOVW*) e algoritmos de aprendizado de máquina tradicionais combinados com características extraídas manualmente. Por fim, a rede neural convolucional ResNet (HE et al., 2016) obteve melhores resultados quando comparado a todas as outras abordagens.

Apesar de amplamente estudado, o reconhecimento de imagens pornográficas permanece sendo um problema desafiador da Visão Computacional. Diversos fatores podem ser enumerados como responsáveis por essa dificuldade, tais como: (i) a não existência de um consenso na definição do que é pornografia, (ii) a ausência de um cenário/local padrão, (iii) os diferentes níveis de luminosidade nas imagens, (iv) as incontáveis posições e quantidade de indivíduos e (iv) a alta similaridade entre algumas imagens pornográficas e imagens rotineiras/cotidianas (e.g. trajes de banho e esportivo, esportes de contato) (WANG; JIN; TAN, 2016).

Por fim, foram sumarizadas na Tabela 4.1 as principais características, assim como os resultados, de cada um dos estudos relacionados ao reconhecimento de pornografia adulta referenciados nesta seção.

Tabela 4.1: Principais características e resultados dos estudos relacionados ao reconhecimento de pornografia adulta referenciados nesta seção.

Estudo	Detecção de Pele	Aprendizado de Máquina Tradicional	Aprendizado Profundo	Transferência de Aprendizado	Aumento de Dados	Base de Dados	Resultado
Ap-Apid (2005)	X					Privada	Rev.: 94,32% TFP: 5,04%
Polastro e Eleutério (2010)	X					COMPAQ	Rev.: 94,00% Prec.: 88,40%
Platzer, Stuetz e Lindorfer (2014)	X	X				COMPAQ+ Privada	Ac.: 91,90% Rev.: 65,70% Prec.: 39,80% TFP: 6,40%
Huang e Kong (2016)			X			Privada	Ac.: 90,19%
Zhou et al. (2016)	X	X	X	X		Privada	Prec.: 97,20%
Wang, Jin e Tan (2016)			X	X		NPDI+ Privada	Ac.: 98,81%
Li et al. (2016)			X			Privada	Ac.: 92.04%
Medina e Palladino (2017)	X	X				-	-
Surendran e Stephen (2017)	X	X	X			Privada	Ac.: 97,03%
Ou et al. (2017)			X			NPDI DMCV SPD Sensitive	Ac.: 85,30% Ac.: 81,40% Ac.: 95,40% Ac.: 97,80%
Surinta e Khamket (2019)	X	X	X			TI-UNRAM	Ac.: 88,00%

Fonte: Autor.

4.2 Estimação de Idade por Meio de Reconhecimento Facial

A evolução de metodologias aplicadas à estimação de idade de imagens faciais tem sido um dos problemas mais desafiadores no campo da análise facial. Essa dificuldade se dá principalmente pelo processo de envelhecimento não uniforme nos seres humanos, que depende de diversos fatores como: (i) genética, (ii) alimentação, (iii) práticas esportivas, (iv) incidência solar, (v) bem estar mental, entre outros. Essas variáveis acabam fazendo com que dois indivíduos com idades distintas possuam aparência similar quanto à idade, assim como dois indivíduos com a mesma idade apresentem diferença em sua aparência etária. Esse tipo de estimativa de idade pode ser classificado em dois tipos: (i) biológica, no qual a idade real do ser humano é predita e a (ii) aparente, cuja idade de saída é baseada nas suposições de um grupo de indivíduos, por meio da aparência do sujeito (RONDEAU; ALVAREZ, 2018).

Os primeiros estudos na área lidavam com o problema utilizando duas etapas sequenciais. Na primeira acontecia a extração das características mais relevantes das imagens de entrada, que poderiam ser locais (e.g. rugas da testa, contorno dos olhos, bochechas) ou globais. Essas características eram organizados na forma de vetor, servindo como representação da imagem para a próxima etapa. O aprendizado de máquina era utilizado na etapa seguinte para mapear o vetor de características com o rótulo de saída, no caso, a idade real (XIA et al., 2020). Esses primeiros trabalhos utilizavam métodos básicos de aprendizado de máquina e colocavam o problema como uma regressão ou uma classificação (NAM et al., 2020).

Essas abordagens baseadas na extração manual de características apresentam uma série de desvantagens: (i) a seleção das características de maneira não automatizada geralmente é uma tarefa difícil, demorada e entediante; (ii) o vetor de representação não é capaz de capturar relacionamentos hierárquicos entre as características; e (iii) o design não segue uma estratégia ponta-a-ponta (LI et al., 2020). Em relação ao aprendizado de máquina, ao apresentar o problema como uma tarefa de classificação, o modelo ignora a relação existente entre as idades, visto que o referido problema apresenta classes que podem apresentar maior ou menor similaridade entre si. Dessa forma, os erros cometidos pelo modelo são penalizados com o mesmo peso, independentemente de quão longe o erro da predição está do valor verdadeiro. Além disso, para treinar um classificador de n-classes, as idades reais são

geralmente discretizadas utilizando uma determinada resolução r (geralmente $r = 1$), implicando em classes com dados desbalanceados que podem influenciar no desempenho final do modelo (ROTHE; TIMOFTE; GOOL, 2015).

Com a expansão das aplicações baseadas em aprendizado profundo na Visão Computacional, diversos trabalhos vêm sendo propostos utilizando essa técnica para a estimativa de idade que, além de obter melhores resultados comparados a técnicas tradicionais, não necessita realizar a extração manual das características das imagens, como observado na pesquisa de Castrillón-Santana et al. (2018). O referido trabalho foca na diferenciação entre faces de crianças e adultos, comprovando a superioridade de uma rede neural convolucional rasa (com apenas três camadas convolucionais) quando comparada a uma máquina de vetores de suporte utilizando uma extensa combinação de características extraídas manualmente.

Entretanto, para que os modelos baseados em aprendizado profundo não se sobreajustem (*overfitting*), se faz necessário o uso de uma grande quantidade de dados na etapa de treinamento. Baseado nessa necessidade, a pesquisa de Oliveira et al. (2016) propôs a criação de dados artificiais, com intuito de aumentar a quantidade de imagens utilizadas nessa etapa. Esse aumento de dados, implementado por meio de variações de expressões faciais, baseadas em *Active Appearance Models* (AAM), aprimoraram o desempenho final da estimativa de idade facial de um modelo baseado em redes neurais convolucionais.

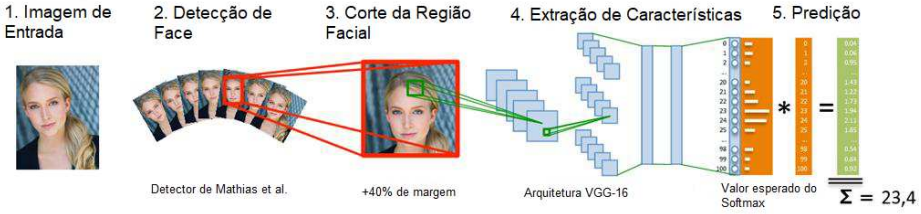
Uma das pesquisas precursoras da área a utilizar aprendizado profundo, a *Deep EXpectation (DEX)* (ROTHE; TIMOFTE; GOOL, 2015), foi a vencedora do desafio ChaLearn LAP 2015 na categoria de estimativa de idade aparente, tendo concorrido contra mais de 115 equipes. O referido estudo utilizou como base a rede neural convolucional VGG-16 (SIMONYAN; ZISSERMAN, 2014) já pré-treinada com os dados da competição *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) (RUSSAKOVSKY et al., 2015a). Visando melhorar a estimação das idades, a rede neural foi retreinada com a *IMDB-WIKI dataset*, uma base de dados extraída da internet pelos próprios autores. A base de dados em questão contém cerca de meio milhão de imagens possuindo faces e suas respectivas idades, sendo a maior base de dados para estimativa de idade biológica baseada em faces. Por fim, foi criado um conjunto (*ensemble*) contendo 20 VGG-16 retreinadas com os dados aumentados em dez vezes da base de dados da competição, a *ChaLearn dataset*. Foi adotado o detector de faces proposto por Mathias et al. (2014). Uma contribuição interessante desse trabalho

foi o uso de uma distribuição de probabilidade como rótulo, baseada na idade real da face, para treinar a rede. Dessa forma, na etapa de testes a idade prevista era calculada por meio da soma ponderada aplicada à última camada da rede, conforme mostrado na Equação (4.1).

$$E(O) = \sum_{i=0}^{100} y_i o_i \tag{4.1}$$

em que $O = \{o_0, o_1, \ldots, o_{100}\}$ é a saída de 101 dimensões da rede, representando probabilidades da função *Softmax*, e y_i são os anos discretos correspondentes a cada classe $i \in [0, 100]$. Embora o foco de seu trabalho fosse a estimativa da idade aparente, isso desencadeou o desenvolvimento de outros métodos para estimativa da idade aparente e real. O esquema do modelo em questão pode ser visualizado na Figura 4.2.

Figura 4.2: Arquitetura possuindo detecção de face e predição da idade facial por meio de rede neural convolucional.



Fonte: Adaptada de Rothe, Timofte e Gool (2015).

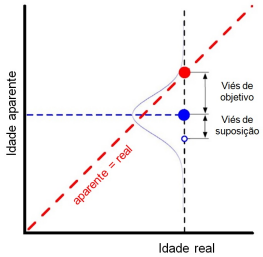
A pesquisa de Antipov et al. (2016) obteve a primeira colocação na segunda etapa do desafio *ChaLearn LAP* (ESCALERA et al., 2016). Os autores se basearam na estratégia utilizada pelo campeão da etapa anterior (ROTHE; TIMOFTE; GOOL, 2015), fazendo uso da mesma rede neural convolucional VGG-16 (SIMONYAN; ZISSERMAN, 2014) já pré-treinada com os dados da competição ILSVRC (RUSSAKOVSKY et al., 2015a) e retreinada com uma versão melhorada da base de dados *IMDB-WIKI dataset* (ROTHE; TIMOFTE; GOOL, 2015). Entretanto utilizou-se uma estratégia distinta, fazendo uso de um conjunto (*ensemble*) de 11 VGG-16 treinadas com os dados da competição para estimar a idade aparente e caso a estimativa fosse menor que 12 anos, a imagem passaria por outro conjunto (*ensemble*) de VGG-16 treinada apenas com imagens de faces de crianças entre 0 e 12 anos.

Um modelo de regressão, denominado *DEX residual*, foi adicionado ao modelo *Deep EXpectation (DEX)* (ROTHE; TIMOFTE; GOOL, 2015), a fim de minimizar a diferença entre as predições para os verdadeiros valores (AGUSTSSON et al., 2017). Esse estudo tam-

bém apresentou a base de dados APPA-REAL, a primeira base de dados de idade facial com rótulos de idade real e aparente. Além disso, sua proposta foi avaliada em ambas as categorias de estimativa de idade, real e aparente. Dessa forma, foi possível concluir por parte dos autores que tais abordagens não eram totalmente independentes entre si. A partir desse estudo, pôde-se concluir que prever a idade aparente é uma tarefa mais fácil em comparação à predição da idade real, e que o uso de dados advindos da idade aparente pode potencialmente melhorar a estimativa da idade real.

É importante notar que as suposições de idade aparente podem ter um viés significativo em relação à idade real. Esse viés pode ser dividido em: (i) viés de objetivo, que é inerente à distância entre a idade cronológica de uma pessoa e sua aparência; e (ii) viés de suposição, que é a tendência introduzida pelos humanos ao inferir uma idade. Uma representação gráfica desses vieses pode ser vista na Figura 4.3.

Figura 4.3: O viés de objetivo é a diferença entre a idade real (ponto vermelho) e a média de todas as estimativas de idade (ponto azul). O viés de suposição é a diferença entre uma estimativa específica (ponto branco) e a média de todas as estimativas de idade (ponto azul).



Fonte: Adaptada de Clapes et al. (2018).

Um mecanismo de correção de viés para as previsões de um modelo baseado em redes neurais convolucionais foi introduzido para melhoria do desempenho (CLAPES et al., 2018). O estudo de Jacques Junior et al. (2019) propôs uma abordagem em que a correção de viés foi integrada ao modelo como parte de uma estratégia ponta-a-ponta, em vez de uma etapa de pós-processamento. Um vetor com informações de gênero, etnia, nível de felicidade e uso de maquiagem foi usado durante a fase de treinamento, para capturar a forma como os rostos das pessoas são percebidos pelos humanos.

Dada a natureza ordinal das idades, o uso de funções de perda que aplicam a mesma penalidade a erros de classificação, independentemente de sua distância do valor real, não é apropriado. Para mitigar esse problema, alguns estudos vêm introduzindo novas funções de

perda (LI et al., 2020) ou recorrendo ao aprendizado de distribuição de rótulos (*Label Distribution Learning - LDL*). A ideia do LDL é treinar o modelo com imagens faciais em que cada imagem é rotulada com uma distribuição de probabilidade de suposições de humanos (GAO et al., 2017). Essas distribuições de probabilidade geralmente são modeladas a partir do conjunto de estimativas humanas disponíveis para cada imagem em conjuntos de dados de idade aparente, como a base de dados APPA-REAL. Baseado nessa técnica, porém utilizando uma estimativa de densidade por kernel para rotular cada imagem, o estudo de Rondeau e Alvarez (2018) superou o estado da arte na estimativa de idade aparente. Além do mais, treinou inicialmente uma rede neural convolucional DenseNet (HUANG et al., 2017) com uma versão otimizada da base de dados *IMDB-WIKI* (ROTHE; TIMOFTE; GOOL, 2015), aplicando a técnica de transferência de aprendizado.

A extensão dessa estratégia para a estimativa da idade real é direta. Uma distribuição normal pode ser parametrizada usando a idade real como a média e um valor fixo (por exemplo, $\sigma = 3$) como o desvio-padrão. No entanto, essa parametrização ingênua apresentou resultados piores em comparação com um modelo *baseline* simples, como um classificador de n-classes em (RONDEAU; ALVAREZ, 2018).

Ademais, combinações ponderadas de diferentes funções de perda também foram exploradas em estudos voltados para a estimativa de idade (PAN et al., 2018; ZHANG et al., 2019; LIU et al., 2020). A ideia é encontrar o melhor conjunto de pesos para múltiplas funções de perda (e.g. regressão, entropia cruzada, divergência de Kullback-leibler), a fim de minimizar o erro de predição. Esses estudos têm mostrado que combinações ponderadas de funções de perda produzem melhores resultados quando comparadas a modelos que utilizaram uma única função de perda.

Dessa forma, a detecção de material pornográfico que contenha crianças e/ou adolescentes pode ser ampliada por meio do uso dessa técnica de estimação de idade em faces em imagens. O uso dessa tecnologia não tem implicações apenas na esfera criminal, mas também nas áreas comerciais, pois é de interesse das empresas em detectar, censurar e reportar atividades ilegais em suas plataformas (JUNG; MAKHIJANI; MORLOT, 2017).

Por fim, foram sumarizadas na Tabela 4.2 as principais características, assim como os resultados, de cada um dos estudos relacionados à estimativa de idade facial referenciados nesta seção.

Tabela 4.2: Principais características e resultados dos estudos relacionados à estimativa de idade facial referenciados nesta seção.

Estudo	Aprendizado Profundo	Transferência de Aprendizado	Aumento de Dados	Função de Custo Composta	Idade Real	Idade Aparente	Base de Dados	Resultado
Rothe, Timofte e Gool (2015)	X	X	X			X	LAP 2015	erro-ϵ: 0,2650
Antipov et al. (2016)	X	X	X			X	LAP 2016	erro-ϵ: 0,2411
Agustsson et al. (2017)	X	X	X		X	X	APPA-REAL	Real EMA: 5,296 Aparente EMA: 4,082
Rondeau e Alvarez (2018)	X	X	X		X	X	APPA-REAL	Real EMA: 5,434 Aparente EMA: 3,688
Pan et al. (2018)	X	X		X	X	X	MORPH-II FG-NET LAP 2016	EMA: 2,16 EMA: 2,68 erro-ϵ: 0,2867
Jacques Junior et al. (2019)	X	X			X	X	APPA-REAL	Real EMA: 7,356 Aparente EMA: 6,131
Zhang et al. (2019)	X	X		X	X		MORPH-II FG-NET	EMA: 2,75 EMA: 2,95
Liu et al. (2020)	X	X		X	X	X	MORPH-II FG-NET LAP 2016	EMA: 2,56 EMA: 2,98 erro-ϵ: 0,2850

Fonte: Autor.

4.3 Reconhecimento de Pornografia Infanto-juvenil

Antes da popularização do aprendizado profundo na Visão Computacional, os estudos voltados para a classificação de imagens obrigatoriamente necessitavam desenvolver estratégias para extrair informações que representassem as imagens da melhor maneira possível, não sendo diferente para o problema de detecção de pornografia infanto-juvenil. Um dos estudos precursores nesta área (POLASTRO; ELEUTÉRIO, 2010) utilizava regiões de píxeis de pele como representação da imagem e simples regras estáticas para classificar se determinada imagem continha pornografia infanto-juvenil ou não.

Com o passar dos anos, o aprendizado de máquina passou a figurar na classificação de imagens como um todo, refletindo nos métodos adotados para a detecção de pornografia infanto-juvenil. A pesquisa de Ulges e Stahl (2011), utilizando como classificador uma máquina de vetores de suporte, mostrou que o levantamento de características utilizando a técnica de saco de palavras visuais (*bag-of-visual-words* - *BOVW*) apresentou melhores resultados quando comparado a um modelo *baseline* que utilizava apenas características baseadas em píxeis de pele.

O detector de pornografia infantil proposto por Sae-Bae et al. (2014) é composto por dois módulos principais: (i) um detector de pornografia e (ii) um detector de faces infantis. Para que haja a detecção de imagens pornográficas, as imagens passam por um pré-processamento que ajusta balanço de branco das imagens, diminuindo a sensibilidade às variações de iluminação e em seguida são detectados os píxeis de pele por meio de uma máquina de vetor de suporte treinada com dados no espaço de cor RGB. Para o levantamento das características, além dos píxeis de pele, um detector de faces é utilizado para construir um vetor de tamanho 11. São utilizadas proporções de píxeis de pele, distribuição das regiões de pele e tamanho das faces na imagem. Na sequência, também é utilizada uma máquina de vetor de suporte para classificar as imagens como pornografia ou não. Tendo sido classificada como pornografia, o detector de faces infantis é acionado para inferir se a imagem pornográfica possui criança(s) ou não. Mais uma vez foi utilizada uma máquina de vetor de suporte para a classificação, tendo feito uso de um vetor de características de tamanho 66, esse composto por distâncias entre diversos componentes faciais (e.g. olhos, nariz, boca, bochechas). Por fim, a abordagem proposta atingiu 74,19% de acurácia na detecção de pornografia infantil.

A pesquisa de Schulze et al. (2014) analisou seu modelo para diferenciação de imagens em três diferentes cenários: (i) imagens não ofensivas versus pornografia adulta, (ii) imagens não ofensivas versus pornografia infantil e (iii) pornografia adulta versus pornografia infantil. Os autores propuseram o uso de características de baixo-nível (i.e. correlogramas de cores, píxeis de pele, palavras visuais) e de médio nível (i.e. análise de sentimento). A classificação do modelo se deu mediante utilização de uma máquina de vetores de suporte. Foi possível comprovar que o uso combinado dos conjuntos de caraterísticas atingiu melhores resultados comparado ao seu uso de maneira isolada, diminuindo a taxa de erros de 17% para 10%.

Devido à ausência de dados para treinamento de modelos que possam detectar imagens dessa natureza, o estudo de Yiallourou, Demetriou e Lanitis (2017) construiu uma base de dados sintéticos simulando imagens contendo pornografia infanto-juvenil. Inicialmente, as faces são detectadas, utilizando um algoritmo baseado em *Haar cascade*, para que na sequência a idade e o sexo de cada face sejam estimados. De posse dessas informações, é gerado um vetor de características de tamanho cinco, contendo as seguintes informações: (i) presença de criança, (ii) número de pessoas, (iii) diversidade etária, (iv) proporção de gênero e (v) nível de iluminação do ambiente. Por fim, o vetor de características em questão foi utilizado para treinar regressor para classificar as imagens como: (a) adequadas, (b) neutras ou (c) inadequadas, atingindo uma acurácia de 48% na etapa de testes.

O trabalho de Gangwar et al. (2017), influenciado pela expansão do aprendizado profundo na Visão Computacional, realizou um comparativo entre cinco diferentes modelos para a detecção de pornografia infanto-juvenil. Salienta-se que os modelos originalmente foram desenvolvidos para a detecção de pornografia adulta. Na fase experimental foi utilizada uma base de dados contendo 5.000 imagens, dentre elas 2.500 retratando pornografia infantil e as demais normais, sem qualquer indício de pornografia. Os dois primeiros modelos eram baseados em detecção de pele, no qual o classificador define se determinada imagem é pornográfica ou não de acordo com um limiar baseado na quantidade de píxeis de pele da imagem. O terceiro modelo era baseado em um descritor de imagem, no qual é gerado vetor de características que representa cada imagem e posteriormente é categorizada por um classificador. O quarto modelo é baseado em uma versão reduzida da rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), possuindo apenas quatro camadas convolucionais e uma rede totalmente conectada com uma camada de 1024 neurônios conectada

ao classificador *softmax*. Esse modelo utilizou transferência de aprendizado para inicializar seus parâmetros. Por fim, o último modelo trata-se de uma rede neural residual profunda com 50 camadas desenvolvida pela Yahoo! (MAHADEOKAR; PESAVENTO, 2016), tendo apresentado a melhor acurácia dentre todos os modelos.

O estudo de Vitorino et al. (2018) propõe uma metodologia em duas camadas que tem como base o uso da rede neural convolucional GoogLeNet (SZEGEDY et al., 2015) previamente treinada com 1,2 milhões de imagens subdivididas em mil categorias, essas sem qualquer relação com pornografia, tampouco infanto-juvenil. Na primeira é realizado o retreinamento da rede, aplicando a técnica de transferência de aprendizado (*transfer learning*), ou seja, aproveitando os pesos da rede preexistente. Foram utilizadas cerca de 200.000 imagens pornográficas e não pornográficas. A mesma técnica foi utilizada na segunda camada, contudo, o retreinamento foi realizado utilizando uma base de dados com aproximadamente 59.000 imagens possuindo conteúdo pornográfico infanto-juvenil e conteúdo lícito. Os autores atestaram que o uso da técnica de transferência de aprendizado potencializou os resultados, tanto para a detecção de pornografia adulta, quanto para a detecção de pornografia infanto-juvenil. Por fim, atestaram que o trabalho proposto apresentou melhor acurácia que os trabalhos envolvendo diversas técnicas, como detecção de pele, saco de palavras visuais (*bag-of-visual-words - BOVW*) e até mesmo outra categoria de rede neural convolucional.

Com o intuito de avaliar sua abordagem, em parceria com a Polícia Federal do Brasil, Macedo, Costa e Santos (2018) desenvolveram uma base de dados contendo pornografia infantil (até 13 anos). A referida base consiste em 2.168 imagens, dentre essas, lícitas e ilícitas. Sua arquitetura é baseada em duas ramificações. A primeira tem como objetivo detectar a existência de pornografia nas imagens por meio de uma ferramenta de moderação de imagens disponibilizada gratuitamente pela Yahoo! (MAHADEOKAR; PESAVENTO, 2016). Essa ferramenta é baseada na arquitetura ResNet-50 (HE et al., 2016), tendo sido pré-treinada com a base de dados ImageNet (RUSSAKOVSKY et al., 2015a) e ajustada com um conjunto de dados proprietário contendo imagens seguras e inapropriadas. A segunda ramificação, acionada caso a imagem tenha sido classificada como pornográfica, tem o papel de estimar a idade de todas as faces contidas na imagem. Esse modelo, proposto pelos autores, é capaz de classificar a idade facial em oito grupos (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53 e 60+). Foi utilizada uma rede neural convolucional VGG-16 (SIMONYAN; ZISSERMAN, 2014)

previamente treinada também com os dados da base de dados ImageNet (RUSSAKOVSKY et al., 2015a) e ajustada com a base de dados facial Adience (LEVI; HASSNER, 2015). Por fim, a proposta de Macedo, Costa e Santos (2018) foi capaz de distinguir de maneira correta entre imagens lícitas (apropriadas) e ilícitas (pornografia infantil) em 79,84% das predições.

O trabalho de Jung, Makhijani e Morlot (2017) tem como objetivo a detecção de vídeos com conteúdo pornográfico infanto-juvenil, entretanto, trata-os como imagens, visto que utiliza os quadros de maneira isolada. Para tal, o autor fez uso de três redes neurais convolucionais (i.e. VGG-16 (SIMONYAN; ZISSERMAN, 2014), ResNet (HE et al., 2016) e Inception-v4 (SZEGEDY et al., 2016)). As redes em questão já foram treinadas com diversas classes de imagens, havendo o uso da técnica de transferência de aprendizado com o retreinamento apenas das últimas camadas com conteúdo pornográfico. Para as imagens preditas como pornográficas, a estimativa de idade real em faces é utilizada como critério único na tomada de decisão de pornografia infanto-juvenil. Inicialmente, a face é detectada por meio da técnica aplicada por Redmon et al. (2016) e, em seguida, os autores adotaram o modelo proposto por Levi e Hassner (2015), que utiliza uma rede neural convolucional com arquitetura bastante similar a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) para classificar as faces em oito subgrupos de idades (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+). Para tal, foi utilizada a base de dados *Face Image Project* (HASSNER, 2014).

Apesar de não ter sido especificamente desenvolvida para esse fim, a detecção de pornografia acaba exercendo um papel fundamental no auxílio à detecção da pornografia infanto-juvenil. Diversos trabalhos têm-se utilizado dessa técnica como base para a detecção desse material ilícito (YIALLOUROU; DEMETRIOU; LANITIS, 2017; JUNG; MAKHIJANI; MORLOT, 2017), devido à ausência de material pois, além de não ser facilmente encontrado, sua posse sem excludentes de ilicitude (e.g. autorização judicial, estrito cumprimento do dever legal) é configurada como crime (YIALLOUROU; DEMETRIOU; LANITIS, 2017; MACEDO; COSTA; SANTOS, 2018). Por fim, pôde-se concluir que a detecção da pornografia infanto-juvenil é uma tarefa muito mais desafiadora quando comparada à detecção de pornografia comum (GANGWAR et al., 2017).

Por fim, foram sumarizadas na Tabela 4.3 as principais características, assim como os resultados, de cada um dos estudos relacionados ao reconhecimento de pornografia infanto-juvenil referenciados nesta seção.

Tabela 4.3: Principais características e resultados dos estudos relacionados ao reconhecimento de pornografia infanto-juvenil referenciados nesta seção.

Estudo	Detecção de Pele	Aprendizado de Máquina Tradicional	Aprendizado Profundo	Transferência de Aprendizado	Aumento de Dados	Estimação de Idade	Resultado
Ulges e Stahl (2011)	X	X					EER: 21,50%
Sae-Bae et al. (2014)	X	X				X	Ac.: 74,19%
Schulze et al. (2014)	X	X					AVP: 97,32% EER: 7,38%
Yiallourou, Demetriou e Lanitis (2017)	X					X	Ac.: 48,00%
Gangwar et al. (2017)	X	X	X	X			Ac.: 87,56%
Vitorino et al. (2018)			X	X	X		Ac.: 86,50%
Macedo, Costa e Santos (2018)			X	X	X	X	Ac.:79,85%

Fonte: Autor.

4.4 Considerações Finais

No capítulo em questão, foi realizada uma revisão bibliográfica abarcando os principais trabalhos relacionados ao trabalho proposto. Em um primeiro momento, foram pesquisados por estudos capazes de reconhecer conteúdo pornográfico convencional de maneira automática, sendo esses baseados no levantamento de características da pele humana ou por meio do uso de redes neurais convolucionais.

Em seguida, foram levantadas pesquisas capazes de estimar a idade de um ser humano por meio das características faciais, etapa de extrema importância para definir se uma imagem com conteúdo pornográfico possui como atores crianças ou adolescentes.

Por fim, foram revisadas pesquisas específicas de reconhecimento de pornografia infanto-juvenil, esses bem menos explorados na literatura, devido à escassez de material dessa natureza, assim como às restrições de posse desse tipo de imagem.

Capítulo 5

Abordagem Proposta para Detecção de Pornografia Infanto-juvenil

Este capítulo descreve a abordagem adotada para a realização da tarefa alvo da pesquisa, a detecção de pornografia infanto-juvenil, além de discriminar o papel e onde encontram-se inseridas as técnicas expostas nos Capítulos 6, 7 e 8 para a arquitetura como um todo.

5.1 Arquitetura Proposta

Visando à realização da detecção de pornografia infanto-juvenil em imagens, tarefa alvo da pesquisa, foi adotada uma arquitetura sequencial composta por dois módulos: (i) um detector de pornografia (Módulo Pornográfico) e (ii) um estimador de idade facial (Módulo Facial). A arquitetura em questão vem sendo adotada em trabalhos anteriores da área (SAE-BAE et al., 2014; JUNG; MAKHIJANI; MORLOT, 2017; MACEDO; COSTA; SANTOS, 2018), entretanto, foram propostos módulos com particularidades inovadoras para a realização da referida tarefa de maneira mais acurada, além do uso de um modelo baseado em aprendizado de máquina para otimizar a determinação da menoridade penal dos indivíduos (Classificador de Menoridade Penal).

Sendo assim, dada uma determinada imagem de entrada de qualquer natureza, a mesma será classificada em uma das seguintes categorias:

- CONTEÚDO LÍCITO: São imagens que NÃO POSSUEM RESTRIÇÕES de serem portadas e/ou compartilhadas de acordo com a legislação brasileira, no que diz res-

peito à pornografia. São imagens sem nenhum conteúdo pornográfico ou que contêm pornografia com todos o indivíduos sendo maiores de idade;

- CONTEÚDO SUSPEITO: São imagens que PODEM APRESENTAR RESTRIÇÃO de serem portadas e/ou compartilhadas de acordo com a legislação brasileira, no que diz respeito à pornografia. São imagens que possuem conteúdo pornográfico dotadas das probabilidades de cada um dos indivíduos identificados pela face possuírem menos de 18 anos de idade;

- CONTEÚDO INDETERMINADO: São imagens pornográficas, dada a abordagem proposta, inviabilizadas de serem categorizadas como CONTEÚDO LÍCITO ou SUSPEITO, pois não possuem qualquer face identificável para que seja estimada a idade facial do(s) indivíduo(s);

O fluxograma completo da arquitetura, assim como onde os módulos encontram-se inseridos, podem ser visualizados na Figura 5.1. Em seguida, é possível verificar o detalhamento das etapas do processo.

1. Entrada de imagem para predição;

2. Submissão da imagem ao Módulo Pornográfico - Classificador: Pornografia vs. Não Pornografia;

3. Se não for detectada pornografia, a imagem é categorizada como CONTEÚDO LÍCITO, caso contrário, será dada continuidade ao processo;

4. Submissão da imagem ao Módulo Facial;

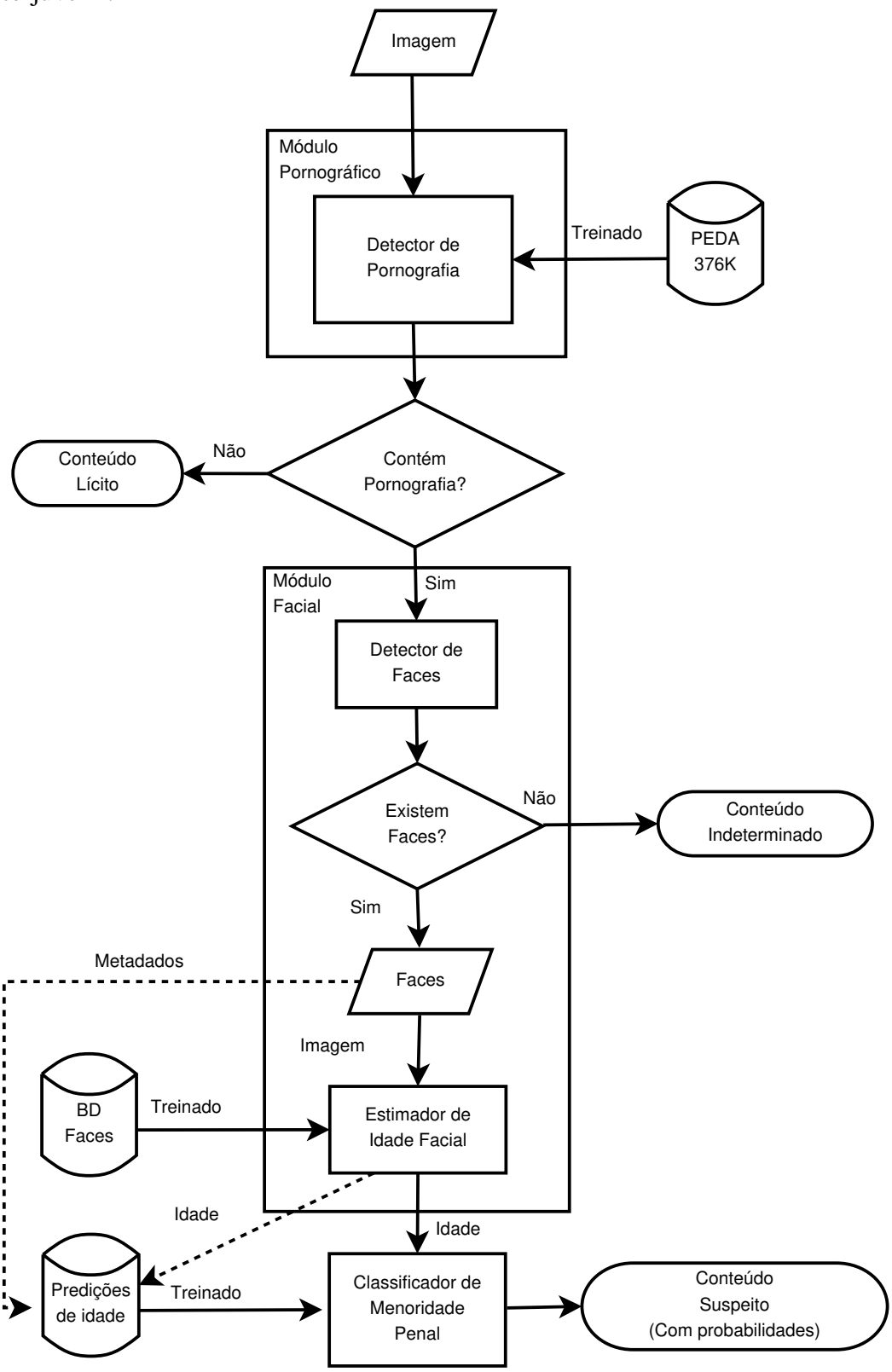
5. Realização da detecção de faces na imagem;

6. Se nenhuma face for identificada, a imagem é categorizada como CONTEÚDO INDETERMINADO, caso contrário, será dada continuidade ao processo;

7. Estimação da idade facial;

8. Classificador de Menoridade Penal que determina as probabilidades das faces detectadas pertencerem a indivíduos com menos de 18 anos.

Figura 5.1: Fluxograma do processo proposto de detecção de imagens contendo pornografia infanto-juvenil.



Fonte: Autor.

5.1.1 Módulo Pornográfico

O Módulo Pornográfico tem a função de classificar imagens de entrada em duas categorias distintas: (i) pornografia e (ii) não pornografia. Foi utilizada a arquitetura proposta em Moreira, Pereira e Alvarez (2020), cuja metodologia é apresentada detalhadamente no Capítulo 6 (Detector de Pornografia Baseado em Aprendizado Profundo e Nova Base de Dados Pornográfica). Em suma, foi utilizada uma estratégia gulosa, em três etapas, para a seleção de uma rede neural convolucional e do ajuste fino de seus hiperparâmetros. A base de dados pornográfica proposta, a *Pornographic and Explicit Database 376K (PEDA 376K)*, foi utilizada para o treinamento, validação e teste da arquitetura.

