

# Study on the impact of the degradation method on the generalization of Super-Resolution models for ALPR

Cristiano L. Oliveira<sup>1</sup>, Leonardo N. Matos<sup>1</sup>, Paulo S. G. de M. Neto<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação (PROCC) – Universidade Federal de Sergipe (UFS) – 49.100-000 – São Cristóvão – SE – Brazil

<sup>2</sup>Centro de Informática – Universidade Federal de Pernambuco (UFPE) – 50.740-560 – Recife – PE – Brazil

cristianoengufs@gmail.com, leonardo@dcomp.ufs.br, psgmn@cin.ufpe.br

**Abstract.** *Given the complexity of variations in scenarios and equipment, it is crucial to employ advanced image enhancement methods in Automatic License Plate Recognition (ALPR). This study examined the impact of different image degradation methods during data synthesis for training models based on the Real-ESRGAN super-resolution architecture. The results showed significantly greater generalization power when using a dataset constructed with a more robust degradation method.*

**Resumo.** *Dada a complexidade das variações nos cenários e nos equipamentos, é fundamental empregar métodos avançados de aprimoramento de imagens no Reconhecimento Automático de Placas de Licença (ALPR). Este estudo analisou o impacto de diferentes métodos de degradação de imagens na síntese de dados para treinamento de modelos baseados na arquitetura de super-resolução Real-ESRGAN. Os resultados demonstraram um poder de generalização significativamente maior ao utilizar um conjunto de dados construído com um método de degradação mais robusto.*

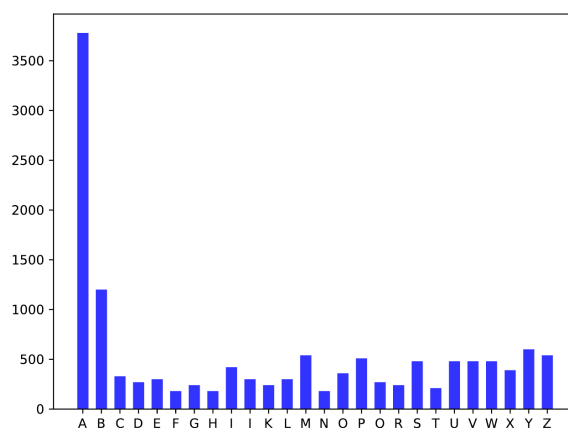
## 1. Introdução

O reconhecimento automático de placas de licença (ALPR - *Automatic License Plate Recognition*) é um sistema inteligente composto normalmente pelas etapas de detecção de placas, segmentação e reconhecimento de caracteres. Esse sistema é capaz de extrair a sequência de caracteres que identifica a placa de um veículo a partir de uma imagem. Em aplicações reais, essas imagens podem ser capturadas em uma ampla variedade de ambientes, condições climáticas, tipos e configurações de câmeras, o que resulta em variações significativas na qualidade das imagens com as quais o sistema deve lidar para realizar o reconhecimento preciso dos caracteres [Laroca et al. 2018].

Os principais desafios enfrentados por aplicações de ALPR estão relacionados principalmente à presença de ruídos nas imagens, que podem surgir por diversos fatores, incluindo a própria natureza de imagens em baixa resolução, que tende a introduzir ruídos indesejados [Sereethavekul and Ekpanyapong 2023]. Isso ocorre porque a placa de um veículo é um objeto pequeno em uma imagem que geralmente inclui vários outros elementos, como o próprio veículo e parte do ambiente circundante [Pan et al. 2023]. Assim, ao recortar a região da imagem onde a placa está localizada para realizar o reconhecimento, os caracteres frequentemente aparecem degradados, com contornos pouco nítidos

ou até mesmo completamente deformados, o que reduz significativamente as chances de sucesso do reconhecimento pelo sistema ALPR. Esse desafio é ainda maior em cenários onde os veículos estão em movimento, pois o desfoque de movimento impacta negativamente a qualidade da imagem [Kim et al. 2024].

Outros fatores que impactam os resultados do ALPR incluem a deformação de perspectiva [Rao et al. 2024], onde os caracteres são alterados devido à falta de paralelismo entre o plano de projeção da lente da câmera e o plano da placa. Isso ocorre quando a câmera registra uma imagem tridimensional em uma representação bidimensional, resultando em deformidades visuais dos objetos. Além disso, os conjuntos de dados utilizados no treinamento [Asaad et al. 2023] também representam um desafio. Frequentemente, a etapa de reconhecimento pode ter desempenho inferior para determinados caracteres, devido à sub-representação de alguns deles no conjunto de dados de treinamento. Isso significa que certas letras ou números aparecem com menor frequência nas imagens utilizadas para treinar o modelo de rede neural, como ilustrado na Figura 1.



