



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

JHONATAN BATISTA DA SILVA

**SUPERRESOLUÇÃO EM SISTEMAS DE VIDEOVIGILÂNCIA:
UM ESTUDO COMPARATIVO**

CAMPINA GRANDE - PB

2020

JHONATAN BATISTA DA SILVA

**SUPERRESOLUÇÃO EM SISTEMAS DE VIDEOVIGILÂNCIA:
UM ESTUDO COMPARATIVO**

**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.**

Orientador: Professor Dr. Herman Martins Gomes.

CAMPINA GRANDE - PB

2020



S586s Silva, Jhonatan Batista da.
Superresolução em sistemas de videovigilância: um estudo comparativo. / Jhonatan Batista da Silva. - 2020.
12 f.

Orientador: Prof. Dr. Herman Martins Gomes.
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. Superresolução de imagens digitais. 2. Redes neurais artificiais profundas. 3. Videovigilância. 4. Visão computacional. 5. Placas de veículos - identificação. 6. Identificação de caracteres - placas de veículos. I. Gomes, Herman Martins. II. Título.

CDU:004(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

JHONATAN BATISTA DA SILVA

**SUPERRESOLUÇÃO EM SISTEMAS DE VIDEOVIGILÂNCIA:
UM ESTUDO COMPARATIVO**

**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:

**Professor Dr. Herman Martins Gomes
Orientador – UASC/CEEI/UFPG**

**Professor Dr. Maxwell Guimarães de Oliveira
Examinador – UASC/CEEI/UFPG**

**Professor Dr. Tiago Lima Massoni
Disciplina TCC – UASC/CEEI/UFPG**

Trabalho aprovado em: 2020.

CAMPINA GRANDE - PB

Superresolução em sistemas de videovigilância: um estudo comparativo

Jhonatan Batista da Silva*
jhonatan.silva@ccc.ufcg.edu.br
Unidade Acadêmica de Sistemas e Computação,
Universidade Federal de Campina Grande
Campina Grande, Paraíba

Herman Martins Gomes
hmg@computacao.ufcg.edu.br
Unidade Acadêmica de Sistemas e Computação,
Universidade Federal de Campina Grande
Campina Grande, Paraíba

RESUMO

Gravações de sistemas de videovigilância podem ser consultadas por sistemas inteligentes com o objetivo de identificação de placas de veículos, faces, entre outros elementos. As imagens encontradas, muitas vezes, não possuem um nível de detalhamento necessário para o reconhecimento automático destes componentes. Esta pesquisa objetiva comparar técnicas que permitam realizar o aprimoramento de imagens de baixa resolução, com foco na identificação de placas de veículos licenciados no Brasil. Para alcançar o objetivo proposto, técnicas do estado-da-arte (superresolução via redes neurais artificiais profundas) são comparadas com técnicas convencionais (a exemplo de redimensionamento com interpolação) com o propósito de verificar a capacidade de melhoria de cada técnica. As avaliações qualitativas e quantitativas foram realizadas sobre os resultados relativos ao reconhecimento dos caracteres presentes nas placas dos veículos com o objetivo de contrastar as técnicas estudadas. Foi concluído, em razão dos resultados produzidos, que as técnicas de superresolução do estado-da-arte superam os métodos convencionais quando aplicadas ao contexto de identificação de elementos em sistemas de videovigilância.

PALAVRAS-CHAVE

Superresolução, redes neurais artificiais profundas, videovigilância.

1 INTRODUÇÃO

Sistemas de videovigilância podem ser implementados em rodovias, ruas, estacionamentos de shoppings, empresas, ou em qualquer lugar que se faça necessário o monitoramento de atividades que envolvam fluxo de veículos [4]. Uma vez que, a forma de identificação de veículos pode ser realizada pela imagem da placa, os sistemas citados são utilizados principalmente para determinar quais veículos estão autorizados ou não a entrarem em um ambiente, abrir cancelas e portões automaticamente, aplicar restrições de trânsito e até mesmo para fins de investigação policial.

Os principais problemas atrelados aos fins supracitados estão relacionados à não identificação dos caracteres impressos nas placas de veículos, bem como a falsa identificação desta informação. Desta forma, o presente trabalho apresenta formas de aprimorar a legibilidade de placas de veículos brasileiras por intermédio do uso de técnicas para o aprimoramento da resolução de imagens digitais.

Superresolução de imagens digitais, como é definido por [3], é o processo de síntese de uma imagem de alta resolução a partir de uma determinada imagem de baixa resolução. Uma imagem pode ser dita de baixa definição devido a sua resolução espacial mais baixa, ou seja, suas dimensões, ou devido ao resultado de alguma

fonte de degradação (como desfoque, ruídos de aquisição ou de transmissão).