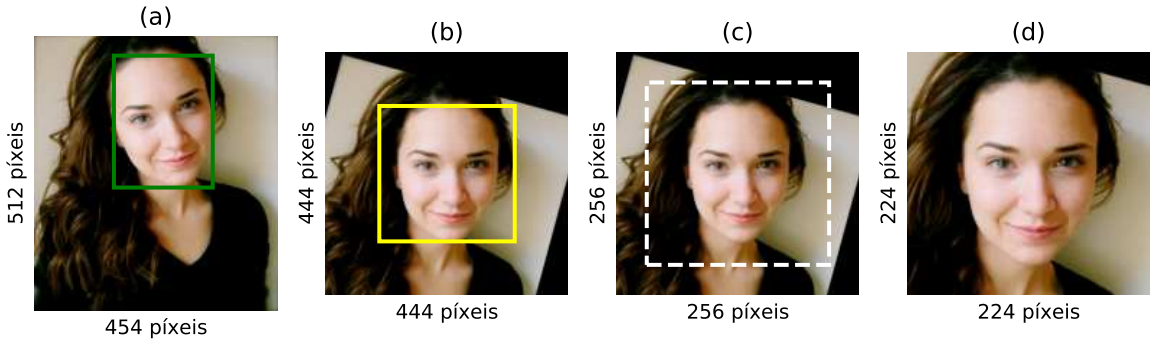
5.1.2 Módulo Facial

O papel do Módulo Facial é de determinar as idades de todas as faces detectadas nas imagens de entrada. Baseado nos estudos de Macedo, Costa e Santos (2018) e Liu et al. (2020), foi utilizado o detector facial MTCNN (ZHANG et al., 2016), que utiliza uma arquitetura em cascata composta por três redes neurais convolucionais para a detecção das faces. Cada uma das etapas refina essa detecção, mas apenas na última etapa são adicionadas características como pontos fiduciais (*landmarks*), nível de confiança e tamanho da face.

Baseado na pesquisa de Rothe, Timofte e Gool (2015), são utilizadas apenas as faces que possuem um valor padrão de confiança (que varia de 0 a 1) igual ou maior que 0,95. A partir de então, as faces passam por um pré-processamento para que possam ser submetidas à estimação de idade. Primeiramente, o retângulo contendo a face é convertido em quadrado (tendo como base o maior lado) e as faces são rotacionadas para que os olhos se alinhem horizontalmente. Leva-se em consideração, além da área detectada da face, 40% adicionais de margem da imagem. Por fim, a imagem é redimensionada para 256×256 píxeis e cortada centralmente em 224×224 píxies, obedecendo ao tamanho padrão da rede neural utilizada, como pode ser visto na Figura 5.2.

De posse das faces pré-processadas, para realizar a estimativa da idade, foi realizado um estudo comparativo entre as técnicas propostas e modelos já consolidados no estado da arte da estimação de idade real por meio de faces. Toda a metodologia encontra-se discriminada de maneira detalhada no Capítulo 7 (Aprimorando a Estimativa de Idade Real a Partir de

Figura 5.2: Etapas do pré-processamento das faces detectadas. A etapa (a) mostra o retângulo em verde que representa a face detectada na imagem original. A etapa (b) mostra a imagem já rotacionada com os olhos alinhados horizontalmente e a mudança para o quadrado em amarelo representando a face detectada. A etapa (c) mostra o redimensionamento para 256×256 píxeis e a área de corte central de 224×224 píxeis, representado pelo quadrado branco serrilhado. A etapa (d) mostra a face pré-processada pronta para ser inserida na rede neural convolucional.



Fonte: Autor.

Dados de Idade Aparente).

5.1.3 Classificador de Menoridade Penal

A estimação de idade obtida pelo Módulo Facial facilmente pode categorizar um indivíduo como menor ou maior de idade, por meio do uso de um limiar (igual a 18) que separe as referidas classes. Entretanto diante da grande responsabilidade da determinação da menoridade penal de um indivíduo na seara da Perícia Criminal, que é capaz de condenar ou inocentar suspeitos, foi proposta a utilização da probabilidade de uma determinada face pertencer a um indivíduo menor de idade em detrimento do uso desse resultado binário (i.e.menor de idade e maior de idade).

Sendo assim, foi realizado um estudo, em que sua metodologia é descrita de maneira detalhada no Capítulo 8 (Classificador de Menoridade Penal), que propõe avaliar os resultados obtidos pelo método intrínseco ao Módulo Facial e uma proposição, baseada em aprendizado de máquina, que visa aprimorar o resultado final dessa classificação utilizando predições de idade do Módulo Facial e metadados das faces.

5.2 Considerações Finais

Neste capítulo, é exposta a arquitetura adotada para a detecção de pornografia infanto-juvenil, sendo descritas as etapas e o fluxograma de todo o processo. Ademais, foram descritos e contextualizados na arquitetura os Módulos Pornográfico e Facial, assim como o Classificador de Menoridade Penal, que tem como objetivo otimizar o resultado final das probabilidades inferidas a cada face de pertencer a um indivíduo menor de idade.

Capítulo 6

Detector de Pornografia Baseado em Aprendizado Profundo e Nova Base de Dados Pornográfica

Este capítulo, de conteúdo publicado na *2020 International Joint Conference on Neural Networks* (MOREIRA; PEREIRA; ALVAREZ, 2020), aborda a necessidade e as características de uma base de dados voltada para o uso de redes neurais profundas, mais especificamente no âmbito da detecção de pornografia. Sendo assim, é apresentada a base de dados pornográfica *Pornographic and Explicit Database 376K (PEDA 376K)*, em que é descrito todo o procedimento e embasamento da sua construção. Além do mais, são descritas as bases de dados mais utilizadas na literatura que se encontram disponíveis. Também é proposta uma nova abordagem capaz de padronizar os resultados dos serviços de moderação de imagens do estado da arte, viabilizando a comparação entre esses serviços e a abordagem proposta, treinada com os dados da PEDA 376K.

6.1 Introdução

A rápida expansão do mundo digital apresenta desafios complexos nos domínios forense e de segurança digital. Em particular, a ampla disponibilidade de mídia pornográfica na Internet é um grande problema para serviços que procuram prevenir a exposição desse tipo de material para públicos inadequados e/ou indesejados ou, principalmente, automatizar a

detecção de material ilícito, especificamente, pornografia infanto-juvenil (RAAIJMAKERS, 2019; PEREZ et al., 2017).

Técnicas tradicionais para detectar conteúdo pornográfico, como o uso de listas negras de nomes de arquivos ou URLs (NIAN et al., 2016) já não são mais aplicáveis. A Visão Computacional e as tecnologias de aprendizado profunda tornaram-se cruciais para essa tarefa, mudando o foco dos metadados para os conteúdos das mídias (LI et al., 2016).

Embora as tecnologias recentes de aprendizado profundo sejam muito poderosas em aplicações de visão computacional, pesquisadores e engenheiros precisam lidar com a presença de subjetividade em seus modelos, visto a existência de uma linha tênue que separa a definição pornografia e não pornografia, tornando difícil, mesmo para humanos, chegar a um consenso sobre essa interpretação. Por exemplo, um único arquivo de mídia pode ser considerado pornografia ou não por dois indivíduos diferentes (PUTRO; ADJI; WINDURATNA, 2015).

Devido a essa subjetividade, parte dos serviços disponíveis para detecção de mídia inapropriada (*Not-Safe-For-Work - NSFW*) não é capaz de inferir claramente se uma mídia detém conteúdo pornográfico ou não. Em geral, dada uma determinada imagem de entrada, os serviços de moderação de imagens retornam um conjunto de probabilidades, deixando a responsabilidade da decisão final para os usuários. Além disso, esses serviços não compartilham publicamente suas bases de dados de treinamento.

Além do mais, as redes neurais profundas exigem, se não pré-treinadas adequadamente, uma grande quantidade de dados na etapa de treinamento para que possam ser atingidos resultados satisfatórios. Essa quantidade de informação requerida apresenta relação com a profundidade e, consequentemente, com o número de parâmetros que a rede em questão possui. Muitas vezes, torna-se impraticável realizar experimentos com esse tipo de rede neural, visto que não é comum encontrar bases de dados disponíveis com essa quantidade de dados (milhares ou até milhões) para treinar esse tipo de modelo (NIAN et al., 2016).

Essa dificuldade mostra-se ainda maior com relação ao uso de imagens para detecção de pornografia. Apesar da vasta quantidade de imagens dessa natureza disponível na Internet, não existe uma base de dados confiável, estruturada e com um quantitativo de dados necessário disponível para ser utilizada em experimentos desse tipo, pois a grande maioria das bases de dados utilizadas nos trabalhos contidos na literatura é produzida pelos próprios

autores (WANG; JIN; TAN, 2016; NIAN et al., 2016; HUANG; KONG, 2016; ZHOU et al., 2016; LI et al., 2016; CHASE; HE; HEGAZY, 2017) e acabam não sendo disponibilizadas pelos mesmos (GANGWAR et al., 2017). Dessa forma, as poucas bases de dados de imagens pornográficas disponíveis não são adequadas para o cenário em questão.

Sendo assim, foi proposta uma base de dados de imagens pornográficas, a *Pornographic and Explicit Database 376K (PEDA 376K)*, que foi cuidadosamente rotulada usando um conjunto de regras bem definidas para determinar se uma imagem é pornográfica ou não. De posse da referida base de dados, foram realizados experimentos extensivos envolvendo o treinamento de redes convolucionais e ajuste fino de seus hiperparâmetros para detecção de pornografia. Por fim, os resultados obtidos foram comparados com cinco serviços de moderação de imagens inseridos no estado da arte. No geral, as contribuições desse capítulo incluem:

- Uma nova base de dados de imagens (PEDA 376K), contendo mais de 376.000 imagens rotuladas em duas categorias: (i) pornografia e (ii) não pornografia, em que foi utilizada uma definição objetiva de pornografia, minimizando a subjetividade da categorização das imagens;
- Aplicação de uma estratégia gulosa, objetivando a obtenção do melhor cenário envolvendo uma rede neural convolucional e seus hiperparâmetros, para a detecção de pornografia, utilizando o novo conjunto de dados PEDA 376K;
- Uma abordagem para transformar resultados probabilísticos dos serviços de moderação de imagens em decisões binárias, viabilizando a comparação dos resultados entre si e a abordagem proposta.

6.2 Bases de Dados Pornográficas na Literatura

Poucas são as bases de dados na literatura que apresentam conteúdo pornográfico. Ademais, muitas vezes esses conjuntos de dados não possuem uma quantidade suficiente de imagens para o uso em aprendizado profundo e/ou não possuem rótulos confiáveis das imagens com relação às categorias. A seguir são discriminadas algumas das principais bases de dados pornográficas disponíveis.

6.2.1 AIIA-PID4 Pornographic Data Set

A *AIIA-PID4 pornographic data set* (KARAVARSAMIS et al., 2013) é uma base de dados de imagens bem estruturada, estando dividida em quatro classes: 1) pornográfica; 2) biquíni; 3) pele e 4) não pele, conforme discriminado na Tabela 6.1. Contudo apresenta algumas falhas de categorização, ou seja, imagens classificadas como pornográficas de maneira equivocada, assim como o oposto. Além do mais, para redes neurais profundas, o quantitativo de apenas 12.770 imagens na maioria das vezes não é suficiente.

Tabela 6.1: Quantidade de imagens em cada categoria da AIIA-PID4 pornographic data set.

Categoria	**Quantidade**
Pornográfica	1.900
Biquíni	4.742
Pele	1.160
Não pele	4.968
Total	12.770

Fonte: Adaptada de Karavarsamis et al. (2013).

6.2.2 NPDI Pornography-800

A *NPDI Pornography-800* (AVILA et al., 2013) é uma base de dados de vídeos pornográficos, porém, também é utilizada para a detecção de imagens desse gênero. Foram utilizadas aproximadamente 80 horas de 400 vídeos pornográficos e 400 vídeos não pornográficos. Para a categoria pornográfica, foram extraídos de sites específicos diversos vídeos dos mais variados gêneros e etnias, conforme a Tabela 6.2. Para a categoria não pornográfica, foram extraídos 200 vídeos aleatórios (denominados como fáceis) e 200 vídeos com consulta textual como: “praia”, “luta livre”, “natação” (denominados como difíceis).

Tabela 6.2: Distribuição étnica dos vídeos pornográficos da NPDI Pornography-800.

Etnia	**% dos vídeos**
Asiáticos	16%
Negros	14%
Brancos	46%
Multiétnico	24%

Fonte: Adaptada de Avila et al. (2013).

A base de dados foi pré-processada, segmentando os vídeos em quadros-chave que resumem o conteúdo de determinado trecho do vídeo. Embora existissem maneiras sofisticadas

de escolher o quadro-chave, optou-se por selecionar o quadro intermediário de trechos dos vídeos. Por fim, a base de dados é composta por 16.727 imagens. Na Tabela 6.3 evidencia-se a proporção de imagens por vídeo de cada categoria.

Tabela 6.3: Distribuição da quantidade de vídeos, horas e imagens por vídeo de cada categoria da NPDI Pornography-800.

Categoria	**Vídeos**	**Horas**	**Imagens por vídeo**
Pornográfico	400	57	15,6
Não pornográfico (“fácil”)	200	11,5	33,8
Não pornográfico (“difícil”)	200	8,5	17,5
Todos os vídeos	800	77	20,6

Fonte: Adaptada de Avila et al. (2013).

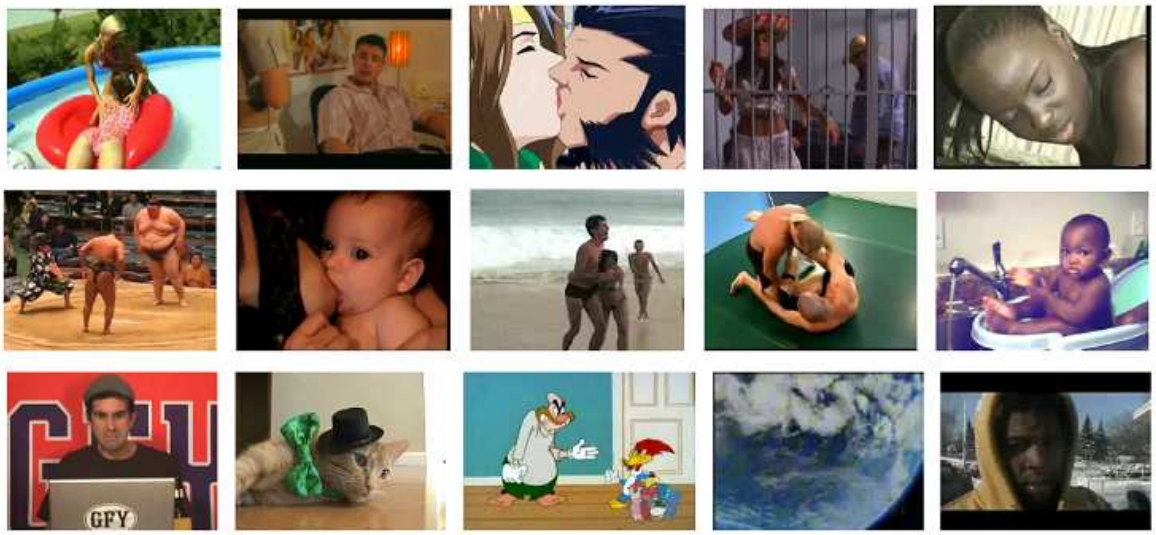
Apesar de possuir uma enorme quantidade de vídeos analisados, a base é formada por uma quantidade não tão grande de imagens que representam os referidos vídeos, o que acaba não sendo proveitoso para o uso em redes neurais profundas. Além do mais, as imagens ditas como representativas muitas vezes não refletem o real conteúdo do vídeo. Na Figura 6.1, mostram-se em suas linhas imagens de vídeos pornográficos, não pornográficos difíceis e não pornográficos fáceis, respectivamente. Conforme pode ser observado, os exemplos da primeira linha foram retirados de vídeos pornográficos, contudo, estão fora do contexto e não apresentam esse tipo de imagem.

O trabalho de Wang, Jin e Tan (2016) corrobora com essa assertiva. Em seu estudo, a referida base de dados foi utilizada e 1.198 das 6.387 imagens ditas pornográficas foram removidas (cerca de 19%) após uma análise não automática, por serem consideradas mal classificadas.

6.2.3 NPDI Pornography-2K

A *NPDI Pornography-2K* (MOREIRA et al., 2016) é uma versão estendida da *NPDI Pornography-800* (AVILA et al., 2013). Essa nova versão utilizou aproximadamente 140 horas de vídeos, dentre esses, mil pornográficos e mil não pornográficos. Os vídeos não pornográficos foram adquiridos de maneira similar à versão anterior (AVILA et al., 2013), balanceando os exemplos considerados “fáceis” e “difíceis”. Entretanto houve diferença quanto a extração dos vídeos considerados pornográficos, não se restringindo apenas aos sites especializados. Também foram exploradas redes sociais de vídeo de propósito geral,

Figura 6.1: Imagens representativas dos vídeos de cada uma das categorias da NPDI Pornography-800. A primeira linha retrata as imagens oriundas dos vídeos pornográficos. As demais linhas ilustram as imagens provenientes dos vídeos não pornográficos, sendo a segunda linha dos exemplos "difíceis" e a terceira linha dos exemplos "fáceis".



Fonte: Extraída de Avila et al. (2013).

em que facilmente foram encontrados vídeos com esse conteúdo. Portanto, o conteúdo pornográfico dessa nova versão acaba sendo mais variado, possuindo conteúdo profissional e amador, além de vários gêneros de pornografia e de diferentes etnias.

Apesar das melhorias dessa nova versão, a base de dados acaba recaindo nos mesmos problemas da sua antecessora (AVILA et al., 2013), não possuindo uma grande quantidade de imagens e que, por muitas vezes, acabam não representando de maneira fiel o rótulo que recebe, como pode ser visualizado na Figura 6.2.

6.3 *Pornographic and Explicity Database - PEDA 376K*

Percebendo-se a dificuldade na obtenção de uma base de dados pornográfica que atendesse os requisitos necessários para um bom funcionamento com redes neurais profundas, foi proposta a criação de uma base de dados estruturada, confiável e detentora de uma considerável quantidade de imagens para que o modelo baseado na arquitetura em questão conseguisse diferenciar imagens pornográficas de imagens não pornográficas.

Figura 6.2: Imagens representativas dos vídeos de cada uma das categorias da NPDI Pornography-2K. A primeira linha retrata as imagens oriundas dos vídeos pornográficos. As demais linhas ilustram as imagens provenientes dos vídeos não pornográficos, sendo a segunda linha dos exemplos "difíceis" e a terceira linha dos exemplos "fáceis".

Fonte: Extraída de Moreira et al. (2016).

6.3.1 Desenvolvimento da Base de Dados

O desenvolvimento de uma base de dados com os requisitos expostos não é uma tarefa simples, principalmente no que diz respeito à grande quantidade de imagens desejada. Visando dirimir essa dificuldade, percebeu-se que alguns trabalhos utilizaram técnicas de cópia automatizada de imagens (*scraping*) de sítios da Internet, principalmente redes sociais capazes de separar seu conteúdo por tópicos (CHASE; HE; HEGAZY, 2017; MAHADEOKAR; PESAVENTO, 2016).

Um desses sítios é o Reddit[1]. Trata-se de uma agregação de notícias sociais on-line e um fórum na Internet, contando com mais de 540 milhões de visitantes por mês, 70 milhões de envios e 700 milhões de comentários. Essa rede social acaba desempenhando bem esse papel de separação de conteúdo, pois possui tópicos individuais, denominados *subreddits*, que são frequentemente dedicados a um tema específico, como comida, espaço e até mesmo pornografia (CHASE; HE; HEGAZY, 2017).

Da Obtenção das Imagens

Munido da estratégia e das informações de como proceder no desenvolvimento ágil de uma grande base de dados de imagem (CHASE; HE; HEGAZY, 2017; MAHADEOKAR; PESAVENTO, 2016), buscou-se o maior numero de tópicos (*subreddits*) da rede social *Reddit*,

[1] https://www.reddit.com/.

tanto aqueles relacionados à pornografia (IRANIANGENIUS, 2018; MIKESIZZ, 2018a), quanto temas rotineiros sem relação alguma com o meio pornográfico (MIKESIZZ, 2018b).

Procurou-se diversificar ao máximo o conteúdo de ambas as classes. Com relação aos tópicos pornográficos, foram utilizados tópicos abrangendo diversas modalidades. A seguir são expostas as modalidades abarcadas, assim como alguns exemplos de tópicos relacionados.

- Ângulo: *selfies* (femalepov, normalnudes, Nude_Selfie, ratemynudebody, realsexyselfies), traseiro (lovetowatchyouleave);
- Biotipo: atlético (athleticgirls, hardbodies, fitgirls, FitNakedGirls), curvo (curvy, gonewildcurvy), gestante (preggoporn), gordo (BBW, chubby, GoneWildplus), semelhante à adolescente (fauxbait, twinks), tatuado (Hotchickswithtattoos);
- Cenário: interno (ChangingRooms), externo (gwpublic, publicflashing, publicplug, realpublicnudity), natureza (NakedAdventures, NotSafeForNature);
- Etnia: asiática (AsianAmericanPorn, asiangirlswhitecocks, AsianHotties, AsianNSFW, asianporn, AsiansGoneWild, bustyasians, funsizedasian, juicyasians, nextdoorasians, realasians, virtualgeisha), branca (blonde, ginger, palegirls, redheads, snowwhites), indiana (IndianBabes, indiansgonewild), coreana (NSFW_Korea), filipina (phgonewild,) japonesa (japanpornstars, NSFW_Japan), latina (latinacuties, latinas, latinasgw), negra (Afrodisiac, darkangels, ebony, gonewildcolor, WomenOfColor), várias (blackchickswhitedicksm, damngoodinterracial);
- Gênero: Homens (Beardsandboners, BHMGoneWild, DadsGoneWild, GuysFromBehind, ladybonersgw), mulheres (bikinibridge, bodyperfection, christiangirls, GirlsHumpingThings, girlsinschooluniforms, girlsinyogapants, girlskissing, girlswithglasses, GirlswithNeonHair, justhotwomen, shorthairchicks), transexuais (gonewildtrans, sissies, tgifs, Tgirls, traps);
- Idade: Jovens (18_19, gonewild18, just18, legalteens, legalteensXXX, missalice_18, PetiteGoneWild), adultos (GoneWild), maduros (AgedBeauty, gonewild30plus, milf, realmoms), idosos;

- Modalidade sexual: anal (anal, analgw, masterofanal, painal), oral (blowjobs, blowjobsandwich, deepthroat, distension, facesitting, oralcreampie, throatbarrier)

- Número de indivíduos: dupla (gonewildcouples, GWCouples, mmgirls, twingirls), trio (blowjobsandwich), grupo (funwithfriends, gangbang, groupofnudegirls);

- Objetos fálicos: canetas (buttsharpies), dildos (baddragon, buttplug, suctiondildos), diversos (insertions);

- Órgãos específicos: ânus/nádegas (alteredbuttholes, ass, asshole, AssholeBehindThong, assholegonewild, assinthong, asstastic, bigasses, booty, BubbleButts, buttsandbarefeet, celebritybutts, cosplaybutts, cutelittlebutts, facedownassup, frogbutt, frostedbholes, HungryButts, paag, pawg), pênis (cock, foreskin, hugedicktinychick, massivecock, omgbeckylookathiscock, ratemycock, softies, ThickDick, WhiteAndThick), seios (amazingtits, bigareolas, bigboobsgonewild, BigBoobsGW, boltedontits, bonermaterial, boobbounce, boobies, boobs, breastenvy, burstingout, BustyPetite, engorgedveinybreasts, ghostnipples, homegrowntits, hugeboobs, mycleavage, nipples, smallboobs, thehangingboobs, tinytits, titfuck, torpedotits, voluptuous), testículos (balls), vaginas (celebritypussy, creampies, godpussy, hairypussy, LabiaGW, lipsthatgrip, moundofvenus, pelfie, pussy, pussymound, rearpussy, simps).

- Qualidade: amador (amateur, amateurcumsluts, amateurgirlsbigcocks, CollegeAmateurs, nsfw_amateurs, realgirls), profissional (AidraFox_XXX, AlexisTexas, Anjelica_Ebbi, AvaAddams, c0rtanablue, dillion_harper, emilybloom, evalovia, GiannaMichaels, gillianbarnes, Holly_Peers, JadaStevens, JaydenJaymes, KatyaClover, Kawaiikitten, keriberry_420, KimmyGranger, KyliePage, lanarhoades, LiaraRoux, LiaraRoux, lucypinder, miakhalifa, miamalkova, nicoleaniston, remylacroix, rileyreid, sophiedee, tessafowler);

- Sexualidade: Heterossexual (JustStraightSex), Homossexual (AlphaMalePorn, gaybrosgonewild, GayChubs, gayporn, lesbians, Men2Men, TotallyStraight).

No que diz respeito às categorias não pornográficas, foi traçada a mesma estratégia, buscando-se diversificar ao máximo, como se percebe a seguir.

- Animais: cães (Bulldogs), coelhos (Rabbits), gatos (cats, MEOW_IRL), geral (AnimalPorn, mlem), lobos (wolves), papagaios (parrots), pássaros (birdpics), porco-espinho (Hedgehog), ratos (RATS);
- Arte: alternativa (alternativeart), animes (anime), capas de álbuns musicais (AlbumArtPorn), geral (art, artporn);
- Comidas: geral (food, foodporn, shittyfoodporn), dieta vegetariana (PlantBasedDiet, vegetarian);
- Esportes: automobilismo (formula1, nascar), basquete (CollegeBasketball, nba), bodybuilding (bodybuilding, FitAndNatural), críquete (Cricket), futebol (mls, soccer), futebol americano (nfl), geral (sports), levantamento de peso (weightlifting), mma (mma), olimpíadas (olympics), pescaria (Fishing), radicais (AdrenalinePorn), róquei (hockey), rúgbi (rugbyunion), tenis (tennis);
- Lugares: abandonados (AbandonedPorn), arquitetura (architectureporn), astros (astrophotography), cidades (cityporn), cômodos (AmateurRoomPorn, roomporn);
- Memes: geral (AdviceAnimals, dankmemes, funny, me_irl, memes, PrequelMemes, wholesomememes);
- Natureza: céu (skyporn), geral (earthporn, NatureIsFuckingLit), jardins (IndoorGarden, SavageGarden), rural (ruralporn);
- Objetos: armas de fogo (gunporn), cabos de rede (cableporn), capas de álbuns musicais (AlbumArtPorn), charutos (cigars), cigarros (Cigarettes), mapas (Map_Porn, MapPorn), máquinas (MachinePorn);
- Pessoas: cabelos (Hair), cabelos encaracolados (curlyhair), faces (hittableFaces), geral (HumanPorn, RoastMe), homens (ladyboners), imagens antigas (OldSchoolCool), mulheres (gentlemanboners, goddesses, prettygirls, sexyhair), mulheres de biquíni (bikinis);
- Veículos: aviões de guerra (WarplanePorn), bicicletas (bikecommuting), carros (carporn), motos (bikesgonewild), navios de guerra (WarshipPorn), tanques de guerra (TankPorn).

Salienta-se o uso de diversos tópicos considerados difíceis para melhor treinamento do modelo, tais como indivíduos utilizando trajes de banho e praticando esportes, visto que visualmente são semelhantes à pornografia devido à pouca roupa (grande exposição de pele) e contato físico.

Esse alto nível de diversificação tem como objetivo retratar da maneira mais fiel a pornografia nos dias de hoje, pois todas essas variáveis fazem parte da realidade atual. Além do mais, a alta variabilidade de ambientes faz com que o modelo aprenda as características corretas a serem aprendidas, não recaindo no problema da classificação do "cachorro e do lobo", em que ao final, a categorização se dava a partir das características inerentes ao cenário (i.e. grama e neve, respectivamente) e não do animal em si (RIBEIRO; SINGH; GUESTRIN, 2016).

Após o levantamento de todos os tópicos (*subreddits*), foi utilizado o software de código aberto RipMe[2] para a obtenção das imagens de maneira automatizada. Foi necessária a alteração do código fonte para a realização do *download* apenas de imagens, excluindo vídeos e gifs. Na sequência, foi utilizado um *script* (*batch file*) com uma linha de comando para o *download* das imagens de cada tópico.

Da Análise e Seleção das Imagens

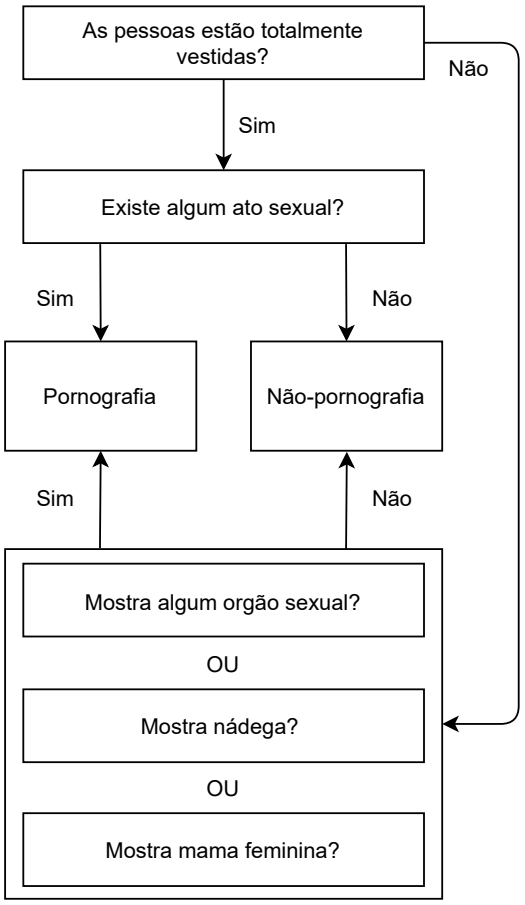
Mesmo se tratando de tópicos relacionados à pornografia, muitas das imagens não se enquadravam como tal pois, por diversas vezes, existiam imagens ilustrando: (i) logotipos, *printscreen* de diálogos ou arquivo não encontrado; (ii) apenas faces ou imagens dos atores em situações normais; ou (iii) fotos iniciais de ensaios fotográficos, cujo modelo aparece em diversas situações trajando vestes completas.

A distinção das imagens referentes aos casos supramencionados é bastante simples, entretanto, determinar o limiar entre uma imagem sensual e uma imagem pornográfica não é uma tarefa trivial, sendo muito difícil julgar essa diferença, pois muitas vezes imagens com conteúdo subjetivo podem ser classificadas de maneira distinta, de acordo com a observação de diferentes indivíduos (PUTRO; ADJI; WINDURATNA, 2015). Portanto, foi utilizado um critério objetivo, adotado na pesquisa de (WANG; JIN; TAN, 2016), em que se classifica como imagem pornográfica aquela que visualmente contém pessoa(s) nua(s), mostrando ex-

[2]https://github.com/RipMeApp/ripme

plicitamente pelo menos uma parte do corpo particular exposta, incluindo mama feminina, nádegas, órgãos sexuais feminino ou masculino; ou em atos sexuais, independentemente das vestimentas. Uma representação gráfica do processo de decisão pode ser visto na Figura 6.3.

Figura 6.3: Fluxograma do processo de tomada de decisão para rotular uma imagem como pornográfica ou não.



Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Dessa forma, cada uma das imagens referentes aos tópicos (*subreddit*) pornográficos foram analisadas e revisadas de maneira não automática (manual) por meio do uso de miniaturas. No caso de dúvida, a imagem era aberta individualmente para uma análise criteriosa. Aquelas que não se enquadravam como pornografia, de acordo com os motivos expostos, foram removidas. Por não se tratarem de pornografia, algumas dessas imagens foram aproveitadas para a categoria não pornográfica.

No que diz respeito aos tópicos não pornográficos, foi realizada uma inspeção manual menos cautelosa, visto a baixa probabilidade de haver imagens pornográficas nesses tópicos. As exceções foram os tópicos não pornográficos que continham pessoas, que receberam o

mesmo cuidado aplicado aos tópicos pornográficos.

Em seguida, foram removidos todos os arquivos duplicados em ambas as categorias, levando em consideração o seu tamanho e código *hash*. Por fim, a base de dados resultante foi constituída por 376.034 imagens, sendo 150.940 pornográficas (40,14%) e 225.094 (59,86%) não pornográficas.

Da Estrutura da Base de Dados

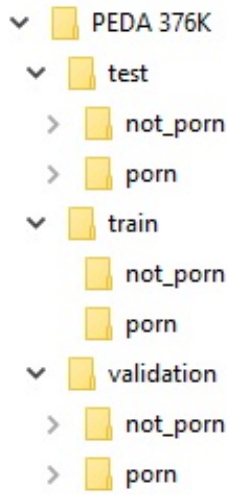
Na etapa de *download* das imagens, foram destinadas duas pastas no diretório raiz da base de dados, uma para as imagens pornográficas e outra para as não pornográficas. Em cada uma das referidas pastas, existiam subpastas referentes a cada um dos tópicos (*subreddit*) da rede social Reddit.

Após concluído o download de todos os arquivos, cada uma das imagens foi renomeada com o seguinte padrão de nomenclatura: [X]_[subreddit]_[número].jpg, em que [X] é igual a P quando a imagem for pornográfica e N quando for não pornográfica; [subreddit] é o nome do respectivo tópico e [número] é a posição da imagem na subpasta referente ao tópico atual. Por exemplo, uma imagem não pornográfica de posição 134 do tópico (*subreddit*) "*food*" receberá a nomenclatura "N_food_134.jpg". Dessa forma, cada imagem é identificada de maneira única, facilitando a visualização da sua categoria e a que tópico (*subreddit*) pertencia, auxiliando uma análise posterior de imagens mal classificadas, a fim de verificar em que tópicos essas imagens estariam contidas.

Depois de renomeadas, as imagens são reorganizadas em apenas dois diretórios temporários: pornográfico e não pornográfico. Por fim e de maneira definitiva, as imagens são distribuídas para as etapas de treinamento, validação e testes em três diretórios principais: (i) *train*, (ii) *validation* e (iii) *test*, respectivamente. Em cada um dos diretórios principais existirão duas subpastas, *porn* (contendo as imagens pornográficas) e *not_porn* (contendo imagens não pornográficas), conforme mostrado na Figura 6.4.

De maneira aleatória, 95% das imagens de cada tipo (pornográfica e não pornográfica) são destinadas para a etapa de treinamento e 2,5% para cada uma das etapa de validação e testes. Salienta-se que, mesmo sendo uma base de dados desbalanceada, as etapas de validação e de testes se comportarão de maneira balanceada, pois a quantidade de arquivos de imagens destinados para essas etapas foi calculada baseado no montante total, como pode

Figura 6.4: Estrutura dos arquivos da base de dados PEDA 376K.



Fonte: Autor.

ser visualizado na Tabela 6.4.

Tabela 6.4: Distribuição numérica e percentual da quantidade de imagens para cada uma das etapas (i.e treinamento, validação e teste).

