



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

VINÍCIUS JORGE PEREIRA DA SILVA

**APRENDIZAGEM DE MÁQUINA APLICADA AO
MONITORAMENTO DA PRESENÇA DE ANIMAIS EM
RESERVAS NATURAIS DE AMBIENTES INDUSTRIAIS**

CAMPINA GRANDE - PB

2021

VINÍCIUS JORGE PEREIRA DA SILVA

**APRENDIZAGEM DE MÁQUINA APLICADA AO
MONITORAMENTO DA PRESENÇA DE ANIMAIS EM
RESERVAS NATURAIS DE AMBIENTES INDUSTRIAIS**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientadora: Professora Dra. Joseana Macêdo Fachine Régis de Araújo.

CAMPINA GRANDE - PB

2021



S586a Silva, Vinícius Jorge Pereira da.

Aprendizagem de máquina aplicada ao monitoramento da presença de animais em reservas naturais de ambientes industriais. / Vinícius Jorge Pereira da Silva - 2021.

11 f.

Orientadora: Profa. Dra. Joseana Macêdo Fechine Régis de Araújo.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Aprendizagem de máquina. 2. Monitoramento de animais. 3. Processamento digital de imagem. 4. Rede neural de aprendizagem profunda. 5. Identificação de animais. 6. Visão computacional. I. Araújo, Joseana Macêdo Fechine Régis de. II. Título.

CDU:004.932(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

VINÍCIUS JORGE PEREIRA DA SILVA

**APRENDIZAGEM DE MÁQUINA APLICADA AO
MONITORAMENTO DA PRESENÇA DE ANIMAIS EM
RESERVAS NATURAIS DE AMBIENTES INDUSTRIAIS**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

**Professora Dra. Joseana Macêdo Fachine Régis de Araújo
Orientadora – UASC/CEEI/UFCG**

**Professor Dr. Leandro Balby Marinho
Examinador – UASC/CEEI/UFCG**

**Professor Tiago Lima Massoni
Professor da Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 25 de MAIO de 2021.

CAMPINA GRANDE - PB

ABSTRACT

Several Brazilian industries consume natural resources on a daily basis. Some of them have their own regions of environmental reserves. In this context, these industries continually aim to provide protection for their resources. However, a complicating factor present in most of these spaces is animals. Your habits sometimes lead to damage to these areas. In this sense, this article aims to monitor the presence of animals in natural reserve environments in industrial environments. For this purpose, digital image processing techniques were used to identify the presence of animals in these environments. Due to the momentary impossibility of acquiring images in an industrial environment, the work presents a proof of concept, with the acquisition of data from a public database. Deep learning networks were used to identify animal images. The presented approach enabled the identification of animals of different sizes and species, with relevant accuracy in comparison with other similar studies. These results allow to validate the proposed methodology, enabling its installation in an industrial environment, in order to monitor the presence of animals.

Aprendizagem de Máquina Aplicada ao Monitoramento da Presença de Animais em Reservas Naturais de Ambientes Industriais

Vinicius Jorge Pereira da Silva
Unidade Acadêmica de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil
vinicius.jorge.silva@ccc.ufcg.edu.br

Joseana Macêdo Fechine Régis de Araújo
Unidade Acadêmica de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil
joseana@computacao.ufcg.edu.br

RESUMO

Diversas indústrias brasileiras consomem diariamente recursos naturais. Algumas delas possuem suas próprias regiões de reservas ambientais. Nesse contexto, essas indústrias objetivam continuamente prover proteção para os seus recursos. Contudo, um complicador presente na maioria desses espaços, são os animais. Seus hábitos por vezes acarretam na danificação dessas áreas. Nesse sentido, este artigo tem como proposta, o monitoramento da presença dos animais nos ambientes de reservas naturais de ambientes industriais. Para tanto, foram utilizadas técnicas de processamento digital de imagens para identificação da presença de animais nesses ambientes. Em virtude da impossibilidade momentânea de aquisição de imagens em um ambiente industrial, o trabalho apresenta uma prova de conceito, com aquisição dos dados a partir de uma base de dados pública. Para identificação das imagens de animais foram utilizadas redes de aprendizagem profunda. A abordagem apresentada possibilitou a identificação de animais de diferentes portes e espécies, com acurácia relevante em comparação com outras pesquisas similares. Esses resultados permitem validar a metodologia proposta, possibilitando a sua instalação em um ambiente industrial, de forma a realizar o monitoramento da presença de animais¹.