**Figura 1. Distribuição de caracteres no conjunto de dados UFPR-ALPR.**

[Laroca et al. 2018]

Para cada desafio encontrado nas aplicações de ALPR, existem métodos e técnicas específicas para abordá-los. No contexto deste trabalho, foi explorado o método de super-resolução, amplamente utilizado para resolver problemas de baixa resolução, ruídos e desfoques causados por limitações ópticas [Wang et al. 2021a]. Um exemplo notável é o Real-ESRGAN [Wang et al. 2021b], que não apenas aumenta a resolução das imagens, mas também trata simultaneamente os ruídos e desfoques. Contudo, alguns estudos tratam a baixa resolução e a baixa qualidade separadamente, utilizando diferentes modelos para cada problema, com a premissa de melhorar o resultado final ao isolar o domínio das imagens de baixa resolução do domínio das imagens ruidosas e desfocadas [Kim et al. 2024, Hamdi et al. 2021]. No entanto, essa abordagem aumenta o tempo de execução do modelo final, o que pode limitar a aplicabilidade em sistemas de ALPR que exigem processamento em tempo real [Pattanaik and Balabantaray 2023].

No trabalho atual, especificamente, buscou-se avaliar a capacidade de generalização dos modelos de super-resolução, analisando o impacto da forma como as imagens de baixa resolução são sintetizadas para criar o conjunto de dados utilizado no treinamento do modelo Real-ESRGAN, cuja arquitetura foi projetada para lidar com um espaço de degradação mais amplo. No caso do ALPR, como as placas de veículos são fa-

bricadas seguindo padrões específicos, o espaço de degradação pode não ser tão extenso quanto o exigido para cenas com maior complexidade de detalhes e variações. Diante disso, é possível que uma arquitetura menor e mais simples do que a Real-ESRGAN ainda apresente bons resultados para aplicações de ALPR.

É importante destacar que, ao treinar um modelo de super-resolução, as imagens de alta resolução são geralmente capturadas por câmeras de alta qualidade, que oferecem detalhes nítidos dos objetos na cena, considerando o contexto específico da aplicação. Em contraste, as imagens de baixa resolução são sintetizadas e utilizadas para treinar o modelo generativo, com o objetivo de ensinar o modelo a mapear imagens degradadas para sua versão de alta resolução. Essa sintetização é necessária porque é difícil alcançar um alinhamento preciso entre imagens de alta e baixa resolução no mundo real, mesmo quando se usa duas câmeras próximas para capturar a mesma cena, devido principalmente ao atraso no registro da imagem por uma câmera em relação à outra.

A construção de um conjunto de dados emparelhado, que inclui imagens de alta resolução e suas correspondentes de baixa resolução da mesma cena, exige a aplicação de métodos de degradação eficazes. Caso contrário, o modelo pode apresentar baixa capacidade de generalização devido à presença de "lacunas" no domínio das imagens sintetizadas em comparação com o domínio natural das imagens degradadas [Chen et al. 2022]. Além disso, a super-resolução é intrinsecamente desafiadora, pois uma imagem de baixa resolução pode corresponder a várias versões em alta resolução [Abdelaziz et al. 2021]. Isso torna a implementação do modelo em aplicações reais particularmente difícil. Portanto, este trabalho destaca a importância do impacto do método de degradação utilizado, comparando um método mais simples com um mais robusto, e como isso afeta os resultados finais do modelo de super-resolução.

## 2. Trabalhos relacionados

Em muitos estudos sobre super-resolução, a metodologia de degradação das imagens de alta resolução para criação dos conjuntos de dados é um aspecto central ou secundário no processo para o treinamento dos modelos. O objetivo é aprimorar a capacidade dos modelos de lidar com imagens reais em baixa resolução, buscando através da metodologia de degradação simular as condições degradadoras que essas imagens podem encarar no registro ou transmissão das imagens.

Em [Hamdi et al. 2021], um modelo CycleGAN foi utilizado para tradução de estilo, onde as imagens de alta resolução foram primeiramente redimensionadas para um quarto de suas dimensões originais utilizando interpolação bicúbica. Essas imagens redimensionadas foram então processadas pelo modelo de tradução para introduzir ruídos e mudar a imagem de domínio, levando-as para um domínio de imagens degradadas. O estudo envolveu dois treinamentos da arquitetura D\_GAN\_ESR proposta pelos autores: um utilizando um conjunto de dados com degradação apenas por desfoque de movimento, e o outro empregando o modelo de tradução de estilo, que inclui ruídos de câmera analógica, ruídos de cores e de movimento. Os resultados mostraram que o modelo que foi treinado com o conjunto de dados com tradução de estilo produziu imagens de saída visualmente mais agradáveis.