	Total	Treinamento (95%)	Validação (2,5%)	Teste (2,5%)
Imagens	376.034	338.434	9.400	9.400
Pornográficas	150.940	141.540	4.700	4.700
Não pornográficas	225.094	215.694	4.700	4.700

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

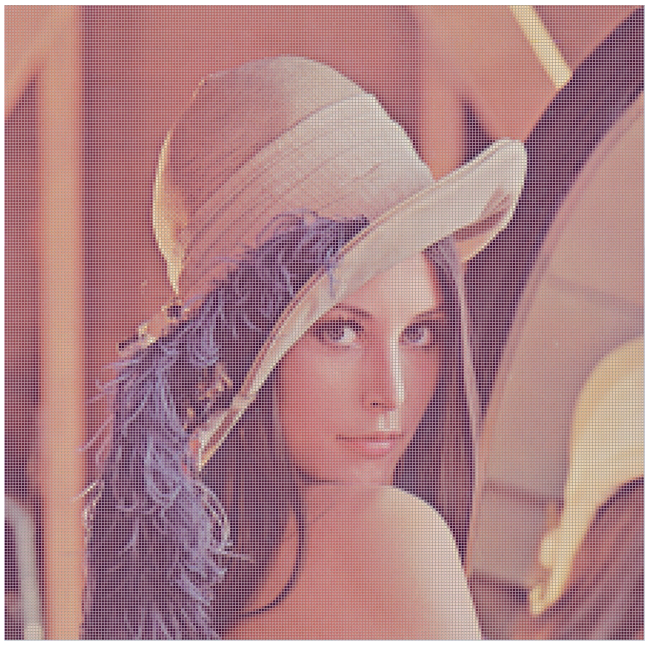
Do Formato das Imagens

Infelizmente, ainda não é viável utilizar imagens em alta resolução como dados de entrada nas redes neurais convolucionais, por questões de desempenho de processamento e memória. Atualmente, a maioria das redes neurais profundas utilizadas para reconhecimento de imagens vem utilizando o tamanho de entrada padrão 224×224 píxeis ou próximo disso. Sendo assim, todas as imagens foram redimensionadas para a referida resolução, conforme pode ser visualizado na Figura 6.5.

6.4 Metodologia

Com o intuito de avaliar o desempenho de uma rede neural profunda treinada/ajustada com a base de dados proposta, foi viabilizada a comparação entre a arquitetura proposta e os

Figura 6.5: Representação de uma imagem contendo 224 píxeis de altura por 224 píxeis de largura.



Fonte: Adaptada de Po (2001).

moderadores de imagens inapropriadas, inseridos no estado da arte, que encontravam-se disponíveis para uso.

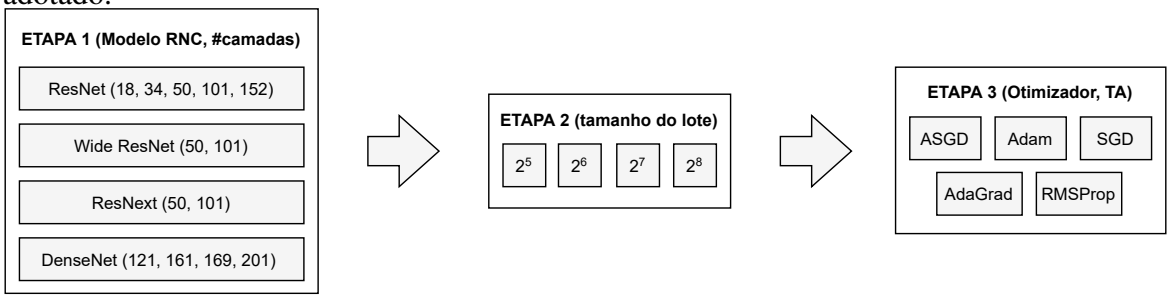
6.4.1 Modelo Proposto Baseado em Redes Neurais Convolucionais

Baseando-se na enorme transição de modelos de detecção de pornografia para abordagens de aprendizado profundo (HUANG; KONG, 2016; ZHOU et al., 2016; WANG; JIN; TAN, 2016; LI et al., 2016; OU et al., 2017; SURINTA; KHAMKET, 2019), foi selecionada uma gama de redes convolucionais para o treinamento do modelo utilizando a base de dados proposta, a PEDA 376K.

Tendo em vista a complexidade do treinamento de redes neurais profundas, aliado ao grande número de combinações para o ajuste fino de seus hiperparâmetros, foi definida uma estratégia que minimiza o tempo despendido nessa tarefa, por meio da não exploração de todo o espaço de busca possível. A metodologia adotada está representada na Figura 6.6 e é composta por três etapas principais, a seleção (i) da rede neural convolucional; (ii) do tamanho do lote (*batch*) e (iii) do otimizador e sua taxa de aprendizado, de maneira gulosa, na referida ordem. Essa estratégia, mesmo não garantindo a melhor configuração possível, via-

biliza a realização do ajuste fino dos hiperparâmetros em modelos baseados em aprendizado profundo, que geralmente utilizam uma grande quantidade de dados na etapa de treinamento.

Figura 6.6: Etapas realizadas para explorar o espaço de busca de hiperparâmetros. A taxa de aprendizado (TA) é dada por $\frac{2^n}{m}$, em que n e m valores são diferentes para cada otimizador adotado.



Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Antes da etapa de treinamento utilizando a base de dados PEDA 376K, todos os modelos foram inicializados com os pesos do pré-treinamento da base de dados ImageNet (DENG et al., 2009). Trata-se de um enorme banco de dados visual, baseado na estrutura de taxonomia WordNet (MILLER, 1998), que foi criado para uso em classificação de imagens e reconhecimento de objetos. A versão original tem mais de 20.000 categorias e 14.000.000 de imagens, mas a versão reduzida (com "apenas" mil categorias e aproximadamente 1.200.000 imagens) se tornou popular em 2010, após a primeira versão do ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (RUSSAKOVSKY et al., 2015b). Atualmente, essa versão reduzida é amplamente utilizada para aprimorar os resultados de modelos baseados em aprendizado profundo por meio da técnica de transferência de aprendizado.

Em cada uma das etapas, os modelos foram treinados durante 30 épocas com parada antecipada (*Early Stopping*). O treinamento era interrompido se não houvesse melhoria na acurácia dos dados de validação por cinco épocas consecutivas. Como configuração inicial, foi utilizado o otimizador AdaGrad com uma taxa de aprendizado de 0,01 e um tamanho de lote (*batch*) igual a 128.

Selecionando a Arquitetura

Com base nos resultados do ILSVRC (RUSSAKOVSKY et al., 2015b), foram selecionadas quatro arquiteturas, cada uma com um número variável de camadas (discriminado entre pa-

rênteses). Todos os modelos são compatíveis com o tamanho da imagem (224×224 píxeis) da base de dados PEDA 376K.

- ResNet (18, 34, 50, 101, 152) (HE et al., 2016): Essa arquitetura foi desenvolvida para solucionar o problema do esvaecimento dos gradientes (*gradient vanishing*) nas camadas mais profundas. A solução se deu pela inclusão de um bloco residual composto por duas camadas convolucionais conectadas também por uma conexão de salto. Essa rede neural convolucional foi desenvolvida em 2015 e venceu o desafio ILSVRC no mesmo ano;

- Wide ResNet (50, 101) (ZAGORUYKO; KOMODAKIS, 2016): Trata-se de uma variação da ResNet, desenvolvida em 2017. A duplicação do número de canais em cada bloco aprimorou a rede neural convolucional predecessora, minimizando o gargalo ocorrido anteriormente pela escassez de canais;

- ResNext (50, 101) (XIE et al., 2016): Essa abordagem aplica uma estratégia de divisão-transformação-agregação a já renomada arquitetura ResNet. Essa nova técnica divide o caminho da camada convolucional em uma determinada cardinalidade C, gerando novos caminhos com a mesma estrutura que são somados no final. Os autores garantiram o segundo lugar na ILSVRC 2016;

- DenseNet (121, 161, 169, 201) (HUANG et al., 2017): Essa arquitetura propôs a utilização de um conjunto de camadas denominadas "blocos densos", em que cada camada obtêm entradas adicionais oriundas das camadas predecessoras. Essa proposição resulta em $\frac{L(L+1)}{2}$ conexões entre as camadas, em vez das L conexões como se vê nas redes neurais profundas tradicionais com L camadas. Foi desenvolvida em 2016 e recebeu o prêmio de melhor artigo na CVPR (*Conference on Computer Vision and Pattern Recognition*) 2017.

Selecionando o Tamanho do Lote (*batch*)

Este hiperparâmetro, o tamanho do lote, define o número de amostras a serem utilizadas por vez no treinamento do modelo. O tamanho do lote é geralmente definido baseado em potência de dois (2^n), devido a fatores de desempenho relacionados à arquitetura de computadores,

com o objetivo de obter maior vantagem de processamento (KANDEL; CASTELLI, 2020). Com base nessa premissa, foi explorada uma gama de valores para o tamanho do lote, em que $n = \{5, 6, 7, 8\}$.

Selecionando o Otimizador e a Taxa de Aprendizado

Foram adotados cinco métodos de otimização para a avaliação dos modelos: três métodos adaptativos: AdaGrad, RMSProp e Adam; e dois não adaptativos, Gradiente Estocástico Descendente (*Stochastic Gradient Descent - SGD*) e Gradiente Estocástico Descendente Assíncrono (*Asynchronous Stochastic Gradient Descent - ASGD*). Os métodos adaptativos convergem mais rápido comparado aos não adaptativos e por isso vêm se tornando muito populares para o treinamento de redes neurais profundas. Essa melhoria se dá pelo uso do histórico de iterações na otimização local (WILSON et al., 2017).

Para ajustar a taxa de aprendizado, foi utilizada a função $T(m, n) = \frac{2^n}{m}$ com base na pesquisa de Wilson et al. (2017). Essa abordagem usa um intervalo diferente em uma potência de dois dividido por um número específico para cada otimizador. Os diferentes valores para m e n para cada caso são mostrados na Tabela 6.5. Se porventura a melhor acurácia atingida nos dados de validação se desse por uso de uma taxa de aprendizado em um extremo dos intervalos, essa faixa de valores era expandida até que a referida situação não acontecesse.

Tabela 6.5: Faixa de valores da taxa de aprendizado. Para cada otimizador, um conjunto diferente de valores de taxa de aprendizado foi explorado. Cada valor é definido por $\frac{2^n}{m}$.

Otimizador	m	n
AdaGrad	10	$\{-2, -1, 0, 1, 2\}$
RMSProp	100	$\{-5, -4, -3, -2, -1, 0, 1, 2\}$
Adam	100	$\{-6, -5, -4, -3, -2, -1, 0, 1\}$
SGD	1	$\{-2, -1, 0, 1, 2\}$
ASGD	1	$\{-2, -1, 0, 1, 2\}$

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

6.4.2 Moderadores de Imagens Inapropriadas

A crescente disseminação de mídia na Internet, principalmente por meio das redes sociais, resultou no aumento do número de empresas voltadas para a filtragem de conteúdo. A maioria desses serviços encontra-se hospedado na nuvem e tem como objetivo identificar

automaticamente se uma determinada imagem é adequada ou não, de acordo com seus próprios critérios. Esses serviços de filtragem geralmente não indicam explicitamente se uma imagem é pornográfica ou não. Na maioria das vezes, retornam um conjunto de probabilidades relativas a diversas categorias de imagem (e.g. adulto, explícito, nudez, nudez parcial, picante, sugestivo, trajes de banho).

Nesse contexto, ao utilizar esse tipo de serviço, a decisão de rotular uma imagem como pornográfica ou não é atribuída ao usuário. Ademais, a não padronização das saídas desses serviços impossibilita realizar uma comparação direta e justa entre seus resultados. Com o intuito de dirimir a não padronização inerente a esses serviços, foi incorporada uma árvore de decisão no topo de cada serviço, transformando suas saídas em decisões binárias, inferindo se a imagem trata-se de pornografia ou não. Os experimentos foram conduzidos utilizando dois conjuntos de dados diferentes, PEDA 376K e RedLight (ALVAREZ, 2012) para comparar todos os serviços. Foram considerados cinco serviços diferentes na etapa experimental.

- **Amazon Rekognition**[3]**:** Cada predição retorna vinte e dois valores de probabilidade para uma única imagem de entrada. Há um total de dezoito rótulos divididos em quatro categorias. Para cada categoria, a predição inclui uma probabilidade para cada rótulo e o valor máximo desses rótulos como sendo a probabilidade da categoria em si. Foram descartadas as duas categorias que não apresentavam relação com pornografia: "Violência" e "Visualmente perturbador". Sendo assim, foram utilizadas: "Nudez explícita" e "Sugestiva". Por fim, foram utilizados apenas dez rótulos, seis relativos à primeira categoria (nudez, nudez masculina, nudez feminina, atividade sexual, nudez ilustrada ou atividade sexual e brinquedos para adultos) e quatro relativas à segunda categoria (roupa de banho feminina ou roupa íntima, roupa de banho masculina ou cueca, nudez parcial, roupas transparente), para um total de doze probabilidades.

- **Clarifai**[4]**:** Esse serviço possui dois módulos de filtragem de conteúdo: "NSFW" e "Moderação". O primeiro módulo retorna a probabilidade de uma imagem de entrada possuir conteúdo pornográfico. O segundo módulo retorna cinco probabilidades relacionadas à filtragem de moderação: "Seguro", "Explícito", "Sugestivo", "Violento" e

[3]https://aws.amazon.com/pt/rekognition
[4]https://www.clarifai.com/models/not-safe-for-work-image-recognition

"Drogas". Por não apresentar relação com pornografia, as duas últimas categorias foram descartadas. No total, foram considerados apenas quatro valores de probabilidade.

- **Google Vision**[5]**:** Fazendo uso de uma abordagem diferente das demais, o serviço fornecido pelo Google retorna valores qualitativos (categorias) em vez de quantitativos (probabilidades). Os valores possíveis são: muito improvável, improvável, possível, provável e muito provável, que são codificados como valor numérico 1, 2, 3, 4 e 5, respectivamente. Esse serviço oferece cinco categorias, das quais usamos apenas duas: "Adulto" e "Picante". As outras três (Simulação, Médico e Violência) não eram relacionadas à pornografia.

- **Microsoft Azure**[6]**:** É especificamente voltado apenas para a moderação de conteúdo pornográfico. Para uma imagem de entrada, o serviço retorna dois valores de probabilidade: "Adulto" e "Picante". Dentre os serviços analisados, é o único que também retorna um valor booleano para cada um dos rótulos, indicando explicitamente se a imagem possui conteúdo pornográfico ou picante.

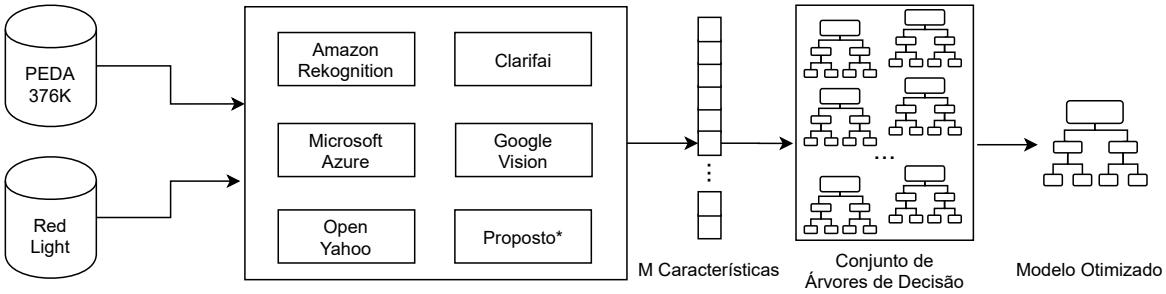
- **Yahoo! NSFW (MAHADEOKAR; PESAVENTO, 2016):** Trata-se de uma rede neural profunda, disponibilizada de maneira gratuita, treinada para a detecção de imagens inapropriadas. O modelo retorna um valor de probabilidade contínuo que indica, baseado em um limiar a ser escolhido, se uma determinada imagem de entrada é pornográfica ou não.

Padronização dos Moderadores de Imagens Inapropriadas

Foi proposta a padronização dos resultados gerados pelos diferentes serviços de moderação de imagens, com o objetivo de viabilizar a comparação entre os referidos serviços e a arquitetura proposta. Por meio do uso de árvores de decisão, os vetores de probabilidades resultantes dos serviços de moderação de imagens foram padronizados, sendo transformados em um decisão binária: (i) pornografia ou (ii) não pornografia. Os experimentos foram realizados usando dois conjuntos de dados diferentes, conforme mostrado na Figura 6.7.

[5]https://cloud.google.com/vision/
[6]https://azure.microsoft.com/en-in/services/cognitive-services/content-moderator/

Figura 6.7: O diagrama ilustra o uso de árvores de decisão para transformar as saídas dos serviços de moderação de imagens em decisões binárias.



*A abordagem proposta é otimizada apenas sobre o conjunto de dados RedLight.
Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

PEDA 376K

Inicialmente, foi definido um modelo padrão servindo como *baseline*. Utilizou-se o conjunto de validação da base de dados PEDA 376K para obter as saídas dos serviços de moderação de imagens ($h(x)$), normalizando os valores para o intervalo [0, 1] quando necessário. Nesse modelo, nenhuma árvore de decisão foi utilizada para a classificação das imagens. Essa inferência se deu por meio do uso de um simples limiar. Para cada um dos serviços, a probabilidade da característica que melhor representava a pornografia foi selecionada. Se o referido valor fosse maior que 0,5, a imagem era considerada pornográfica, caso contrário, como não pornográfica. Essa regra não foi aplicada ao serviço oferecido pela Microsoft Azure, pois esse é o único que já fornecia essa saída binária.

Os modelos otimizados se basearam em árvores de decisão treinadas com os dados resultantes de cada serviço de moderação de imagens. Esses resultados foram obtidos por meio da submissão das imagens contidas na base de dados PEDA 376K. Isso fez com que esses modelos se adequassem à definição objetiva de pornografia adotada por essa base de dados, dando maior isonomia à analise experimental, pois por se tratar de um aspecto que carrega subjetividade, cada serviço pode apresentar diferentes definições para pornografia.

Foram treinadas $M + \left\lceil \frac{M-1}{M} \right\rceil$ árvores de decisão para cada um dos serviços, em que M é o tamanho do seu vetor de probabilidades resultante. Ou seja, uma árvore de decisão binária foi treinada com cada uma das probabilidades resultantes individualmente e uma árvore final foi treinada utilizando o vetor de probabilidades por completo. Os experimentos foram conduzidos com validação cruzada de cinco vezes e, para evitar o su-

perajustamento (*overfitting*), o número mínimo de amostras necessário para dividir um nó interno foi variado usando uma faixa exponencial. Esse intervalo foi dado por 2^n, em que $n = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Para cada serviço e suas respectivas árvores de decisão, o conjunto de testes foi usado para avaliar o desempenho. É importante salientar que os experimentos foram executados por duas vezes, alternando os conjuntos de validação e teste a fim de obter resultados mais confiáveis.

RedLight

Visando à exclusão de qualquer possibilidade de resultados tendenciosos introduzidos pelo uso da base de dados proposta (PEDA 376K), foi realizado um segundo conjunto de experimentos utilizando uma base de dados externa. A base de dados RedLight (ALVAREZ, 2012) contém um conjunto de imagens pornográficas e não pornográficas que são divididas em diversas subcategorias. Entretanto essas informações relativas ao subtipo de cada imagem foi desconsiderada, sendo apenas utilizados os rótulos de pornografia e não pornografia. Ademais, todas as imagens ilegíveis ou duplicadas foram removidas, resultando em um conjunto de 25.616 imagens (10.223 pornográficas e 15.393 não pornográficas). A base de dados foi dividida em seis partições para realizar combinações distintas de conjuntos de treinamento (5 partições) e teste (1 partição). A acurácia foi calculada para todos os experimentos. De forma semelhante aplicada à base de dados PEDA 376K, para a RedLight, coletamos os resultados dos modelos *baseline* e os modelos otimizados, treinados com árvores de decisão. Nesse caso, também foram treinadas árvores de decisão com base na rede neural convolucional proposta. Essa etapa não foi necessária quando utilizada a base de dados PEDA 376K, visto que a referida rede foi treinada com os mesmos dados.

Os recursos computacionais utilizados para o treinamento dos modelos propostos nesta seção de metodologia incluiu: uma estação de trabalho equipada com 64GB de memória ram, dois processadores Intel Xeon E5 2,10 GHz e quatro unidades de processamento gráfico Pascal Titan X 12GB. Toda a análise experimental, assim como seus resultados, encontram-se discriminados detalhadamente no Capítulo 9 (Análise Experimental e Resultados), especificamente no item “9.3.2 Detecção de Pornografia Adulta e Análise Comparativa com Serviços de Moderação de Conteúdo”.

6.5 Considerações Finais

Foi discutida nesse capítulo a necessidade de uma base de dados de imagens pornográficas com características específicas voltadas para seu uso em redes neurais convolucionais. Primeiramente, discorreu-se sobre a importância das referidas especificidades da base de dados para esse tipo de rede neural e da inexistência/indisponibilidade no meio acadêmico.

Em seguida, foi realizada uma revisão bibliográfica das principais bases de dados pornográficas disponíveis na literatura, discriminando seus aspectos positivos e negativos e, em seguida, foi apresentada a PEDA 376K, uma nova base de dados para auxiliar o desenvolvimento de pesquisas e práticas na tarefa de detecção automática de imagens pornográficas. Essa base de dados é composta por mais de 376K imagens e fornece divisões pré-determinadas de treinamento, validação e testes, permitindo uma futura reprodutibilidade experimental. Além disso, as imagens foram cuidadosamente rotuladas utilizando uma definição objetiva de pornografia, na tentativa de minimizar a subjetividade existente nesse ambiente.

Munido dessa nova base de dados pornográfica, foi proposta uma configuração de hiperparâmetros e arquitetura de redes neurais convolucionais com o objetivo de compor o Módulo Pornográfico exposto no Capítulo 5 (Arquitetura Proposta para Detecção de Pornografia Infanto-Juvenil). Também foi proposta uma abordagem de aprendizado de máquina utilizando árvores de decisão para comparar o modelo proposto com serviços de moderação de imagens existentes, em diferentes cenários e bases de dados.

Capítulo 7

Aprimorando a Estimativa de Idade Real a Partir de Dados de Idade Aparente

Neste capítulo, serão apresentadas novas técnicas para a realização da estimativa de idade real em faces. As abordagens baseiam-se primordialmente em dados de idade real para o treinamento do modelo, entretanto, o uso de informações de idade aparente foram propostas para o aprimoramento do resultado final da estimativa de idade real.

7.1 Introdução

A análise de imagens faciais, por meio da Visão Computacional, tem atraído cada vez mais a atenção da comunidade acadêmica, assim como da indústria. A disponibilidade de grandes conjuntos de dados, equipamentos computacionais cada vez mais poderosos e novos métodos de aprendizado profundo tem alavancado essa área de pesquisa. Atualmente, a análise facial vem ampliando seu leque de atuações, como idade, gênero, etnia, emoção e expressão (CARLETTI et al., 2019). As aplicações mais recentes de análise facial vão muito além das tarefas tradicionais de detecção e reconhecimento no domínio da segurança e privacidade. Hoje, por meio dessas novas áreas de atuação, é possível atuar em domínios até então não explorados pela análise facial, como gerenciamento de clientes e análise demográfica (XIA et al., 2020).

No campo da análise facial, o problema da estimativa da idade tem sido historicamente um dos mais desafiadores (AGUSTSSON et al., 2017). A abordagem tradicional para esse problema foca na estimativa da idade real, ou seja, dada uma imagem de face humana, o ob-

jetivo é estimar a idade cronológica do indivíduo em questão. Essa estimativa de idade real é particularmente difícil devido ao processo de envelhecimento não uniforme em humanos, que depende de vários fatores como genética, padrões de ingestão de alimentos, práticas de atividades esportivas, incidência solar, bem-estar mental, entre outros (RONDEAU; ALVAREZ, 2018). Abordagens recentes também exploraram a estimativa da idade aparente, cujo objetivo é estimar quão velho determinado indivíduo aparenta ser. Conjuntos de dados específicos disponíveis publicamente, como APPA-REAL (AGUSTSSON et al., 2017), permitem que métodos de aprendizado de máquina aprendam a partir de imagens faciais rotuladas com idades reais e aparentes. Em geral, os rótulos de idade aparente são calculados como a média de um conjunto de "suposições humanas". Estimar a idade aparente é uma tarefa menos difícil, uma vez que as imagens são rotuladas baseadas na aparência dos indivíduos em vez da sua idade cronológica (CLAPES et al., 2018).

Sendo assim, propõe-se aproveitar os dados de idade aparente para melhorar a estimativa de idade real. Foram lançadas várias contribuições para treinar redes neurais convolucionais a fim de minimizar o erro da estimativa de idade real. No geral, as contribuições desse capítulo, em específico, incluem:

- Uma função de perda que penaliza o modelo com base nas probabilidades das classes incorretas, em oposição à função de perda de entropia cruzada, que penaliza o modelo baseado apenas na probabilidade da classe correta. Dessa forma, é possível aplicar um "classificador justo" que potencializa função de perda de acordo com a diferença entre a predição e o valor verdadeiro;

- Um novo método capaz de gerar uma distribuição gaussiana personalizada para cada idade real (Gaussiana Dinâmica), por meio do uso de um regressor, baseado em informações de idade aparente e;

- Uma nova abordagem que combina duas funções de perda, de dados de idade real e aparente, para melhorar a tarefa alvo de treinar um estimador de idade real.

7.2 Bases de Dados

Nesta seção, são descritas as bases de dados faciais IMDB-WIKI (ROTHE; TIMOFTE; GOOL, 2015) e APPA-REAL (AGUSTSSON et al., 2017), utilizadas para a realização do estudo relativo à estimativa de idade real por meio de faces.

IMDB-WIKI

Essa base de dados foi lançada em 2015 com o objetivo preencher uma lacuna existente na área de pesquisa de estimativa de idade. Até então, era muito difícil ter acesso a uma grande quantidade de imagens faciais rotuladas com idades cronológicas. Todas as imagens e respectivas datas de nascimento foram extraídas automaticamente das páginas *Internet Movie Database (IMDB)*[1] e *Wikipedia*[2], totalizando 523.051 imagens contendo faces de 20.284 diferentes indivíduos. Devido à rotulagem automatizada e do enviesamento das idades, dado que artistas geralmente apresentam aspecto jovial, não representando a população geral como um todo, a referida base de dados não é considerada confiável para a estimativa da idade real, no entanto, é amplamente utilizada para a aplicação da técnica de transferência de aprendizado.

APPA-REAL

Essa base de dados foi lançada em 2017, sendo a primeira a possuir faces rotuladas com idades reais e aparentes. Os dados foram coletados por meio da plataforma AgeGuess[3], que é um site baseado em gamificação que permite aos usuários: (i) enviar imagens faciais rotuladas com idade real e/ou (ii) adivinhar a idade de faces enviadas por outros usuários. De posse desses dados, os autores puderam construir um conjunto de dados de faces contendo 7.591 imagens com rótulos de idade real e aparente associados, divididos em três subconjuntos de tamanhos diferentes: (i) treinamento (4.113 imagens), (ii) validação (1.500 imagens) e (iii) teste (1.978 imagens). A base de dados é muito confiável devido ao grande número de suposições de idade (quase 38 por imagem) e pela variação das condições de iluminação, poses faciais, etnia, expressão e qualidade da imagem. Dessa forma, o ambiente não controlado

[1] http://www.imdb.com
[2] https://www.wikipedia.org/
[3] https://www.ageguess.org/

permite que o conjunto de dados seja uma boa aproximação do que acontece no mundo real.

7.3 Métodos

Nesta seção, são descritas sete diferentes abordagens utilizadas na análise experimental, sendo três já consolidadas na literatura e quatro novas abordagens propostas. Todos os métodos são baseados em redes neurais convolucionais, entretanto, apresentam particularidades inerentes a cada abordagem. Além disso, os métodos apresentados a seguir podem ser agrupados em duas categorias principais, métodos padrão e métodos aprimorados, como pode ser visualizado na Tabela 7.1. Ambas as categorias propõem a predição da idade real de humanos baseados em faces, considerando uma faixa discreta de idades de tamanho n igual a 101, que varia de 0 até 100.

Tabela 7.1: Métodos utilizados para estimativa de idade real.

Método	Tipo	Proposto?
Classificador Multiclasse	padrão	não
Regressor	padrão	não
Classificador Justo	padrão	sim
Gaussiana Estática	aprimorado	não
Gaussiana Dinâmica	aprimorado	sim
COURA	aprimorado	sim
DCOURA	aprimorado	sim

Fonte: Autor.

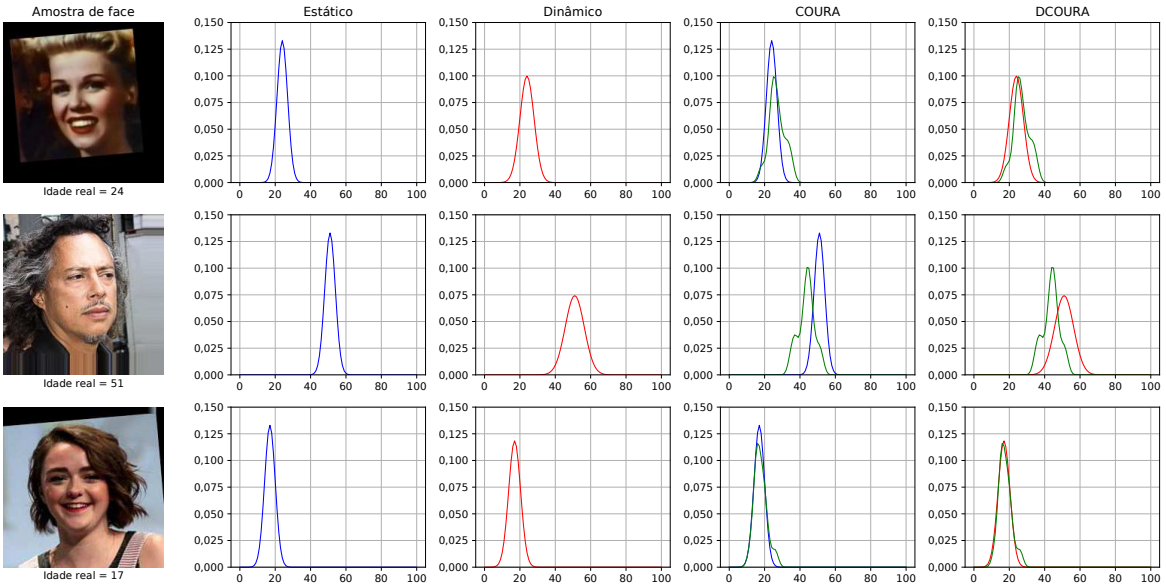
1. *métodos padrão*, que modelam a estimativa de idade real como um mapeamento $\mathcal{X} \rightarrow \mathcal{Y}$. Os dados de treinamento são dados por m pares $(\mathbf{x_i}, \mathbf{y_i})$ em que $\mathbf{x_i} \in \mathcal{X}$ e $\mathbf{y_i} \in \mathcal{Y}$. Os dados de entrada $\mathcal{X}$ são compostos por tensores (imagens de faces), e a saída unidimensional $\mathcal{Y}$ pode ser contínua ou discreta, dependendo se um regressor ou um classificador foi utilizado, respectivamente. Sendo assim, os métodos dessa categoria não usam dados de idade aparente para melhorar a estimativa da idade real, representada por $\hat{y}_i$.

2. *métodos aprimorados*, que são semelhantes aos métodos padrão, mas cada instância $\mathbf{y_i} \in \mathcal{Y}$ é uma distribuição discreta de probabilidade, ou seja, um vetor em que cada elemento representa a probabilidade da face pertencer a idade em questão. A

Figura 7.1 mostra alguns exemplos desses pares $(\mathbf{x_i}, \mathbf{y_i})$. Os métodos nessa categoria requerem uma função de perda que possa calcular a distância entre a verdadeira distribuição de probabilidade $\mathbf{y}$ e a saída $\hat{\mathbf{y}}$ da rede neural convolucional. A função de perda baseada na Divergência de Kullback-leibler atende a esse requisito, sendo amplamente utilizada nesse contexto (GAO et al., 2017; RONDEAU; ALVAREZ, 2018) e pode ser calculada de acordo com a Equação (7.1):

$$L_{KL}(y, \hat{y}) = \frac{1}{m} \sum_{i=0}^{m} \sum_{j=0}^{n} y_i^{(j)} \cdot \left(\ln \frac{y_i^{(j)}}{\hat{y}_i^{(j)}} \right) \tag{7.1}$$

Figura 7.1: A primeira coluna expõe amostras de faces da base de dados APPA-REAL. As demais colunas mostram as distribuições discretas de probabilidade referentes a cada método. As distribuições gaussianas estáticas e dinâmicas são representadas pelas curvas azul e vermelha, respectivamente. A curva verde ilustra a distribuição de probabilidade baseada na estimativa de densidade por kernel.



Fonte: Autor.

7.3.1 Classificador Multiclasse

Um Classificador Multiclasse realiza a predição das idades reais por meio da discretização da faixa de idade em n resoluções de tamanho w (geralmente $w = 1$), resultando em um classificador de n-classes. Para esse método, foi utilizada a entropia cruzada (*cross-entropy*) como função de perda (COX, 1958), como pode ser visto na Equação (7.2).

$$L_{EC}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=0}^{n} -y_i^{(j)} \cdot \log \hat{y}_i^{(j)} \tag{7.2}$$

7.3.2 Regressor

A estimativa de idade real pode ser apresentada naturalmente como um problema de regressão com uma única saída contínua que pode ser arrendondada para a idade inteira mais próxima. Nessa abordagem direta, foi utilizada uma função de perda baseada no erro quadrático médio (*mean squared error*), conforme mostrado na Equação (7.3).

$$L_{EQM}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{7.3}$$

7.3.3 Classificador Justo

Com o intuito de minimizar uma limitação importante do Classificador Multiclasse na estimativa de idade, foi apresentada uma nova função de perda. O Classificador Multiclasse não considera qualquer relação de distância entre as classes. Sendo assim, foi proposta uma função de custo, discriminada na Equação (7.4), que penaliza o modelo baseado nas probabilidades inferidas às classes incorretas, o oposto ao aplicado na entropia cruzada, que penaliza o modelo considerando apenas a probabilidade inferida à classe correta. Ademais, o erro é potencializado de acordo com a distância entre a classe predita e a verdadeira classe.

$$L_{JUSTO}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| \sum_{j=0}^{n} (1 - y_i^{(j)}) \exp(\hat{y}_i^{(j)}) \tag{7.4}$$

7.3.4 Gaussiana Estática

O principal objetivo de aprender a prever distribuições de probabilidades, em vez das classes propriamente ditas, é minimizar as classificações equivocadas devido à relação existente entre as classes vizinhas, no caso, a idade real. Dessa forma, para aplicar a distribuição de probabilidade como rótulo no problema da estimativa de idade real é preciso parametrizar uma distribuição gaussiana para cada imagem do conjunto de treinamento. Essa distribuição tem como média a idade real e um desvio-padrão fixo indicado por um parâmetro σ (geralmente $\sigma = 2$ ou $\sigma = 3$) (RONDEAU; ALVAREZ, 2018; ROTHE; TIMOFTE; GOOL, 2015).

Uma vez que esses rótulos baseados em distribuição de probabilidade são gerados, uma rede neural convolucional pode ser treinada utilizando como função de perda a Divergência de Kullback-leibler (KULLBACK; LEIBLER, 1951).

Visto que o modelo é treinado utilizando distribuições de probabilidades estáticas como rótulos, sua camada final é composta por N probabilidades referentes à faixa discreta de idades. A predição da idade não é dada pela categoria com maior probabilidade, como acontece usualmente, mas sim mediante o somatório dos produtos entre probabilidades e as referidas classes (idades), como pode ser observado na Equação (7.5)

$$E(O) = \sum_{i=0}^{100} y_i o_i \tag{7.5}$$

em que $O = \{o_0, o_1, \ldots, o_{100}\}$ é a saída de 101 dimensões da rede, representando probabilidades da função *Softmax*, e y_i são os anos discretos correspondentes a cada classe $i \in [0, 100]$.

