



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Generalização de Modelos de Super-Resolução em ALPR: Um estudo sobre o impacto da degradação sintética no desempenho em domínios realistas

Dissertação de Mestrado

Cristiano Lima Oliveira



São Cristóvão – Sergipe

2026

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Cristiano Lima Oliveira

Generalização de Modelos de Super-Resolução em ALPR: Um estudo sobre o impacto da degradação sintética no desempenho em domínios realistas

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de mestre em Ciência da Computação.

Orientador(a): Dr. Leonardo Nogueira Matos
Coorientador(a): Dr. Paulo Salgado Gomes de Mattos Neto

São Cristóvão – Sergipe

2026

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE**

O48g Oliveira, Cristiano Lima.
Generalização de modelos de super-resolução em ALPR: um estudo sobre o impacto da degradação sintética no desempenho em domínios realistas / Cristiano Lima Oliveira; orientador Leonardo Nogueira Matos. – São Cristóvão, SE, 2026.
48 f. : il.

Dissertação (mestrado em Ciência da Computação) –
Universidade Federal de Sergipe, 2026.

1. Computação. 2. Imagens digitais. 3. Processamento de imagens.
4. Modelagem. I. Matos, Leonardo Nogueira, orient. II. Título.

CDU 004.45

Resumo

Sistemas de Reconhecimento Automático de Placas Veiculares (ALPR) enfrentam desafios ao processar imagens de baixa resolução, nas quais a área da placa pode representar menos de 0,3% do total da captura, comprometendo o reconhecimento óptico de caracteres (OCR). A super-resolução (SR) emerge como técnica promissora para reconstrução de detalhes textuais antes do OCR; contudo, a literatura carece de estudos sistemáticos sobre como o protocolo de degradação sintética, usualmente limitado à interpolação bicúbica com desfoque Gaussiano, afeta a capacidade de generalização dos modelos para condições reais de captura. Esta dissertação investiga duas questões: (1) o impacto da complexidade do protocolo de degradação sintética na generalização *cross-domain* de modelos SR; e (2) como a escolha da função de perda perceptual afeta a reconstrução semântica de texto em placas veiculares. Para isso, foram treinados modelos baseados em três arquiteturas estado da arte — RealESRGAN, ESRGAN e LPR-RSR-EXT — em cinco protocolos de degradação progressivamente complexos, totalizando 30 configurações avaliadas em 150 cenários cruzados, com desempenho medido pela taxa de reconhecimento de caracteres por meio do modelo YOLOv8. Os resultados evidenciam uma relação não-monotônica entre complexidade de degradação e generalização: modelos treinados com dois estágios de degradação reconheceram, em testes fora de domínio, até 18 vezes mais caracteres do que modelos treinados apenas com desfoque Gaussiano, ao passo que protocolos com três ou quatro estágios reduziram a acurácia média em até 59%, indicando destruição de características semânticas por degradação excessiva. Em relação à função de perda perceptual, verificou-se dependência arquitetural: ESRGAN com OCR Loss obteve 18,1% de correspondência completa da placa no protocolo de um estágio, contra 7,4% com VGG Loss, mas colapsou completamente (0,0% de acurácia) com dois ou mais estágios de degradação; RealESRGAN com OCR Loss alcançou média de 1,22 caracteres reconhecidos, contra 3,56 com VGG Loss; e LPR-RSR-EXT, especializado em texto, falhou com VGG Loss (0,0%) mas operou com OCR Loss (1,86 caracteres médios). VGG Loss demonstrou robustez consistente em todas as condições avaliadas, consolidando-se como escolha padrão segura independentemente da arquitetura ou do protocolo de degradação adotado.

Palavras-chave: Super-resolução. Reconhecimento Automático de Placas Veiculares. Redes Adversárias Generativas. Modelagem de Degradação. Perda Perceptual. Generalização *Cross-Domain*.