Ao aplicar uma função de degradação em uma imagem de alta resolução qualquer, é possível obter uma imagem com menor nível de detalhes. A questão de pesquisa é a seguinte: seria possível realizar a operação inversa à degradação? No caso ideal, sim. Se conhecermos a função exata de degradação, aplicando sua inversa à imagem de baixa definição, é possível recuperar a imagem de alta resolução.

Contudo, a função de degradação geralmente não é conhecida. A estimativa direta da função de degradação inversa é um problema sem solução única. Apesar disso, as técnicas de Aprendizagem de Máquina Profunda, como foi introduzido por Dong et al. [5], vem provando serem adequadas a este contexto de problema.

Dentre as técnicas conhecidas e utilizadas globalmente, é possível citar as técnicas que podem ser denominadas por técnicas convencionais. As técnicas convencionais surgiram antes da explosão de uso das redes neurais convolucionais que, em seu estado atual, são utilizadas para melhor solucionar a maioria dos problemas relacionados à visão computacional, inclusive, como será apresentado mais adiante, o problema de aprimoramento de resolução de imagens. Em relação às técnicas convencionais, as mais populares, usadas frequentemente para este fim, são as técnicas de Interpolação Bilinear e Interpolação Bicúbica (para fins de referência, consultar o livro de Gonzalez e Woods [8]), em que a segunda técnica tem historicamente apresentando resultados superiores à primeira. Por outro lado, as técnicas fundamentadas em redes neurais, são soluções apresentadas em trabalhos recentes, sendo notável a crescente melhoria nos resultados observados.

As soluções da classe de redes neurais analisadas no presente documento são implementações baseadas em uma arquitetura de redes neurais inicialmente projetada por Ian Goodfellow [9] conhecidas como GAN (*Generative Adversarial Network*). O processo de aprendizado ocorre de tal forma que duas redes neurais competem entre si na forma de um jogo de soma zero, em que o ganho de um agente é a perda de outro.

Adiante, é feita uma revisão sobre pesquisas que contribuíram para o campo de superresolução de imagens digitais, citando as abordagens convencionais mais comuns, e como as técnicas de superresolução via redes neurais artificiais profundas colaboraram para o avanço desta área. Além disso, é explanado a técnica que integra o estado-da-arte no contexto de recuperação de placas de veículos. O estudo comparativo entre as técnicas convencionais e as técnicas via redes neurais conta com um experimento detalhado na Seção 5, que consiste nas avaliações sobre os resultados visuais e numéricos.

2 OBJETIVOS

Este estudo tem como objetivo investigar o emprego de técnicas de superresolução de imagens digitais no contexto de imagens contendo placas de veículos obtidas por câmeras de videovigilância e inferir conclusões de seus resultados comparativamente a técnicas mais tradicionais. Entre as técnicas analisadas estão sistemas de aprendizado de máquina conhecidos como GAN (Generative Adversarial Network), especificamente os métodos SRGAN [11] e ESRGAN [13] detalhados na Seção 4 e a técnica convencional de interpolação bicúbica. As soluções investigadas são avaliadas a partir análise visual subjetiva e de métricas numéricas, inferidas dos resultados de recuperação dos caracteres das placas de veículos contidas nas imagens, geradas pelos métodos de aprimoramento de resolução de imagem citados. A extração do conteúdo das placas de veículos das imagens foi realizada utilizando a ferramenta de OCR (*Optical Character Recognition*) Tesseract [1].

3 PESQUISAS RELACIONADAS

De uma forma geral, as abordagens para lidar com o problema de superresolução de imagens podem ser divididas em três categorias: 1) as abordagens baseadas em interpolação, 2) as abordagens baseadas em reconstrução e 3) as abordagens baseadas em aprendizagem. Os algoritmos baseados em interpolação perduraram por muito tempo por serem abordagens simples e eficientes.

A tarefa de superresolução de imagens via redes neurais artificiais é um campo de pesquisa relativamente novo quando levamos em consideração que o método pioneiro proposto por Dong et al. [5] surgiu apenas em 2014, já apresentando resultados superiores aos resultados conquistados por métodos convencionais. Nos anos seguintes, surgiram várias pesquisas que propuseram diferentes abordagens utilizando redes neurais artificiais, sendo importante destacar a pesquisa de Zhang et al. [14], onde foi introduzida uma arquitetura de rede neural artificial significativamente mais complexa, na qual foi apresentado o mecanismo RDB (*residual dense block*) para extrair características locais por meio de densas camadas convolucionais fortemente conectadas. O RDB permitiu ainda conexões diretas, durante a fase de treinamento, do estado do RDB anterior para todas as camadas do RDB atual por meio de um mecanismo de memória contínua.