7.3.5 Gaussiana Dinâmica

A aplicação de um valor único de desvio-padrão σ para a formação de uma distribuição gaussiana para todas as idades vai de encontro com a não padronização do processo de envelhecimento humano (RONDEAU; ALVAREZ, 2018). Baseado nessa não uniformidade, foram criadas distribuições gaussianas específicas para cada idade. Um regressor foi ajustado para aprender a relação de cada idade real j (variável independente) e seu desvio-padrão correspondente $\sigma^{(j)}$ usando todas as suas estimativas de idades aparentes. Dessa forma, as distribuições gaussianas foram construídas, de maneira personalizada, utilizando como desvio-padrão um valor inerente a cada idade. O objetivo de usar um regressor é reter essa relação aprendida para previsões futuras, nas quais as estimativas humanas por imagem podem não existir. Como na Gaussiana Estática, com rótulos baseados em distribuição de probabilidade, pode ser utilizada a Divergência de Kullback-leibler como função de perda no treinamento da rede neural convolucional.

7.3.6 COURA and DCOURA

A combinação de funções de perda é uma prática comum para obter melhores resultados na estimativa da idade aparente e real (PAN et al., 2018; ZHANG et al., 2019; LIU et al., 2020). Até onde se sabe, não existem tentativas na literatura de fundir funções de perda de dados de idade real e aparente para melhorar a estimativa de idade real. Dessa forma, foi proposto o uso de duas funções de perda distintas, uma de cada tipo de estimativa de idade, sendo ambas com rótulos baseados em distribuição de probabilidade. Por fim, a função de perda final é dada pela soma ponderada das duas funções parciais de perda L_1 e L_2, por meio do uso do hiperparâmetro ajustável λ, conforme mostrado na Equação (7.6).

$$L = \lambda \cdot L_1 + (1 - \lambda) \cdot L_2 \tag{7.6}$$

O primeiro método, denominado COURA, funde duas funções de perda baseadas na Divergência de Kullback-leibler. Foram utilizadas as funções de perda adotadas no (i) método proposto da Gaussiana Estática e na (ii) metodologia proposta por Rondeau e Alvarez (2018), que alcançou o estado da arte na estimativa de idade aparente. Nesse método de estimativa de idade aparente, para cada imagem do conjunto de treinamento, uma estimativa de densidade por kernel é formada de acordo com um conjunto de suposições de idade por humanos, com o objetivo de definir um rótulo baseado em uma distribuição de probabilidade discreta. O método DCOURA diferencia-se do método COURA por utilizar como primeira função de perda a adotada no método proposto Gaussiana Dinâmica, em vez da adotada no método Gaussiana Estática.

7.4 Metodologia

Nesta seção, inicialmente é descrita métrica de avaliação adotada, assim como a justificativa da sua utilização. Em seguida, são detalhados os protocolos utilizados para a realização das etapas de treinamento e avaliação dos modelos adotados.

7.4.1 Métrica de Avaliação

Para estimativa de idade aparente, o erro-ϵ (ϵ-*error*) é uma métrica de avaliação comumente utilizada em competições renomadas como *ChaLearn Looking At People* (ESCALERA et al., 2017). Porém, no contexto atual, a informação da idade aparente é utilizada apenas como forma de melhorar a estimativa da idade real. Portanto, se fez necessário apenas avaliar a estimativa da idade real, utilizando o Erro Médio Absoluto (*Mean Absolute Error - MAE*) (WILLMOTT; MATSUURA, 2005). Essa métrica de avaliação é dada pela média das diferenças absolutas entre cada uma das m idades previstas $\hat{y}_i$ e seu valor real correspondente y_i, conforme mostrado na Equação (7.7):

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{7.7}$$

7.4.2 Protocolo Experimental

Com o objetivo de obter melhores resultados, foi utilizada a técnica de transferência de aprendizado em duas etapas, semelhante ao pré-treinamento realizado em Rondeau e Alvarez (2018). A primeira etapa envolve tirar proveito das representações de características aprendidas com a ImageNet (RUSSAKOVSKY et al., 2015b), uma base de dados de 1,2 milhões de imagens. Os *frameworks* de aprendizado profundo mais modernos, como *PyTorch*[4], possuem redes neurais convolucionais integradas já pré-treinadas com a referida base de dados. A segunda etapa consistiu em ajustar a rede neural convolucional com a base de dados IMDB-WIKI. Nessa etapa, o conjunto de dados foi usado parcialmente porque algumas imagens não apresentavam faces ou tinham mais de uma face detectada na imagem. Depois do descarte dessas imagens, o conjunto de dados da versão filtrada continha 224.573 imagens.

Todas as imagens foram normalizadas utilizando os valores médios dos píxeis e desvios-padrão da base de dados ImageNet. Em seguida, as imagens foram redimensionadas para 256×256 píxeis e recortadas centralmente em 224×244 píxeis. Ademais, aplicou-se a técnica de aumento de dados por meio de inversões horizontais aleatórias. Por fim, a base de dados foi dividida em dois subconjuntos: (i) treinamento, possuindo 90% do total de imagens, e validação com 10%.

[4]https://pytorch.org/

Todos os modelos foram treinados fazendo uso do otimizador Adam, um tamanho de lote (*batch*) 128 e uma taxa de aprendizado inicial de 0,001. Essa taxa foi monitorada e era multiplicada por 0,1 sempre que o erro de validação não diminuía em cinco épocas consecutivas. O processo de treinamento cessou depois que o erro de validação não diminuiu por 20 épocas consecutivas.

Os experimentos foram conduzidos com o objetivo de realizar uma avaliação comparativa entre os seis métodos descritos na seção anterior. Vale ressaltar que a aplicação da transferência de aprendizado, usando a base de dados IMDB-WIKI, replicou a técnica utilizada em cada método, a fim de evitar vieses indesejados.

A base de dados APPA-REAL foi adotada como fim, devido à necessidade de investigar a influência da idade aparente na estimativa da idade real. Esse conjunto de dados fornece rótulos de idade aparente e real para as imagens faciais. Foram utilizadas exatamente as mesmas divisões (treinamento, validação e teste) propostas por Agustsson et al. (2017), com o intuito de reproduzir o mesmo cenário.

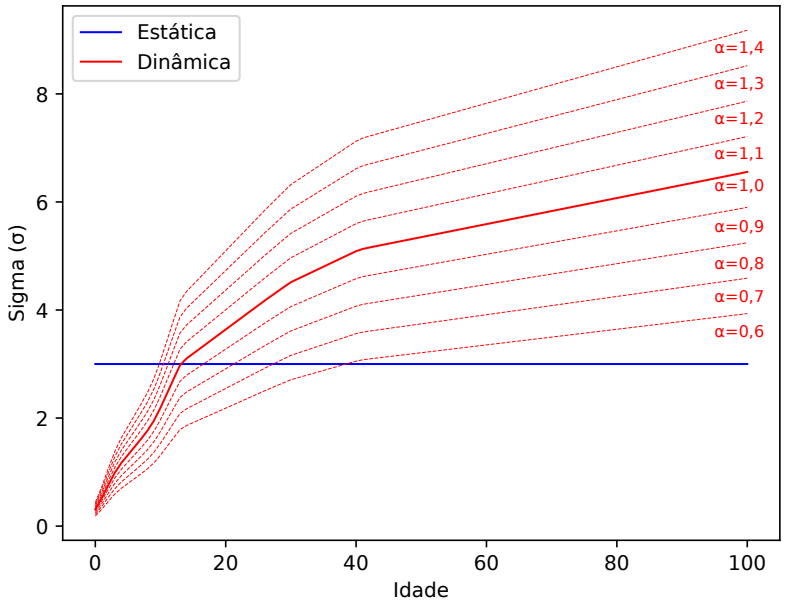
Devido ao número reduzido de parâmetros, eficiência computacional e excelentes resultados no desafio ImageNet, a rede neural convolucional DenseNet (HUANG et al., 2017) com 161 camadas foi adotada como arquitetura em todos os experimentos. Os modelos em cada método foram treinados e ajustados seguindo o mesmo protocolo utilizado para o pré-treinamento, exceto com relação à taxa de aprendizado inicial, que foi definida em 0,0005.

Ao treinar a abordagem Gaussiana Estática, escolhemos o valor de $\sigma = 3$ para o desvio-padrão fixo, com base em Rondeau e Alvarez (2018). Para a abordagem Gaussiana Dinâmica, foi treinada em separado uma rede neural *perceptron* multicamadas (*Multilayer Perceptron - MLP*) de seis camadas com 80, 70, 60, 40, 40, 20 neurônios por camada, respectivamente.

Essa rede neural perceptron multicamadas, dada baixa complexidade do problema, foi escolhida para assumir o papel de um regressor que aprende a prever o desvio-padrão correto para um valor de idade real de entrada, baseado em dados de idade aparente. A rede é treinada usando como entrada uma idade real de uma imagem e tem como respectivo rótulo o desvio-padrão de todas estimativas de idade aparente atribuídas àquela mesma imagem do conjunto de dados APPA-REAL. Após um conjunto de experimentos preliminares, o modelo foi treinado utilizando o otimizador Adam, um lote completo (*full batch*) e uma taxa de

aprendizado inicial de 0,001, que monitorada, foi multiplicada por 0,9 sempre que o erro de validação não diminuía em 20 períodos consecutivos. O processo cessou após 1000 épocas. Um hiperparâmetro ajustável α foi introduzido para dimensionar o desvio-padrão regredido σ. A Figura 7.2 mostra um gráfico das variações dos valores σ previstos, para diferentes idades reais, sob a influência de diferentes valores de α.

Figura 7.2: Curvas que representam os valores de σ para a Gaussiana estática (em azul) e dinâmica (em vermelho), em relação aos valores da idade real. As curvas vermelhas tracejadas são variações da curva quando dimensionadas por diferentes valores de α.



Fonte: Autor.

Os recursos computacionais utilizados para o treinamento dos modelos propostos nesta seção de metodologia incluiu: uma estação de trabalho equipada com 64GB de memória ram, dois processadores Intel Xeon E5 2,10 GHz e quatro unidades de processamento gráfico Pascal Titan X 12GB. Toda a análise experimental, assim como seus resultados, encontram-se discriminados detalhadamente no Capítulo 9 (Análise Experimental e Resultados), especificamente no item “9.3.4 Estimação de Idade Real Facial e Análise Comparativa com Pesquisas do Estado da Arte”.

7.5 Considerações Finais

Neste capítulo, com o intuito de compor o Módulo Facial exposto no Capítulo 5 (Arquitetura Proposta para Detecção de Pornografia Infanto-Juvenil), foram propostos novos métodos para melhorar a previsão da idade real (cronológica) de faces, principalmente por meio do uso de dados de idade aparente. Ademais, foram descritas as bases de dados utilizadas, assim como a metodologia experimental e métrica de avaliação adotadas.

Capítulo 8

Classificador de Menoridade Penal

Neste capítulo, é justificada a necessidade, no âmbito da Perícia Criminal, de que nas imagens classificadas como pornográficas as faces detectadas recebam a probabilidade de pertencerem a indivíduos menores de 18 anos. Sendo assim, foi proposto um Classificador de Menoridade Penal que propõe otimizar os resultados referentes a essas probabilidades.

8.1 Introdução

Dada a impossibilidade, tendo como base os estudos inseridos no estado da arte desta área, de se estimar com certeza absoluta a idade de um indivíduo por meio de sua face e a grande responsabilidade, na seara da Perícia Criminal, de inferir se determinado indivíduo possui menos de 18 anos em imagens pornográficas (que é capaz subsidiar a condenação ou absolvição suspeitos), foi proposto que essa determinação não seja realizada de maneira estrita (i.e. menor de idade, maior de idade), mas sim por meio de probabilidades relativas às faces detectadas, apontando a chance dessas pertencerem a indivíduos menores de dezoito anos. Dessa forma, a técnica proposta auxiliará o Perito Criminal a tomar sua decisão, podendo ainda se apoiar em outros elementos oriundos do exame pericial como um todo (e.g. nomenclatura de arquivos, diálogos em aplicativos, histórico de navegação).

8.2 Métodos

Mesmo o Módulo Facial estando apto a apresentar a probabilidade de um indivíduo possuir menos de dezoito anos, por meio do somatório de probabilidades individuais das idades obtido na camada final da rede, foi apresentada uma nova proposta com o intuito de aprimorar o resultado final. Trata-se do Classificador de Menoridade Penal, apresentando dois tipos: (a) o Simples e (b) o Composto. As referidas abordagens encontram-se descritas em detalhes em seguida.

8.2.1 Somatório de Probabilidades

O Módulo Facial, responsável por predizer as idades inerentes às faces detectadas, tem como resultado uma distribuição de probabilidade discreta de, dada uma face, essa pertencer a uma idade contida na faixa de idades no intervalo fechado [0, 100]. Sendo assim, o cálculo da probabilidade P de uma face predita pertencer a um indivíduo com menos de dezoito anos é dado pelo somatório das primeiras 18 saídas da rede neural convolucional, conforme a Equação 8.1.

$$P(O) = \sum_{i=0}^{17} o_i \tag{8.1}$$

em que $O = \{o_0, o_1, \ldots, o_{17}\}$ é a saída das primeiras 18 dimensões da rede, que representam as probabilidades da função *Softmax*.

8.2.2 Classificador de Menoridade Penal

Dada a baixa complexidade do problema, baseado nos dados de entrada e saída do modelo, foi adotada uma rede neural perceptron multicamadas, contendo cinco camadas escondidas com cem neurônios cada. A camada final é composta por dois neurônios que, por meio de uma função *SoftMax*, determina a probabilidade de inferência para cada uma das seguintes categorias: (i) menor de idade ou (ii) maior de idade.

Os dados de entrada utilizados para o treinamento do Classificador de Menoridade Penal são as idades preditas das faces pelo Módulo Facial da arquitetura proposta no Capítulo 7 (Aprimorando a Estimativa de Idade Real a Partir de Dados de Idade Aparente) e os res-

pectivos metadados dessas faces (para o tipo Composto dessa abordagem), discriminados ainda neste capítulo. Os rótulos utilizados para treinamento são as verdadeiras categorias de cada uma dessas faces (i.e. menor de idade, maior de idade). Na etapa de teste, utilizando o mesmo tipo de dados usados no treinamento, o Classificador de Menoridade Penal retorna a probabilidade dessa face pertencer a uma pessoa com menos de dezoito anos.

Para o tipo **Simples**, o vetor de características é composto apenas pela idade. No tipo **Composto**, o vetor de características varia de tamanho, pois além da idade, é composto por metadados oriundos da face detectada. Os referidos metadados, obtidos por meio do uso do detector facial MTCNN (ZHANG et al., 2016), encontram-se descritos na sequência e ilustrados na Figura 8.1.

- Confiança: valor variando de zero a um ([0,1]) que quantifica o nível de confiança do objeto detectado ser realmente uma face.

- Tamanho: dadas as coordenadas da face detectada, é possível determinar seu tamanho em píxeis. Trata-se de um aspecto importante, pois caso o tamanho da face não seja grande o suficiente para o padrão de entrada da rede neural convolucional, a mesma é expandida, perdendo qualidade.

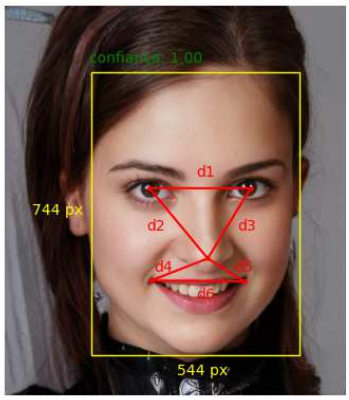
- Distância entre pontos fiduciais (*landmarks*): dada as coordenadas dos pontos fiduciais (i.e. olho esquerdo, olho direito, nariz, canto esquerdo da boca, canto direito da boca), foi possível calcular seis distâncias normalizadas, essas descritas na Tabela 8.1. Por meio do uso das referidas distâncias é possível inferir a rotação das faces detectadas (YANG et al., 2015), assim como particularidades inerentes a diferentes faixas etárias (HAMMOND et al., 2020).

Tabela 8.1: Distâncias calculadas entre pontos fiduciais da face (*landmarks*).

Distância	Ponto Fiducial A	Ponto Fiducial B
d1	olho esquerdo	olho direito
d2	olho esquerdo	nariz
d3	olho direito	nariz
d4	canto esquerdo da boca	nariz
d5	canto esquerdo da boca	nariz
d6	canto direito da boca	canto esquerdo da boca

Fonte: Autor.

Figura 8.1: Metadados inerentes às faces detectadas. Em verde, a confiança do objeto detectado ser uma face. Em amarelo, o retângulo com a área da face detectada. Em vermelho, as distâncias entre os pontos fiduciais.



Fonte: Autor.

8.3 Metodologia

Nesta seção, inicialmente é descrita métrica de avaliação adotada. Em seguida, são detalhados os protocolos utilizados para a realização das etapas de treinamento e avaliação dos modelos propostos.

8.3.1 Métrica de Avaliação

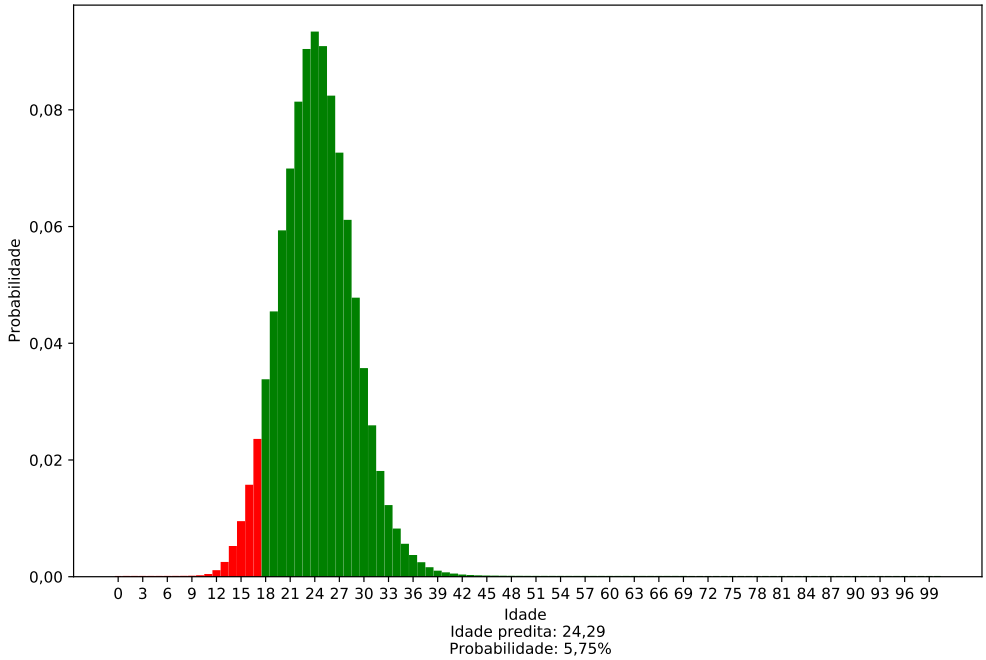
Utilizou-se como métrica de avaliação o erro médio absoluto (*Mean Absolute Error - MAE*) entre as m probabilidades inferidas $\hat{y}_i$ e as verdadeiras classes y (1 para menor de 18 anos e 0 para maior ou igual a 18 anos), como pode ser visto na Equação 8.2.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{8.2}$$

8.3.2 Protocolo Experimental

Para a abordagem do Somatório de Probabilidades, as imagens em questão foram submetidas à técnica de maneira direta, visto que os valores para o cálculo da probabilidade já encontrava-se incorporados à camada final Módulo Facial, bastando realizar a soma das probabilidades e calcular o erro médio. A Figura 8.2 mostra um exemplo de distribuição de probabilidade discreta de uma predição de idade facial e, por meio de cores, expõe quais probabilidades individuais devem ser somadas para atingir a probabilidade final.

Figura 8.2: Exemplo de distribuição de probabilidade resultante de uma predição de idade facial real utilizando a técnica do Somatório de Probabilidades. As barras em vermelho representam as probabilidades referentes às idades menores que dezoito anos. As barras em verde representam as probabilidades referentes às idades maiores ou iguais a dezoito anos.

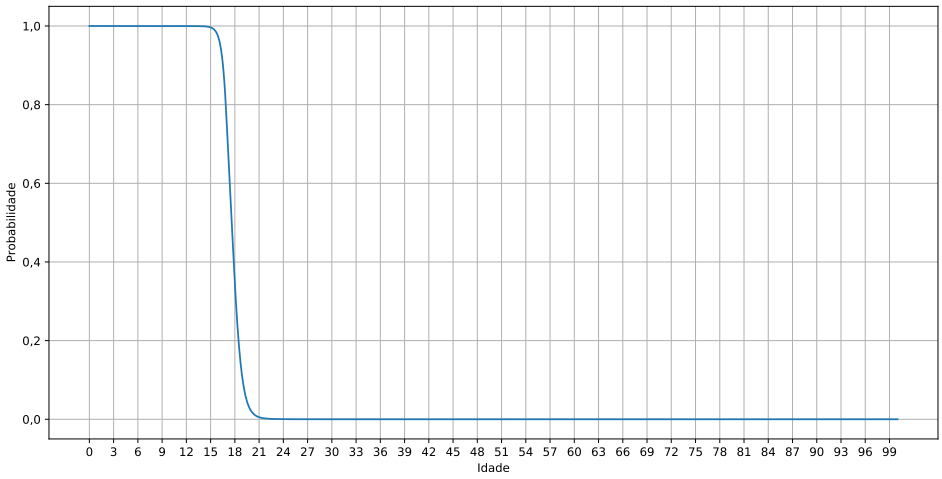


Fonte: Autor.

Com relação ao Classificador de Menoridade Penal Simples, foi necessária a realização do treinamento da rede neural adotada. Após um conjunto de experimentos preliminares, essa etapa se deu utilizando 200 épocas com parada antecipada (*Early Stopping*). O treinamento era interrompido se não houvesse melhoria na acurácia dos dados de validação por dez épocas consecutivas. Como configuração padrão, foi utilizado o otimizador Adam com uma taxa de aprendizado de 0,001 e um tamanho de lote (*batch*) igual a 200. Com relação à disposição dos dados para o treinamento e teste do modelo, foi aplicada a técnica de validação cruzada utilizando dez subconjuntos (*10-fold*), em que nove eram destinados à etapa de treinamento/validação (de proporção 90%/10%) e um destinado à etapa de teste. Por fim, o resultado final se deu pela média erro médio absoluto dos subconjuntos de teste nos dez diferentes cenários. A Figura 8.3 mostra a curva de probabilidades gerada pelo classificador do tipo Simples na etapa de testes, quando variado o valor de entrada (idade) no intervalo fechado $[0, 100]$.

Seguindo o mesmo protocolo experimental do Classificador de Menoridade Penal Sim-

Figura 8.3: Curva que retrata por meio da técnica do Classificador de Menoridade Penal Simples, dada uma idade previamente estimada, a probabilidade de o indivíduo ser menor de idade.



Fonte: Autor.

ples, foram criados seis modelos adicionais para o Classificador de Menoridade Penal Composto, referentes a cada uma das combinações dos metadados utilizados no vetor de características (i.e. Idade + Confiança, Idade + Tamanho, Idade + Distâncias, Idade + Confiança + Tamanho, Idade + Confiança + Distâncias, Idade + Confiança + Tamanho + Distâncias).

Os recursos computacionais utilizados para o treinamento dos modelos propostos nesta seção de metodologia incluiu: uma estação de trabalho equipada com 32GB de memória ram, uma processador Intel Xeon E5 2,60 GHz e uma unidade de processamento gráfico NVIDIA GeForce GTX 1080 Ti 11GB. Toda a análise experimental, assim como seus resultados, encontram-se discriminados detalhadamente no Capítulo 9 (Análise Experimental e Resultados), especificamente no item “9.3.6 Classificador de Menoridade Penal”.

8.4 Considerações Finais

Dada a impossibilidade de se atestar a idade facial com a certeza absoluta que a Perícia Criminal exige, foi proposto o uso de probabilidades para determinar as chances de que um indivíduo, dada sua face, possua menos de 18 anos. Mesmo o Módulo Facial tendo a possibilidade de apresentar seus resultados nesses moldes, foi proposta uma nova abordagem baseada em aprendizado de máquina que visa aprimorar essa probabilidade final.

Capítulo 9

Análise Experimental e Resultados

Este capítulo descreve os experimentos realizados visando à validação da arquitetura de detecção de pornografia infanto-juvenil proposta, assim como dos seus módulos. Ademais, é justificada a métrica de avaliação utilizada para uma comparação mais justa dos modelos. Também é detalhada a construção do conjunto de dados de pornografia infanto-juvenil utilizado. Por fim, foi realizada uma análise de desempenho computacional visando discriminar os tempos de processamento da abordagem proposta.

9.1 Base de Dados Privada de Pornografia Infanto-juvenil

Visando à análise experimental do modelo proposto para a detecção de conteúdo pornográfico contendo crianças e/ou adolescentes, ou seja, indivíduos menores de 18 anos, foi elaborada uma base de dados constituída com imagens dessa natureza.

Da Obtenção das Imagens

Para a construção da base de dados em questão, foram utilizados os arquivos de imagem obtidos dos dispositivos de armazenamento referentes ao exame pericial que resultou no Laudo 01.03.02.072017.18612. O laudo elaborado pelo Perito Oficial Criminal Danilo Coura Moreira tinha como objetivo verificar a existência de conteúdo pornográfico envolvendo crianças e/ou adolescentes.

Da Análise e Seleção das Imagens

Para determinar a existência de pornografia nas imagens, foi aplicado o mesmo protocolo utilizado para categorizar as imagens da base de dados proposta de pornografia adulta (PEDA 376K). Utilizou-se o critério adotado por Wang, Jin e Tan (2016) para determinar se uma a imagem detinha teor pornográfico. A imagem é classificada como tal caso possua pessoa(s) nua(s), mostrando explicitamente pelo menos uma parte do corpo particular exposta, incluindo mama feminina, nádegas, órgãos sexuais feminino ou masculino; ou em atos sexuais, independente das vestimentas.

A determinação de que uma imagem pornográfica se categoriza como pornografia infanto-juvenil se dá pela simples presença de pelo menos uma criança ou adolescente na imagem, mesmo que suas partes íntimas estejam preservadas e não participe de atos sexuais. Por exemplo, a existência de uma criança ao lado de um adulto com sua partes íntimas expostas. Uma representação gráfica do processo de decisão, baseado no utilizado na construção da base de dados PEDA 376K, pode ser vista na Figura 9.1.

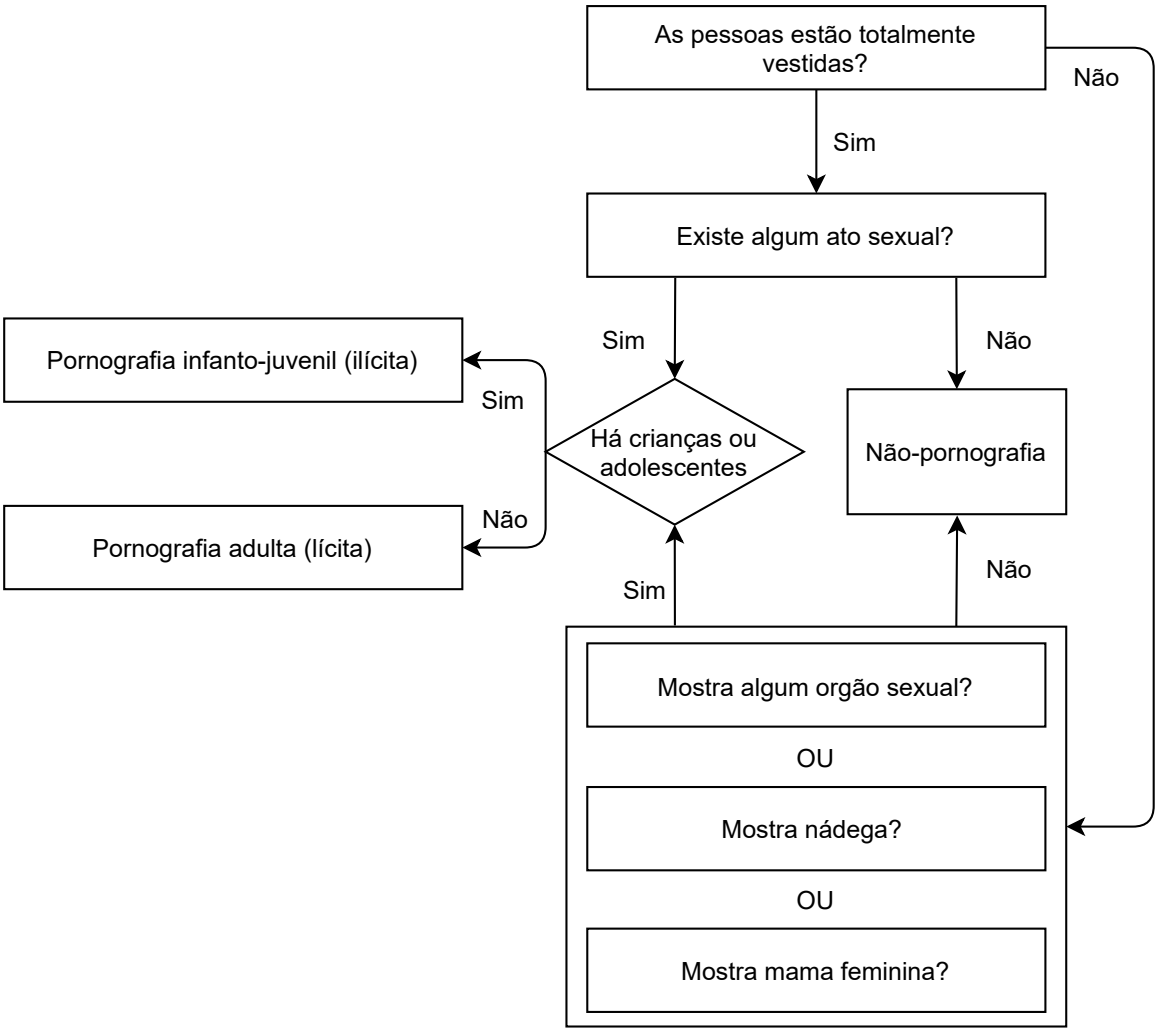
Dado que algumas das imagens fazem parte de um ensaio fotográfico completo, muitas vezes as imagens iniciais não possuem teor pornográfico, estando os atores vestidos ou as imagens retratando apenas faces. Sendo assim, imagens contendo crianças ou adolescentes nesse contexto não pornográfico também foram utilizadas na categoria de imagens não pornográfica.

Devido a características peculiares inerentes às crianças, principalmente a ausência de maturação sexual nesses indivíduos, não houve dificuldade para identificá-las nas imagens pornográficas. Além do mais, parte dos diretórios em que encontravam-se as imagens em questão detinham nomes que sugeriam esse tipo de material, como “child” (criança), “infant” (infantil), “XXyo” (XX anos de idade, em que XX era menor ou igual a 12).

Entretanto, diferenciar de maneira precisa adolescentes e jovens adultos é uma tarefa complexa e ainda não atingida pelo estado da arte da Visão Computacional. Essa dificuldade se dá devido à heterogeneidade da maturação humana que é influenciada principalmente pela genética, mas também por fatores externos como padrões de ingestão de alimentos, práticas de atividades esportivas, incidência solar, bem-estar mental, entre outros (RONDEAU; ALVAREZ, 2018).

Mesmo diante das dificuldades enumeradas, algumas imagens puderam ser classificadas

Figura 9.1: Fluxograma do processo de tomada de decisão para rotular uma imagem como pornografia infanto-juvenil ou não.



Fonte: Autor.

com juvenis também pela existência de nomes nos diretórios que as imagens foram encontradas, como “teen” (adolescente), “lolit” (lolita), “XXyo” (XX anos de idade, em que XX era maior que 12 e menor que 18).

As demais imagens pornográficas que não puderam claramente ser classificadas como juvenis, foram determinadas como tal, salvo exceções patentes da presença de apenas adultos, partindo do pressuposto que o infrator era aficionado por imagens contendo pornografia infanto-juvenil, visto que o crime foi claramente caracterizado devido à existência das demais imagens ilícitas já citadas. Ademais, as imagens encontravam-se agrupadas em diretórios adjacentes aos das imagens confirmadas como ilícitas, sugerindo serem do mesmo tipo.

Essa indução tem o objetivo de maximizar o conjunto de testes, visando uma maior confiabilidade do resultado final, sem qualquer intenção de influenciar positivamente o desempenho do modelo proposto. Pelo contrário, a inserção de possíveis falsos positivos resultaria na diminuição da acurácia. Além do mais, mesmo tendo sido inseridos como verdadeiros positivos, adolescentes são mais difíceis de serem classificadas corretamente como menores de idade, em comparação quando crianças são submetidas a essa classificação, o que também leva a uma diminuição da acurácia do resultado final. Em suma, a construção da base de dados visa retratar de maneira fiel a realidade dos dispositivos submetidos aos exames periciais relativos a crimes de violência sexual contra crianças e adolescentes.

Da Estrutura da Base de Dados

Após a análise, as imagens foram distribuídas em dois diretórios: (i) porn e (ii) not_porn. Cada uma das imagens foi renomeada com o seguinte padrão de nomenclatura: PEDO_porn (X).jpg e PEDO_not_porn (X).jpg, em que X é um contador de um até quantidade de imagens de cada categoria, para as imagens possuindo pornografia infanto-juvenil e para as imagens contendo crianças e/ou adolescentes em situações cotidianas, respectivamente. Por fim, foram obtidas 14.080 imagens contendo pornografia infanto-juvenil e 352 imagens não pornográficas retratando crianças e/ou adolescentes.

Do Formato das Imagens

Diferentemente da base de dados PEDA 376K, as imagens foram mantidas em seu tamanho original. Isso se deu pela necessidade de averiguar a existência de faces nas mesmas para o módulo proposto de estimativa de idade facial, pois o redimensionamento comprometeria o desempenho do detector de faces adotado.

Da Junção com a Base de Dados PEDA 376K

Visto que o trabalho proposto lida com imagens não pornográficas, pornográficas lícitas (adulta) e pornografia ilícita (infanto-juvenil), as bases de dados foram utilizadas em conjunto com o objetivo de validar os módulos, assim como a arquitetura como um todo. A Tabela 9.1 expõe a quantidade de imagens pornográficas e não pornográfica de cada uma das

bases de dados utilizadas para a avaliação do modelo (i.e. imagens não utilizadas previamente para treinamento/validação), assim como o quantitativo da junção das mesmas.

Tabela 9.1: Quantidade de imagens pornográficas e não pornográficas por base de dados e unificada.

Base de Dados	Tipo	
	Não pornografia	Pornografia
PEDA 376K (adulta)	4.700	4.700
Infanto-juvenil	352	14.080
PEDA 376K (adulta) + Infanto-juvenil	5.052	18.780

Fonte: Autor.

Dada a junção das referidas bases de dados, é possível visualizar na Tabela 9.1 de maneira detalhada a quantidade de imagens pornográficas, de acordo com a presença de face e licitude da imagem (infanto-juvenil ou adulta) e também a quantidade de imagens não pornográficas.

Tabela 9.2: Descrição detalhada dos tipos de imagens da união das bases de dados infanto-juvenil e PEDA 376K.

Tipo			Quantidade de Imagens
Não Pornografia			5.052
Pornografia	Infanto-juvenil	Com face	9.629
		Sem face	4.451
	Adulta	Com face	1.990
		Sem face	2.710

Fonte: Autor.

9.2 Métrica de Avaliação

O uso correto das métricas de avaliação é fundamental para analisar de maneira adequada o desempenho de um determinado classificador. Muitas vezes a adoção de uma métrica de maneira isolada ou incorreta pode não refletir a real eficácia do classificador (MAHMOODI; SAYEDI, 2016).

O ponto inicial da avaliação se dá pela construção da matriz de confusão, como mostrada na Tabela 9.2. Essa representa o número de ocorrências das quatro possíveis combinações entre a condição verdadeira e a prevista de cada imagem (AL-MOHAIR; SALEH; SUANDI, 2015b). Dessa forma, a mesma apresenta tamanho 2x2, tendo como elementos nas posições:

- 0,0: Verdadeiro Positivo (VP). Referente ao número de imagens positivas corretamente classificadas;
- 0,1: Falso Negativo (FN). Referente ao número de imagens positivas erradamente classificadas;
- 1,0: Falso Positivo (FP). Referente ao número de imagens negativas erradamente classificadas;
- 1,1: Verdadeiro Negativo (VN). Referente ao número de imagens negativas corretamente classificadas.