Palavras-chave

Monitoramento de Animais, Processamento Digital de Imagem, Rede Neural de Aprendizagem Profunda.

1. INTRODUÇÃO

Um dos pilares do Produto Interno Bruto (PIB) nacional que mais movimentava bens na economia é o setor industrial brasileiro. Responsável por cerca de 21,6% do PIB [4], o setor tem peso expressivo na balança comercial, nos investimentos em pesquisas e desenvolvimento e na geração de empregos.

Contudo, a administração e a preservação de regiões com reservas naturais, em ambiente industrial, ainda é considerada um dos desafios enfrentados pelo setor. Nesse cenário, uma das razões que acentua a situação é o controle e monitoramento das regiões na indústria. Assim, tem-se como um problema a deterioração das reservas, causadas muitas vezes devido à ação da presença de agentes externos, comumente animais.

Diante do exposto, o objetivo do trabalho ora descrito é realizar a identificação desses agentes, por meio de análises e processamento digital de imagens, de forma a viabilizar estudos sobre os locais e os períodos em que ocorre a presença desses animais na área. O monitoramento da região impactará diretamente no controle e na proteção do ambiente, ampliando a otimização dos recursos e eliminando possíveis desperdícios de manutenção.

2. TRABALHOS RELACIONADOS

No campo da identificação de animais, pode-se mencionar *Fahad* [3], que apresenta um modelo para classificação de animais utilizando técnicas de aprendizagem de máquina, como *Support Vector Machine (SVM)* e *Machine Learning Solution (MLS)*.

Schindler [11], por sua vez, mostra um estudo de reconhecimento de ações e de animais em ambientes naturais noturnos, com aprendizagem de máquina, utilizando redes neurais profundas. Há também trabalhos como o de *Burghardt* [1], que demonstra um estudo de reconhecimento de animais selvagens utilizando algoritmos de *Face Detection*. Assim como, *Schindler and Kumar* [9] apresentam um método de classificação de animais, utilizando técnicas de aprendizagem de máquinas, com segmentação de imagens e rede neural probabilística.

¹ Os autores retêm os direitos, ao abrigo de uma licença Creative Commons Atribuição CC BY, sobre todo o conteúdo deste artigo (incluindo todos os elementos que possam conter, tais como figuras, desenhos, tabelas), bem como sobre todos os materiais produzidos pelos autores que estejam relacionados ao trabalho relatado e que estejam referenciados no artigo (tais como códigos fonte e bases de dados). Essa licença permite que outros distribuam, adaptem e evoluam seu trabalho, mesmo comercialmente, desde que os autores sejam creditados pela criação original.

3. FUNDAMENTAÇÃO TEÓRICA

De acordo com *Schindler* [11], câmeras escondidas apresentam várias vantagens sob o monitoramento dos animais, por não serem invasivas e não terem efeito sobre o comportamento do animal. Além de ser útil na observação das mudanças do habitat natural e da biodiversidade com o passar do tempo.

Um sistema para reconhecimento automático de imagens consiste no reconhecimento de padrões, sendo constituído pelas fases de treinamento e teste (ou classificação). Na fase de treinamento, faz-se necessário o pré-processamento, que se refere ao tratamento inicial dos dados para padronização das imagens e ajustes de especificação, a extração de características e a construção de padrões de referência (por meio de um classificador). Na fase de classificação, repete-se o pré-processamento e a extração de características. Em seguida, os padrões de teste são comparados com os padrões de referência e, por meio de uma regra de decisão, as imagens são classificadas. Em alguns casos, pode ser necessário o pós-processamento, para realizar correções de possíveis defeitos gerados pelo próprio modelo.

Esses tipos de sistemas utilizam técnicas computacionais, para que seja possível conseguir boas taxas de performance de máquinas, com a intenção de alcançar o reconhecimento de padrões dentro de uma imagem. Ambas as abordagens apresentadas por *Schindler* [11] obtiveram sucesso apreciável em seu propósito, e serviram de incentivo para a realização deste trabalho.

Em seu trabalho, *Schindler* [11] propõe a utilização de redes neurais convolucionais clássicas com aprendizagem profunda na construção de um classificador de animais em ambientes noturnos e de um sistema de reconhecimento de ações, e realiza um estudo comparativo utilizando as redes *Mask R-CNN* [6] e a família *3D ResNet*, *MC ResNet* e $(2 + 1)D$ *ResNet* [7].