Em maio de 2017, Ledig et al. [11] propôs um método baseado em uma rede neural artificial do tipo GAN, nomeado SRGAN, que, segundo os autores foi o primeiro método capaz de produzir imagens foto-realísticas com fator de 4X de superresolução. As contribuições do método SRGAN, no momento de sua publicação, foram as de: 1) estabelecer um novo estado da arte no contexto de superresolução de imagens e 2) Introduzir uma nova função de perda atrelada ao treinamento da rede neural artificial chamada *Perceptual Loss* que substituiu a muito utilizada MSE (*mean squared error*).

Com o objetivo de aprimorar a solução proposta pelo método SRGAN, Wang e Yu et al. [13], propuseram o ESRGAN (*Enhanced SRGAN*), também um método baseado em redes neurais artificiais muito similar ao proposto anteriormente em SRGAN - em que, segundo os autores de ESRGAN, os resultados produzidos casualmente eram acompanhados de artefatos indesejados - mas com algumas adições e melhorias na arquitetura e nas funções de perda

da rede neural. SRGAN e ESRGAN são as duas soluções baseadas em redes neurais avaliadas na seção experimental do presente trabalho.

3.1 Estado da arte em recuperação de placas de veículos

O método proposto por Wu Liu et al. [12], integra o estado da arte quando se trata de superresolução de imagens de placas de veículos. Aquele trabalho propôs uma solução baseada em redes neurais e também implementa uma rede neural do tipo GAN, mas com algumas particularidades. O método proposto parte de uma abordagem conhecida como MSR (*Multi-Image Super-Resolution*) e depende de mais de uma imagem de entrada para gerar a saída, que constitui de uma imagem de resolução aprimorada SR. A rede neural desenvolvida por Wu Liu foi denominada DP-GAN. No entanto, como pode ser percebido na Figura 1, existem vários pré-requisitos para que o método funcione como foi proposto:

- Parte da premissa que a placa do veículo que se deseja identificar seja capturada em diferentes momentos e câmeras; isso inclui variações de desfoque, resolução, oclusão, iluminação e ângulo de captura.
- Depende da segmentação manual da localização da placa do veículo na imagem.
- Usa um mecanismo de busca que recupera imagens que supostamente contém a imagem de consulta, onde consulta é a imagem da placa derivada da imagem original.

Resumindo o funcionamento do método de superresolução aplicado a placas de veículos proposto no artigo *Beyond human-level license plate super-resolution with progressive vehicle search and domain priori GAN* [12], é possível destacar as seguintes contribuições: 1) foi projetada uma estrutura de super resolução de placas de veículos que ficou representada como uma das primeiras tentativas de utilizar a informação adicional entre as imagens de placas de veículos pesquisadas e capturadas em diferentes câmeras e momentos e 2) foi proposto o modelo de rede neural artificial DP-GAN, que combina o conhecimento prévio da distribuição de dados e o conhecimento do domínio de placas de veículos para gerar as imagens de placas de veículos de alta resolução e correspondentes espaciais.

4 ARQUITETURAS DAS SOLUÇÕES BASEADAS EM GANS AVALIADAS

Como foi observado na pesquisa de Liu et al. [12], mesmo não sendo a única arquitetura de rede neural artificial utilizada para este fim, a arquitetura GAN se prova compatível com o problema de superresolução de imagens. Nesta seção é apresentada uma visão um pouco mais técnica sobre ambos os métodos baseados em GANs avaliados (SRGAN e ESRGAN).

Uma vez que o método ESRGAN [13] é uma extensão do método SRGAN [11] o entendimento do primeiro está atrelado ao entendimento do segundo. Na Figura 2 é apresentada uma visão detalhada da arquitetura das duas redes neurais que compõem o método. Durante a fase de treinamento a rede neural geradora G tem o objetivo de enganar a rede neural discriminadora D, gerando imagens de alta resolução a partir de uma imagem de baixa resolução passada como entrada para o treinamento da rede. A rede neural D, por sua vez, tem o papel de diferenciar entre imagens geradas por G e imagens