Tabela 9.3: Matriz de Confusão.

		Predição	
		Positivo	Negativo
Verdade	Positivo	VP	FN
	Negativo	FP	VN

Fonte: Autor.

Dada a existência de desbalanceamento de dados na realização dos experimentos deste capítulo, a acurácia propriamente dita (vide Equação (9.1)) não é a métrica de avaliação mais adequada a ser adotada, visto que não daria a mesma importância as classes positivas e negativas (AL-MOHAIR; SALEH; SUANDI, 2015b).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \tag{9.1}$$

Apesar de o F-score, conforme a Equação (9.4), lidar melhor com o desbalanceamento dos dados por meio do uso da média harmônica entre a Precisão (vide Equação (9.2)) e a Revocação (vide Equação (9.3)), essa métrica de avaliação desconsidera os verdadeiros negativos, que é fundamental no âmbito da Perícia Criminal, pois também se faz necessário analisar a inocência de suspeitos, verificando se determinada imagem não contém pornografia infanto-juvenil.

$$Precisão = \frac{VP}{VP + FP} \tag{9.2}$$

$$Revocação = \frac{VP}{VP + FN} \tag{9.3}$$

$$F\text{-}score = \frac{2 \times Precisão \times Revocação}{Precisão + Revocação} \tag{9.4}$$

Dessa forma, visando dar a mesma importância aos verdadeiros positivos e verdadeiros negativos, foi adotada como métrica de avaliação a acurácia ponderada, conforme a Equação (9.7). Essa métrica é calculada pela média da taxa de verdadeiros positivos (vide Equação (9.5)) e verdadeiros negativos (vide Equação (9.6)).

$$Taxa\ de\ verdadeiros\ positivos\ (TVP) = \frac{VP}{VP + FN} \tag{9.5}$$

$$Taxa\ de\ verdadeiros\ negativos\ (TVN) = \frac{VN}{VN + FP} \tag{9.6}$$

$$Acurácia\ ponderada = \frac{TVP + TVN}{2} \tag{9.7}$$

9.3 Análise Experimental

Com o objetivo de avaliar a arquitetura proposta como um todo, assim como os módulos adotados de maneira isolada, foram realizados diversos experimentos, esses divididos em sete seções discriminadas em detalhes na sequência.

9.3.1 Aprendizado de Máquina Tradicional vs. Aprendizado Profundo na Detecção de Pornografia

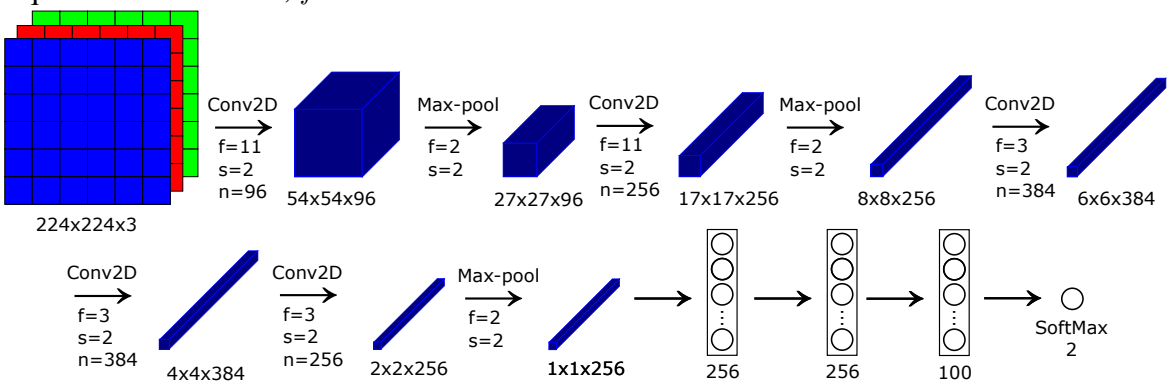
Foram realizados experimentos utilizando aprendizado profundo visando à realização de uma análise comparativa com os trabalhos propostos em Moreira e Fechine (2018a) e em Moreira e Fechine (2018b), descritos detalhadamente no Apêndice A (Detectores de Conteúdo Impróprio Baseado em Aprendizado de Máquina Tradicional).

Dessa forma, utilizou-se uma versão simplificada da rede neural convolucional AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), como pode ser visualizado na Figura 9.2. Apesar de possuir a mesma quantidade de camadas de convolução, a arquitetura possui uma quantidade reduzida de parâmetros, exigindo menos processamento, assim como uma

menor quantidade de memória (MYDATAHACK, 2018).

Foram utilizadas as configurações padrões do modelo, não tendo sido realizado o ajuste fino dos seus hiperparâmetros. Os referidos hiperparâmetros estão explicitados na Tabela 9.4. Salienta-se que os experimentos foram reproduzidos de maneira isonômica, utilizando as mesmas proporções de dados para treinamento, validação e teste utilizados em Moreira e Fechine (2018a) e Moreira e Fechine (2018b).

Figura 9.2: Arquitetura baseada na rede neural convolucional AlexNet, em que n representa a quantidade de filtros, f o seu tamanho e s o *stride* adotado.



Fonte: Autor.

Tabela 9.4: Parâmetros utilizados na rede neural convolucional baseada na AlexNet.

Parâmetro	**Valor**
Função de perda	Erro quadrático médio
Métrica de avaliação	Acurácia
Tamanho do batch	16
Épocas de Treinamento	40
Otimizador	Gradiente descendente estocástico
Taxa de aprendizado	0,01
Decaimento da taxa de aprendizado	0,000001
Momentum	0,9
Uso do momentum nesterov	Sim

Fonte: Autor.

Os resultados finais podem ser visualizados nas Tabela 9.5 e Tabela 9.6, que comparam a acurácia final de um modelo simples baseado em aprendizado profundo com os modelos baseados em aprendizado de máquina tradicional propostos nos trabalhos de Moreira e Fechine (2018a) e Moreira e Fechine (2018b), respectivamente.

Com base nos resultados obtidos, foi possível afirmar que o modelo baseado em aprendizado profundo, mesmo com as limitações de utilizar um quantidade reduzida de dados e

Tabela 9.5: Comparativo da acurácia entre o modelo baseado em aprendizado profundo e o modelo baseado em aprendizado de máquina tradicional proposto em Moreira e Fechine (2018a), considerando um intervalo de confiança de 95%.

Modelos	Acurácia $\pm$ margem de erro
Aprendizado profundo	93,56% $\pm$ 1,35%
Aprendizado de máquina tradicional (MOREIRA; FECHINE, 2018a)	93,64% $\pm$ 1,34%

Fonte: Autor.

Tabela 9.6: Comparativo da acurácia entre o modelo baseado em aprendizado profundo e o modelo baseado em aprendizado de máquina tradicional proposto em Moreira e Fechine (2018b), considerando um intervalo de confiança de 95%.

Modelos	Acurácia $\pm$ margem de erro
Aprendizado profundo	96,96% $\pm$ 1,42%
Aprendizado de máquina tradicional (MOREIRA; FECHINE, 2018b)	97,67% $\pm$ 1,25%

Fonte: Autor.

de não realizar um ajuste fino de seus hiperparâmetros, foi capaz de apresentar resultados equivalentes aos apresentado pelos modelos baseados em aprendizado de máquina tradicional em Moreira e Fechine (2018a) e Moreira e Fechine (2018b), levando em consideração um intervalo de confiança de 95%.

9.3.2 Detecção de Pornografia Adulta e Análise Comparativa com Serviços de Moderação de Conteúdo

Influenciado pelo resultado do subitem anterior e também pelo levantamento bibliográfico relativo ao uso do aprendizado profundo na classificação de imagens como um todo e, principalmente, na detecção de imagens pornográficas, decidiu-se utilizar um modelo baseado em aprendizado profundo em vez de modelos baseados em aprendizado de máquina tradicional.

Baseado na metodologia exposta no Capítulo 6 (Detector de Pornografia Baseado em Aprendizado Profundo e Nova Base de Dados Pornográfica), os experimentos foram conduzidos com o objetivo de decidir que arquitetura, tamanho do lote (*batch*), otimizador e taxa de aprendizado utilizar.

Arquitetura de Rede Neural Convolucional

A Tabela 9.7 mostra o desempenho referente aos dados de treinamento e validação para a seleção da rede neural convolucional e a quantidade de camadas adotadas. Sendo assim, a Densenet-121 foi adotada para as etapas subsequentes por ter atingido a maior acurácia no conjunto de validação, 98,5%, juntamente com a Densenet-169. O fator de desempate escolhido para descartar a Densenet-169 foi a menor complexidade do modelo.

Tabela 9.7: Acurácias referentes aos dados de treinamento e validação de cada modelo e seu respectivo número de camadas.

Modelo	Camadas	Treinamento	Validação
ResNet	18	$99,5\%$	$97,6\%$
	34	$99,4\%$	$97,7\%$
	50	$98,8\%$	$97,4\%$
	101	$98,7\%$	$97,3\%$
	152	$99,2\%$	$97,7\%$
Wide ResNet	50	$97,6\%$	$97,2\%$
	101	$99,3\%$	$97,4\%$
ResNext	50	$99,2\%$	$98,1\%$
	101	$99,4\%$	$97,8\%$
DenseNet	121	$99,4\%$	**98,5%**
	161	$99,5\%$	$98,3\%$
	169	$99,6\%$	**98,5%**
	201	$98,6\%$	$98,2\%$

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Tamanho do Lote (*batch*)

O objetivo dessa etapa foi encontrar o tamanho de lote que proporcionasse os melhores resultados, como pode ser visto na Tabela 9.8. Percebeu-se que o modelo treinado com um tamanho de lote (*batch*) igual a 128 apresentou a melhor acurácia: 98,5%. Dessa forma, foi mantida a Densenet-121 com um tamanho de lote igual 128 para a última etapa da estratégia gulosa.

Otimizador e Taxa de Aprendizado

Nesta última etapa, foram avaliados cinco otimizadores e suas respectivas taxas de aprendizado. A Tabela 9.9 mostra o desempenho final do treinamento e validação. Sendo assim, a melhor configuração entre arquiteturas e hiperparâmetros encontrada é composta por uma

Tabela 9.8: Acurácias correspondentes aos dados de treinamento e validação quando variado o tamanho do lote (*batch*).

Tamanho do lote (batch)	Treinamento	Validação
32	$99,2\%$	$98,1\%$
64	$99,4\%$	$98,4\%$
128	$99,4\%$	**98,5%**
256	$99,6\%$	$98,4\%$

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

rede neural convolucional Densenet-121 com tamanho de lote (*batch*) igual a 128, treinada utilizando um otimizador SGD e uma taxa de aprendizado igual a 2^{-8}. Essa configuração atingiu 99,2% de acurácia utilizando os dados do conjunto de validação. Em seguida, o modelo foi analisado utilizando os dados do conjunto de teste, atingindo a acurácia de 99,1%. Ressalta-se que os dados do conjunto de teste nunca foram usados nas etapas de treinamento e/ou validação.

Tabela 9.9: Acurácias inerentes aos dados de treinamento e validação variando o otimizador. Ademais, é apresentada melhor taxa de aprendizado para cada otimizador.

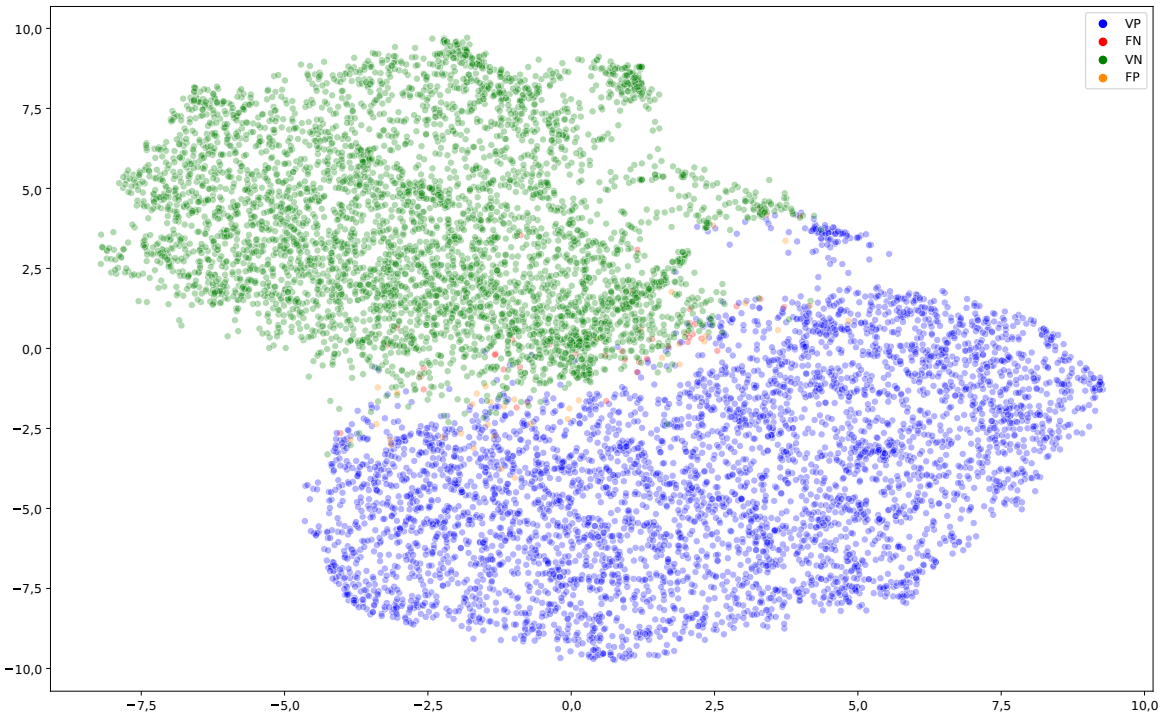
Otimizador	Treinamento	Validação	Taxa de Aprendizado
AdaGrad	100%	$98,6\%$	$\frac{2^{-4}}{10}$
RMSProp	$99,8\%$	$99,1\%$	$\frac{2^{-7}}{100}$
Adam	$99,8\%$	$99,1\%$	$\frac{2^{-7}}{100}$
SGD	$99,9\%$	**99,2%**	2^{-8}
ASGD	$99,9\%$	$99,0\%$	2^{-5}

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Para analisar o comportamento da arquitetura proposta no conjunto de dados PEDA 376K, a rede neural convolucional recebeu como entrada todas as imagens do conjunto de teste e em seguida foram extraídas as ativações da última camada convolucional. Como a dimensionalidade de cada conjunto de características é de 1.568, foi utilizada uma combinação de duas técnicas de redução de dimensionalidade: (i) a Análise de Componentes Principais (*Principal Component Analysis - PCA*) e (ii) o t-SNE (*t-Distributed Stochastic Neighbor Embedding*) (MAATEN; HINTON, 2008) para visualizar todas as instâncias em duas dimensões, conforme mostrado na Figura 9.3. É possível visualizar a formação de dois conjuntos bem definidos, exceto por algumas instâncias classificadas incorretamente que representam menos de um por cento do total.

Ademais, foi realizada uma análise comparativa entre os resultados da arquitetura pro-

Figura 9.3: O comportamento da arquitetura proposta utilizando a base de testes PEDA 376K, sob a ótica da projeção t-SNE. Os verdadeiros positivos e negativos são representados pelos conjuntos de pontos em azul e verde, respectivamente. Os poucos pontos laranja e vermelhos representam as imagens classificadas incorretamente, respectivamente os falsos positivos e negativos.



Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

posta para detecção de pornografia e cinco serviços de moderação de imagens, conforme metodologia exposta no Capítulo 6 (Detector de Pornografia Baseado em Aprendizado Profundo e Nova Base de Dados Pornográfica).

Para os conjuntos de experimentos realizados utilizando ambas as bases de dados (i.e. PEDA 376K e RedLight), as maiores acurácias foram atingidas pelos modelos otimizados que utilizaram o vetor de probabilidades completo para o treinamento das árvores de decisão. De acordo com a Tabela 9.10, a abordagem proposta de otimização do modelo *baseline* aumentou a acurácia de todos os serviços de moderação de imagens. Ao considerar apenas a base de dados PEDA 376K, a rede neural convolucional proposta superou todos os serviços de moderação de imagens, mesmo considerando as versões otimizadas propostas. A segunda seção na Tabela 9.10 expõe o desempenho dos serviços de moderação de imagens e da rede convolucional proposta ao usar a base de dados RedLight. A arquitetura proposta supera

quatro dos cinco serviços, considerando tanto os modelos *baseline*, quanto suas respectivas versões otimizadas. O serviço oferecido pela Amazon obteve a maior acurácia nesse cenário.

Tabela 9.10: Acurácias dos modelos *baseline* e otimizado por meio do uso de árvores de decisão.

Base de Dados	Modelo	*Baseline*	Otimizado
PEDA 376K	Proposto	**99,1%**	-
	Amazon Rekognition	96,9%	97,9%
	Clarifai	96,4%	97,4%
	Google Vision	96,5%	96,7%
	Microsoft Azure	93,6%	95,8%
	Yahoo! NSFW	95,1%	96,1%
RedLight (ALVAREZ, 2012)	Proposto	95,2%	95,6%
	Amazon Rekognition	**95,3%**	**96,5%**
	Clarifai	92,6%	93,0%
	Google Vision	94,7%	95,1%
	Microsoft Azure	94,4%	95,5%
	Yahoo! NSFW	94,2%	94,2%

Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Em seguida, foi calculada a acurácia ponderada para a arquitetura proposta e para cada serviço de moderação de imagem, como pode ser observado na Tabela 9.11. Esse cálculo baseou-se no número de predições realizadas para cada conjunto de dados, considerando 18.800 imagens da base de dados PEDA 376K e 25.616 da base de dados RedLight.

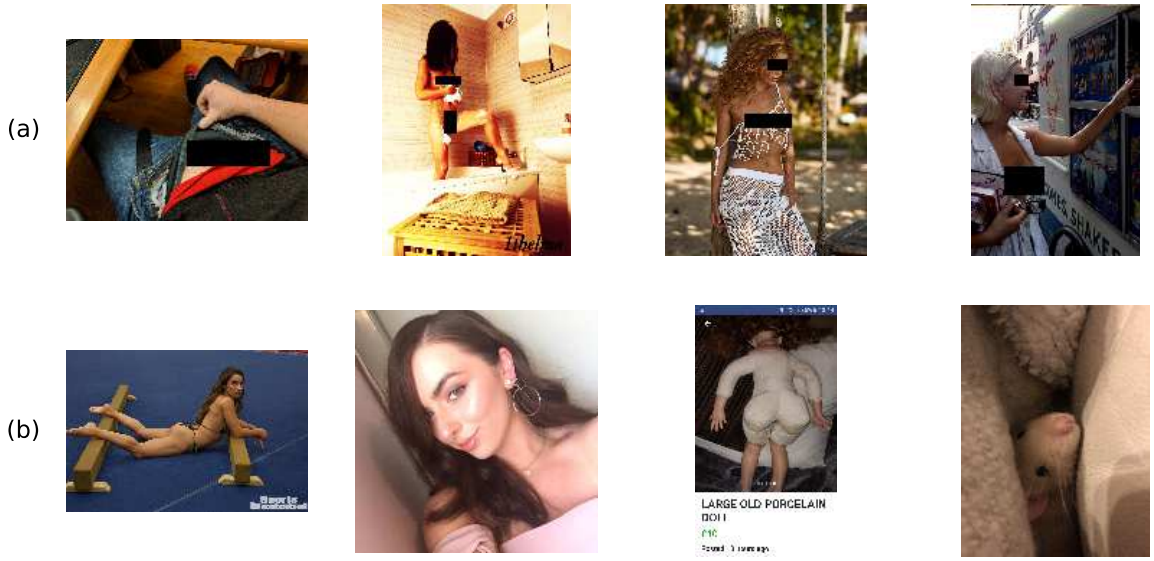
Tabela 9.11: Acurácia ponderada do módulo pornográfico proposto e dos serviços de moderação de imagens, utilizando conjuntamente as bases de dados PEDA 376K e RedLight, considerando um intervalo de confiança de 95%.

Modelo	Acurácia ponderada $\pm$ margem de erro
Proposto	**97,1% $\pm$ 0,16%**
Amazon Rekognition	**97,1% $\pm$ 0,16%**
Clarifai	$94,9\% \pm 0{,}20\%$
Google Vision	$95,8\% \pm 0{,}19\%$
Microsoft Azure	$95,6\% \pm 0{,}19\%$
Yahoo! NSFW	$95,0\% \pm 0{,}20\%$

Fonte: Autor.

Além disso, de acordo com da Figura 9.4, é possível constatar a superioridade da arquitetura proposta por meio da exibição de uma amostra aleatória de imagens classificadas incorretamente por todos os serviços de moderação de imagens e corretamente pela abordagem proposta.

Figura 9.4: Imagens classificadas incorretamente por todos os modelos otimizados que foram classificados corretamente pelo modelo proposto usando a base de dados PEDA 376K. A primeira linha (a) mostra imagens pornográficas erroneamente classificadas como não pornográficas (Falsos negativos - FN). A segunda linha (b) exibe imagens não pornográficas erroneamente classificadas como pornográficas (Falsos Positivos - FP).



Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Por fim, foi realizada uma análise não automatizada (manual) das imagens mal classificadas pela abordagem proposta utilizando a base de dados PEDA 376K. Das 86 imagens classificadas de maneira equivocada, 46 imagens não pornográficas foram classificadas como pornografia (falsos positivos) e 40 imagens pornográficas foram categorizadas como não pornografia (falsos negativos).

Infelizmente, não foi identificado um padrão responsável pela má classificação, entretanto, percebeu-se que os falsos positivos apresentaram majoritariamente pessoas com pouca roupa ou grande exposição de pele. Constatou-se também a presença de animais com genitália exposta, desenhos ou bonecos sem roupa, imagens com texturas semelhantes à pele humana e imagens em escala de cinza. Além de também apresentar imagens em escala de cinza, os falsos negativos retratavam em sua maioria imagens com pessoas vestidas quase na sua totalidade, exceto por mostrarem alguma região íntima, como mama feminina, nádegas ou genitália.

9.3.3 Desempenho do Módulo Pornográfico na Detecção de Pornografia Infanto-juvenil

Para a análise dos resultados obtidos pelo Módulo Pornográfico (arquitetura proposta para a detecção de pornografia adulta) na detecção de pornográfica em um contexto geral (i.e. pornografia adulta e pornografia infanto-juvenil), foi realizada uma análise comparativa entre os resultados já obtidos na detecção de pornografia adulta e um novo cenário que abrange também pornografia infanto-juvenil, em que foram adicionadas às categorias não pornográfica e pornográfica 352 imagens retratando crianças e/ou adolescentes em contexto não pornográfico e 14.080 imagem apresentando conteúdo pornográfico infanto-juvenil, respectivamente (vide linha três da Tabela 9.1). Dessa forma, por meio da Tabela 9.12 é possível visualizar a acurácia ponderada obtida em cada um dos casos.

Tabela 9.12: Comparativo entre os diferentes cenários utilizados para avaliação do Módulo Pornográfico proposto para diferenciação entre imagens não pornográficas e pornográficas, considerando um intervalo de confiança de 95%.

Cenário	Acurácia ponderada $\pm$ margem de erro
1 - Pornografia Adulta	**99,09% $\pm$ 0,16%**
2 - Pornografia Geral	98,47% $\pm$ 0,19%

Fonte: Autor.

Ademais, ainda considerando a detecção de pornografia no contexto geral (i.e. pornografia adulta e pornografia infanto-juvenil), realizou-se uma análise comparativa entre o desempenho do Módulo Pornográfico proposto e o serviço de moderação de conteúdo Yahoo! NSFW (MAHADEOKAR; PESAVENTO, 2016), utilizado em pesquisas relevantes para a detecção de imagens dessa natureza (GANGWAR et al., 2017; MACEDO; COSTA; SANTOS, 2018). O resultado dessa análise comparativa pode ser observado na Tabela 9.13. A utilização desse serviço nesse cenário foi viabilizada dada sua disponibilidade para *download*. Em contrapartida, os demais serviços de moderação de conteúdo, por estarem hospedados na nuvem, não puderam ser analisados devido à impossibilidade legal do envio de arquivos contendo pornografia infanto-juvenil.

Salienta-se que, baseado no trabalho de Macedo, Costa e Santos (2018), foi utilizado o limiar predefinido τ = 0,3 para identificar se a probabilidade prevista pelo Yahoo! NSFW indica ou não pornografia na imagem. A imagem é considerada não pornográfica se a probabilidade for inferior a τ, caso contrário, é considerada pornográfica.

Tabela 9.13: Comparativo entre o Módulo Pornográfico proposto e o Yahoo! NSFW para diferenciação entre imagens não pornográficas e pornográficas em geral (adulta e infanto-juvenil), considerando um intervalo de confiança de 95%.

Modelo	Acurácia ponderada ± margem de erro
Módulo Pornográfico (proposto)	**98,47% ± 0,19%**
Yahoo! NSFW	91,45% ± 0,35%

Fonte: Autor.

Considerando um intervalo de confiança de 95%, verificou-se que o Módulo Pornográfico no contexto de pornografia geral (i.e. pornografia adulta e infanto-juvenil) apresentou desempenho ligeiramente inferior quando comparado ao contexto de apenas pornografia adulta. Dado os intervalos dos dois cenários, respectivamente 98,28%-98,66% e 98,83%-99,25%, percebe-se que por 0,27% não é possível afirmar o Módulo Pornográfico atua de maneira igual em ambos os cenários.

Entretanto, ainda levando em consideração o intervalo de confiança de 95% e o contexto de pornografia geral (i.e. pornografia adulta e infanto-juvenil), verificou-se a superioridade do Módulo Pornográfico em relação ao serviço de moderação de conteúdo Yahoo! NSFW.

Visando esmiuçar os resultados obtidos nesta etapa, foi exposta na Tabela 9.3.3 a matriz de confusão do módulo pornográfico no contexto de pornografia geral (i.e. pornografia adulta e infanto-juvenil).

Tabela 9.14: Matriz de confusão do módulo pornográfico proposto no contexto de pornografia geral (i.e. pornografia adulta e infanto-juvenil).

		Predição	
		Pornografia Geral	Não Pornografia
Verdade	Pornografia Geral	18499 (98,51%)	280 (1,49%)
	Não Pornografia	79 (1,56%)	4973 (98,44%)

Fonte: Autor.

9.3.4 Estimação de Idade Real Facial e Análise Comparativa com Pesquisas do Estado da Arte

Com base na metodologia apresentada no Capítulo 7 (Aprimorando a Estimativa de Idade Real a Partir de Dados de Idade Aparente), os experimentos foram conduzidos com o objetivo de avaliar qual método apresentou menor erro médio absoluto na estimativa de idade real facial.

Como esperado, e com base em observações de trabalhos anteriores, os métodos aprimorados (i.e. Gaussiana Estática, Gaussiana Dinâmica, COURA e DCOURA) superaram os métodos padrão (i.e. Classificador Multiclasse, Regressor e Classificador Justo). Ressalta-se que, com a intenção de obter resultados mais precisos e confiáveis, todos os experimentos foram repetidos 10 vezes para cada método. A Tabela 9.15 apresenta um resumo do resultado experimental, por meio das médias do erro médio absoluto de cada método.

Tabela 9.15: Médias do erro médio absoluto de cada método, considerando uma confiança de 95%.

Método	Erro absoluto médio $\pm$ margem de erro
Classificador Multiclasse	12,851$\pm$0,158
Regressor	5,184$\pm$0,100
Classificador Justo	5,776$\pm$0,106
Gaussiana Estática	5,093$\pm$0,099
Gaussiana Dinâmica	4,856$\pm$0,097
COURA	5,076$\pm$0,099
DCOURA	**4,847$\pm$0,097**

Fonte: Autor.

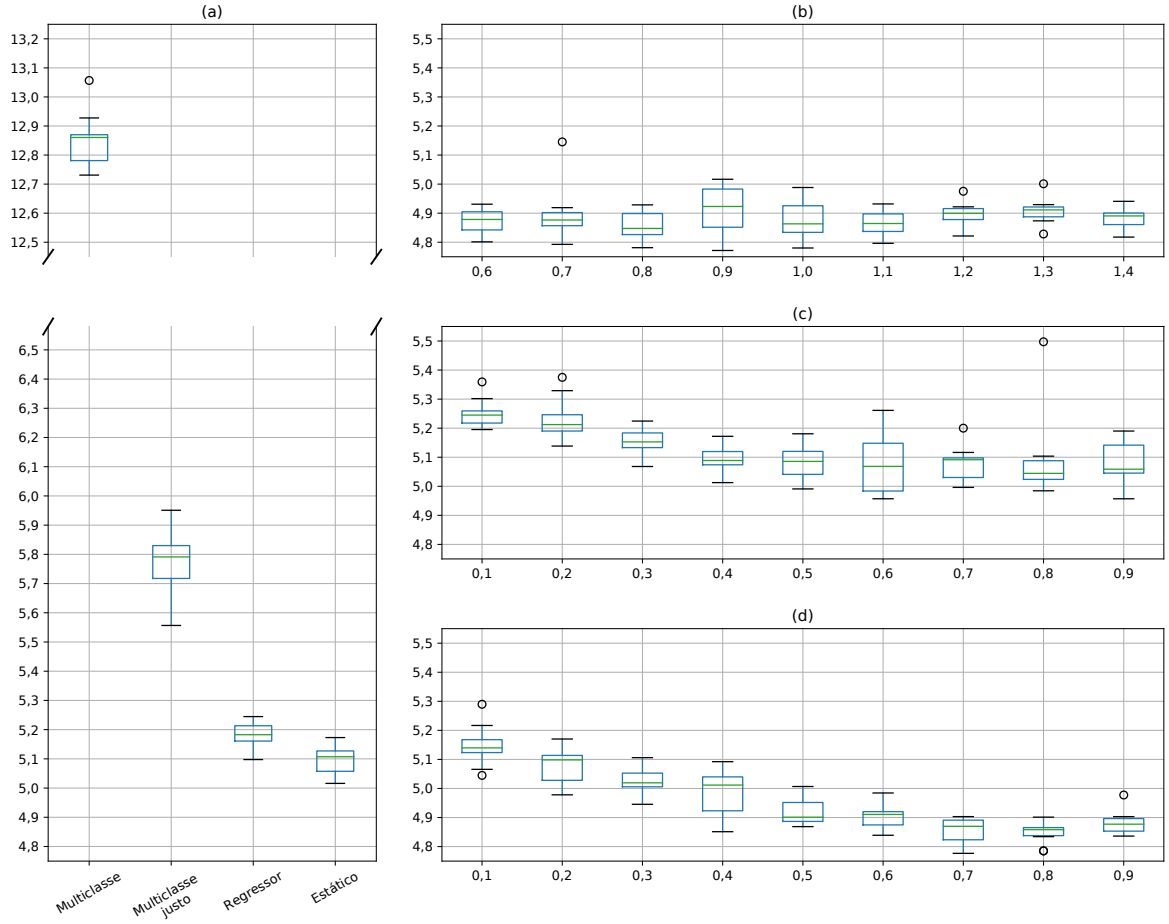
Utilizando o erro médio absoluto como métrica de avaliação, o Classificador Multiclasse apresentou a pior média, 12,851. O Classificador Justo foi capaz de melhorar a metodologia de classificação de múltiplas classes, alcançando uma média de 5,776, mas não foi o suficiente para superar o Regressor que atingiu a pontuação média de 5,184. Embora o método Gaussiano Estático tenha apresentado o pior resultado entre os métodos aprimorados, ainda apresentou melhor resultado comparado aos métodos padrão, com uma média de 5,093, entretanto, considerando uma confiança de 95%, apresentou resultado equivalente ao Regressor. A parte (a) na Figura 9.5 mostra os diagramas de caixa correspondentes aos resultados relativos a esses quatro métodos.

O método Gaussiano Dinâmico melhorou os resultados alcançados pelo método Gaussiano Estático, atingindo uma média de 4,856. Esse resultado foi obtido definindo o valor de α em 0,8. Os resultados com todas as variações de α podem ser vistos na parte (b) da Figura 9.5.

O método COURA, considerando uma confiança de 95%, mostrou-se equivalente aos métodos Gaussiano Estático e Regressor, obtendo uma média de 5,076 usando λ igual a 0,6. Por outro lado, o método DCOURA superou todos os métodos, alcançando uma média de 4,847, entretanto, considerando uma confiança de 95%, se equipara ao método Gaussiano

Dinâmico. Esse resultado foi obtido usando os hiperparâmetros λ e α iguais a 0,8. Os resultados com todas as variações de λ podem ser vistos nas partes (c) e (d) da Figura 9.5.

Figura 9.5: Diagramas de caixa ilustrando os erros absolutos médios, no eixo y, para cada um dos métodos testados. A parte (a) ilustra os diferentes comportamentos do Classificador Multiclasse, Classificador Justo, Regressor e Gaussiana Estática. Todos esses métodos não dependem de hiperparâmetros adicionais. A parte (b) mostra os diagramas de caixa para o método Gaussiano Dinâmico, variando os valores do hiperparâmetro σ no eixo x. As partes (c) e (d) mostram diagramas de caixa para os métodos COURA e DCOURA, respectivamente, variando os valores do hiperparâmetro λ no eixo x.



Fonte: Autor.

Os melhores métodos propostos, o Gaussiano Dinâmico e DCOURA, também foram comparados com outros métodos encontrados na literatura. Em todos os casos, a tarefa alvo é a estimativa da idade real, usando a base de dados APPA-REAL. Conforme mostrado na Tabela 9.16, os referidos métodos propostos superaram todos os demais estudos inseridos no estado da arte, reduzindo em até aproximadamente 8,5% o menor erro alcançado até então.

Tabela 9.16: Comparativo entre os métodos do estado da arte na em estimativa de idade real usando o conjunto de dados APPA-REAL e os métodos propostos DCOURA e Gaussiana Dinâmica, considerando uma confiança de 95%.

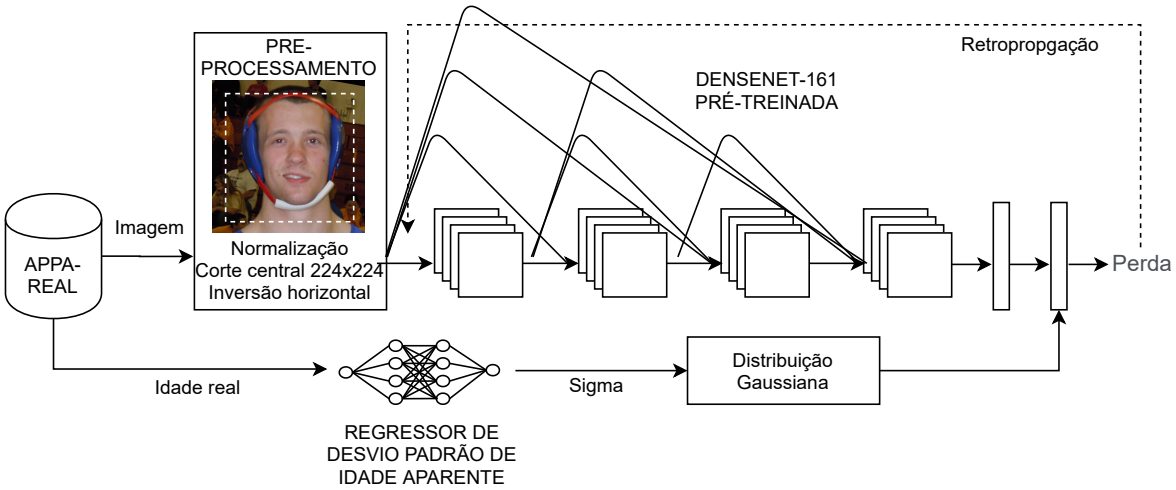
Modelo	Erro absoluto médio ± margem de erro
Clapes et al. (2018)	13,577 ± 0,162
Jacques Junior et al. (2019)	7,356 ± 0,119
Bešenic, Ahlberg e Pandzic (2019)	6,260 ± 0,110
Rondeau e Alvarez (2018)	5,434 ± 0,103
Singh et al. (2018)	5,423 ± 0,103
Agustsson et al. (2017)	5,296 ± 0,101
Gaussiano Dinâmico (Proposto)	**4,856 ± 0,097**
DCOURA (Proposto)	**4,847 ± 0,097**

Fonte: Autor.