Abstract

Automatic License Plate Recognition (ALPR) systems face challenges when processing low-resolution images, in which the license plate area may represent less than 0.3% of the total capture, compromising optical character recognition (OCR). Super-Resolution (SR) emerges as a promising technique for reconstructing textual details prior to OCR; however, the literature lacks systematic studies on how synthetic degradation protocols — typically limited to bicubic interpolation with Gaussian blur — affect the generalization capacity of SR models to real-world capture conditions. This dissertation investigates two questions: (1) the impact of synthetic degradation protocol complexity on the cross-domain generalization of SR models; and (2) how the choice of perceptual loss function affects the semantic reconstruction of text on license plates. To this end, models based on three state-of-the-art architectures — RealESRGAN, ESRGAN, and LPR-RSR-EXT — were trained on five progressively complex degradation protocols, yielding 30 model configurations evaluated across 150 cross-domain scenarios, with performance measured by the character recognition rate using the YOLOv8 model. The results reveal a non-monotonic relationship between degradation complexity and generalization: models trained with two degradation stages recognized up to 18 times more characters in out-of-domain tests than models trained solely with Gaussian blur, whereas protocols with three or four stages reduced average accuracy by up to 59%, indicating destruction of semantic features due to excessive degradation. Regarding the perceptual loss function, architectural dependency was observed: ESRGAN with OCR Loss achieved 18.1% full plate match under the one-stage protocol, compared to 7.4% with VGG Loss, but collapsed completely (0.0% accuracy) with two or more degradation stages; RealESRGAN with OCR Loss reached an average of 1.22 recognized characters, against 3.56 with VGG Loss; and LPR-RSR-EXT, specialized in text, failed with VGG Loss (0.0%) but operated with OCR Loss (1.86 average characters). VGG Loss demonstrated consistent robustness across all evaluated conditions, establishing itself as a safe default choice regardless of the architecture or degradation protocol adopted.

Keywords: Super-resolution. Automatic License Plate Recognition. Generative Adversarial Networks. Degradation Modeling. Perceptual Loss. Cross-Domain Generalization.

Lista de ilustrações

| | |
|--|----|
| Figura 1 – Sequência de módulos que fazem parte de um sistema de ALPR | 11 |
| Figura 2 – Imagem de veículo (800×600 pixels) | 13 |
| Figura 3 – Recorte da placa (65×26 pixels) | 13 |
| Figura 4 – Visão geral da geração de dados sintéticos no RealESRGAN | 22 |
| Figura 5 – Diagrama do fluxo experimental integrado, destacando a preparação dos dados (Nós 1 e 2), o treinamento com adaptações arquiteturais (Nós 3 e 4) e a avaliação cruzada (Nó 5). | 36 |
| Figura 6 – Mapa de calor — Taxa de Reconhecimento Completo (%) por configuração treino × teste. Eixo vertical: protocolo de treinamento; eixo horizontal: protocolo de teste. Verde indica maior taxa de reconhecimento; vermelho indica menor. | 42 |
| Figura 7 – Mapa de calor — Acurácia Média de Caracteres (escala 0–7) por configuração treino × teste. Eixo vertical: protocolo de treinamento; eixo horizontal: protocolo de teste. Verde indica maior número médio de caracteres reconhecidos; vermelho indica menor. | 43 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Distribuição de conteúdos duplicados, rejeitados e aceitos por base de busca | 26 |
| Tabela 2 – Síntese cronológica dos trabalhos relacionados aceitos | 27 |
| Tabela 3 – Mapeamento entre fases do experimento e seções analíticas | 35 |
| Tabela 4 – Resultados médios de PSNR, SSIM, MSE, LPIPS e tempo de execução (ms) por modelo e conjunto de teste. | 40 |
| Tabela 5 – Desempenho do OCR em função do modelo de super-resolução e do método de degradação. | 40 |
| Tabela 6 – Configurações ótimas por arquitetura — síntese baseada em desempenho <i>cross-domain</i> . Desempenho reportado como acurácia média de caracteres nos 5 domínios de teste. | 47 |