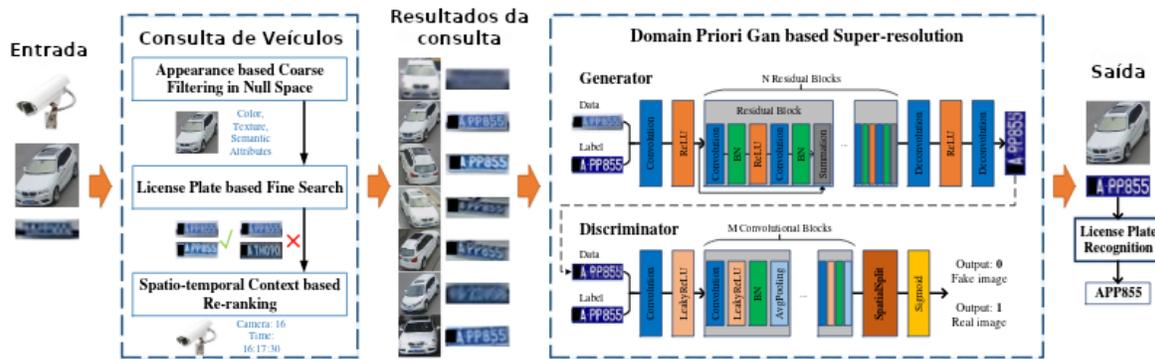


Figura 1: Visão geral da proposta de super-resolução de placas de veículos múltiplas com pesquisa progressiva de veículos e domain priori GAN. Fonte: [12].

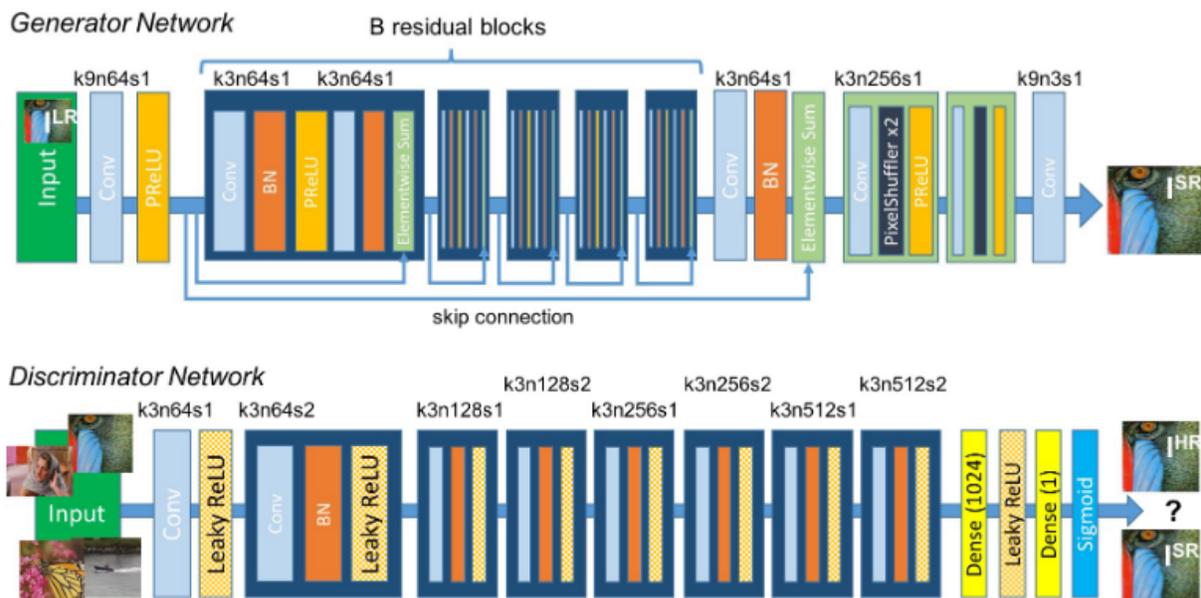


Figura 2: Arquitetura das redes neurais artificiais G e D implementadas no método SRGAN. Fonte: [11]

de alta resolução que serviram de base para originar as imagens em baixa resolução utilizadas por G. Esse é o funcionamento de um GAN em síntese. Nas camadas internas de G, são incorporados blocos que seguem a arquitetura proposta por Gross e Wilber em Training and investigating residual nets [10], e também foi utilizado no trabalho de Zhang et al. [14]. Especificamente, são usadas duas camadas convolucionais com pequenos núcleos 3×3 e 64 mapas de características seguidos por camadas de batch-normalization.

Dessa forma, as alterações adicionadas por ESRGAN, no contexto arquitetural, estão atreladas aos blocos residuais incorporados ao modelo GAN apresentado na Figura 2. A esta alteração foi atribuído o conceito RRDB (*Residual in Residual Dense Block*) e sua

estrutura pode ser entendida através da Figura 3. O uso dos blocos RRDB tornam a arquitetura do método ESRGAN mais profunda e complexa do que a arquitetura do método SRGAN, baseados na observação, declarada pelos autores, de que mais complexidade significa melhores resultados. Além disso, a remoção de camadas BN (*batch-normalization*) provou aumentar o desempenho e reduzir a complexidade computacional.