9.3.5 Estimação de Idade Real Facial Aplicada à Menoridade Penal

O método proposto da Gaussiana Dinâmica, por apresentar (i) flexibilidade de poder ser treinado com faces rotuladas apenas com idade real (apesar de ter sido aprimorada com dados de idade aparente) e (ii) resultado equivalente ao também proposto método DCOURA, considerando um intervalo de confiança de 95%, foi escolhido para compor o Módulo Facial, responsável pela estimação de idade real facial aplicada à menoridade penal. A Figura 9.6 ilustra toda a arquitetura da etapa de treinamento desse método.

Figura 9.6: Fluxograma da fase de treinamento do método da Gaussiana Dinâmica.



Fonte: Autor.

Visando, simultaneamente, analisar e melhorar o desempenho do modelo adotado a estimação de menoridade penal, foram realizados quatro experimentos em sequência que, de maneira gulosa, mantém as configurações dos melhores resultados atingidos.

I - Ampliação da quantidade de imagens da base de dados APPA-REAL para treinamento do modelo

Foi utilizada uma nova distribuição de dados para o treinamento e a validação dos dados, com o intuito de verificar se o aumento no número de imagens na etapa de treinamento influenciaria positivamente nos testes. Portanto, as 7.591 imagens passaram pelo mesmo pré-processamento utilizado no Módulo Facial da arquitetura proposta no Capítulo 5. As imagens brutas são submetidas a um detector de faces MTCNN (ZHANG et al., 2016), em que somente as faces com confiança igual ou superior a 0,95 são aceitas. Caso haja mais de uma face, é escolhida a com maior área. Em seguida a face é rotacionada para alinhamento horizontal dos olhos e tem a sua área ampliada em 40%, como pode ser visto na Figura 5.2. Por fim, foram obtidas 7.464 imagens que foram utilizadas para o novo treinamento, utilizando validação cruzada com dez subconjuntos (*10-fold*), ou seja, 90% dos dados para treinamento e 10% dos dados para validação.

Sendo assim, com o objetivo de atestar a presença de menores de idade em imagens pornográficas, realizou-se comparativo entre: (i) o resultado obtido pelo modelo originalmente treinado e validado, respectivamente, com 4.113 e 1.978 imagens, seguindo o protocolo adotado em Agustsson et al. (2017); e (ii) a média dos resultados obtidos por cada um dos dez modelos oriundos da validação cruzada utilizando as 7.464 imagens. Os dados utilizados para os testes em questão são oriundos da junção das bases de dados PEDA 376K e da base de dados pornográfica infanto-juvenil.

Para a realização de tal experimento, se fez necessário o uso de imagens que apresentassem pelo menos uma face. Sendo assim, das 18.780 imagens pornográficas utilizadas para avaliação dos modelos, como pode ser observado na Tabela 9.1, 11.619 imagens possuem faces, em que 9.629 retratam crianças e/ou adolescentes e 1.990 mostram apenas adultos, conforme a Tabela 9.1.

Desse universo de 11.619 imagens pornográficas contendo faces, foi utilizada a parcela de 50% dos dados referentes à validação da determinação de menoridade penal, ajustando assim o Módulo Facial. Os 50% remanescentes foram utilizados no teste final da arquitetura de detecção de pornografia infanto-juvenil, como pode ser visto na Tabela 9.17.

Baseado nos resultados obtidos e utilizando uma confiança de 95%, conforme a Tabela 9.18, foi possível afirmar que o uso de uma maior quantidade de imagens da mesma

Tabela 9.17: Distribuição das imagens pornográficas nos conjuntos de validação e teste.

	Validação (50%)		Teste (50%)		Total (100%)	
Imagens pornográficas com faces	Inf.-juvenil	Adulto	Inf.-juvenil	Adulto	Inf.-juvenil	Adulto
	4.815	995	4.814	995	9.629	1.990

Fonte: Autor.

base de dados APPA-REAL para o treinamento do modelo aumentou a acurácia ponderada da determinação de menoridade penal das imagens pornográficas contendo faces.

Tabela 9.18: Comparativo entre o uso da distribuição original do conjunto de dados APPA-REAL e do uso com maior quantidade de dados para treinamento para a determinação de menoridade penal em faces, considerando uma confiança de 95%.

Dados utilizados da APPA-REAL	Acurácia ponderada $\pm$ margem de erro
Distribuição original	83,06% $\pm$ 0,30%
Treinamento 90% - Validação 10%	**84,98% $\pm$ 0,29%**

Fonte: Autor.

II - Análise Comparativa Entre o Método Gaussiana Dinâmico e Um Classificador de Duas Classes

Para a realização da análise comparativa em questão, foi utilizado para fins de teste o conjunto de validação das imagens pornográficas com faces, conforme Tabela 9.17, como realizado no experimento anterior. Os modelos a serem comparados são: (i) o modelo Gaussiana Dinâmica (ii) e um classificador de duas classes (i.e. menor de idade e maior de idade), utilizando como função de perda a entropia-cruzada. Salienta-se que, visando à isonomia do experimento, ambos os modelos utilizaram a mesma rede neural convolucional e pré-treinamento com a base de dados IMDB-WIKI, assim como a nova e bem sucedida distribuição de treinamento/validação para a APPA-REAL obtida no experimento anterior. Dado o uso da validação cruzada com dez subconjuntos (*10-fold*), foram calculadas as médias das acurácias ponderadas dos testes realizados nos dez modelos para cada uma das técnicas, como podem ser observadas na Tabela 9.19 na sequência.

De acordo com os resultados obtidos e considerando uma confiança de 95%, ficou evidente a superioridade do método proposto Gaussiana Dinâmica quando comparado a um classificador de duas classes, utilizando entropia cruzada.

Tabela 9.19: Comparativo entre o método proposto Gaussiana Dinâmica e o método de classificador de duas classes, para a determinação de menoridade penal em faces, considerando uma confiança de 95%.

Técnica	Acurácia ponderada ± margem de erro
Gaussiana Dinâmica	**84,98% ± 0,29%**
Classificador de duas classes	75,82% ± 0,35%

Fonte: Autor.

III - Uso de Outras Bases de Dados Faciais e Suas Combinações

Com o objetivo de verificar se o uso de mais de uma base de dados facial apresentaria melhoria na determinação da menoridade penal, foi proposta a utilização de uma combinação de bases que apresentassem rótulos de idade real. Sendo assim, além da já utilizada APPA-REAL, foram utilizadas as bases de dados FGNET e UTKFace.

Para a captura das faces de cada base de dados, foi aplicado o mesmo protocolo de pré-processamento utilizado no Experimento I dessa seção. Em seguida, para cada cenário oriundo da combinação, as faces obtidas de cada base de dados foram agrupadas e randomizadas. Para cada um dos cenários, foram criados dez modelos resultantes da validação cruzada com dez subconjuntos (*10-fold*), em que o resultado final se deu por meio da média dos resultados obtidos por cada um dos dez modelo. Os experimentos foram realizados utilizando como dados de teste o conjunto de validação das imagens pornográficas com faces, conforme Tabela 9.17, como realizado no Experimento I.

De acordo com a Tabela 9.20 e considerando um intervalo de confiança de 95%, o uso de uma combinação de base de dados de faces não foi capaz de aprimorar a acurácia ponderada relativa a determinação de menoridade penal por meio de faces.

Tabela 9.20: Comparativo entre a combinação de base de dados de faces, para a determinação de menoridade penal em faces, considerando uma confiança de 95%.

Combinação das Base de Dados	Acurácia ponderada ± margem de erro
APPA-REAL	**84,98% ± 0,29%**
FGNET	78,54% ± 0,33%
UTKFace	83,09% ± 0,30%
APPA-REAL + FGNET	84,67% ± 0,29%
APPA-REAL + UTKFace	84,16% ± 0,30%
FGNET + UTKFace	84,25% ± 0,30%
APPA-REAL + FGNET + UTKFace	84,53% ± 0,29%

Fonte: Autor.

IV - Análise Comparativa Entre o Método Gaussiana Dinâmico e o Estado da Arte

Para fins de comparação com outras pesquisas inseridas no estado da arte na estimativa de idade real em faces, foi utilizado o modelo gratuito disponibilizado pela empresa Spectro[1]. Trata-se de uma empresa que utiliza as mais avançadas técnicas de Inteligência Artificial aplicadas à Visão Computacional. Foi mantida a nova proporção bem sucedida de dados proposta para treinamento do modelo utilizando a base de dados APPA-REAL. Assim como nos experimentos anteriores, a proporção de 90% dos dados para treinamento e 10% para validação foi realizada por meio do uso da validação cruzada com dez subconjuntos. A obtenção do resultado final se deu pela média das acurácias ponderadas de cada um dos dez modelos.

Para a realização do experimento, foi utilizado como dados de teste o conjunto de validação das imagens pornográficas com faces, conforme Tabela 9.17, como realizado no Experimento I. Por meio dos resultados expostos na Tabela 9.21, é possível afirmar que técnica Gaussiana Dinâmica, considerando uma confiança de 95%, apresentou melhores resultados em comparação ao modelo proposto pela Spectro.

Tabela 9.21: Comparativo entre o método proposto Gaussiana Dinâmica e o modelo disponibilizado pela Spectro, para a determinação de menoridade penal em faces, considerando uma confiança de 95%.

Técnica	Acurácia ponderada ± margem de erro
Gaussiana Dinâmica	**84,98% ± 0,29%**
Spectrico	72,75% ± 1,14%

Fonte: Autor.

Com objetivo de promover uma análise mais detalhada dos resultados obtidos nesta etapa, foi exposta na Tabela 9.3.5 a matriz de confusão do módulo facial que diferencia menores e maiores de idade em imagens pornográficas.

Tabela 9.22: Matriz de confusão, referente aos dados de validação, do módulo facial que diferencia menores e maiores de idade em imagens pornográficas.

		Predição	
		Menor de Idade	Maior de Idade
Verdade	Menor de Idade	3968 (82,46%)	846 (17,57%)
	Maior de Idade	124 (12,46%)	871 (87,54%)

Fonte: Autor.

[1] http://spectrico.com/

Por fim, foi possível também inferir, de maneira indireta, que o modelo proposto apresentou melhores resultados quando comparados a outras empresas também inseridas no estado da arte na estimativa de idade real por meio de faces. Essa indução se deu com base na avaliação comparativa realizada pela Spectro que, como pode ser observado na Tabela 9.23, afirma ter apresentado o menor erro médio na estimativa de idade real facial dentre os modelos das empresas envolvidas.

Tabela 9.23: Avaliação comparativa realizada pela Spectro com demais empresas inseridas no estado da arte na estimativa de idade real por meio de faces.

Empresa	Erro médio estimado (em anos)
Spectrico	7,1
Microsoft	8,5
Skybiometry	7,2
Sighthound	8,1
Eyedea	8,0
VisageCloud	9,1

Fonte: Adaptada de http://spectrico.com/pricing-age-gender.html.

9.3.6 Classificador de Menoridade Penal

Fundamentado na metodologia apresentada no Capítulo 8 (Classificador de Menoridade Penal), os experimentos foram conduzidos com o objetivo de avaliar se o método proposto era capaz de aprimorar os resultados do Módulo Facial, no que diz respeito à probabilidade de uma determinada face pertencer a um indivíduo menor de idade.

Com relação aos dados, foi utilizado no experimento o conjunto de testes das imagens pornográficas com faces (5.809 imagens), conforme a Tabela 9.17. Salienta-se que esses dados não foram utilizados em nenhum momento para treinamento ou validação da arquitetura proposta no Capítulo 5, o que enviesaria o resultado final. A Tabela 9.24 mostra os resultados obtidos nas abordagens, considerando um intervalo de confiança de 95%.

A partir desses resultados e levando em consideração um intervalo de confiança de 95%, foi possível concluir que o Classificador de Menoridade Penal Simples foi capaz de diminuir o erro absoluto médio em relação à abordagem intrínseca ao Módulo Facial, o Somatório de Probabilidades. Além disso, a proposta do Classificador de Menoridade Penal Composto, utilizando os metadados da face detectada (i.e. confiança, tamanho da face e distâncias entre

Tabela 9.24: Erro médio absoluto (EMA) e margem de erro em cada uma das abordagens propostas, considerando um intervalo de confiança de 95%.

Características adotadas	EMA $\pm$ margem de erro
Somatório de Probabilidades	0,2058 $\pm$ 0,0117
CMP Simples (Idade)	0,1502 $\pm$ 0,0100
CMP Composto (Idade + Confiança)	0,1496 $\pm$ 0,0099
CMP Composto (Idade + Tamanho)	0,1300 $\pm$ 0,0092
CMP Composto (Idade + Distâncias)	0,1424 $\pm$ 0,0097
CMP Composto (Idade + Confiança + Tamanho)	0,1270 $\pm$ 0,0092
CMP Composto (Idade + Confiança + Distâncias)	0,1453 $\pm$ 0,0098
CMP Composto (Idade + Confiança + Tamanho + Distâncias)	**0,1243 $\pm$ 0,0091**

Fonte: Autor.

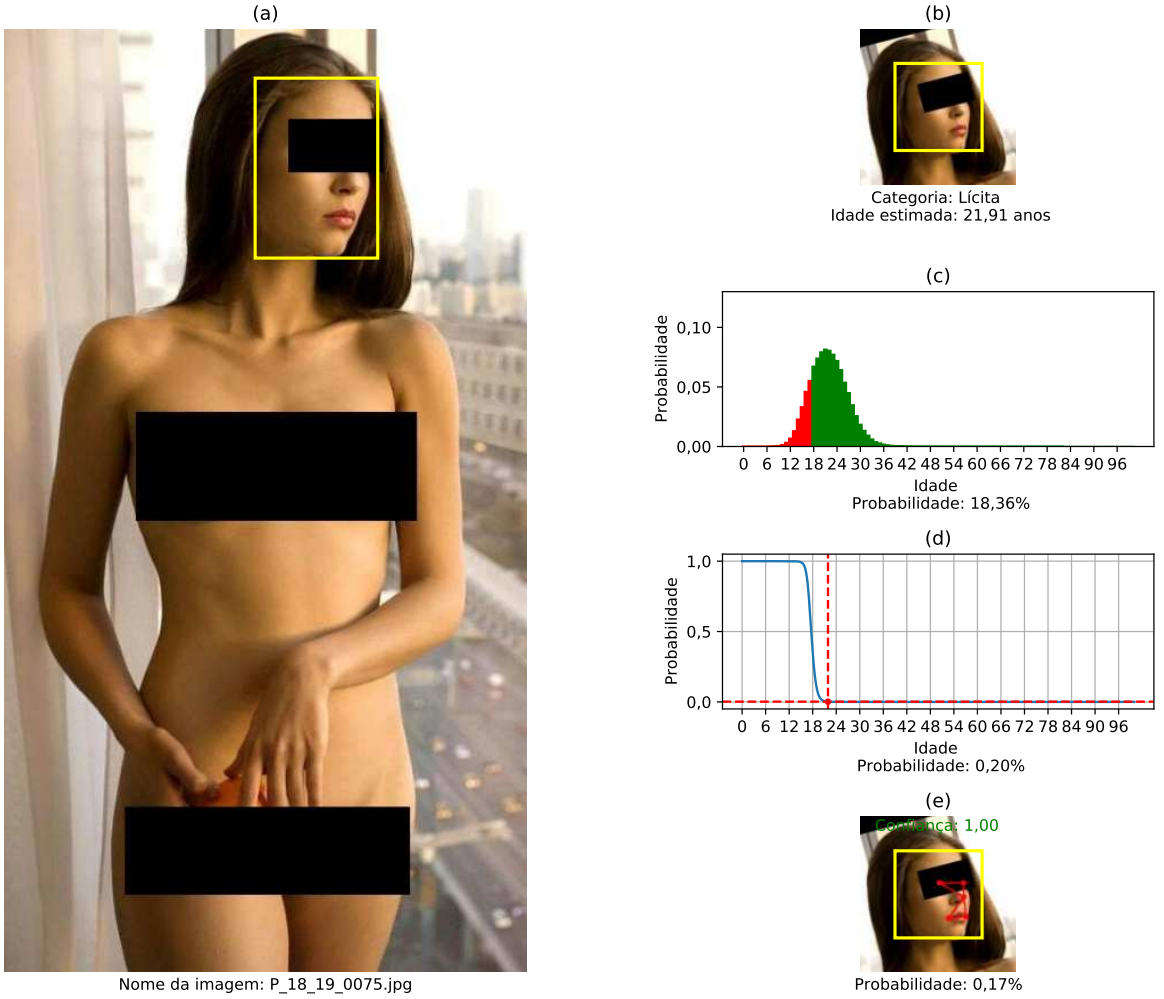
pontos fiduciais) apresentou um erro absoluto médio ainda menor quando comparado ao Somatório de Probabilidades.

As Figuras 9.7, 9.8, 9.9 e 9.10 mostram imagens submetidas à classificação de menoridade penal e as probabilidades de cada abordagem (i.e. Somatório de Probabilidades, Estimadores de Menoridade Penal Simples e Composto) pertencerem a indivíduos menores de idade.

9.3.7 Detecção de Pornografia Infanto-juvenil

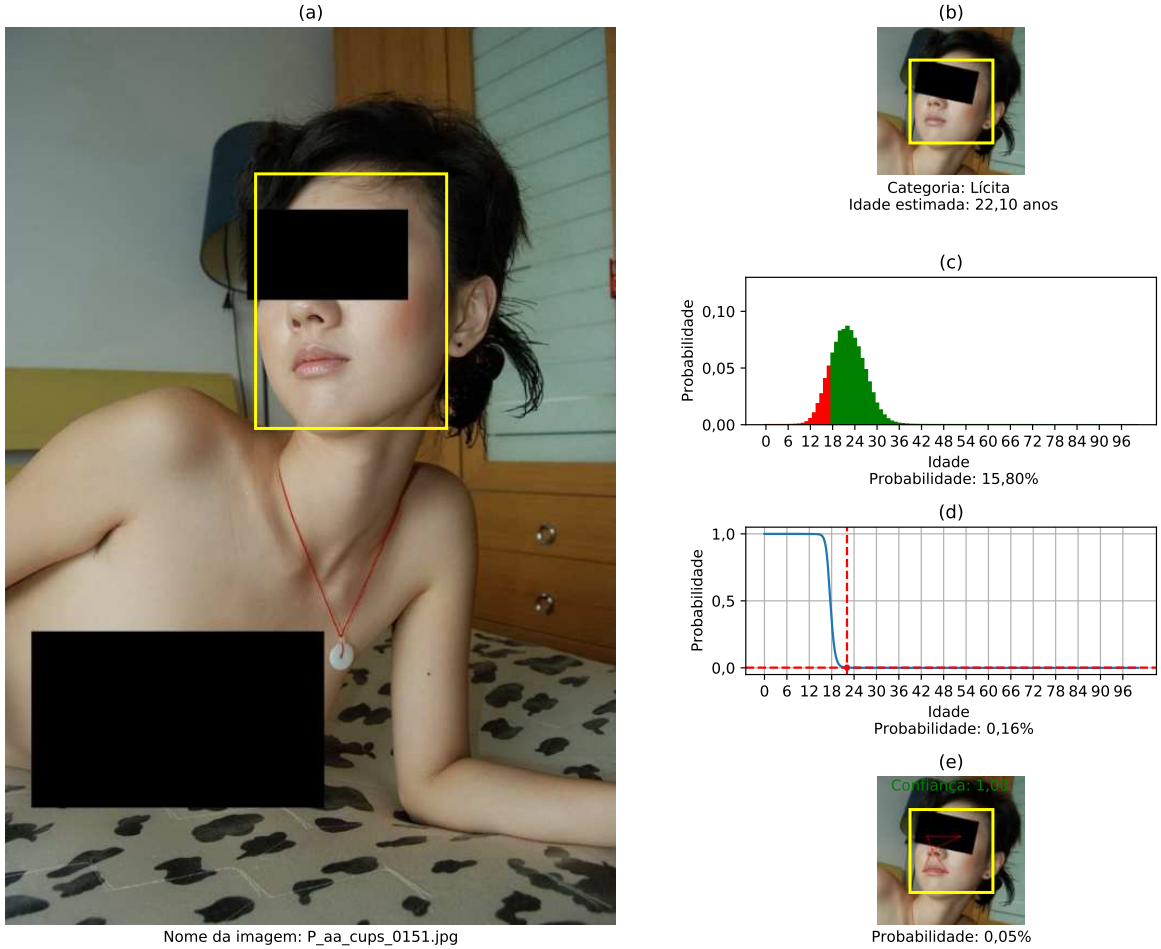
Para a realização da avaliação final da arquitetura proposta, foram determinadas duas categorias: (i) conteúdo lícito, composto por imagens não pornográficas e pornografia adulta com faces e (ii) conteúdo ilícito, composto apenas por imagens pornográficas infanto-juvenil com faces. A categoria lícita foi formada por 5.052 imagens não pornográficas de ambas as bases de dados (vide Tabela 9.1) e 995 imagens de pornografia adulta, do conjunto de testes das imagens pornográficas contendo faces (vide Tabela 9.17). A categoria ilícita foi formada por 4.814 imagens contendo pornografia infanto-juvenil, oriundas do conjunto de testes das imagens pornográficas contendo faces (vide Tabela 9.17). A distribuição detalhada das imagens das categorias lícitas e ilícitas pode ser observada na Tabela 9.25.

Figura 9.7: Em (a) é mostrada uma imagem, retratando indivíduo maior de idade, submetida à estimação de idade, por meio do reconhecimento facial. Em (b) é ilustrada a face detectada já pré-processada. Em (c), (d) e (e) são expostas as probabilidades obtidas por cada uma das abordagens de o indivíduo possuir menos de dezoito anos (Somatório de Probabilidades, Classificador de Menoridade Penal Simples e Composto, respectivamente). Salienta-se que as imagens originais não possuem tarjas de preservação de identidade e intimidade, utilizadas apenas para exposição neste trabalho.



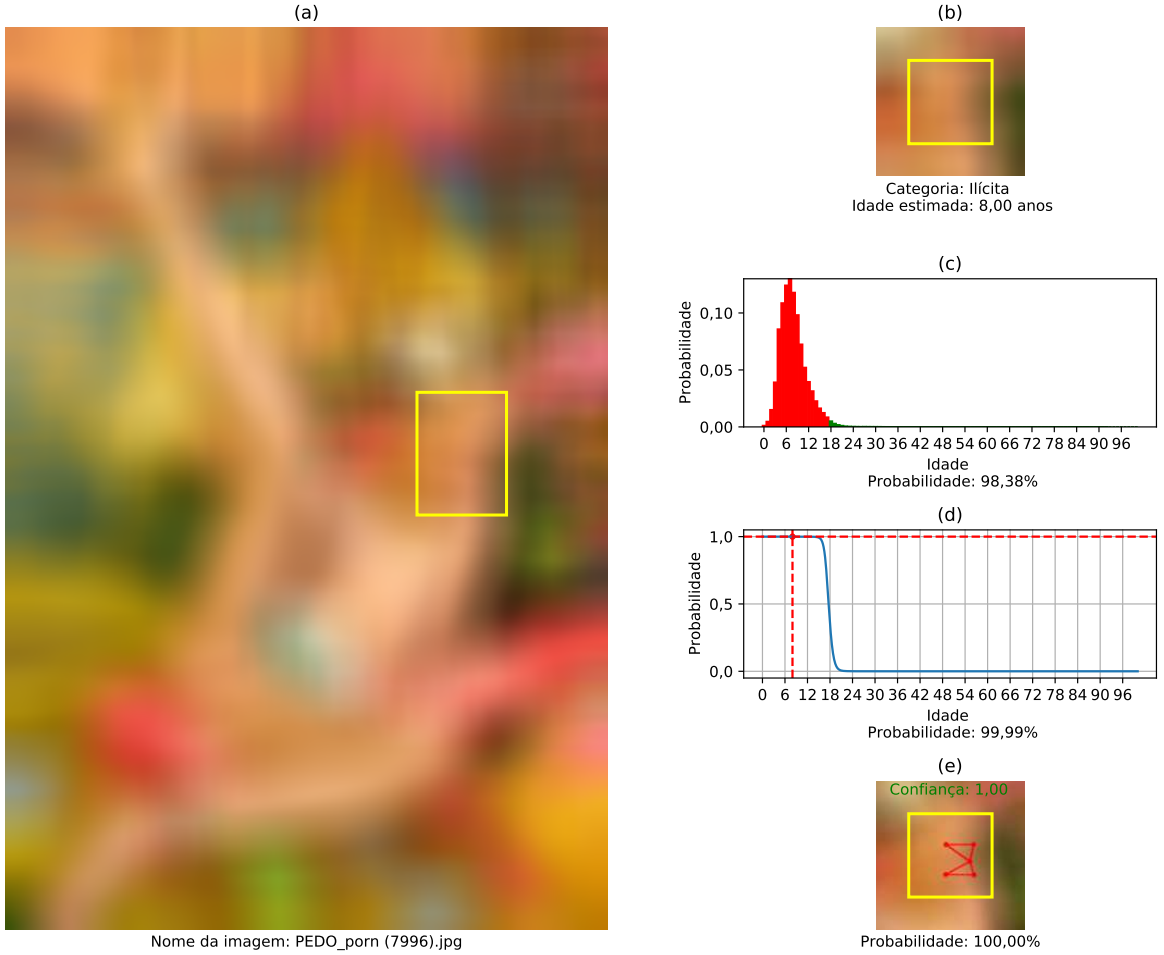
Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Figura 9.8: Em (a) é mostrada uma imagem, retratando indivíduo maior de idade, submetida à estimação de idade, por meio do reconhecimento facial. Em (b) é ilustrada a face detectada já pré-processada. Em (c), (d) e (e) são expostas as probabilidades obtidas por cada uma das abordagens de o indivíduo possuir menos de dezoito anos (Somatório de Probabilidades, Classificador de Menoridade Penal Simples e Composto, respectivamente). Salienta-se que as imagens originais não possuem tarjas de preservação de identidade e intimidade, utilizadas apenas para exposição neste trabalho.



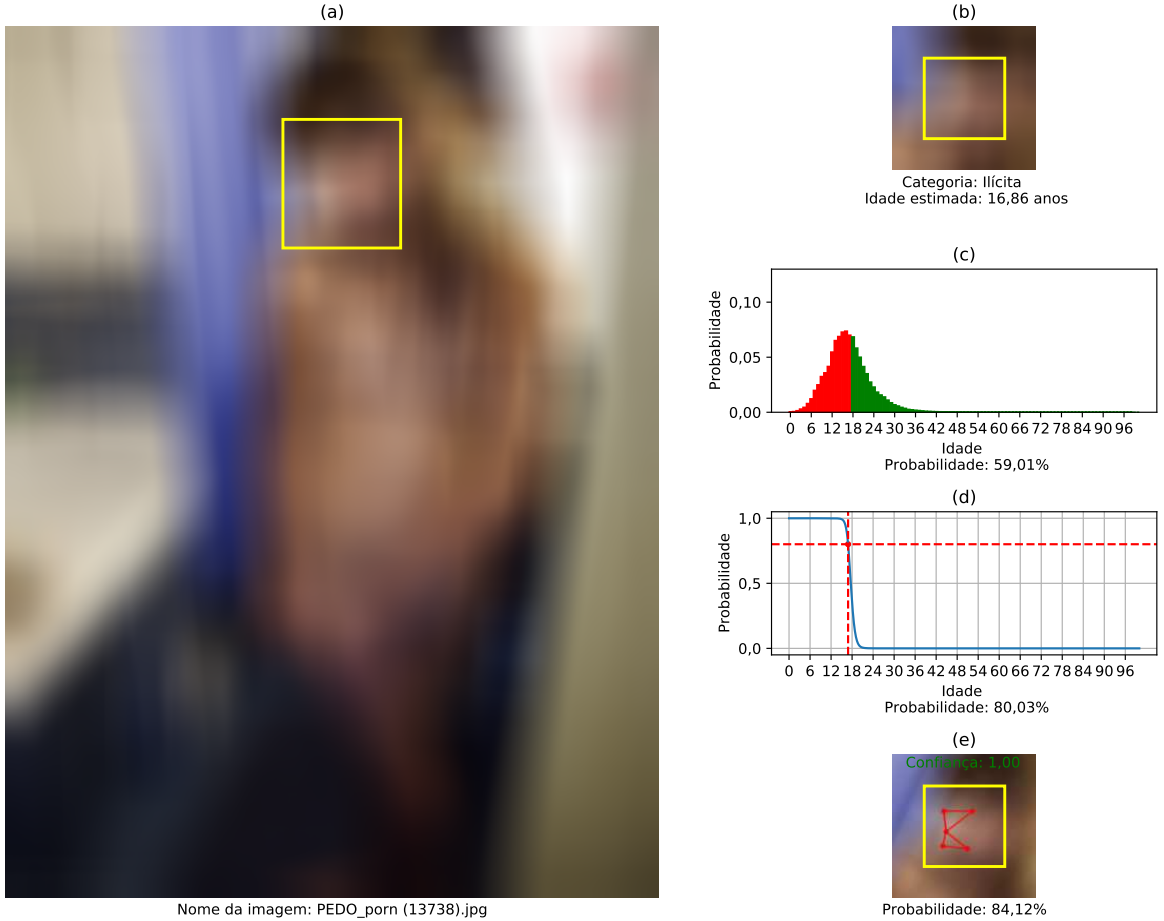
Fonte: Adaptada de Moreira, Pereira e Alvarez (2020).

Figura 9.9: Em (a) é mostrada uma imagem, retratando indivíduo menor de idade, submetida à estimação de idade, por meio do reconhecimento facial. Em (b) é ilustrada a face detectada já pré-processada. Em (c), (d) e (e) são expostas as probabilidades obtidas por cada uma das abordagens de o indivíduo possuir menos de dezoito anos (Somatório de Probabilidades, Classificador de Menoridade Penal Simples e Composto, respectivamente). Salienta-se que as imagens foram borradas por questões legais e com intuito de preservar a identidade e a intimidade das crianças e adolescentes.



Fonte: Autor.

Figura 9.10: Em (a) é mostrada uma imagem, retratando indivíduo menor de idade, submetida à estimação de idade, por meio do reconhecimento facial. Em (b) é ilustrada a face detectada já pré-processada. Em (c), (d) e (e) são expostas as probabilidades obtidas por cada uma das abordagens de o indivíduo possuir menos de dezoito anos (Somatório de Probabilidades, Classificador de Menoridade Penal Simples e Composto, respectivamente). Salienta-se que as imagens foram borradas por questões legais e com intuito de preservar a identidade e a intimidade das crianças e adolescentes.



Fonte: Autor.

Tabela 9.25: Distribuição das imagens das categorias lícita e ilícita para avaliação da arquitetura proposta.

Categoria	Lícita		Ilícita
Tipo	Não pornografia	Pornografia adulta	Pornografia infanto-juvenil
Quantidade de Imagens	5.052	995	4.814

Fonte: Autor.

Dado que a arquitetura proposta não determina de maneira estrita se uma imagem contém pornografia infanto-juvenil ou não, mas sim por meio de probabilidades de acordo com as faces detectadas (pelos motivos já apresentados), foi necessário adaptar a arquitetura proposta para viabilizar a análise comparativa com os estudos inseridos no estado da arte, visto que os estudos apresentam resultados binários (i.e conteúdo lícito, conteúdo ilícito). Sendo assim, apenas para a realização deste experimento, arquitetura proposta não utilizou o Classificador de Menoridade Penal, verificando apenas as idades estimadas pelo Módulo Facial e retornando um resultado binário (i.e conteúdo lícito, conteúdo ilícito), como pode ser observado no fluxograma contido na Figura 9.11.

Mesmo com as adaptações, não foi possível realizar uma comparação totalmente isonômica da pesquisa proposta com outras inseridas no estado da arte da detecção de pornografia infanto-juvenil. O impedimento legal de possuir e/ou compartilhar imagens contendo crianças e/ou adolescentes em situações pornográficas dificulta a criação de uma base de dados disponível com esse conteúdo para a reprodutibilidade experimental.

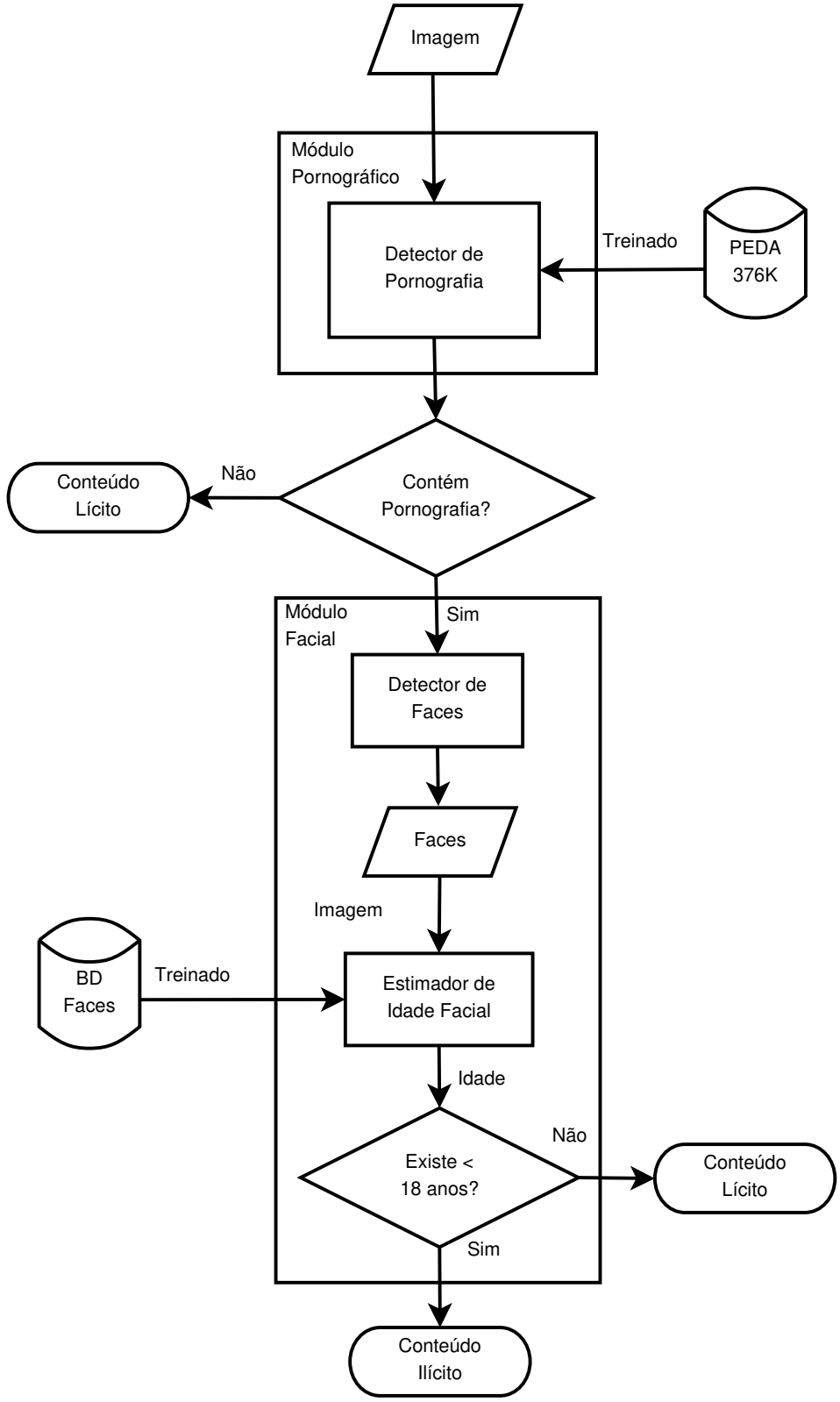
Ainda assim, foi realizada uma análise comparativa com pesquisas inseridas no estado da arte que utilizaram a mesma métrica de avaliação e categorias adotadas pela abordagem proposta. A Tabela 9.26 mostra os resultados atingidos por cada estudo, sendo possível afirmar que o modelo proposto supera as pesquisas inseridas no estado da arte, considerando um intervalo de confiança de 95%.