Em [Abdelaziz et al. 2021], foi criado um conjunto de dados contendo 1.000 imagens capturadas com uma câmera digital. As versões em baixa resolução dessas imagens

foram geradas por meio da redução de dimensões utilizando interpolação bicúbica, além da adição aleatória de desfoques e ruídos. Por outro lado, [Pourhadi et al. 2022] redimensionou as imagens de alta resolução para uma resolução menor que a resolução de baixa resolução alvo e, em seguida, ajustou a resolução para as dimensões finais de baixa resolução. O objetivo foi capacitar o modelo a lidar com diferentes qualidades de imagem.

Em [Nascimento et al. 2022], para o treinamento do modelo proposto, foram criados quatro subconjuntos de imagens degradadas com base em intervalos definidos pela métrica de Índice de Medida da Similaridade Estrutural (SSIM - *Structural Similarity Index Measure*), que avalia a similaridade estrutural entre uma imagem e uma referência, sendo que um valor de SSIM igual a 1 indica estruturas idênticas. Os subconjuntos foram classificados nos seguintes intervalos:  $[0,00;0,10]$ ,  $[0,10;0,25]$ ,  $[0,25;0,50]$  e  $[0,50;0,75]$ . Para alcançar esses níveis de SSIM, foram aplicados ruídos gaussianos aleatórios de forma iterativa. Durante o treinamento, as imagens de cada subconjunto foram combinadas em um único conjunto. Em um trabalho mais recente, [Nascimento et al. 2023] focou na degradação do conjunto de dados de treinamento apenas no intervalo de SSIM entre 0 e 0,1.

[Kim et al. 2024] construiu um conjunto de dados emparelhado denominado LBLP, que consiste em 4.568 imagens de treinamento e 990 imagens de validação. Dentro desse conjunto, 2.284 imagens de treinamento e 495 imagens de validação foram degradadas por meio da redução das dimensões para um quarto do tamanho original, além da inserção de desfoque de movimento. É importante ressaltar que, neste trabalho, foi utilizada uma nova arquitetura chamada Joint-IRLPRNet, que combina a arquitetura AFA-Net com um modelo de reconhecimento. A fusão ocorre quando o modelo de reconhecimento é alimentado com as características extraídas pela AFA-Net, também proposta pelos autores. A AFA-Net aborda a super-resolução separando o tratamento da imagem de baixa resolução em duas subredes durante a etapa de pré-restauração: uma que lida com desfoque e ruídos, e outra que se concentra no aumento da resolução. Além disso, são utilizados quatro módulos de atenção para compor as características extraídas de cada subrede antes do módulo de reconstrução da imagem. Também foi proposta uma função de perda que relaciona a perda da restauração com a perda do reconhecimento dos caracteres da placa. Embora essa abordagem fim-a-fim apresente resultados superiores em comparação ao uso da AFA-Net junto com a WPOD-Net, existem limitações, como o fato de o modelo ser considerado pesado devido à dependência do tamanho dos modelos nas subredes que tratam individualmente a imagem de baixa resolução durante a pré-restauração.

Em [Yang et al. 2024], com o objetivo de minimizar o viés causado pelos métodos de degradação convencionais, os autores propuseram um modelo pragmático de degradação. Este modelo utiliza *isotropic Gaussian fuzzy kernel* e *anisotropic Gaussian fuzzy kernel*, argumentando que *kernels fuzzy* lineares e uniformes não capturam adequadamente a complexidade da degradação no mundo real. Para aumentar ainda mais a robustez do método, os autores incorporaram ruídos de compressão JPEG, ruído Poisson, ruído gaussiano e ruído *speckle*.

Diferente de outros trabalhos na área, [Pan et al. 2024] focou especificamente na metodologia de degradação das imagens, desenvolvendo um método denominado *n-Stage*. Este método permite aplicar uma sequência de elementos de degradação, como desfoque, redimensionamento, ruídos, compressão JPEG, efeito *ringing* e efeitos de sensores de

câmeras, por  $n$  vezes. A arquitetura proposta, chamada LPSRGAN, é projetada para lidar com um amplo espaço de degradação. No entanto, o estudo considera três estágios no método  $n$ -Stage, justificando que menos de três estágios não gera um impacto significativo nas taxas de reconhecimento, enquanto mais de três estágios pode levar à perda completa das informações dos caracteres (Figura 2).