Funções de perda otimizadas foram contribuições propostas por ambos trabalhos (SRGAN e ESRGAN). O conceito de *Perceptual Loss* foi introduzido no método SRGAN, e seu funcionamento é o de uma função de perda que avalia uma solução com relação às características perceptualmente relevantes. Resumidamente o

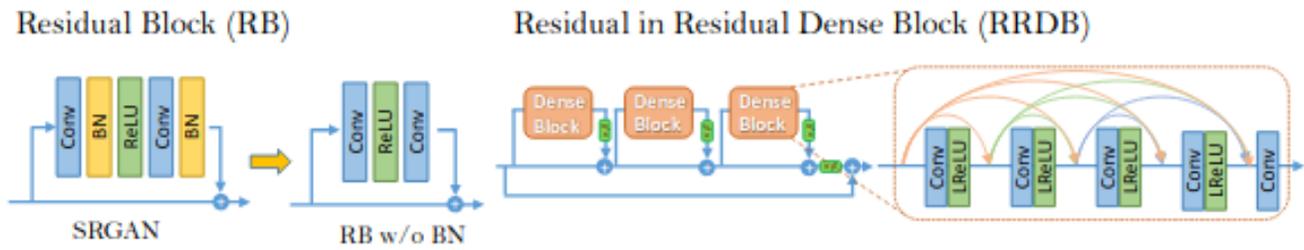


Figura 3: Alteração proposta nos blocos residuais RB introduzindo RRDB implementada no método ESRGAN. Fonte: [13]

Perceptual Loss é a soma ponderada de duas funções de perda: a do conteúdo gerado e a perda adversarial, que, por sua vez, é o tipo de função de perda utilizada por GANs. A contribuição do método ESRGAN para o *Perceptual Loss*, se dá pelo fato de o uso de recursos acontecer antes das camadas de ativação, ao contrário do que acontece no método SRGAN, no qual acontece depois das camadas de ativação.

de dados inteiro contém um total de 35 Gigabytes. Existem 3.595 amostras de treinamento, 2.360 amostras de teste e 705 amostras de validação, totalizando 6.660 imagens com 8.683 placas de 815 veículos diferentes. Além das imagens, é fornecido juntamente ao dataset rotulagens de posição e texto contido de uma placa para cada uma das 6.660 imagens. Na Figura 4 a estão reunidas amostras das imagens que compõem o dataset.



Figura 4: Amostras de imagens que compõem o conjunto de dados.

5 AVALIAÇÃO EXPERIMENTAL

A avaliação experimental teve como propósito comparar técnicas de superresolução de imagens e avaliá-las de forma qualitativa e quantitativa, por intermédio de análise subjetiva e métricas obtidas sobre o resultado do OCR aplicado aos recortes de placas de veículos contidas nas imagens.

5.1 Conjunto de dados

O conjunto de dados selecionado para os experimentos, chamado *Sense-ALPR Database* [7], conta com 6.660 imagens de câmeras de videovigilância, todas com veículos, onde é possível ter a visualização de pelo menos uma placa de veículo. Este dataset foi criado para ser usado na pesquisa de Gonçalves et al. [7]. Seu objetivo original era o de auxiliar os pesquisadores a avaliar problemas de reconhecimento automático de placas brasileiras. As imagens do conjunto de dados estão em Full-HD e estão disponíveis no formato Portable Network Graphics (PNG) com 1920 × 1080 pixels cada. O conjunto

5.2 Detalhes do treinamento

Para ser possível o treinamento de uma rede neural de superresolução, é necessário que para cada imagem X em baixa resolução utilizada durante o treinamento, exista uma imagem Y em alta resolução de tal forma que X seja uma imagem derivada de Y através de um processo de redução de resolução. O fator de redução da resolução adotado nos experimentos deste artigo foi de 1/4 (ou seja, X tem dimensões que são 1/4 das dimensões de Y) resultando em imagens de resolução 480×270. Este mesmo fator foi adotado nos artigos em que foram introduzidos os métodos SRGAN e ESRGAN.

Ambas as redes neurais (SRGAN e ESRGAN) foram treinadas em GPU utilizando as implementações disponíveis na biblioteca ISR de Cardinale [6] com código fonte em Python utilizando o framework Tensorflow 2.0. O modelo de GPU utilizado foi a Nvidia GTX 1660 com 6GB de memória RAM GDDR5. As demais configurações da máquina utilizada são: Processador AMD Ryzen 7 2700, 16GB de memória RAM DDR4 e sistema operacional Ubuntu 20.04 LTS. Os dados ficam armazenados em um HD SATA de 500GB.