A seguir, uma descrição de parâmetros utilizados para compor o conjunto de características, bem como classificadores

(Redes Neurais Profundas) utilizados na construção de um sistema para reconhecimento de imagens.

3.1 Conjunto de Características

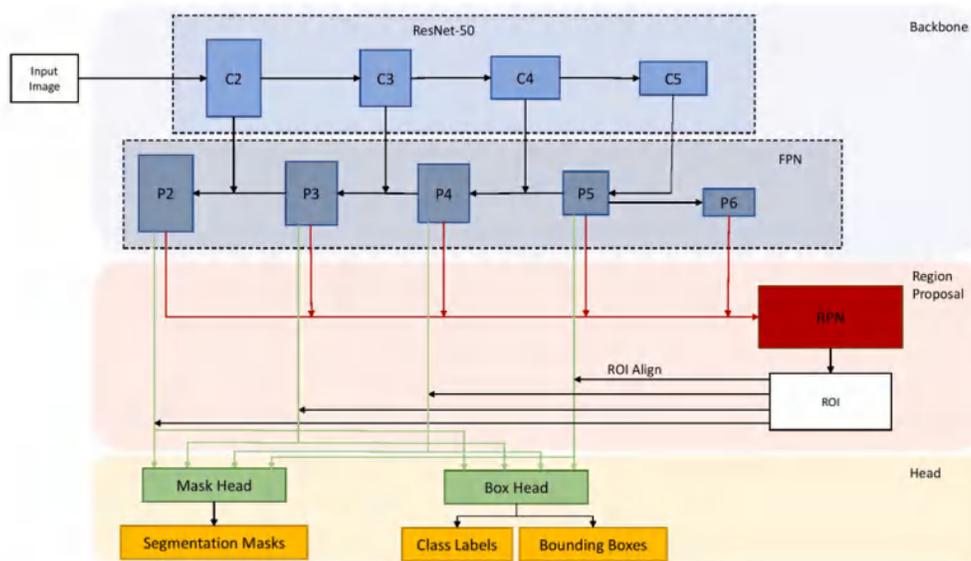
Para o processamento das imagens, de maneira adequada, torna-se necessário realizar sucessivas transformações na imagem para obter mais facilmente as características existentes.

Dentre as presentes, estão a manipulação de Histograma, com a normalização dos níveis em um modelo *BGR* (Sistema de cores em que a ordem das áreas são invertidas, com o Azul ocupando a área mais significativa), a utilização de 5 camadas de informações Convolucionais com 11 dimensões e 96 saídas, Normalização de Valores Centrais da imagem (*Local Response Normalization, LRN*) sob as regiões de entrada, utilizando como parâmetros os valores: $\alpha = 0,0001$ e $\beta = 0,75$ em cada iteração, resultando em áreas com 5 canais distintos de informações. Esse conjunto de características alimentará a rede, para que essa possa conseguir adquirir o aprendizado necessário e apresentar predições ao longo do processo.

3.2 Redes Neurais Profundas

Dentre os tipos de Redes Neurais profundas que se torna cada vez mais utilizada para o reconhecimento de objetos, destaca-se a *Mask R-CNN*, cuja sua arquitetura é composta por três partes, como mostra a Figura 1: *Backbone*, ou *Rede de transporte*, que é o principal canal responsável por gerenciar todo o tráfego de informações iniciais da rede. A *Region Proposal Network (RPN)*, que é encarregada por prever os limites dos objetos analisados em cada posição da imagem. E o *Head*, que indicará a região detectada associada a uma máscara e uma classificação.

Figura 1. Modelo arquitetural da rede Mask R-CNN.



Fonte: Figura extraída de [11].

Na primeira etapa da rede, o *Backbone* é constituído por uma rede ResNet-50, que produzirá 4 camadas de informações que serão utilizadas por uma *Feature Pyramid Network* (FPN), que utiliza uma arquitetura *top-down* para extrair recursos da imagem que serão utilizados para detectar os objetos de interesse nas etapas seguintes.

Na segunda etapa, com as regiões de propostas de interesse geradas, a *RPN* segue criando novas Regiões com os Objetos de Interesse (ROI) a partir dos dados gerados pela *FPN* na primeira etapa.