Tabela 9.26: Análise comparativa entre a arquitetura proposta para detecção de pornografia infanto-juvenil e demais trabalhos relacionados, considerando uma confiança de 95%.

Modelo	Acurácia $\pm$ margem de erro
Proposto	**90,30% $\pm$ 0,56%**
Vitorino et al. (2018)	86,50% $\pm$ 0,48%
Macedo, Costa e Santos (2018)	79,84% $\pm$ 1,70%
Sae-Bae et al. (2014)	74,19% $\pm$ 8,4%

Fonte: Autor.

Figura 9.11: Fluxograma do processo de detecção de imagens contendo pornografia infanto-juvenil sem o Classificador de Menoridade Penal.



Fonte: Autor.

Por fim, por meio da matriz de confusão exposta na Tabela 9.3.7, foram discriminados de maneira detalhada os resultados da abordagem proposta para detecção de pornografia infanto-juvenil, que diferencia imagens lícitas de ilícitas.

Tabela 9.27: Matriz de confusão, referente aos dados de teste, da abordagem proposta para detecção de pornografia infanto-juvenil, que diferencia imagens lícitas de ilícitas.

		Predição	
		Imagem Ilícita	Imagem Lícita
Verdade	Imagem Ilícita	3975 (82,57%)	839 (17,43%)
	Imagem Lícita	214 (3,54%)	5833 (96,46%)

Fonte: Autor.

9.4 Análise de Desempenho Computacional

Com o intuito de avaliar o tempo despendido para a detecção de pornografia infanto-juvenil por meio da abordagem proposta, foi realizada uma análise de desempenho computacional que avaliou o tempo médio demandado em etapas específicas. Essa análise também contemplou dois diferentes cenários para processamento dos dados: (i) uso exclusivo do processador (*Central Processing Unit - CPU*) e (ii) uso aliado a uma unidade de processamento gráfico (*Graphics Processing Unit - GPU*). A Tabela 9.28 mostra o tempo médio, em segundos, requerido para cada etapa avaliada, de acordo com o tipo de imagem classificada e do uso ou não da unidade de processamento gráfico (GPU).

Ademais, baseado nos tempos médios de execução de cada etapa da abordagem proposta, foi realizado o cálculo de estimativa de tempo necessário para a classificação das imagens em um exame pericial real. Sendo assim, utilizou-se como exemplo o exame pericial exposto no estudo de Polastro e Eleutério (2010), em que um dispositivo de armazenamento contendo aproximadamente 300.000 imagens apresentou apenas 148 imagens contendo pornografia. Ou seja, existiam 299.852 imagens não pornográficas e 148 imagens pornográficas.

Utilizando apenas o processador (CPU) para o processamento dos dados, o tempo estimado T_{CPU}, em segundos, para a realização da tarefa em questão pode ser visualizado na Equação (9.8). O tempo estimado T_{GPU} para a realização da mesma tarefa, utilizando também uma unidade de processamento gráfico (GPU) pode ser visto na Equação (9.9).

$$T_{CPU} = (299.852 \times 0,165676s) + (148 \times 1,477341s) = 49.897,0700s \quad (9.8)$$

$$T_{GPU} = (299.852 \times 0,002726s) + (148 \times 0,242491s) = 853,2781s \quad (9.9)$$

Tabela 9.28: Análise de desempenho computacional expondo o tempo médio, em segundos, requerido para cada etapa avaliada, de acordo com o tipo de imagem classificada e do uso ou não da unidade de processamento gráfico (GPU).

Etapa \ Cenário		CPU	GPU
Módulo Pornográfico		0,165676s	0,002726s
Módulo Facial	Detecção da Face	0,937685s	0,181911s
	Rotação da Face	0,052484s	0,052484s*
	Estimação da Idade Real	0,322523s	0,005355s
Classificador de Menoridade Penal		0,000016s	0,000016s*
Abordagem Completa	Imagem Não Pornográfica	0,165676s	0,002726s
	Imagem Pornográfica	1,477341s	0,242491s

*Foi utilizado o processador (CPU) para a realização dessa etapa.

Fonte: Autor.

Podemos constatar que, apesar de não ser essencial na etapa de predição (teste), o uso de uma unidade de processamento gráfico (GPU) reduziu a estimativa da realização da referida tarefa de aproximadamente 13 horas e 52 minutos para pouco mais de 14 minutos, viabilizando a realização de exames periciais complexos em um curto espaço de tempo.

9.5 Considerações Finais

Inicialmente, foi possível verificar que uma abordagem simples baseada em aprendizado profundo, mesmo com as limitações do uso reduzido de dados e sem o ajuste fino dos hiperparâmetros, conseguiu resultados equivalentes aos atingidos por modelos bem ajustados baseados em aprendizado de máquina tradicional na detecção de imagens pornográficas. Esse resultado foi crucial para determinar o uso de aprendizado profundo para a detecção de

imagens dessa natureza, visto que o uso de uma rede neural profunda complexa, com uma grande quantidade de dados para treinamento e um ajuste fino bem executado traria melhores resultados, baseado no levantamento bibliográfico realizado.

Foi proposta então a criação da *Pornographic and Explicit Database 376K* (MOREIRA; PEREIRA; ALVAREZ, 2020), uma base de dados contendo quase 151 mil imagens pornográficas e mais de 225 mil imagens não pornográficas, em que foi utilizado um critério extremamente objetivo para a categorização das imagens. O resultado obtido, se igualando a um e superando quatro de cinco serviços de moderação de imagens inseridos no estado da arte, validou a importância do uso de uma grande quantidade de dados bem classificados e de um modelo bem ajustado, visto que foi utilizada uma rede neural convolucional já consolidada, sem nenhuma alteração específica para a detecção de pornografia.

O Módulo Pornográfico proposto, treinado apenas com pornografia adulta, foi capaz de atingir uma acurácia ponderada de 98,47% na diferenciação entre pornografia e não pornografia, mesmo também tendo sido utilizadas imagens contendo crianças e adolescentes. Um resultado superior à abordagem proposta pela Yahoo! NSFW, que quando submetido às mesmas imagens de teste, atingiu uma acurácia ponderada de 91,45%.

Com o intuito de identificar crianças e adolescentes nas imagens preditas como pornográficas, foi proposto o Módulo Facial, que tem a função estimar a idade facial de um determinado indivíduo. Esse módulo baseia-se na técnica proposta no Capítulo 7 (Aprimorando a Estimativa de Idade Real a Partir de Dados de Idade Aparente). Foi possível então comprovar que a técnica proposta atingiu melhores resultados que as pesquisas inseridas no estado da arte na estimativa de idade real que utilizaram a base de dados facial APPA-REAL.

Confirmado o potencial da técnica proposta para estimativa de idade facial, essa técnica foi utilizada para a determinação da menoridade penal ou não dos indivíduos que tiveram suas faces detectadas nas imagens pornográficas. Em seguida, verificou-se que a utilização de uma maior quantidade de dados na etapa de treinamento traria melhores resultados, visto que não mais era necessário o uso dos conjuntos de treino, validação e teste pré-determinados em Agustsson et al. (2017) para comparação com outros trabalhos.

Apesar de ter sido comprovada a superioridade da técnica Gaussiana Dinâmica em comparação ao Classificador Multiclasse, tratava-se de um cenário teoricamente mais complexo, em que existiam 101 classes, uma para cada idade variando de zero a cem ([0,100]). O novo

cenário contemplava apenas duas classes: (i) menor de dezoito anos e (ii) maior ou igual a dezoito anos. Portanto, foi realizada uma análise experimental entre esses dois métodos e foi confirmada a superioridade da técnica Gaussiana Dinâmica em relação ao classificador de duas classes.

Cogitou-se também a utilização de outras bases de dados faciais disponíveis para ampliar a quantidade de imagens utilizadas do treinamento do modelo proposto. Dessa forma, questionou-se se o uso de alguma das combinações de três bases de dados faciais (APPA-REAL, FGNET e UTKFace) atingiria maior acurácia ponderada quando comparada ao uso isolado da base de dados APPA-REAL. Por fim, verificou-se que nenhuma das combinações foi capaz de superar a acurácia ponderada já atingida.

Mesmo tendo mostrado resultados superiores em relação a outros trabalhos inseridos no estado da arte que utilizaram a base de dados APPA-REAL, sentiu-se a necessidade de realizar uma análise comparativa neste novo cenário em que o objetivo é determinar a menoridade penal por meio da estimativa de idade facial. O modelo baseado em aprendizado profundo disponibilizado pela empresa Spectro que, de acordo com a Tabela 9.23, apresenta menor erro médio dentre as empresas citadas, foi submetido a uma análise experimental utilizando o mesmo conjunto de teste aplicado no Módulo Facial. Sendo assim, verificou-se que o Módulo Facial proposto apresentou maior acurácia ponderada quando comparado ao modelo proposto pela empresa Spectro.

Ademais, mesmo não sendo uma comparação totalmente isonômica, dada a não padronização das imagens utilizadas para a avaliação de cada uma das pesquisas, o resultado atingido pela abordagem proposta mostrou-se superior aos alcançados por outros estudos inseridos no estado da arte na detecção de pornografia infanto-juvenil.

Por fim, foi realizada uma análise de desempenho computacional da abordagem proposta. Essa análise foi capaz de, por meio da medição dos tempos de execução de diferentes etapas do processo, mostrar a importância do uso de unidades de processamento gráfico na etapa de predição (teste). A adoção desse recurso de hardware diminuiu em mais de 58 vezes o tempo de execução total do processo.

Capítulo 10

Considerações Finais

Neste trabalho de tese, foi apresentada uma arquitetura que tem como objetivo detectar imagens pornográficas e, dada a existência de faces, inferir a probabilidade de haver menores de idade nessas imagens. Essa abordagem possibilita direcionar o Perito Criminal na conclusão sobre a existência ou não de imagens contendo pornografia infanto-juvenil em dispositivos de armazenamento submetidos a exame pericial.

No estudo em questão, foram levantadas Questões de Pesquisa que nortearam o desenvolvimento da técnica proposta. A Questão de Pesquisa **QP1**, questiona a possibilidade da construção de um modelo que supere o estado da arte na detecção de pornografia infanto-juvenil sem o uso de imagens dessa natureza. A resposta dessa questão está estritamente relacionada às respostas das Questões de Pesquisa **QP2** e **QP3**.

A Questão de Pesquisa **QP2**, que questiona a possibilidade do desenvolvimento de um modelo com resultado compatível ao estado da arte na detecção de imagens pornográficas, foi respondida por meio de análise experimental atestando que o modelo proposto para a detecção de pornografia se igualou a um dos serviços de moderação de imagens inseridos no estado da arte e superou outros quatro. Dessa forma, a resposta para a Questão de Pesquisa **QP2** é **SIM**.

A Questão e Pesquisa **QP3**, que questiona a possibilidade do desenvolvimento de um modelo com resultado superior ao estado da arte na estimação de idade real facial, foi respondida por meio de análise experimental atestando que o modelo proposto para a estimação de idade real em faces superou todas as pesquisas do estado da arte avaliadas nesta tese que utilizaram a base de dados APPA-REAL. Dessa forma, a resposta para a Questão de Pesquisa

QP3 é **SIM**.

Mesmo diante dos resultados positivos das Questões de Pesquisa **PQ2** e **PQ3**, foi necessário realizar uma análise diante dos cenários específicos do problema: (i) o desempenho do detector de pornografia quando também submetido a imagens contendo pornografia infanto-juvenil e (ii) o desempenho do estimador de idade facial quando submetido à determinação de menoridade penal.

Apesar do desempenho do Módulo Pornográfico não ser o mesmo quando também submetido a imagens contendo pornografia infanto-juvenil, verificou-se que, por uma diferença de 0,27% e considerando um intervalo de confiança de 95%, não foi possível afirmar que o Módulo Pornográfico atua de maneira igual quando submetido apenas a imagens de pornografia adulta. Entretanto, mostrou-se que o desempenho do referido módulo nesse cenário de pornografia geral (i.e. pornografia adulta e infanto-juvenil) atingiu uma acurácia ponderada de 98,47%, superando o modelo inserido no estado da arte proposto pela Empresa Yahoo! em mais de 7%.

Com relação ao Módulo Facial, comprovou-se a superioridade dos métodos propostos DCOURA e Gaussiana Dinâmica em relação aos estimadores de idade real facial contidos na literatura que utilizam a base de dados APPA-REAL. Sendo assim, o método da Gaussiana Dinâmica foi adotado para estimar a menoridade penal baseado nas faces dos indivíduos nas imagens pornográficas. O método proposto apresentou superioridade quando comparado a modelos inseridos no estado da arte também nesse novo cenário.

Porém, mesmo superando o estado da arte na estimação de idade real em faces, não foi possível determinar de maneira exata se um indivíduo é menor de idade, o que é fortemente recomendado no âmbito da Pericia Criminal, dada as importantes consequências do resultado final (i.e. incriminar ou inocentar um suspeito). Sendo assim, foi levantada a Questão de Pesquisa **QP4**, que questiona como determinar a presença ou não de crianças e/ou adolescentes nas imagens pornográficas, por meio do uso de uma probabilidade em vez de uma decisão estritamente binária.

Foi realizada então uma análise comparativa entre a abordagem já existente no Módulo Facial, um somatório das probabilidades individuais de cada idade já existente na saída da rede, e a proposta do Classificador de Menoridade Penal. A abordagem proposta minimizou o erro médio absoluto das predições sobre menoridade penal das faces detectadas dos

indivíduos. Dessa forma, foi possível além de responder como **SIM** a Questão de Pesquisa **QP4**, aprimorar o resultado final. Por fim, mesmo não sendo possível realizar uma análise comparativa isonômica, visto que cada pesquisa utilizou um conjunto de imagens distinto para realização da sua etapa de teste, foi possível afirmar que o trabalho proposto superou estudos inseridos no estado da arte na detecção de pornografia infanto-juvenil.

Baseado no exposto, visto que a arquitetura proposta utiliza os referidos módulos em série, é possível responder como **SIM** a Questão de Pesquisa **QP1**, que questiona se é possível detectar, no nível do estado da arte, pornografia infanto-juvenil em imagens sem o uso de imagens dessa natureza para a construção do modelo.

10.1 Limitações da Abordagem

Baseado nas questões éticas de uso de imagens nas técnicas de Inteligência Artificial, assim como direitos autorais e de imagens, não foi possível disponibilizar a base de dados pornográfica *Pornographic and Explicit Dataset 367K* (PEDA 376K). Essa limitação, além de inviabilizar uma enorme quantidade de imagens para treinamento de modelos em pesquisas voltadas para a detecção de pornografia, impossibilita também a sua reprodutibilidade experimental, que tem como objetivo comparar e validar resultados alcançados.

Por questões legais, especificamente por meio da Lei N° 11.829, de 25 de novembro de 2008 (BRASIL, 2008) que altera a redação original do Estatuto da Criança de do Adolescente (ECA) (BRASIL, 1990), a impossibilidade de disponibilizar a base de dados contendo imagens pornográficas infanto-juvenil trouxe as mesmas limitações destacadas no parágrafo anterior (treinamento de modelos e reprodutibilidade experimental) para os estudos voltados especificamente para essa área.

Constatou-se que, mesmo utilizando técnicas inseridas no estado da arte para a detecção de pornografia e estimação de idade real baseada em faces, a detecção e rotação das imagens das faces trouxe algum prejuízo no desempenho computacional da abordagem como um todo.

Ademais, observou-se uma lacuna importante com relação à detecção de imagens pornográficas sem a existência de face humana. A abordagem proposta não é capaz de estimar se imagens dessa natureza retratam crianças e/ou adolescente, sendo classificada como

"CONTEÚDO INDETERMINADO" para posterior análise de um Perito Oficial Criminal. Essa fragilidade se deu pela inexistência de recursos computacionais adequados para o treinamento de modelos baseados em aprendizado profundo nas dependências do Núcleo de Criminalística de Campina Grande - PB. A não existência de uma permissão legal para transportar as imagens contendo pornografia infanto-juvenil ali existentes, tampouco transmiti-las via Internet impossibilitou que um modelo treinado com imagens ilícitas pudesse ser treinado em outro ambiente detentor de recursos adequados. Por fim, salienta-se que os modelos baseados em aprendizado profundo, tanto para a detecção de pornografia adulta, quanto para a estimação de idade real em faces, dada a inexistência de qualquer impedimento de portar imagens lícitas, foram treinados em locais alheios ao Núcleo de Criminalística de Campina Grande - PB.

10.2 Sugestões Para Pesquisas Futuras

Com o objetivo de minimizar as limitações encontradas na abordagem proposta, assim como explorar possibilidades reveladas pelo estudo em questão, foram sugeridas as seguintes pesquisas futuras:

- Dada a impossibilidade da reprodutibilidade experimental por outros estudos, visto que não foi possível compartilhar a base de dado *Pornographic and Explicit Dataset 367K* (PEDA 376K), por questões éticas e de direitos autorais, e a base de dados contendo pornografia infanto-juvenil, por questões legais, propõe-se a criação de um serviço capaz de receber modelos previamente treinados para que os mesmos sejam submetidos a uma avaliação experimental utilizando o mesmo conjunto de imagens de teste utilizados nesta pesquisa;

- Baseado no desempenho computacional observado na detecção e rotação das imagens das faces para estimação de idade real, sugere-se a (i) realização de um estudo que avalie, além da acurácia aliada, o desempenho computacional de métodos para detecção de face e o (ii) uso da unidade de processamento gráfico (GPU) para o rotacionamento das faces;

- Fundamentado na impraticabilidade da atual proposta em determinar a presença de

menores de idade em imagens pornográficas que não contêm qualquer face humana, sugere-se a criação de um módulo complementar treinado apenas com imagens pornográficas sem faces, estas divididas em duas categorias: (i) lícitas (pornografia adulta) e (ii) ilícitas (pornografia infanto-juvenil). Dessa forma, o referido módulo seria capaz de categorizar como pornografia adulta ou pornografia infanto-juvenil as imagens que atualmente recebem o rótulo de "CONTEÚDO INDETERMINADO";

- Com o objetivo de aprimorar a rotulação das imagens em ambas as bases de dados criadas, sugere-se que essas sejam reanalisada por mais de um categorizador, sendo especificamente um Perito Oficial Criminal quando tratar da base de dados contendo pornografia infanto-juvenil;

- Visto a não utilização da técnica de aumento de dados nos modelos destinados à detecção e pornografia, dada a possibilidade de descaracterização de imagens consideradas pornográficas por meio do uso de algumas técnicas (i.e. corte, rotação), sugere-se que sejam utilizados nesses modelos o aumento de dados sem restrições nas imagens não pornográficas e o uso técnicas nas imagens pornográficas que não sejam capazes de alterar sua categoria (i.e. variação de cores, rotações sem cortes, espelhamento);

- Ampliar a análise de desempenho computacional, não se atendo apenas ao desempenho do modelo adotado, mas realizando uma análise comparativa contemplando diferentes modelos baseados em aprendizado profundo;

- Utilizar um detector de faces capaz de especificar a posição da face e/ou disponibilizar uma maior quantidade de pontos fiduciais com o objetivo de aprimorar o resultado final do Classificador de Menoridade Penal proposto;

- Estender a abordagem proposta para a detecção de conteúdo pornográfico infanto-juvenil em vídeos, analisando a possível existência de uma relação temporal entre os quadros mais significativos por meio do uso de uma Rede Neural Recorrente.

10.3 Trabalhos Realizados

Ao longo do período de realização do doutorado, o aluno Danilo Coura Moreira produziu publicações de trabalhos relevantes em conferências e periódicos, além de ter realizado o Programa de Doutorado-sanduíche no Exterior no período entre agosto de 2019 e maio de 2020 na *University of Rhode Island* - URI, nos Estados Unidos da América.

Em 2018, foi publicado como sendo primeiro autor, no XVIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (QUALIS B3), o estudo "*A Forensic Nudity Detector Based on Machine Learning*" que, baseado em detecção de pele e em aprendizado de máquina tradicional, propôs a diferenciação entre imagens ofensivas (pornográficas e sensuais) e imagens não ofensivas (de conteúdo geral) (MOREIRA; FECHINE, 2018a).

Ainda em 2018, foi publicado como sendo primeiro autor, na *International Joint Conference on Neural Networks* (QUALIS A1), o estudo "*A Machine Learning-based Forensic Discriminator of Pornographic and Bikini Images*" que, baseado em detecção de pele e em aprendizado de máquina tradicional, propôs a diferenciação entre imagens contendo pornografia adulta e imagens de mulheres trajando biquíni (MOREIRA; FECHINE, 2018b).

Em 2020, foi publicado como sendo primeiro autor, na *International Joint Conference on Neural Networks* (QUALIS A1), o estudo "*A Novel Dataset for Deep-learning Based Porn-detectors*" que propôs a construção de uma grande base de dados pornográfica com imagens criteriosamente classificadas entre pornografia e não pornografia. Ademais, sugeriu-se o uso de uma estratégia gulosa em três etapas para a definição da melhor Rede Neural Convolucional, assim como seus hiperparâmetros, para tratar do problema de detecção de imagens pornográficas. Por fim, os resultados de serviços de moderação de imagens foram padronizados com o objetivo de viabilizar a comparação entre os resultados obtidos e esses serviços inseridos no estado da arte da detecção de pornografia (MOREIRA; PEREIRA; ALVAREZ, 2020).

Em 2020, foi realizada uma parceria entre o departamento de Ciência da Computação da Universidade Federal de Campina Grande (UFCG) e os departamentos de Biologia e de Ciência da Computação e Estatística da *University of Rhode Island* (URI), em que foi publicado como sendo segundo autor, no periódico *PLoS Neglected Tropical Diseases* (QUALIS

B2), o estudo "Delimiting Cryptic Morphological Variation Among Human Malaria Vector Species Using Convolutional Neural Networks", que, por meio de aprendizado profundo, foi capaz de distinguir diferentes espécies de mosquitos, mesmo algumas até então tendo sido diferenciadas apenas por meio de exames laboratoriais. Apesar de o foco do estudo em questão não ter relação com o objeto de estudo do doutorado, foram utilizadas técnicas de Inteligência Artificial aplicada à Visão Computacional bastante semelhantes, o que chancela a menção da publicação deste estudo (COURET et al., 2020).

Em 2021, foi submetido e encontrando-se ainda em análise, como sendo primeiro autor, para a *International Joint Conference on Neural Networks* (QUALIS A1), o estudo "*Improving Real Age Estimation from Apparent Age Data*" que, propõe novas técnicas para o aprimoramento da estimativa de idade real, utilizando distribuições de probabilidades discretas, dados de idade aparente e funções de custo compostas.

Bibliografia

AGUSTSSON, E.; TIMOFTE, R.; ESCALERA, S.; BARO, X.; GUYON, I.; ROTHE, R. Apparent and real age estimation in still images with deep residual regressors on appa-real database. In: IEEE. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. [S.l.], 2017. p. 87–94.

AL-MOHAIR, H. K.; SALEH, J. M.; SUANDI, S. A. Hybrid human skin detection using neural network and k-means clustering technique. *Applied Soft Computing*, Elsevier, v. 33, p. 337–347, 2015.

AL-MOHAIR, H. K.; SALEH, J. M.; SUANDI, S. A. Hybrid human skin detection using neural network and k-means clustering technique. *Applied Soft Computing*, Elsevier, v. 33, p. 337–347, 2015.

ALVAREZ, S. P. *RedLight an efficient illicit image detection application for law enforcement*. [S.l.]: University of Rhode Island, 2012.

ANTIPOV, G.; BACCOUCHE, M.; BERRANI, S.-A.; DUGELAY, J.-L. Apparent age estimation from face images combining general and children-specialized deep learning models. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2016. p. 96–104.

AP-APID, R. An algorithm for nudity detection. In: *Proceedings of the 5th Philippine Computing Science Congress*. Cebu City, Philippines: [s.n.], 2005.

AVILA, S.; THOME, N.; CORD, M.; VALLE, E.; ARAÚJO, A. D. A. Pooling in image representation: The visual codeword point of view. *Computer Vision and Image Understanding*, Elsevier, v. 117, n. 5, p. 453–465, 2013.

BASILIO, J. A. M.; TORRES, G. A.; PÉREZ, G. S.; MEDINA, L. K. T.; MEANA, H. M. P. Explicit image detection using ycbcr space color model as skin detection. In: *Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2011. (AMERICAN-MATH'11/CEA'11), p. 123–128. ISBN 978-960-474-270-7. Disponível em: <http://dl.acm.org/citation.cfm?id=1959666.1959689>.

BEŠENIC, K.; AHLBERG, J.; PANDZIC, I. S. Unsupervised facial biometric data filtering for age and gender estimation. In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*. [S.l.: s.n.], 2019. p. 209–217.

BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006.

BRANCATI, N.; PIETRO, G. D.; FRUCCI, M.; GALLO, L. Human skin detection through correlation rules between the ycb and ycr subspaces based on dynamic color clustering. *Computer Vision and Image Understanding*, Elsevier, v. 155, p. 33–42, 2017.

BRASIL. Decreto-lei n. 2.848, de 7 de dezembro de 1940. In: *Código Penal*. Brasília, DF: [s.n.], 1940.

BRASIL. Decreto-lei n. 3.689, de 3 de outubro de 1941. In: *Código de Processo Penal*. Brasília, DF: [s.n.], 1941.

BRASIL. Lei n. 8.069, de 13 de julho de 1990. In: *Estatuto da Criança e do Adolescente*. Brasília, DF: [s.n.], 1990.

BRASIL. Lei n. 10.764, de 12 de novembro de 2003. In: *Código Penal*. Brasília, DF: [s.n.], 2003.

BRASIL. Lei n. 11.829, de 25 de novembro de 2008. In: *Altera a Lei n. 8.069, de 13 de julho de 1990 - Estatuto da Criança e do Adolescente, para aprimorar o combate à produção, venda e distribuição de pornografia infantil, bem como criminalizar a aquisição e a posse de tal material e outras condutas relacionadas à pedofilia na internet*. Brasília, DF: [s.n.], 2008.

CAPPELLARI; VERONEZI, M. S. A pedofilia na pós-modernidade: um problema que ultrapassa a cibercultura. *Em Questão*, v. 11, p. 67–82, 2005. ISSN 1807-8893. Acessado em 05 de abril de 2019. Disponível em: <http://www.redalyc.org/articulo.oa?id=465645952005>.

CARLETTI, V.; GRECO, A.; PERCANNELLA, G.; VENTO, M. Age from faces in the deep learning revolution. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, 2019.

CASTRILLÓN-SANTANA, M.; LORENZO-NAVARRO, J.; TRAVIESO-GONZÁLEZ, C. M.; FREIRE-OBREGÓN, D.; ALONSO-HERNANDEZ, J. B. Evaluation of local descriptors and cnns for non-adult detection in visual content. *Pattern Recognition Letters*, Elsevier, v. 113, p. 10–18, 2018.

CHASE, T.; HE, R.; HEGAZY, K. Cs 231n final project: Deep visual learning of reddit images. 2017.

CLAPES, A.; BILICI, O.; TEMIROVA, D.; AVOTS, E.; ANBARJAFARI, G.; ESCALERA, S. From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2018. p. 2373–2382.

COMASCHI, F. *Robust online face detection and tracking*. Tese (Doutorado) — PhD thesis, Technische Universiteit Eindhoven, 2016.

COURET, J.; MOREIRA, D. C.; BERNIER, D.; LOBERTI, A. M.; DOTSON, E. M.; ALVAREZ, M. Delimiting cryptic morphological variation among human malaria vector species using convolutional neural networks. *PLOS Neglected Tropical Diseases*, Public Library of Science San Francisco, CA USA, v. 14, n. 12, p. e0008904, 2020.

COX, D. R. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 20, n. 2, p. 215–232, 1958.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255.

DESHPANDE, A. *A Beginner's Guide To Understanding Convolutional Neural Networks*. 2016. <https://adeshpande3.github.io/A-Beginner\%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>. Acessado em 18 de abril de 2018.

DONG, M.; YIN, L.; DENG, W.; GUO, J.; XU, W. A computationally efficient algorithm for building statistical color models. In: IEEE. *2012 IEEE International Conference on Multimedia and Expo Workshops*. [S.l.], 2012. p. 402–407.

DU, Y.; CAI, Z.; GUAN, X.; LI, Q. Almost optimal skin detection approach within the gaussian framework. *Optical Engineering*, International Society for Optics and Photonics, v. 51, n. 2, p. 027007, 2012.

ELEUTÉRIO, P. M. da S.; MACHADO, M. P. *Desvendando a computação forense*. 1. ed. São Paulo: Novatec Editora, 2011.

ESCALERA, S.; BARÓ, X.; ESCALANTE, H. J.; GUYON, I. Chalearn looking at people: A review of events and resources. In: IEEE. *2017 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2017. p. 1594–1601.

ESCALERA, S.; TORRES, M. T.; MARTINEZ, B.; BARÓ, X.; ESCALANTE, H. J.; GUYON, I.; TZIMIROPOULOS, G.; CORNEOU, C.; OLIU, M.; BAGHERI, M. A. et al. Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2016. p. 1–8.

ESPÍNDULA, A. et al. *Local de Crime: isolamento e preservação, exames periciais e investigação criminal*. [S.l.]: Brasília: Alberi Espindula, 2007.

FELZENSZWALB, P. F.; GIRSHICK, R. B.; MCALLESTER, D.; RAMANAN, D. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 32, n. 9, p. 1627–1645, 2010.

FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Pictorial structures for object recognition. *International journal of computer vision*, Springer, v. 61, n. 1, p. 55–79, 2005.

GANGWAR, A.; FIDALGO, E.; ALEGRE, E.; GONZÁLEZ-CASTRO, V. Pornography and child sexual abuse detection in image and video: A comparative evaluation. In: *8th*

International Conference on Imaging for Crime Detection and Prevention (ICDP 2017). [S.l.: s.n.], 2017. p. 37–42.

GAO, B.-B.; XING, C.; XIE, C.-W.; WU, J.; GENG, X. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 6, p. 2825–2838, 2017.

GENG, Z.; ZHUO, L.; ZHANG, J.; LI, X. A comparative study of local feature extraction algorithms for web pornographic image recognition. In: IEEE. *2015 IEEE International Conference on Progress in Informatics and Computing (PIC)*. [S.l.], 2015. p. 87–92.

GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. [S.l.]: O'Reilly Media, 2019.

GHAHRAMANI, Z. Unsupervised learning. In: SPRINGER. *Summer School on Machine Learning*. [S.l.], 2003. p. 72–112.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1.

GREENBERGER, E.; JOSSELSON, R.; KNERR, C.; KNERR, B. The measurement and structure of psychosocial maturity. *Journal of Youth and Adolescence*, Springer, v. 4, n. 2, p. 127–143, 1975.

GRUS, J. *Data science from scratch: first principles with python*. [S.l.]: O'Reilly Media, 2019.

GUPTA, V.; SHARMA, D. A study of various face detection methods. *International Journal of Advanced Research in Computer and Communication Engineering*, v. 3, n. 5, p. 6694–6697, 2014.

HAMMOND, E. N. A.; ZHOU, S.; CHENG, H.; LIU, Q. Improving juvenile age estimation based on facial landmark points and gravity moment. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 18, p. 6227, 2020.

HASSNER, T. *Face Image Project*. 2014. <https://talhassner.github.io/home/projects/Adience/Adience-data.html>. Acessado em 22 de fevereiro de 2019.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770–778, 2016.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708.

HUANG, Y.; KONG, A. W. K. Using a cnn ensemble for detecting pornographic and upskirt images. In: IEEE. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. [S.l.], 2016. p. 1–7.

IRANIANGENIUS. */r/ListOfSubreddits*. 2018. <https://www.reddit.com/r/ListOfSubreddits/wiki/nsfw>. Acessado em 15 de outubro de 2018.

JACQUES JUNIOR, J. C.; OZCINAR, C.; MARJANOVIC, M.; BARÓ, X.; ANBARJAFARI, G.; ESCALERA, S. On the effect of age perception biases for real age regression. In: IEEE. *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. [S.l.], 2019. p. 1–8.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112.

JIA, Y.; SHELHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R.; GUADARRAMA, S.; DARRELL, T. Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*. [S.l.: s.n.], 2014. p. 675–678.

JONES, M. J.; REHG, J. M. Statistical color models with application to skin detection. *International Journal of Computer Vision*, Springer, v. 46, n. 1, p. 81–96, 2002.

JUNG, J.; MAKHIJANI, R.; MORLOT, A. Combining cnns for detecting pornography in the absence of labeled training data. 2017.

KANDEL, I.; CASTELLI, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT express*, Elsevier, v. 6, n. 4, p. 312–315, 2020.

KARAVARSAMIS, S.; NTARMOS, N.; BLEKAS, K.; PITAS, I. Detecting pornographic images by localizing skin rois. *International Journal of Digital Crime and Forensics (IJDCF)*, v. 5, p. 39–53, 01 2013.

KHAN, R.; HANBURY, A.; STÖTTINGER, J.; BAIS, A. Color based skin classification. *Pattern Recognition Letters*, Elsevier, v. 33, n. 2, p. 157–163, 2012.

KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, v. 160, p. 3–24, 2007.

KOVAC, J.; PEER, P.; SOLINA, F. Human skin color clustering for face detection. In: *The IEEE Region 8 EUROCON 2003. Computer as a Tool.* [S.l.: s.n.], 2003. v. 2, p. 144–148 vol.2.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. USA: Curran Associates Inc., 2012. (NIPS'12), p. 1097–1105. Disponível em: <http://dl.acm.org/citation.cfm?id=2999134.2999257>.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The annals of mathematical statistics*, JSTOR, v. 22, n. 1, p. 79–86, 1951.

LEVI, G.; HASSNER, T. Age and gender classification using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2015. p. 34–42.

LI, D.; MA, X.; REN, Y.; TENG, S.-W. Rectified softmax loss with all-sided cost sensitivity for age estimation. *IEEE Access*, IEEE, v. 8, p. 32551–32563, 2020.

LI, K.; XING, J.; LI, B.; HU, W. Bootstrapping deep feature hierarchy for pornographic image recognition. In: *2016 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2016. p. 4423–4427.

LIU, H.; SUN, P.; ZHANG, J.; WU, S.; YU, Z.; SUN, X. Similarity-aware and variational deep adversarial learning for robust facial age estimation. *IEEE Transactions on Multimedia*, IEEE, 2020.

LOPES, A. P.; AVILA, S. E. de; PEIXOTO, A. N.; OLIVEIRA, R. S.; ARAÚJO, A. d. A. A bag-of-features approach based on hue-sift descriptor for nude detection. In: IEEE. *2009 17th European Signal Processing Conference*. [S.l.], 2009. p. 1552–1556.

LU, D.; WENG, Q. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, Taylor & Francis, v. 28, n. 5, p. 823–870, 2007.

MA, B.; ZHANG, C.; CHEN, J.; QU, R.; XIAO, J.; CAO, X. Human skin detection via semantic constraint. In: ACM. *Proceedings of International Conference on Internet Multimedia Computing and Service*. [S.l.], 2014. p. 181.

MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. Nov, p. 2579–2605, 2008.

MACEDO, J.; COSTA, F.; SANTOS, J. A. dos. A benchmark methodology for child pornography detection. In: IEEE. *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2018. p. 455–462.