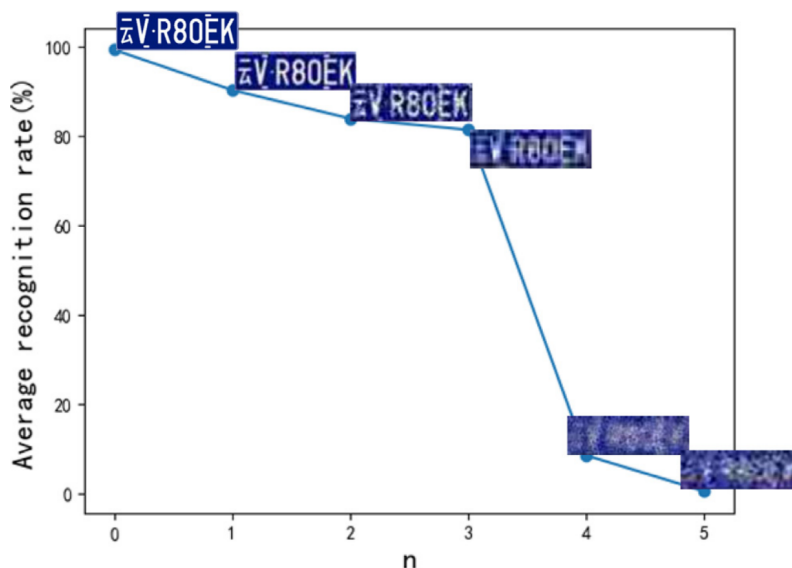


Figura 2. Taxa de reconhecimento médio em função do número de estágios do método de degradação  $n$ -Stage.

[Pan et al. 2024]

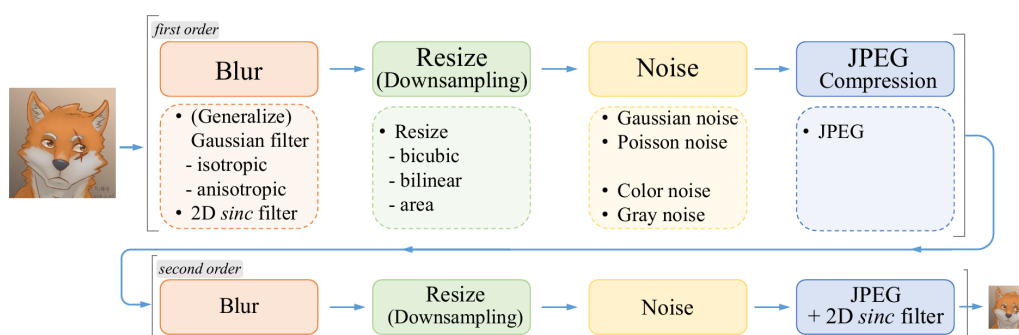
### 3. Metodologia

Para avaliar o impacto da sintetização das imagens de baixa resolução nos resultados dos modelos de super-resolução, foi primeiramente definida a arquitetura a ser utilizada. Optou-se pela Real-ESRGAN, que é empregada em trabalhos recentes [Pan et al. 2024] como a base para arquiteturas e métodos de sintetização mais robustos e generalistas. Embora a Real-ESRGAN seja menos complexa comparada a soluções mais modernas, sua estrutura menor remete a um bom custo-benefício no contexto de ALPR. Isso se deve ao seu equilíbrio entre tempo de execução e capacidade de generalização, o que é crucial para sistemas de ALPR que frequentemente precisam operar em tempo real.

Considerando isso, foram criados dois conjuntos de dados derivados do RodoSol-SR [Nascimento et al. 2023], que por sua vez é baseado no conjunto RodoSol-ALPR. O RodoSol-ALPR contém 20.000 imagens de veículos, distribuídas em 5.000 imagens de carros com placas brasileiras, 5.000 de motocicletas com placas brasileiras, 5.000 de carros com placas no padrão Mercosur e 5.000 de motocicletas com placas no padrão Mercosur. No processo de construção do RodoSol-SR, composto por 10000 imagens de alta resolução e 10000 imagens de baixa resolução, as placas dos veículos nas imagens do RodoSol-ALPR foram extraídas e retificadas.