Por fim, na última etapa do processo, o *Head* é responsável por gerar uma região delimitadora do objeto (*Bounding Box*), junto com uma etiqueta de classificação e uma máscara de segmentação do objeto na imagem, como é mostrado na Figura 2.

Figura 2. Exemplo da saída gerada pela Mask R-CNN.



Fonte: Figura extraída de [11].

A *Bounding box* e a classificação são estimadas utilizando as camadas da rede totalmente conectadas, enquanto que a máscara, por sua vez, é obtida utilizando apenas parte delas, para não perder as informações espaciais do objeto analisado durante a transformação.

3.3 3D ResNet, MC ResNet e (2 + 1)D ResNet

A família de variantes das redes ResNet utiliza aprendizagem residual com convoluções espacotemporal. O processamento ocorre em cima de cliques com 8 a 16 frames por segundo. Tem como informação intermediária, a saída do bloco residual z_i , que serve de entrada para alimentar a rede durante seu treinamento, e que é computado conforme Equação 1.

$$z_i = z_{i-1} + F(z_{i-1}, \theta_i). \quad (1)$$

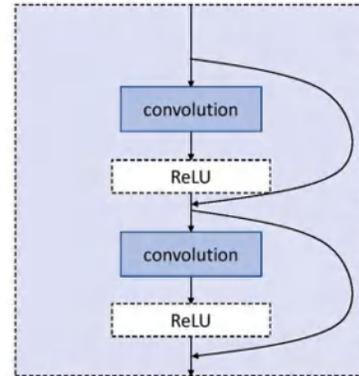
Um bloco residual consiste em duas camadas convolucionais consecutivas e Unidades Lineares Retificadas (ReLU) de funções de ativação após cada camada. Todas as variantes da *ResNet* consistem em 5 blocos seguidos uns dos outros, em que o primeiro bloco é o único que apresenta apenas uma camada convolucional em seu bloco residual.

A função F apresentada na Equação 1 representa estes passos, para uma entrada com os pesos θ_i que, acrescentado ao bloco anterior, é responsável por gerar o novo resíduo.

Na Figura 3 é apresentado, de maneira visual, o processo para a obtenção dos demais blocos residuais gerados pela rede.

Após esse processamento, um vetor com 512 características é gerado. Este vetor é utilizado como entrada para uma última camada totalmente conectada, para assim, produzir a classificação da instância analisada.

Figura 3. Modelo visual do processo de obtenção do bloco residual.



Fonte: Figura extraída de [11].

4. METODOLOGIA

A metodologia adotada consistiu na aquisição da base de dados, seguida do estudo e implementação da aprendizagem de máquina para pré-processamento, extração de características e construção do classificador para reconhecimento de imagens de animais.

Foram utilizadas bibliotecas de software de visão computacional e *machine learning* para os processos de extração de características, construção de padrões e classificação das imagens.

4.1 Conjunto de Dados

Para o desenvolvimento do trabalho, optou-se, inicialmente, pela utilização de uma base de dados com imagens já disponíveis de animais em reservas ambientais disponibilizada por *Tekeli, U.* [12]. Esta base possui 2.500 imagens, das quais foram selecionadas 450 imagens que possuem a presença bem definida de animais. Foram criados dois subconjuntos, um para treinamento composto por 80% dos dados, que representa 361 imagens e um subconjunto de teste, com 89 imagens (20% do total).

Durante as análises preliminares, verificou-se que a quantidade de imagens da base de dados não era suficiente para o treinamento de uma rede neural convolucional. Redes dessa magnitude exigem uma grande quantidade de dados, para que consiga extrair informações precisas sobre o objeto a ser analisado. Assim, tornou-se necessário a aquisição de uma nova base de dados.

A nova base de dados utilizada foi composta do conjunto de imagens da *ImageNet* (versão 2012). Essa base tem tamanho significativo e é representativa para o contexto do trabalho, sendo bastante popular no desenvolvimento de vários sistemas de classificação de imagens, dado o uso em competição mundial na área, contando com mais de 14 milhões de imagens [10]. Na Figura 4, são apresentados exemplares das imagens utilizadas no trabalho ora descrito.

Figura 4. Exemplares de imagens do conjunto de dados selecionados.



Fonte: *ImageNet 2012* [10].