Lista de abreviaturas e siglas

| | |
|----------|--|
| UFS | Universidade Federal de Sergipe |
| ANPR | Reconhecimento Automático de Placas Numéricas - <i>Automatic Number Plate Recognition</i> |
| ALPR | Reconhecimento Automático de Placas de Licença - <i>Automatic License Plate Recognition</i> |
| OCR | Reconhecimento Óptico de Caracteres - <i>Optical Character Recognition</i> |
| GPU | Unidade de Processamento Gráfico - <i>Graphics Processing Unit</i> |
| AFF-Net | Rede de Fusão de Características Adaptativa - <i>Adaptive Feature Fusion Network</i> |
| WPOD-Net | Rede de Detecção de Objeto Plano Deformado - <i>Warped Planar Object Detection Network</i> |
| CNN | Rede Neural Convolucional - <i>Convolutional Neural Network</i> |
| MPRNet | Rede de Restauração Progressiva Multi-Estágio - <i>Multi-Stage Progressive Image Restoration</i> |
| LR | Baixa Resolução - <i>Low-Resolution</i> |
| HR | Alta Resolução - <i>High-Resolution</i> |
| GAN | Rede Generativa Adversária - <i>Generative Adversarial Network</i> |
| SRCNN | Rede Neural Convolucional de Super-Resolução - <i>Super-Resolution Convolutional Neural Network</i> |
| SRGAN | Rede Generativa Adversária de Super-Resolução - <i>Super-Resolution Generative Adversarial Network</i> |
| MSE | Erro Quadrático Médio - <i>Mean Squared Error</i> |
| PSNR | Relação Sinal-ruído de Pico - <i>Peak Signal-to-noise Ratio</i> |
| SSIM | Medida do Índice de Similaridade Estrutural - <i>Structural Similarity Index Measure</i> |
| ESRGAN | Rede Generativa Adversária de Super-Resolução Aprimorada - <i>Enhanced Super Resolution Generative Adversarial Network</i> |

| | |
|-------|---|
| RRDB | Bloco de Densidade de Recuperação Residual - <i>Residual-in-Residual Dense Block</i> |
| STISR | Super-Resolução em Imagens de Texto em Cena - <i>Scene Text Image Super-Resolution</i> |
| FID | Distância de Incepção de Fréchet - <i>Fréchet Inception Distance</i> |
| SCC | Coefficiente de Correlação Espacial - <i>Spatial Correlation Coefficient</i> |
| VIF | Fidelidade da Informação Visual - <i>Visual Information Fidelity</i> |
| LPIPS | Similaridade Aprendida de Patches de Imagem Perceptual - <i>Learned Perceptual Image Patch Similarity</i> |
| mAP | média das Médias de Precisão - <i>mean Average Precision</i> |
| ANN | Rede Neural Artificial - <i>Artificial Neural Network</i> |
| SISR | Super-Resolução de Imagem Única - <i>Single Image Super-Resolution</i> |
| LPR | Reconhecimento de Placa de Licença - <i>License Plate Recognition</i> |
| VDSR | Super-Resolução Muito Profunda - <i>Very Deep Super Resolution</i> |
| YOLO | Você Só Olha uma Vez - <i>You Only Look Once</i> |
| RMSE | Raíz Quadrada do Erro-médio - <i>Root Mean Squared Error</i> |
| SCSR | Super-Resolução de Caractere Único - <i>Single Character Super Resolution</i> |
| JPEG | Grupo de Especialistas em Fotografia Conjunta - <i>Joint Photographic Experts Group</i> |




UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO


Ata da Sessão Solene de Defesa da Dissertação do Curso
de Mestrado em Ciência da Computação-UFS.
Candidato: **CRISTIANO LIMA OLIVEIRA**