No trabalho atual, inicialmente, as imagens de alta resolução foram ajustadas com preenchimento preto nas bordas superior e inferior da placa para torná-las quadradas. Além disso, as imagens foram redimensionadas para se adequarem ao treinamento da

Real-ESRGAN com um fator de escala de 2x, resultando em imagens de alta resolução de 128x128 pixels e imagens de baixa resolução de 64x64 pixels. Após essa etapa, o algoritmo de treinamento da Real-ESRGAN, conforme publicado pelos autores, foi modificado para isolar a parte do processo de degradação (Figura 3), uma vez que, na implementação original, a degradação ocorre durante o treinamento do modelo. Com essa etapa isolada, foram salvas as imagens resultantes da degradação, gerando assim o conjunto de dados com degradação robusta.



**Figura 3. Visão geral da geração de dados sintéticos adotada pela Real-ESRGAN.**  
[Wang et al. 2021b]

As mesmas imagens de alta resolução utilizadas para a construção do conjunto de dados com degradação robusta foram também empregadas para criar o conjunto de dados com degradação simplificada. Neste último, foi aplicada apenas a interpolação bicúbica para reduzir a resolução de 128x128 para 64x64, e desfoque gaussiano para gerar um impacto mais significativo na qualidade da imagem. Essa abordagem está alinhada com as práticas comuns em trabalhos relacionados à super-resolução, como aplicado em [Nascimento et al. 2022] e [Nascimento et al. 2023].

Após a criação dos conjuntos de dados, um modelo Real-ESRGAN foi treinado usando 4000 imagens de treinamento e 2000 imagens de validação do conjunto de dados com degradação mais robusta, resultando no modelo mais generalista. Deste modelo foi esperado melhor capacidade de generalização, devido à maior diversidade no espaço de degradação. Para possibilitar a comparação, foi treinado um segundo modelo Real-ESRGAN com as mesmas 4000 imagens de treinamento e 2000 de validação em alta resolução usadas no treinamento do modelo mais generalista, mas utilizando as imagens de baixa resolução do conjunto de dados com degradação simplificada, resultando no modelo mais especialista.

Com o modelo mais generalista e o modelo mais especialista em mãos, estes foram avaliados através de 4000 imagens de teste do conjunto de dados com degradação robusta e 4000 imagens conjunto de dados com degradação simplificada, sendo consideradas as mesmas 4000 imagens de alta resolução em ambos. Essa avaliação teve como objetivo coletar os resultados das métricas PSNR, SSIM, MSE, LPIPS, desempenho no reconhecimento dos caracteres e o tempo de execução. Com base nesses dados, foi possível construir tabelas comparativas, conforme apresentado na Tabela 1 e na Tabela 2.

## 4. Resultados e discussões

Após a implementação das etapas descritas na metodologia, foram coletados resultados para confirmar ou refutar a hipótese deste trabalho sobre o impacto do método de degradação na qualidade das imagens super-resolvidas produzidas pelos modelos da arquitetura Real-ESRGAN.

Os testes foram realizados em um computador de mesa equipado com um processador AMD Ryzen 7 5700G, 16 GB de memória RAM Kingston DDR4 com frequência de 3200 MHz e um SSD Lenovo SATA de 1 TB.

Foram necessárias 5 horas, 29 minutos e 29 segundos para completar a coleta dos resultados dos quatro casos de teste e gerar um arquivo no formato JSON, contendo as imagens salvas em codificação base64 e os valores obtidos para cada métrica em cada imagem super-resolvida.

### 4.1. Métricas de qualidade

A avaliação inicial foi baseada na qualidade das imagens super-resolvidas por cada modelo em comparação com as imagens de referência em alta resolução. Para realizar essa avaliação, foram consideradas as métricas Relação Sinal-Ruído de Pico (PSNR - *Peak Signal-to-Noise Ratio*), Índice de Medida da Similaridade Estrutural (SSIM - *Structural Similarity Index Measure*) e Erro Quadrático Médio (MSE - *Mean Squared Error*), comumente utilizadas em trabalhos relacionados à super-resolução. Embora essas métricas sejam amplamente adotadas, elas são criticadas por não reproduzirem o que seria a percepção humana de qualidade da imagem, pois avaliam a imagem apenas em termos de diferenças de pixel. Para uma avaliação mais alinhada à percepção humana da qualidade das imagens super-resolvidas, foi incluída também a métrica Similaridade de Patch de Imagem Perceptual Aprendida (LPIPS - *Learned Perceptual Image Patch Similarity*), que calcula a distância entre as características da imagem super-resolvida em relação a imagem de referência. A seguir, são descritas essas métricas para elucidar melhor o objetivo de cada uma:

- **PSNR:** Avalia a qualidade da imagem super-resolvida em relação à imagem de referência em alta resolução, calculando a potência do sinal com base no Erro Quadrático Médio (MSE). Quanto maior o valor de PSNR, melhor a qualidade da imagem [Sereethavekul and Ekpanyapong 2023].
- **SSIM:** Mede a similaridade estrutural entre a imagem super-resolvida e a imagem de referência, quantificando como as informações estruturais são preservadas na imagem super-resolvida [Sereethavekul and Ekpanyapong 2023]. Para essa métrica, quanto mais próximo de 1, maior a similaridade entre as imagens comparadas; quanto mais próximo de 0, maior a diferença entre elas.
- **MSE:** Calcula a média dos quadrados das diferenças entre os valores de pixel da imagem super-resolvida e os valores de pixel da imagem de referência em alta resolução, fornecendo uma medida direta do erro [Sereethavekul and Ekpanyapong 2023]. Quanto menor o erro medido entre as imagens, melhor é o resultado.
- **LPIPS:** Utiliza redes neurais profundas para extrair características semânticas da imagem, comparando a distância semântica entre as características da

imagem super-resolvida e a imagem de referência. Essa métrica avalia a qualidade da imagem de forma mais alinhada à percepção humana [Sereethavekul and Ekpanyapong 2023]. Quanto mais próximo de 0, melhor é o resultado.

## 4.2. Avaliação qualitativa

O arquivo final, que contém todos os dados obtidos através da execução do algoritmo de avaliação dos modelos, foi analisado para extrair os valores médios alcançados em cada métrica para cada caso de teste, conforme apresentado na Tabela 1. Além disso, as imagens resultantes foram salvas, e dois exemplos dessas imagens estão ilustrados na Figura 4.

A Tabela 1 revela, especialmente através da métrica LPIPS, que o modelo mais generalista demonstrou uma capacidade de generalização superior em comparação com o modelo mais especialista. O modelo mais generalista alcançou resultados semelhantes tanto no conjunto de testes com degradação robusta, um subconjunto do conjunto de dados com degradação robusta, quanto no conjunto de testes com degradação simplificada, um subconjunto do conjunto de dados com degradação simplificada. Em contraste, o modelo mais especialista obteve resultados comparáveis ao modelo mais generalista somente no conjunto de testes com degradação simplificada, que possui um espaço de degradação alinhado com o aprendido pelo modelo durante o treinamento.

Além disso, é importante observar o tempo médio de execução para cada combinação de avaliação. Notou-se uma redução de 19,067 ms no tempo de execução ao usar o modelo mais generalista com o conjunto de testes com degradação robusta em comparação com o modelo mais especialista com o conjunto de testes com degradação simplificada. Esta diferença não era esperada, dado que não houve alterações arquiteturais entre o modelo mais generalista e o modelo mais especialista durante o desenvolvimento deste trabalho.

## 4.3. Avaliação quantitativa

Com base nos dados armazenados durante a avaliação de cada caso e utilizando um modelo de Reconhecimento Óptico de Caracteres (OCR - *Optical Character Recognition*) baseado no YOLOv8, treinado com o conjunto de dados UFS-ALPR, foi possível aferir o desempenho do reconhecimento por meio do algoritmo de distância de Levenshtein. A avaliação considerou os casos em que todos os caracteres da placa foram corretamente reconhecidos, em que pelo menos 6 caracteres foram reconhecidos, e em que pelo menos 5 caracteres foram reconhecidos. Com esses resultados, foi possível construir a Tabela 2, que demonstra que o modelo mais generalista obteve um desempenho mais equilibrado ao comparar os dois conjuntos de testes, enquanto o modelo mais especializado apresentou resultados significativos no conjunto de testes com degradação simplificada.

## 5. Conclusão

Como previsto, o modelo treinado com um conjunto de dados emparelhado, utilizando um método de degradação mais robusto para sintetizar as imagens de baixa resolução, apresentou uma maior capacidade de generalização em comparação ao modelo treinado com um método mais simples, tanto em termos de qualidade da imagem



quanto no impacto sobre o desempenho do reconhecimento dos caracteres. Além disso, foi observada uma possível relação entre o método de degradação e o tempo de execução do modelo, mesmo sem alterações arquiteturais, o que será explorado em trabalhos futuros. Também será investigado se a capacidade de generalização do modelo se mantém após a compressão por destilação de conhecimento.