O banco de dados escolhido conta com 1.431.167 imagens, compostas por 1.000 categorias que englobam animais e objetos, separados em três grandes grupos de diferentes tamanhos, conforme apresentado na Tabela 1.

Tabela 1. Configuração do conjunto de dados ImageNet 2012.

	Treinamento	Validação	Teste
Quantidade de imagens	1.281.167	50.000	100.000
Quantidade de imagens por rótulo	Entre 732 e 1.300	50	100
Quantidade de imagens de animais	Entre 351.076 e 445.900	17.150	34.300

Considerando o objetivo do trabalho ora descrito, foi realizada uma filtragem no conjunto de imagens de animais, das mil categorias iniciais, foram selecionados 343 rótulos que

representam apenas animais com diferentes classes, tais como: Ave, Baleia, Cachorro, Camelo, Cervo, Raposa, Urso e etc.

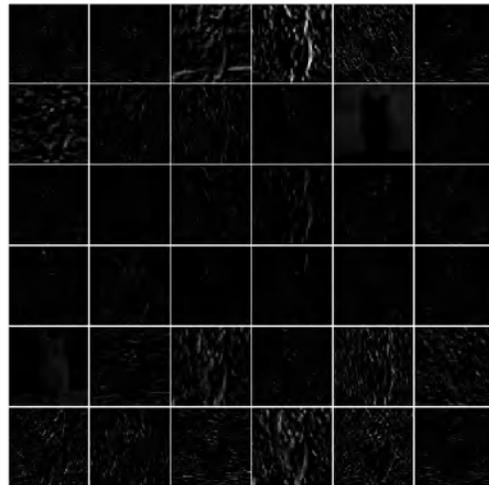
4.2 Pré-Processamento

Para cada imagem do conjunto, foi gerada uma média sobre os pixels, para obter os valores médios dos valores em RGB^2 , foram movidos os canais da imagem para a dimensão mais externa, com a intenção de subtrair o valor médio do conjunto de dados em cada redimensionamento do canal, re-escalando de $[0, 1]$ para $[0, 255]$ em um formato BGR com três canais, além de realizar o processamento da imagem remodelando o seu tamanho para 227×227 pixels.

4.3 Treinamento

Na etapa de treinamento, foi utilizado um modelo bem definido, com algumas camadas específicas para as extrações das informações convolucionais, do conjunto de imagens de treinamento, foi utilizado os *layers*³: *ReLU*, *Pooling*⁴ (Figura 5), *LRN* e *Softmax*⁵, para a extração dos dados normalizados de maneira automática, sobre cada imagem observada no conjunto.

Figura 5. Resultado da extração do *Pooling* durante o processo de treinamento.



Fonte: Figura extraída de *Yangqing Jia et al* [2].

² RGB é a abreviatura de um sistema de cores em que o Vermelho, o Verde e o Azul são combinados com a intenção de reproduzir um espectro cromático.

³ São componentes funcionais separados que interagem de forma sequencial.

⁴ É uma camada de agrupamento comumente utilizada após a aplicação de camadas convolucionais, para ordenar as camadas dentro de uma rede neural.

⁵ Uma função exponencial normalizada para múltiplas dimensões, e geralmente é utilizada como última função de ativação de uma rede neural, para normalizar os resultados e distribuir a probabilidade sobre as classes de saída.

Para a parametrização, foram utilizados os valores distribuídos pela própria *Berkeley*, por ela já possuir um modelo de estudo bem definido e pré-treinado. Sendo assim, a rede foi colocada em treinamento utilizando as camadas de extração de informações mencionadas, e aplicando os valores da distribuição de *Berkeley*, para alimentar a rede durante todo o processo de aprendizagem, e assim, operar por cerca de 310.000 iterações.

4.3.1 Rede Neural

Para a classificação, foi utilizado o *framework* de rede neural de aprendizagem profunda desenvolvido pela *Berkeley AI Research* [5] e por colaboradores da comunidade denominado *Caffe* (O *framework* é distribuído sob licença *BSD* compartilhada entre a Universidade da Califórnia e os respectivos colaboradores).

A escolha desse *framework* foi definida pelo fato de possuir uma arquitetura em que os modelos de configuração são definidos sem codificação permanente, e por dispor de uma comunidade ativa com referências de uso para o conjunto de dados escolhido.