Cada treinamento durou aproximadamente 8 horas e o critério de parada foi configurado como sendo o número de 400 épocas. O script de treinamento para ambas as redes estão disponíveis no github do autor do trabalho¹. Foram utilizadas todas 3.595 imagens disponíveis no conjunto de treinamento e 705 imagens disponíveis no conjunto de validação para os seus respectivos fins. A evolução do treinamento pode ser observada nas curvas de perda no conjunto de validação das redes neurais geradoras de cada método SRGAN e ESRGAN apresentadas na Figura 5, em (a) e (b), respectivamente. Os valores de perda ao final do treinamento de ambas as redes neurais ficaram em entre 0.03 e 0.04, e também é perceptível a convergência da rede neural em torno do número de épocas 200, pois, a partir desse ponto, a variação dos valores de perda é muito pequena.

¹Repositório contendo scripts de treinamento disponível através do link <https://github.com/jhonatanbds/GANS-superresolution>

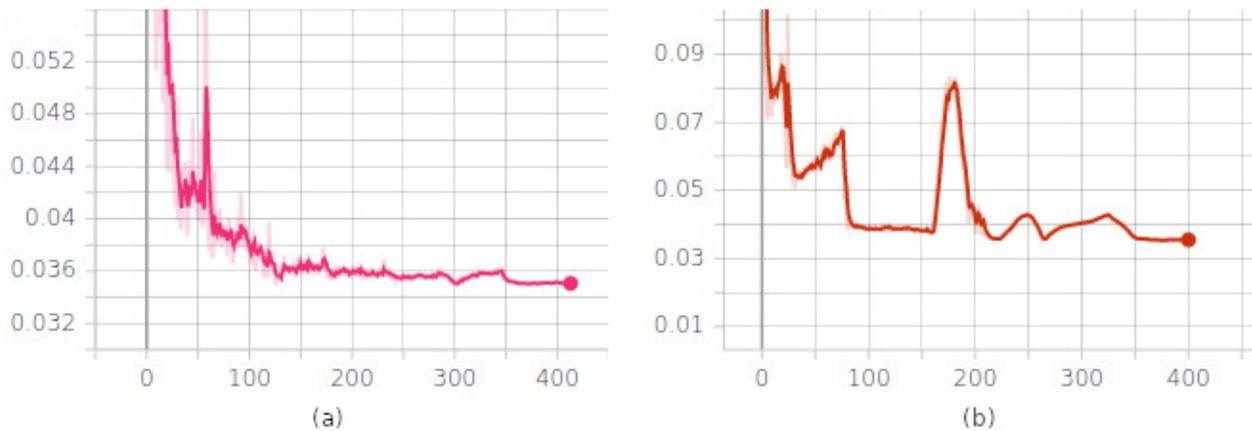


Figura 5: (a) Curva de evolução do valor de perda da rede geradora do método SRGAN e (b) Curva de evolução do valor de perda da rede geradora do método ESRGAN.

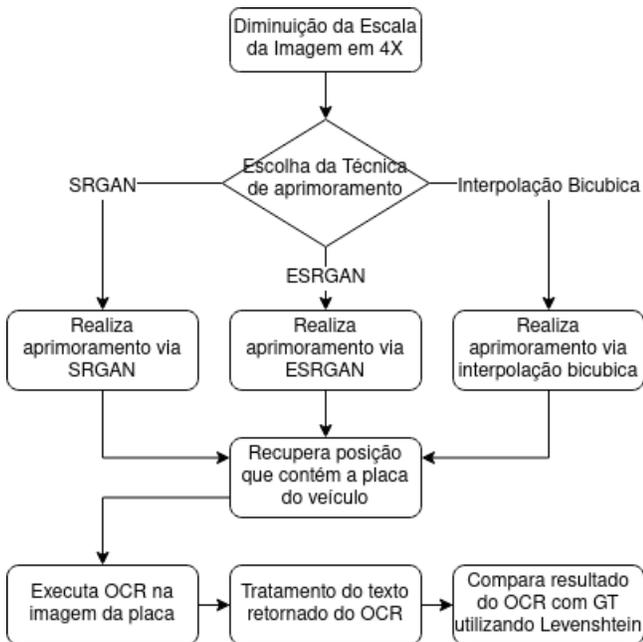


Figura 6: Fluxograma que ilustra o projeto experimental. Fonte: autoria própria.