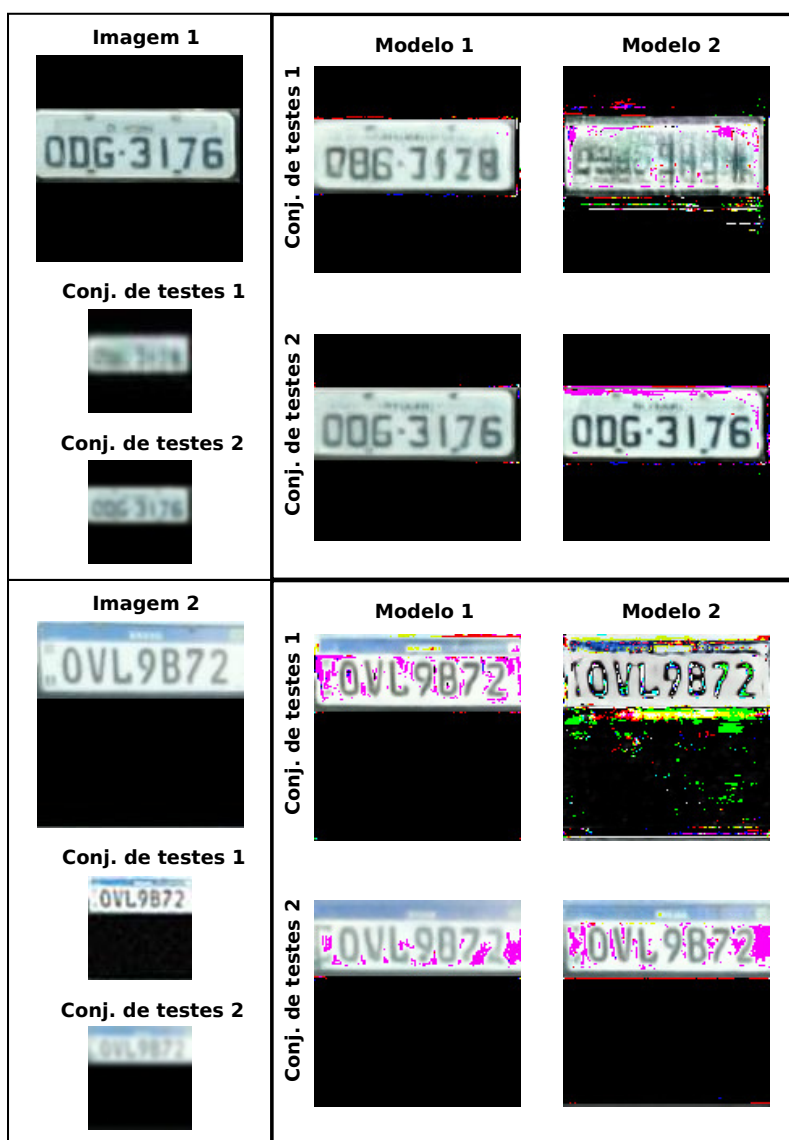


Figura 4. Grade de resultados apresentando as imagens de referência em alta resolução, suas respectivas versões em baixa resolução geradas por cada método de degradação abordado, e as imagens resultantes de cada um dos quatro casos de teste. *Modelo 1* refere-se ao modelo mais generalista; *Modelo 2*, ao modelo mais especializado; *Conj. de testes 1* é o conjunto de testes com degradação robusta; e *Conj. de testes 2* é o conjunto de testes com degradação simplificada.

**Tabela 1.** Valores médios de PSNR, SSIM, MSE, LPIPS e tempo de execução para cada caso experimental proposto, aplicados às 4000 imagens do conjunto de testes com degradação robusta (*Conj. de testes 1*) e às 4000 imagens do conjunto de testes com degradação simplificada (*Conj. de testes 2*). *Modelo 1* refere-se ao modelo mais generalista, enquanto *Modelo 2* se refere ao modelo mais especializado.

	PSNR	SSIM	MSE	LPIPS	Tempo de Execução (ms)
<b>Modelo 1 e Conj. de Testes 1</b>	18,000	0,672	1342,885	0,243	94,203
<b>Modelo 1 e Conj. de Testes 2</b>	18,776	0,707	1235,146	0,237	82,376
<b>Modelo 2 e Conj. de Testes 1</b>	13,196	0,223	3560,135	0,537	80,353
<b>Modelo 2 e Conj. de Testes 2</b>	17,250	0,712	1698,232	0,262	75,136

**Tabela 2.** Avaliação dos casos de teste em termos da capacidade de reconhecimento pelo modelo de OCR. Na coluna "Todos", é indicada a quantidade de placas em que os 7 caracteres foram corretamente reconhecidos. A coluna " $\geq 6$ " refere-se aos casos em que pelo menos 6 caracteres foram reconhecidos, enquanto a coluna " $\geq 5$ " abrange os casos em que pelo menos 5 caracteres foram reconhecidos.