MAHADEOKAR, J.; PESAVENTO, G. Open sourcing a deep learning solution for detecting nsfw images. *Retrieved August*, v. 24, p. 2018, 2016. Disponível em: <https://yahooeng.tumblr.com/post/151148689421/open-sourcing-a-deep-learning-solution-for>.

MAHMOODI, M. R.; SAYEDI, S. M. A comprehensive survey on human skin detection. *International Journal of Image, Graphics & Signal Processing*, v. 8, n. 5, 2016.

MATHIAS, M.; BENENSON, R.; PEDERSOLI, M.; GOOL, L. V. Face detection without bells and whistles. In: . [S.l.: s.n.], 2014. v. 8692.

MEDINA, M. R.; PALLADINO, P. *Pornographic images jacking algorithm*. 2017. <https://github.com/alcuadrado/pija>. Accessed: November 17, 2017.

MIKESIZZ. *redditlist nsfw*. 2018. <http://www.ww.redditlist.com/nsfw>. Acessado em 15 de outubro de 2018.

MIKESIZZ. *redditlist sfw*. 2018. <http://www.ww.redditlist.com/sfw>. Acessado em 15 de outubro de 2018.

MILLER, G. A. *WordNet: An electronic lexical database*. [S.l.]: MIT press, 1998.

MOREIRA, D.; AVILA, S.; PEREZ, M.; MORAES, D.; TESTONI, V.; VALLE, E.; GOLDENSTEIN, S.; ROCHA, A. Pornography classification: The hidden clues in video space–time. *Forensic science international*, Elsevier, v. 268, p. 46–61, 2016.

MOREIRA, D. C.; FECHINE, J. M. A forensic nudity detector based on machine learning. In: SBC. *SBSeg 2018*. [S.l.], 2018. p. 267–280.

MOREIRA, D. C.; FECHINE, J. M. A machine learning-based forensic discriminator of pornographic and bikini images. In: IEEE. *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2018. p. 1–8.

MOREIRA, D. C.; PEREIRA, E. T.; ALVAREZ, M. Peda 376k: A novel dataset for deep-learning based porn-detectors. In: IEEE. *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2020. p. 1–8.

MUHAMMAD, B.; ABU-BAKAR, S. A. R. A hybrid skin color detection using hsv and ycgcr color space for face detection. In: IEEE. *Signal and Image Processing Applications (ICSIPA), 2015 IEEE International Conference on*. [S.l.], 2015. p. 95–98.

MUSTAFA, A. A.; ELBASHIR, A. A.; BABIKIR, S. F. A study of color constancy methods for skin detection. In: IEEE. *Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE), 2015 International Conference on*. [S.l.], 2015. p. 342–347.

MUSTAFAH, Y. M.; AZMAN, A. W. Skin region detector for real time face detection system. In: IEEE. *Computer and Communication Engineering (ICCCE), 2012 International Conference on*. [S.l.], 2012. p. 653–658.

MYDATAHACK. *Building AlexNet with Keras*. 2018. Acessado em 20 de janeiro de 2019. Disponível em: <https://www.mydatahack.com/building-alexnet-with-keras/>.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: FÜRNKRANZ, J.; JOACHIMS, T. (Ed.). *ICML*. [S.l.]: Omnipress, 2010. p. 807–814.

NAM, S. H.; KIM, Y. H.; TRUONG, N. Q.; CHOI, J.; PARK, K. R. Age estimation by super-resolution reconstruction based on adversarial networks. *IEEE Access*, IEEE, v. 8, p. 17103–17120, 2020.

NASCIMENTO FILHO, D. C. d. *Um Serviço para Monitoramento de Qualidade de Dados em Nuvem*. Tese (Doutorado) — Universidade Federal de Campina Grande (UFCG), Maio 2017. Proposta de Tese.

NIAN, F.; LI, T.; WANG, Y.; XU, M.; WU, J. Pornographic image detection utilizing deep convolutional neural networks. *Neurocomputing*, Elsevier, v. 210, p. 283–293, 2016.

OLIVEIRA, Í. de P.; MEDEIROS, J. L. P.; SOUSA, V. F. de; TEIXEIRA JÚNIOR, A. G.; PEREIRA, E. T.; GOMES, H. M. A data augmentation methodology to improve age estimation using convolutional neural networks. In: IEEE. *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2016. p. 88–95.

OU, X.; LING, H.; YU, H.; LI, P.; ZOU, F.; LIU, S. Adult image and video recognition by a deep multicontext network and fine-to-coarse strategy. *ACM Trans. Intell. Syst. Technol.*, ACM, New York, NY, USA, v. 8, n. 5, p. 68:1–68:25, jul. 2017. ISSN 2157-6904. Disponível em: <http://doi.acm.org/10.1145/3057733>.

PAN, H.; HAN, H.; SHAN, S.; CHEN, X. Mean-variance loss for deep age estimation from a face. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 5285–5294.

PEIXEIRO, M. *How to Build a Deep Neural Network Without a Framework*. 2019. <https://mc.ai/how-to-build-a-deep-neural-network-without-a-framework/>. Acessado em 20 de março de 2019.

PEREZ, M.; AVILA, S.; MOREIRA, D.; MORAES, D.; TESTONI, V.; VALLE, E.; GOLDENSTEIN, S.; ROCHA, A. Video pornography detection through deep learning techniques and motion information. *Neurocomputing*, Elsevier, v. 230, p. 279–293, 2017.

PLATZER, C.; STUETZ, M.; LINDORFER, M. Skin sheriff: A machine learning solution for detecting explicit images. In: *Proceedings of the 2Nd International Workshop on Security and Forensics in Communication Systems*. New York, NY, USA: ACM, 2014. (SFCS '14), p. 45–56. ISBN 978-1-4503-2802-9. Disponível em: <http://doi.acm.org/10.1145/2598918.2598920>.

PO, L. *Lenna 97: A complete story of Lenna*. 2001.

POLASTRO, M. d.; ELEUTÉRIO, P. M. da S. Nudetective: A forensic tool to help combat child pornography through automatic nudity detection. In: *2010 Workshops on Database and Expert Systems Applications*. [S.l.: s.n.], 2010. p. 349–353. ISSN 2378-3915.

POLLITT, M. Applying traditional forensic taxonomy to digital forensics. In: SPRINGER. *IFIP International Conference on Digital Forensics*. [S.l.], 2008. p. 17–26.

POUDEL, R. P.; ZHANG, J. J.; LIU, D.; NAIT-CHARIF, H. Skin color detection using region-based approach. *International Journal of Image Processing (IJIP)*, v. 7, n. 4, p. 385, 2013.

PRIJONO, B. *Student Notes: Convolutional Neural Networks (CNN) Introduction*. 2018. <https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/>. Acessado em 21 de abril de 2018.

PUTRO, M. D.; ADJI, T. B.; WINDURATNA, B. Adult image classifiers based on face detection using viola-jones method. In: *2015 1st International Conference on Wireless and Telematics (ICWT)*. [S.l.: s.n.], 2015. p. 1–6.

QIN, Z.; YU, F.; LIU, C.; CHEN, X. How convolutional neural network see the world-a survey of convolutional neural network visualization methods. *arXiv preprint arXiv:1804.11191*, 2018.

RAAIJMAKERS, S. Artificial intelligence for law enforcement: Challenges and opportunities. *IEEE Security & Privacy*, IEEE, v. 17, n. 5, p. 74–77, 2019.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Why should i trust you?: Explaining the predictions of any classifier. In: ACM. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. [S.l.], 2016. p. 1135–1144.

RONDEAU, J.; ALVAREZ, M. Deep modeling of human age guesses for apparent age estimation. In: IEEE. *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2018. p. 01–08.

ROTHE, R.; TIMOFTE, R.; GOOL, L. V. Dex: Deep expectation of apparent age from a single image. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. [S.l.: s.n.], 2015. p. 10–15.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.

SAE-BAE, N.; SUN, X.; SENCAR, H. T.; MEMON, N. D. Towards automatic detection of child pornography. In: IEEE. *2014 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2014. p. 5332–5336.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, IBM, v. 3, n. 3, p. 210–229, 1959.

SANJEEVI, M. *Chapter 4: Decision Trees Algorithms*. 2017. <https://medium.com/deep-math-machine-learning-ai/chapter-4-decision-trees-algorithms-b93975f7a1f1>. Acessado em 10 de fevereiro de 2019.

SCHULZE, C.; HENTER, D.; BORTH, D.; DENGEL, A. Automatic detection of csa media by multi-modal feature fusion for law enforcement support. In: *Proceedings of International Conference on Multimedia Retrieval*. [S.l.: s.n.], 2014. p. 353–360.

SHIH, J.-L.; LEE, C.-H.; YANG, C.-S. An adult image identification system employing image retrieval technique. *Pattern recognition letters*, Elsevier, v. 28, n. 16, p. 2367–2374, 2007.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, Springer, v. 6, n. 1, p. 60, 2019.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. Disponível em: <http://arxiv.org/abs/1409.1556>.

SINGH, K. K.; YU, H.; SARMASI, A.; PRADEEP, G.; LEE, Y. J. Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. *arXiv preprint arXiv:1811.02545*, 2018.

SONKA, M.; HLAVAC, V.; BOYLE, R. *Image processing, analysis, and machine vision*. [S.l.]: Cengage Learning, 2014.

SURENDRAN, A.; STEPHEN, S. Detection of obscene images and ejection of external devices. In: IEEE. *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*. [S.l.], 2017. v. 2, p. 110–113.

SURINTA, O.; KHAMKET, T. Recognizing pornographic images using deep convolutional neural networks. In: IEEE. *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*. [S.l.], 2019. p. 150–154.

SUTTON, R. S.; BARTO, A. G. *Reinforcement learning: An introduction*. [S.l.]: MIT press, 2018.

SWAMINATHAN, S. *Logistic Regression - Detailed Overview*. 2018. <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>. Acessado em 02 de fevereiro de 2019.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *ICLR 2016 Workshop*. [s.n.], 2016. Disponível em: <https://arxiv.org/abs/1602.07261>.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9. ISSN 1063-6919.

SZELISKI, R. *Computer vision: algorithms and applications*. [S.l.]: Springer Science & Business Media, 2010.

TAN, C.; SUN, F.; KONG, T.; ZHANG, W.; YANG, C.; LIU, C. A survey on deep transfer learning. In: SPRINGER. *International conference on artificial neural networks*. [S.l.], 2018. p. 270–279.

TAYLOR, L.; NITSCHKE, G. Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*, 2017.

ULGES, A.; STAHL, A. Automatic detection of child pornography using color visual words. In: IEEE. *2011 IEEE international conference on multimedia and expo*. [S.l.], 2011. p. 1–6.

VECCHIA, E. D. *Perícia Digital da investigação à análise forense*. [S.l.: s.n.], 2014.

VELHO, J. A. et al. Tratado de computação forense–millenium editora. *São Paulo*, 2016.

VIOLA, P.; JONES, M. J. Robust real-time face detection. *International journal of computer vision*, Springer, v. 57, n. 2, p. 137–154, 2004.

VITORINO, P.; AVILA, S.; PEREZ, M.; ROCHA, A. Leveraging deep neural networks to fight child pornography in the age of social media. *Journal of Visual Communication and Image Representation*, v. 50, p. 303 – 313, 2018. ISSN 1047-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1047320317302377>.

WANG, Y.; JIN, X.; TAN, X. Pornographic image recognition by strongly-supervised deep multiple instance learning. In: *2016 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2016. p. 4418–4422.

WIJAYA, I. G. P. S.; WIDIARTHA, I.; UCHIMURA, K.; KOUTAKI, G. Phonographic image recognition using fusion of scale invariant descriptor. In: IEEE. *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*. [S.l.], 2015. p. 1–5.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, v. 30, n. 1, p. 79–82, 2005.

WILSON, A. C.; ROELOFS, R.; STERN, M.; SREBRO, N.; RECHT, B. The marginal value of adaptive gradient methods in machine learning. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. p. 4148–4158.

XIA, M.; ZHANG, X.; WENG, L.; XU, Y. et al. Multi-stage feature constraints learning for age estimation. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 15, p. 2417–2428, 2020.

XIE, S.; GIRSHICK, R. B.; DOLLÁR, P.; TU, Z.; HE, K. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016. Disponível em: <http://arxiv.org/abs/1611.05431>.

XIONG, W.; LI, Q. Chinese skin detection in different color spaces. In: IEEE. *Wireless Communications & Signal Processing (WCSP), 2012 International Conference on*. [S.l.], 2012. p. 1–5.

YANG, H.; MOU, W.; ZHANG, Y.; PATRAS, I.; GUNES, H.; ROBINSON, P. Face alignment assisted by head pose estimation. *arXiv preprint arXiv:1507.03148*, 2015.

YIALLOUROU, E.; DEMETRIOU, R.; LANITIS, A. On the detection of images containing child-pornographic material. In: IEEE. *Telecommunications (ICT), 2017 24th International Conference on*. [S.l.], 2017. p. 1–5.

ZAGORUYKO, S.; KOMODAKIS, N. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

ZANETTI, S.; SOUSA, E.; CARVALHO, D.; BERNARDO, S. Reference evapotranspiration estimate in rio de janeiro state using artificial neural networks. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 12, p. 174–180, 04 2008.

ZHANG, C.; LIU, S.; XU, X.; ZHU, C. C3ae: Exploring the limits of compact model for age estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 12587–12596.

ZHANG, K.; ZHANG, Z.; LI, Z.; QIAO, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, IEEE, v. 23, n. 10, p. 1499–1503, 2016.

ZHOU, K.; ZHUO, L.; GENG, Z.; ZHANG, J.; LI, X. G. Convolutional neural networks based pornographic image classification. In: IEEE. *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*. [S.l.], 2016. p. 206–209.

ZHUO, L.; GENG, Z.; ZHANG, J.; LI, X. guang. Orb feature based web pornographic image recognition. *Neurocomputing*, Elsevier, v. 173, p. 511–517, 2016.

ZUO, H.; HU, W.; WU, O. Patch-based skin color detection and its application to pornography image filtering. In: ACM. *Proceedings of the 19th international conference on World wide web*. [S.l.], 2010. p. 1227–1228.

Apêndice A

Detectores de Conteúdo Impróprio Baseado em Aprendizado de Máquina Tradicional

Neste apêndice, são descritos os modelos propostos por Moreira e Fechine (2018a, 2018b), publicados no XVIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (2018) e na *2018 International Joint Conference on Neural Networks*, respectivamente. As referidas pesquisas tratam da detecção de conteúdo impróprio e pornográfico em imagens por meio do levantamento de características de pele humana e de detecção de face, além do uso de classificadores baseados em aprendizado de máquina tradicional.

A.1 Introdução

Baseado nos anseios da Computação Forense no que diz respeito aos exames periciais relacionados à detecção de conteúdo pornográfico infanto-juvenil, foi desenvolvido um modelo que, em diferentes cenários, é capaz de predizer se determinada imagem possui conteúdo impróprio/pornográfico, utilizando a base dados *AIIA-PID4 pornographic data set* (KARAVARSAMIS et al., 2013).

O primeiro cenário retrata dois tipos de classes a serem diferenciadas: (i) imagens impróprias e (ii) imagens apropriadas. O modelo proposto foi capaz de diferenciar corretamente entre esses dois tipos de imagens em mais de 93% do conjunto de teste (MOREIRA; FE-

CHINE, 2018a).

A segunda situação se configura na distinção entre imagens pornográficas e não pornográficas. As imagens não pornográficas são compostas apenas por mulheres utilizando trajes de banho. Trata-se de uma tarefa mais desafiadora, visto que as imagens, em ambas as classes, geralmente possuem uma característica marcante em comum, uma grande quantidade de píxeis de pele. Mesmo assim, após alguns ajustes no modelo, foi possível obter uma acurácia de quase 97% de acurácia nas predições no conjunto de testes (MOREIRA; FECHINE, 2018b).

O uso dessa nova abordagem possibilita reduzir o montante de imagens a serem inspecionadas pelo perito digital no momento da realização do exame, se fazendo importante pois, de acordo com Platzer, Stuetz e Lindorfer (2014), a realização de uma inspeção manual de imagens, por um longo período de tempo, faz com que haja a diminuição da capacidade da atenção humana, devido à lentidão, monotonia e repetitividade da referida tarefa, o que resulta no despercebimento de imagens ilícitas, aumentando a taxa de falsos negativos e, consequentemente, da acurácia dessa metodologia.

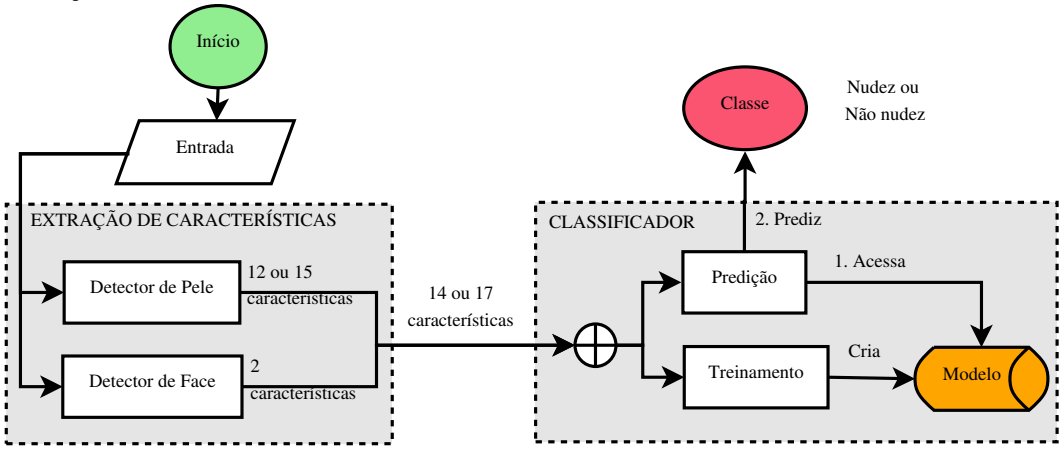
A.2 Arquitetura Proposta

O modelo proposto, que pode ser visualizado na Figura A.1, é formado por duas etapas sequenciais: (i) a extração das características e (ii) o classificador. Apesar dessa arquitetura ser amplamente utilizada na literatura, essa proposição se torna única pela junção de três aspectos: 1) os espaços de cores adotados (RGB e YCbCr); 2) a utilização de características de pele mais simples e em menor quantidade; e 3) o uso do número de faces detectadas e seus tamanhos como características, visando à redução de falsos positivos em retratos.

A.2.1 Extração de Características

A primeira etapa, que tem como objetivo obter as características da imagem, pode ser dividida em dois módulos independentes: (a) detecção e segmentação de pele e (b) detecção de face. As características extraídas dessa etapa atuam como a representação das imagens, fomentando a etapa de classificação, tanto para a criação do modelo, quanto para a predição de uma determinada imagem.

Figura A.1: Diagrama de fluxo do modelo proposto mostrando suas duas fases em detalhes: (i) a extração das características e (ii) o classificador.



Fonte: Adaptada de Moreira e Fechine (2018a).

Detecção e Segmentação de Pele

Utilizou-se um detector de pele híbrido usado por Medina e Palladino (2017), baseado nas regras propostas por Kovac, Peer e Solina (2003) e por Basilio et al. (2011). As referidas técnicas propõem a detecção de píxeis de pele em condições heterogêneas de luminosidade no espaço de cor RGB e originárias de diferentes etnias no espaço de cor YCbCr, respectivamente. Finalmente, para que um píxel seja classificado como de pele, precisa obedecer simultaneamente às equações lógicas dos espaços de cores RGB e YCbCr em seguida.

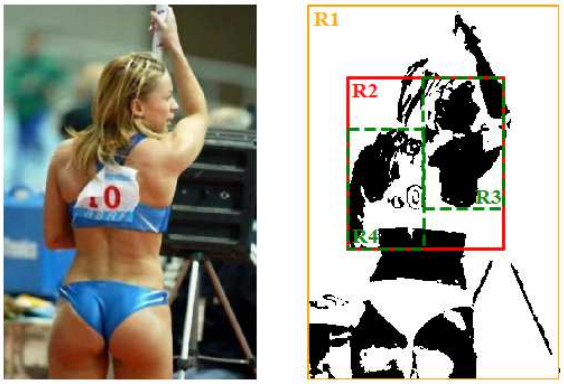
$$(R > 95 \quad \& \quad G > 40 \quad \& \quad B > 20 \quad \&$$
$$max(R,G,B) - min(R,G,B) > 15 \quad \&$$
$$|R - G| > 15 \quad \& \quad R > G \quad \& \quad R > B)$$

$$80 < Cb < 120 \quad \& \quad 133 < Cr < 173$$

Realizada a segmentação da pele, são utilizadas regiões de interesse para a extração das características de cada imagem, conforme mostrado na Figura A.2. A metodologia adotada para selecionar essas regiões foi inspirada primordialmente pela pesquisa de Ap-Apid (2005). O autor classificou imagens entre pornográficas e não pornográficas, utilizando regras estáticas, baseadas principalmente nas maiores regiões de pele das imagens.

Por tratar de um problema mais complexo, pois tem como objetivo diferenciar imagens com maior similaridade entre si (i.e. imagens pornográficas e mulheres com trajes de banho), a pesquisa de Moreira e Fechine (2018b) utiliza cinco regiões de interesse: (i) toda a imagem (R1); (ii) o menor retângulo que contenha as três maiores regiões de pele (R2) e (iii) as três maiores regiões de pele, de maneira individual (R3, R4 e R5). O modelo adotado em Moreira e Fechine (2018a), em que diferencia imagens impróprias de apropriadas, utiliza o mesmo protocolo para o levantamento das regiões de interesse, diferenciando-se apenas por utilizar quatro regiões de interesse, pois utiliza as duas maiores regiões de pele em vez de três, conforme ilustrado na Figura A.1.

Figura A.2: A imagem de entrada antes de extrair suas características de pele em (a). Imagem de entrada quando realizada a detecção dos píxeis de pele, esses representados pela cor preta e as quatro regiões de interesse para extração de características: R1, a imagem por inteiro. R2, o menor retângulo que abarca as duas maiores regiões de pele. R3 e R4, as duas maiores regiões de pele em (b).



Fonte: Extraída de Moreira e Fechine (2018a).

A pesquisa de Ap-Apid (2005) também foi referência para quais características extrair de cada região de interesse. Em cada uma dessas regiões foram consideradas como características da imagem: 1) a sua área total; 2) a quantidade de píxeis de pele e 3) a intensidade média dos píxeis no espaço de cor RGB. Por fim, esse processo resulta em um vetor de características de tamanho 12 e 15 para os modelo utilizados nas pesquisas de Moreira e Fechine (2018a, 2018b), respectivamente.

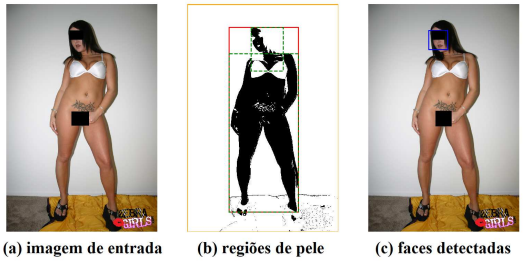
Detecção de Face

A maior parte das imagens consideradas pornográficas e os retratos[1] possuem uma característica determinante em comum para sua classificação: a grande quantidade de exposição de pele. Portanto, sem uma informação relativa à existência ou não de faces na imagem, a diferenciação entre esses dois tipos de imagens se torna mais difícil. Com o intuito de minimizar essas predições equivocadas, foram utilizadas como características: (i) a quantidade de faces detectadas; e (ii) a área total dessas faces, para que o modelo tenha ciência da existência de faces humanas e seus respectivos tamanhos nas imagens. As pesquisas de Platzer, Stuetz e Lindorfer (2014), Zhou et al. (2016) utilizaram a existência de faces e suas áreas para reclassificar imagens preditas de maneira incorreta, entretanto, sem utilizar aprendizado de máquina, tendo sido utilizado apenas um conjunto de regras estáticas.

Para a detecção das faces humanas foi utilizado um classificador baseado no estudo de Viola e Jones (2004). Essa escolha se deu por esse classificador ser recomendado para aplicações que necessitem de um bom desempenho computacional, devido ao grande número de imagens que são verificadas nos exames forenses (ELEUTÉRIO; MACHADO, 2011).

Dessa forma, foram adicionados dois valores, oriundos da detecção de faces, ao vetor de características final. As Figuras A.3 e A.4 ilustram uma imagem pornográfica e outra não pornográfica, respectivamente, e as regiões de interesses aplicadas em Moreira e Fechine (2018a).

Figura A.3: Imagem possuindo conteúdo pornográfico em (a) e suas regiões de interesse para extração de características em (b) e (c).

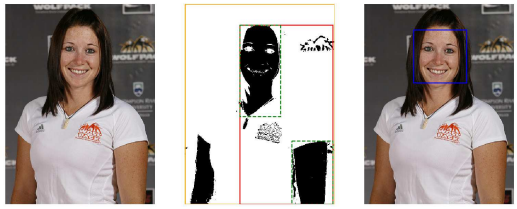


(a) imagem de entrada (b) regiões de pele (c) faces detectadas

Fonte: Adaptada de Moreira e Fechine (2018a).

[1] imagens em que a maior parte do seu conteúdo é uma face humana.

Figura A.4: Imagem sem conteúdo pornográfico (a) e suas regiões de interesse para extração de características em (b) e (c).



(a) imagem de entrada (b) regiões de pele (c) faces detectadas

Fonte: Extraída de Moreira e Fechine (2018a).

A.2.2 Classificador

O classificador tem como funções criar os padrões das categorias na etapa de treinamento e inferir a que classe determinada imagem pertence, em sua fase de teste. Foram adotados quatro classificadores baseados em aprendizado de máquina: (i) baseado em regressão logística, (ii) baseado em rede neural perceptron multicamadas, (iii) baseado em árvore de decisão e (iv) baseado em floresta aleatória.

A.3 Etapa Experimental

Nas etapas de treinamento, validação e teste foi utilizada a base de dados de imagens AIIA-PID4 pornographic data set (KARAVARSAMIS et al., 2013). A referida base de dados possui 12.740 imagens, particionada em quatro categorias: 1) pornográfica; 2) biquíni; 3) pele e 4) não pele.

Para a distinção entre imagens impróprias e apropriadas em Moreira e Fechine (2018a), essas duas novas categorias foram criadas por meio da junção das categorias preexistentes: 1) imagens impróprias: agrupamento das categorias pornográfica e biquíni e 2) imagens apropriadas: agrupamento das categorias pele e não pele. Salienta-se que foram removidas todas as imagens que apresentaram problemas de leitura. Por fim, existiam 6.638 imagens categorizadas como impróprias e 6.102 como apropriadas, sendo os dados relativamente balanceados (52,10% e 47,90%). Para diferenciação entre imagens pornográficas e não pornográficas em Moreira e Fechine (2018b), foram utilizadas apenas as categorias pornográfica e biquíni, respectivamente. Também foram removidas imagens que apresentaram problemas de leitura, utilizando por fim 1.584 imagens da classe pornográfica e 4.014 da classe biquíni.

Foi realizado ajuste fino (*fine-tune*) dos hiperparâmetros regularizadores em cada um dos modelos adotados por meio da técnica de validação cruzada, tendo como objetivo diminuir o erro de generalização. Foi reservado 10% dos dados para a fase de testes que acontece no final do experimento e utilizou-se os demais dados na validação cruzada particionada em cinco blocos (*5-fold*).

Todos os classificadores foram utilizados com a configuração padrão da biblioteca *Python Sci-kit Learn Library*[2]. Para os modelos baseados em regressão logística e rede neural perceptron multicamadas foram utilizados, respectivamente, os regularizadores C e alpha. Os modelos baseados em árvores utilizam como fator de regularização o número mínimo de divisões em um nodo interno ($min\_samples\_split$). A variação dos valores adotados em cada pesquisa pode ser observada nas Tabelas A.1 e A.2.

Tabela A.1: Faixa de valores utilizada para os regularizadores C, alpha $min\_samples\_split$ em Moreira e Fechine (2018a).

Regularizador	Faixa de Valores
C; alpha	0,0001; 0,001; 0,01; 0,1; 1; 10; 100; 1000; 10000; 100000
min_samples_split	2; 4; 8; 16; 32; 64; 128; 256; 512; 1024

Fonte: Adaptada de Moreira e Fechine (2018a).

Tabela A.2: Faixa de valores utilizada para os regularizadores C, alpha $min\_samples\_split$ em Moreira e Fechine (2018b).

Regularizador	Faixa de Valores
C; α	0,0001; 0,001; 0,01; 0,1; 1; 10; 100; 1000
min_samples_split	2; 4; 8; 16; 32; 64; 128; 256

Fonte: Adaptada de Moreira e Fechine (2018b).

Foram utilizadas as métricas de acurácia, f1-score, precisão e revocação para avaliação dos resultados dos modelos em ambas as pesquisas. Entretanto salienta-se que, na etapa de validação cruzada, foram utilizadas métricas distintas em cada pesquisa, devido à diferente distribuição dos dados entre as categorias em cada cenário. Em Moreira e Fechine (2018a), por haver um balanceamento dos dados, foi possível utilizar a acurácia como métrica de avaliação sem nenhum prejuízo. Devido ao desbalanceamento dos dados em Moreira e Fechine (2018b), utilizou-se a métrica f1-score, capaz de lidar melhor com dados que apresentem essa heterogeneidade. Por fim, o melhor classificador adotado foi comparado com os trabalhos de Medina e Palladino (2017) e Karavarsamis et al. (2013) em cada um dos cenários.

[2]https://scikit-learn.org/

A.4 Resultados Obtidos

Para cada uma das pesquisas, foram realizados os experimentos de validação cruzada, de acordo com a faixa de valores dos hiperparâmetros de regularização. Em seguida, foi selecionado para cada classificador o valor do fator de regularização que apresentou maior acurácia na fase de testes, como pode ser observado nas Tabelas A.5 e A.6 para Moreira e Fechine (2018a) nas Tabelas A.7 e A.8 para Moreira e Fechine (2018b).

Por fim, os classificadores foram comparados entre si na fase de testes, por meio das métricas de avaliação precisão, revocação (*recall*), f1-score e acurácia, como pode ser visto nas Tabelas A.3 e A.4, tendo os baseados em floresta aleatória os melhores resultados em ambas as pesquisas (MOREIRA; FECHINE, 2018a, 2018b).

Tabela A.3: Resultados nas métricas de precisão, revocação, f1-score e acurácia de cada classificador fatores de regularização mais bem ajustados.

Métrica / Classificador	Precisão	Revocação	F1-score	Acurácia
Regressão Logística	93,45	86,26	89,71	88,70%
Perceptron Multicamadas	94,64	87,00	90,66	89,72%
Árvore de Decisão	93,15	**93,29**	93,22	92,86%
Floresta Aleatória	**94.94**	93,00	**93,96**	**93,56%**

Fonte: Extraída de Moreira e Fechine (2018a).

Tabela A.4: Resultados nas métricas de precisão, revocação, f1-score e acurácia de cada classificador fatores de regularização mais bem ajustados.

Métrica / Classificador	Precisão	Revocação	F1-score	Acurácia
Regressão Logística	57,64	74,77	65,10	84,11%
Perceptron Multicamadas	91,67	65,67	76,52	85,54%
Árvore de Decisão	88,89	90,14	89,51	94,64%
Floresta Aleatória	**97,92**	**90,97**	**94,31**	**96,96%**

Fonte: Extraída de Moreira e Fechine (2018b).

Tabela A.5: Acurácias resultantes da variação dos respectivos fatores de regularização (C e alpha) nas validações cruzadas da regressão logística e perceptron multicamada.

C; alpha / Classificador	0,0001	0,001	0,01	0,1	1	10	100	1000	10000	100000
REGRESSÃO LOGÍSTICA	89,24%	89,25%	89,34%	89,39%	89,37%	**89,51%**	89,26%	89,49%	89,23%	89,27%
PERCEPTRON MULTICAMADAS	85,34%	87,34%	87,28%	**87,76%**	86,31%	85,91%	87,35%	85,31%	81,80%	78,69%

Fonte: Extraída de Moreira e Fechine (2018a).

Tabela A.6: Acurácias resultantes da variação do fator de regularização (min_split_samples) nas validações cruzadas da árvore de decisão e floresta aleatória.

min_split_samples / Classificador	2	4	8	16	32	64	128	256	512	1024
ÁRVORE DE DECISÃO	92,77%	92,77%	92,90%	**93,14%**	92,99%	92,92%	92,80%	92,35%	91,30%	89,43%
FLORESTA ALEATÓRIA	92,87%	93,11%	93,35%	**93,45%**	93,01%	92,94%	92,94%	91,87%	90,39%	89,45%

Fonte: Extraída de Moreira e Fechine (2018a).

Tabela A.7: Valores de f1-score resultantes da variação dos respectivos fatores de regularização (C e alpha) nas validações cruzadas da regressão logística e perceptron multicamada.

Classifier \ C; alpha	0,0001	0,001	0,01	0,1	1	10	100	1000
REGRESSÃO LOGÍSTICA	71,80	**75,43**	71,93	74,99	73,26	72,53	74,13	72,21
PERCEPTRON MULTICAMADAS	61,04	57,11	70,06	71,90	64,48	65,71	**72,92**	54,40

Fonte: Extraída de Moreira e Fechine (2018b).

Tabela A.8: Valores de f1-score resultantes da variação do fator de regularização (min_split_samples) nas validações cruzadas da árvore de decisão e floresta aleatória.

Classifier \ min_samples_split	2	4	8	16	32	64	128	256
ÁRVORE DE DECISÃO	90,16	**90,50**	90,07	90,34	90,10	89,79	87,69	86,39
FLORESTA ALEATÓRIA	88,74	89,00	89,78	89,33	90,15	**90,71**	88,19	87,35

Fonte: Extraída de Moreira e Fechine (2018b).

A.5 Validação

Os melhores modelos, de ambos os estudos, foram comparados às pesquisas de Medina e Palladino (2017) e Karavarsamis et al. (2013). Para o primeiro, foi reproduzido seu experimento, por meio do uso do código-fonte fornecido pelos autores, fazendo uso das mesmas imagens utilizadas na realização da etapa de testes do nosso trabalho. Já para o segundo trabalho, os resultados foram obtidos por meio dos dados expostos em seu estudo. Foi realizado um comparativo entre a acurácia do melhor modelo proposto (baseado em floresta aleatória) e os referidos trabalhos, como pode ser visualizado na Tabela A.9.

Tabela A.9: Comparação da acurácia dos métodos propostos e os trabalhos relacionados.

Método	Conteúdo Impróprio vs. Conteúdo Apropriado	Pornografia vs. Mulheres de biquíni
Proposto	**93,56**%	**96,96**%
Medina e Palladino (2017)	79,10%	54,11%
Karavarsamis et al. (2013)	85,05%	82,02%

Fonte: Adaptada de Moreira e Fechine (2018a) e Moreira e Fechine (2018b) .

A.6 Considerações Finais

As pesquisas expostas neste capítulo tiveram como objetivo principal aperfeiçoar uma técnica consagrada de detecção de pornografia desenvolvida por Ap-Apid (2005), conhecida como "Um Algoritmo Para a Detecção de Nudez". Essa melhoria se deu pelo treinamento de modelos baseados em aprendizado de máquina, utilizando a técnica de validação cruzada, para classificar as imagens, em vez do simples uso de regras estáticas. Boa parte das características levantadas foram inspiradas pelo referido trabalho Ap-Apid (2005), contudo, foram adicionadas características oriundas da detecção de faces, com objetivo de diminuir o número de falsos positivos em imagens cuja face é predominante.

Os estudos propostos foram capazes de superar as pesquisas de Medina e Palladino (2017), que baseou-se no estudo de Ap-Apid (2005), e de Karavarsamis et al. (2013), que desenvolveu a base de dados utilizada (*AIIA-PID4 pornographic data set*), alcançando 93,56% de acurácia na diferenciação entre imagens impróprias imagens apropriadas e 96,96% de acurácia na classificação de imagens pornográficas e de mulheres trajando biquíni.