5.3 Projeto Experimental

O experimento como um todo, em seu aspecto quantitativo, segue um fluxograma que pode ser melhor visualizado na Figura 6. No primeiro momento temos um conjunto de imagens, que foram apresentadas como conjunto de teste na Subseção 5.1, e então para cada imagem realizamos uma redução de resolução em 1/4. O próximo passo é decidir qual técnica de superresolução utilizar, e, após a aplicação da técnica de superresolução, é recuperado as coordenadas na imagem onde está localizada a placa do veículo para a qual foi

feita a rotulagem do conteúdo da placa. As informações de localização e texto da placa do veículo de cada imagem são providas pelo dataset. No próximo momento, é executado OCR com a ferramenta Tesseract t[1] no recorte da imagem que contém a placa do veículo recuperada no passo anterior. Logo após, é feito um tratamento no resultado retornado pelo OCR que se refere às tarefas de busca pela linha de maior comprimento no texto retornado, que assumimos ser o texto da placa, e remoção de caracteres especiais (todos caracteres não alfa-numéricos). Por fim, é feita uma comparação entre o texto tratado e o *groundtruth* (GT), ou seja, a saída esperada, provido pelo dataset para a imagem a qual executamos o fluxograma. A comparação é feita utilizando a medida de similaridade conhecida como Distância de Levenshtein [2], que, em resumo, considerando uma string S1 e uma segunda string S2, se trata da quantidade de caracteres que é necessário alterar em S1 para que fique exatamente igual à S2.

5.4 Resultados

Os primeiros resultados produzidos ao término do treinamento das duas redes neurais foram gerados a partir do conjunto de teste descrito na Subseção 5.1. Para gerar as imagens do conjunto de teste em resolução aprimorada pelos métodos treinados, primeiro foi necessário a redução de resolução das imagens deste conjunto no fator de 1/4. Na Figura 7 são apresentados recortes de imagens que compõem o conjunto de teste. Os recortes foram feitos para que possam ser visualizadas em mais detalhes as diferenças entre as imagens. Os recortes estão nas dimensões 400x400 para as imagens que estão na resolução 1920x1080 e em 100x100 para a imagem original que possui a resolução de 480x270. Visualmente, é perceptível a melhora entre a imagem original, em baixa resolução, e a imagem aprimorada através da técnica de interpolação bicúbica, a legibilidade da placa do veículo nesse caso não foi prejudicada em nenhuma das imagens. Entretanto, existe uma melhora significativa na definição de detalhes mais finos entre o método convencional de interpolação bicúbica e os métodos baseados em redes neurais. No



Figura 7: Recortes de imagens pertencentes ao conjunto de teste após a aplicação de cada método estudado. Fonte: autoria própria.

Tabela 1: Quantidade de distâncias de Levenshtein, para cada técnica, que foram menores que o valor indicado.

Conjunto de imagens	Qtd. de distâncias < 5	Qtd. de distâncias < 4	Qtd. de distâncias < 3	Qtd. de distâncias < 2
Original	31	19	8	4
Interpolação Bicúbica	196	122	66	27
SRGAN	231	137	73	30
ESRGAN	203	126	60	28

entanto, não é possível notar visualmente alguma diferença entre os métodos SRGAN e ESRGAN nas amostras listadas.

No caso separado na Figura 8, existe uma placa de veículo que não é identificável em baixa resolução, e nem mesmo após a aplicação da técnica de interpolação bicúbica, mas ao utilizar um dos métodos baseados em redes neurais, se torna possível a identificação dos caracteres da placa, que, ao *groundtruth*, foi atribuído o texto "OPJ4900".

A Tabela 1 foi construída a partir dos resultados produzidos pelo cálculo da distância de Levenshtein sobre o resultado do OCR de cada imagem dos conjuntos produzidos por cada uma das técnicas (mais detalhes na Subseção 5.3) mais o conjunto de teste sem nenhum processo de superresolução. Nas colunas, estão registrados os valores que representam a quantidade de distâncias de Levenshtein, para cada técnica, que foram menores que o valor indicado. Através da análise dos valores da tabela, podemos concluir que os melhores resultados são do método SRGAN, seguido dos resultados do método ESRGAN, como era esperado.

Na Figura 9 são exibidos recortes de placas presentes nas imagens do conjunto de teste após a aplicação dos métodos estudados.

É interessante analisar o último caso exibido pois é possível a leitura de diferentes textos nas variadas imagens de placas de veículos produzidas por cada técnica, dependendo da interpretação de quem a lê. Ao verificar o *groundtruth* da imagem em questão nos deparamos com o texto "OWM2434". Sendo assim, é possível concluir que a técnica que mais se aproximou, nesse caso, foi o método ESRGAN.

6 CONSIDERAÇÕES FINAIS

O estudo das técnicas de superresolução de imagens digitais teve intensa atividade nos últimos cinco anos e, no momento da escrita deste artigo, é possível que existam formas mais poderosas para a solução do problema de identificação de elementos, não facilmente visíveis em sistemas de videovigilância, de forma automática. O objetivo de descobrir se as técnicas que foram analisadas auxiliavam na solução do problema proposto foi atingido, e foi obtida uma resposta positiva.