Para o estudo, foi utilizado o modelo distribuído pela Berkeley em cumprimento às instruções de treinamento do modelo ImageNet, em uma replicação do modelo descrito por *AlexNet* [8] com mudanças no processo de agrupamento e de re-iluminação.

4.4 Validação e Teste

Nessa etapa, o processo ocorre semelhante ao explicado na seção anterior, de Treinamento, possuindo como diferença: A mudança dos conjuntos de imagens utilizados para cada tipo, que envolve os grupos de Validação e Testes, a alteração do parâmetro de configuração do *framework* para o modo de teste e a utilização do novo modelo obtido, já treinado para prover as previsões.

5. RESULTADOS

Na Tabela 2, é apresentada uma parcela representativa dos resultados obtidos para algumas classificações realizadas neste estudo. Para todo o conjunto distribuído da *ImageNet*, a rede apresentou uma taxa de acerto global de, aproximadamente, 57,1% para o conjunto de validação, e 56,5% para o grupo de testes. Em relação ao subconjunto analisado, contendo apenas animais, foi obtido uma taxa de acurácia global de 62,29% sobre o conjunto de validação, isso quer dizer que ao categorizar uma figura, a rede neural conseguiu ter êxito na sua previsão em 62,29% das vezes, apresentou um coeficiente de variação de 27,86%.

Se for comparado com a taxa de acerto Top 5, que representa os cinco melhores resultados de probabilidades obtidos pela rede, foi alcançado 85,27% de acerto sob todo o conjunto de animais, e 14,73% de que a previsão inferida possui uma probabilidade menor do que as cinco maiores.

Das 343 categorias de animais consideradas, em 139 delas foi obtida uma taxa de acerto global maior ou igual a 70%, a Figura 6 mostra como exemplo um cachorro da raça Leonberg, no qual sua classe apresentou 84% de acerto. Para 117 classes foram obtidos valores entre 50% e 69%. Em relação ao Top 5, para 327 categorias foram obtidas taxas de acerto maiores ou iguais a 70% e apenas para 16 categorias foram obtidos valores inferiores.

Os resultados obtidos mostram, portanto, que a aprendizagem de máquina adotada para identificação de animais se mostrou eficiente. Observou-se também o esperado, que redes convolucionais, com poder de aprendizagem profunda, exigem uma grande quantidade de artefatos para conseguir extrair informações relevantes sobre as categorias analisadas.

Tabela 2. Parcela representativa dos resultados obtidos pela rede em relação ao conjunto de validação.

Categoria	Taxa de acerto	Taxa Top 5	Qtd. 1° acerto	Qtd. 2° acerto	Qtd. 3° acerto	Qtd. 4° acerto	Qtd. 5° acerto	Erros
Ruddy Turnstone	98%	98%	49	0	0	0	0	1
Proboscis Monkey	96%	100%	48	1	1	0	0	0
Lorikeet	94%	100%	47	1	1	1	0	0
Hamster	92%	96%	46	1	0	1	0	2
Snow Leopard	90%	98%	45	1	2	1	0	1
Lesser Panda	90%	94%	45	1	0	1	0	3

Killer Whale	88%	100%	44	3	3	0	0	0
Zebra	88%	96%	44	2	2	0	0	2
Marmot	86%	92%	43	2	0	1	0	4
Persian Cat	78%	94%	39	2	3	1	2	3
Ice Bear	78%	88%	39	4	0	1	0	6
Lion	76%	90%	38	6	1	0	0	5
Peacock	76%	86%	38	4	1	0	0	7
Tiger	74%	90%	37	6	1	1	0	5
Maltese Dog	72%	90%	36	6	1	0	2	5
Afghan Hound	70%	90%	35	6	3	1	0	5
Llama	66%	78%	33	3	2	0	1	11
Leopard	62%	98%	31	16	2	0	0	1

Fonte. Autoria própria.

Figura 6. Exemplo de um Leonberg reconhecido pela rede com taxa de acerto de 82,07%.



Fonte. ImageNet 2012 [10].

Para avaliar, preliminarmente, o desempenho do sistema de reconhecimento desenvolvido, não apenas para o monitoramento diurno, mas também para o monitoramento noturno, foram realizados testes com imagens noturnas de animais que fazem parte de categorias apresentadas no conjunto da *ImageNet*. Foram alcançadas baixas taxas de acerto global. Dentre as classificações verificadas, a única que apresentou uma taxa acima de 50% foi a Raposa, com 51,52% obtido, para as demais classes, suas taxas variam entre 12% e 21% de acerto, impactando diretamente no valor final da performance. Tais resultados reforçam que o desempenho do sistema para a classificação de imagens noturnas de animais exigirá a utilização de uma base de dados com imagens noturnas de animais para a fase de treinamento, como esperado.

6. CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma prova de conceito, de que a aprendizagem de máquina, alinhado a sistemas de reconhecimento, possui resultados satisfatórios na identificação de

animais, e que podem facilmente ser replicados para a identificação de animais em ambientes noturnos.

Podemos citar como contribuição, a utilização do *framework* de treinamento aplicado, que mais uma vez, se mostrou eficaz em seu propósito, assim como a replicação feita pela Universidade da Califórnia sob o estudo de *AlexNet*, que embasou este estudo. Dessa forma, podemos apontar que a prova de conceito apresentada neste trabalho, reforça os métodos e os resultados obtidos por *A. Krizhevsky*.

6.1 Trabalhos Futuros

No futuro, a identificação de animais em reservas naturais inclina-se a ser algo mais efetivo e produtivo. A obtenção de uma base de dados mais representativa se faz necessária para dar continuidade a esse estudo de monitoramento e classificação. Também vale ressaltar que, a incorporação de novas informações, com características físicas de animais (como a altura, porte físico e entre outras), além da classificação, tende a ampliar as possibilidades de aplicações que possam ser desenvolvidas.

6.2 Limitações

Uma das limitações mais significativas do trabalho, no processo de reconhecimento de animais em ambientes industriais, consiste na falta de conjuntos de dados mais representativos. Os poucos localizados não apresentam uma grande quantidade de imagens relacionadas a esses ambientes, o que torna o treinamento de redes profundas, como a *Caffe*, pouco efetivo, podendo acarretar em baixas taxas de aprendizagem da rede e consequentemente redução da eficácia do sistema de classificação das imagens.

7. REFERÊNCIAS

- [1] Burghardt, T. e Calic, J. e Thomas, B. Tracking Animals in Wildlife Videos Using Face Detection (2004).
- [2] Classification: Instant Recognition with Caffe, <https://nbviewer.jupyter.org/github/BVLC/caffe/blob/master/examples/00-classification.ipynb>.
- [3] Fahad, A. e Abrar, A. e Eiji, K. Animal species classification using machine learning techniques (2019).
- [4] Indústria impulsiona setor produtivo e gera emprego e renda. <https://noticias.portaldaindustria.com.br/noticias/economia/industria-impulsiona-setor-produtivo-e-gera-emprego-e-renda/>
- [5] Jia, Y. e Shelhamer, E. e Donahue, J. e Karayev, S. e Long, J. e Girshick, R. e Guadarrama, S. e Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding (2014).
- [6] Kaiming, H. e Gkioxari, G. e Dollár, P. e Girshick, R. Mask R-CNN (2017).
- [7] Kaiming, H. e Xiangyu, Z. e Shaoqing, R. e Jian, S. Deep residual learning for image recognition (2015).
- [8] Krizhevsky, A. e Sutskever, I. e Hinton, G. E. ImageNet classification with deep convolutional neural networks (2012).
- [9] Kumar Y. H. S. e Manohar, N. e Chethan, H. K. Animal Classification System: A Block Based Approach (2015).
- [10] Russakovsky, O. e Deng, J. e Su, H. e Krause, J. e Satheesh, S. e Ma, S. e Huang, Z. e Karpathy, A. e Khosla, A. e Bernstein, M. e Berg, A. C. e Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge (2015).
- [11] Schindler, F. e Steinhage, V. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques (2021).
- [12] Tekeli, U. e Bastanlar, Y. Elimination of Useless Images from Raw Camera-Trap Data (2019).

Sobre os autores:

Vinicius Jorge Pereira da Silva. Graduando em Ciência da Computação, com 4 anos de experiência em Desenvolvimento de Software. Participou por 1 ano do projeto Guardians - UFCG como Desenvolvedor *Fullstack*. Atuou por 2 anos como estagiário na empresa E.life, como Desenvolvedor *Frontend* de aplicações e serviços e, atualmente, é Desenvolvedor Pleno da mesma empresa.

Joseana Macêdo Fechine Régis de Araújo. Professora orientadora.