	Conj. de Testes Robusto			Conj. de Testes Simplificado		
	Todos	$\geq 6$	$\geq 5$	Todos	$\geq 6$	$\geq 5$
<b>OCR sem super-resolução (quant. de placas dentre as 4000 avaliadas)</b>						
HR	1663	2582	3313	1663	2582	3313
LR	15	38	78	0	0	0
<b>OCR + modelo generalista (quant. de placas dentre as 4000 avaliadas)</b>						
Imagem super-resolvida	69	129	207	113	222	378
<b>OCR + modelo especialista (quant. de placas dentre as 4000 avaliadas)</b>						
Imagem super-resolvida	22	46	76	315	397	522

## 6. Agradecimentos

Este trabalho foi apoiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil (88887.909417/2023-00 - Melhoria do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe (PROCC/UFS)).

## Referências

- Abdelaziz, A. H., Chan, Y. K., and Koo, V. C. (2021). Enhancement for license plate recognition using image super resolution technique. In *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–4. IEEE.
- Asaad, A., Faizabadi, A. R., and Zaki, H. F. M. (2023). Synthetic license plate generation: A novel approach for effective license plate recognition in malaysia. In *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pages 1–8. IEEE.
- Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R. E., and Zhu, C. (2022). Real-world single image super-resolution: A brief review. *Information Fusion*, 79:124–145.
- Hamdi, A., Chan, Y. K., and Koo, V. C. (2021). A new image enhancement and super resolution technique for license plate recognition. *Heliyon*, 7(11).
- Kim, D., Kim, J., and Park, E. (2024). Afa-net: Adaptive feature attention network in image deblurring and super-resolution for improving license plate recognition. *Computer Vision and Image Understanding*, 238:103879.
- Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2018). A robust real-time automatic license plate recognition based on the yolo detector. In *2018 international joint conference on neural networks (ijcnn)*, pages 1–10. IEEE.
- Nascimento, V., Laroca, R., Lambert, J. d. A., Schwartz, W. R., and Menotti, D. (2022). Combining attention module and pixel shuffle for license plate super-resolution. In *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, volume 1, pages 228–233. IEEE.
- Nascimento, V., Laroca, R., Lambert, J. d. A., Schwartz, W. R., and Menotti, D. (2023). Super-resolution of license plate images using attention modules and sub-pixel convolution layers. *Computers & Graphics*, 113:69–76.
- Pan, S., Chen, S.-B., and Luo, B. (2023). A super-resolution-based license plate recognition method for remote surveillance. *Journal of Visual Communication and Image Representation*, 94:103844.
- Pan, Y., Tang, J., and Tjahjadi, T. (2024). Lpsrgan: Generative adversarial networks for super-resolution of license plate image. *Neurocomputing*, page 127426.
- Pattanaik, A. and Balabantaray, R. C. (2023). Enhancement of license plate recognition performance using xception with mish activation function. *Multimedia tools and applications*, 82(11):16793–16815.
- Pourhadi, N., Shafizadeh, B., Eshghi, F., and Kelarestaghi, M. (2022). Yolov5-based alpr improvement using selective sr-gan. In *2022 2nd International Conference on Computing and Machine Intelligence (ICMI)*, pages 1–6. IEEE.
- Rao, Z., Yang, D., Chen, N., and Liu, J. (2024). License plate recognition system in unconstrained scenes via a new image correction scheme and improved crnn. *Expert Systems with Applications*, 243:122878.

- Sereethavekul, W. and Ekpanyapong, M. (2023). Adaptive lightweight license plate image recovery using deep learning based on generative adversarial network. *IEEE Access*, 11:26667–26685.
- Wang, H., Li, Y., Dang, L.-M., and Moon, H. (2021a). Robust korean license plate recognition based on deep neural networks. *Sensors*, 21(12):4140.
- Wang, X., Xie, L., Dong, C., and Shan, Y. (2021b). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914.
- Yang, S., Xie, L., Ran, X., Lei, J., and Qian, X. (2024). Pragmatic degradation learning for scene text image super-resolution with data-training strategy. *Knowledge-Based Systems*, 285:111349.