Uma vez que o foco deste trabalho se prendeu ao processo de superresolução em si e não ao restante do processo de recuperação dos caracteres presentes nas placas de veículos existentes nas imagens que compõem o conjunto de dados, é possível afirmar que

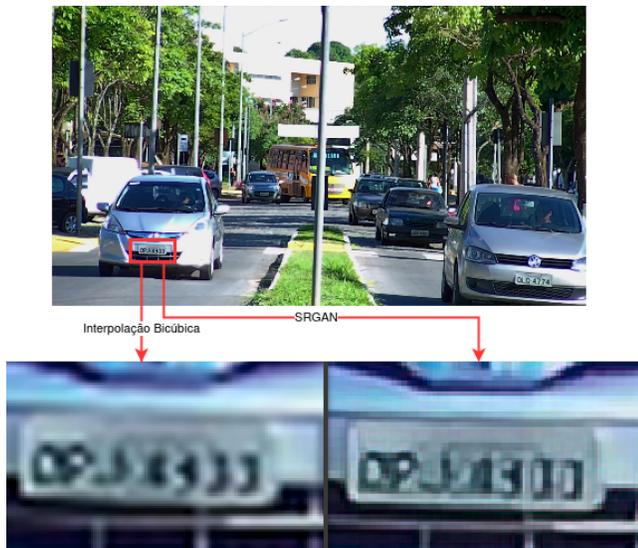


Figura 8: Exemplo de imagem em que o método baseado em redes neurais (SRGAN) torna uma placa de veículo mais legível. Fonte: autoria própria.

o resultado do processo de extração como um todo pode ainda ser melhorado. Dado que existem, na academia, técnicas de pré-processamento de imagens com o objetivo de filtrar ruídos e realçar caracteres em imagens, a tarefa da ferramenta de OCR Tesseract [1] de identificar esses caracteres da imagem pode ser facilitada. Por outro lado, se atribuirmos essa tarefa para um humano, esse resultado poderia ser ainda mais positivo.



Figura 9: Recortes de placas presentes nas imagens do conjunto de teste após a aplicação dos métodos estudados. Fonte: autoria própria.

AGRADECIMENTOS

Agradeço aos meus pais, que apesar de todos problemas que enfrentaram nos últimos dois anos, me deram o suporte para que continuasse, mesmo longe deles, seguindo meu objetivo. Também agradeço a minha esposa, que em todo momento me encorajou e tornou meus dias ruins mais agradáveis. E, por fim, agradeço aos meus colegas de curso e professores que me agregaram todo o conhecimento necessário para a realização deste trabalho.

REFERÊNCIAS

- [1] [n. d.]. tesseract-ocr. <https://github.com/tesseract-ocr>
- [2] 2020. Levenshtein distance. https://en.wikipedia.org/wiki/Levenshtein_distance
- [3] 2020. Super-resolution imaging. https://en.wikipedia.org/wiki/Super-resolution_imaging
- [4] Suporte Active. 2017. Os sistemas de videovigilância inteligentes. <https://www.activesolutions.com.br/blog/os-sistemas-de-videovigilancia-inteligentes/>
- [5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 2 (2016), 295–307. <https://doi.org/10.1109/tpami.2015.2439281>
- [6] Francesco Cardinale et al. 2018. ISR. <https://github.com/idealo/image-super-resolution>.
- [7] Gabriel Resende Gonçalves, Matheus Alves Diniz, Rayson Laroca, David Menotti, and William Robson Schwartz. 2018. Real-time automatic license plate recognition through deep multi-task networks. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 110–117.
- [8] Rafael C. Gonzalez and Gonzalez amp; Woods. [n. d.]. Digital Image Processing, 3rd Edition. <https://www.pearson.com/us/higher-education/program/Gonzalez-Digital-Image-Processing-3rd-Edition/PGM197080.html>
- [9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144. <https://doi.org/10.1145/3422622>
- [10] Sam Gross and Michael Wilber. 2016. Training and investigating residual nets. *Facebook AI Research* 6 (2016).
- [11] Christian Ledig, Lucas Theis, Ferenc Huszár, José Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.
- [12] Wu Liu, Xinchun Liu, Huadong Ma, and Peng Cheng. 2017. Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Prior GAN. *Proceedings of the 2017 ACM on Multimedia Conference - MM 17* (2017). <https://doi.org/10.1145/3123266.3123422>
- [13] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2019. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *Lecture Notes in Computer Science Computer Vision - ECCV 2018 Workshops* (2019), 63–79. https://doi.org/10.1007/978-3-030-11021-5_5
- [14] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. 2018. Residual Dense Network for Image Super-Resolution. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018). <https://doi.org/10.1109/cvpr.2018.00262>