Em 27 dias do mês de janeiro do ano de dois mil e vinte seis, com início às 17hs, realizou-se na Sala de Seminários do PROCC da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidato **Cristiano Lima Oliveira** que desenvolveu o trabalho intitulado: **“Generalização de Modelos de Super-Resolução em ALPR: Um estudo sobre o impacto da degradação sintética no desempenho em domínios realistas”**, sob a orientação do Prof. Dr. **Leonardo Nogueira Matos**. A Sessão foi presidida pelo Prof. Dr. **Leonardo Nogueira Matos** (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, o Dr. **Cleber Zanchettin** (UFPE) e, em seguida, Dr. **Rafael Oliveira Vasconcelos** (PROCC/UFS), na sequência o coorientador Dr. **Paulo Salgado Gomes de Mattos Neto** (UFPE). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando (a) **APROVADO** *“(aprovado/reprovado)”*. Atendidas as exigências da Instrução Normativa 05/2019/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CONEPE), e da Resolução nº 04/2021/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.


Cidade Universitária “Prof. José Aloísio de Campos”, 27 de janeiro de 2026.

Documento assinado digitalmente
 **LEONARDO NOGUEIRA MATOS**
Data: 11/03/2026 13:19:56-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Leonardo Nogueira Matos
(PROCC/UFS)
Presidente

Documento assinado digitalmente
 **RAFAEL OLIVEIRA VASCONCELOS**
Data: 11/03/2026 18:44:49-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Rafael Oliveira Vasconcelos
(PROCC/UFS)
Examinador Interno

Documento assinado digitalmente
 **PAULO SALGADO GOMES DE MATTOS NETO**
Data: 20/03/2026 14:41:39-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Paulo Salgado Gomes de Mattos Neto
(CIN/UFPE)
Examinador Externo ao programa -
coorientador

Documento assinado digitalmente
 **CLEBER ZANCHETTIN**
Data: 20/03/2026 16:08:34-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Cleber Zanchettin (UFPE)
Examinador Externo ao programa

Documento assinado digitalmente
 **CRISTIANO LIMA OLIVEIRA**
Data: 20/03/2026 15:42:46-0300
Verifique em <https://validar.iti.gov.br>

Cristiano Lima Oliveira
Candidato



**UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

UNIVERSIDADE FEDERAL DE SERGIPE

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO - PROCC

Departamento de Computação / UFS, Av. Marcelo Déda Chagas, S/N - Jardim Rosa Elze - Tel. (79) 3194-6353. CEP: 49107-230 - São Cristóvão - Sergipe - Brasil
E-mail: secretaria.pos@dcomp.ufs.br Portal: <http://www.posgraduacao.ufs.br/procc>

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 10 |
| 1.1 | Motivação | 13 |
| 1.2 | Hipóteses do Trabalho | 13 |
| 1.3 | Objetivos | 14 |
| 1.4 | Contribuições | 14 |
| 1.5 | Estrutura da Dissertação | 15 |
| 2 | Fundamentação Teórica | 16 |
| 2.1 | Super-Resolução de Imagens | 17 |
| 2.1.1 | Evolução das Arquiteturas de Super-Resolução | 18 |
| 2.2 | Métricas de Qualidade de Imagem | 19 |
| 2.2.1 | FID – <i>Fréchet Inception Distance</i> | 19 |
| 2.2.2 | PSNR – <i>Peak Signal-to-Noise Ratio</i> | 20 |
| 2.2.3 | SCC – <i>Spatial Correlation Coefficient</i> | 20 |
| 2.2.4 | SSIM – <i>Structural Similarity Index Measure</i> | 20 |
| 2.2.5 | VIF – <i>Visual Information Fidelity</i> | 21 |
| 2.2.6 | LPIPS – <i>Learned Perceptual Image Patch Similarity</i> | 21 |
| 2.3 | Metodologia de Degradação e Ruídos | 22 |
| 2.3.1 | Modelo Clássico de Degradação | 22 |
| 2.3.2 | Modelo de Degradação de Alta Ordem | 23 |
| 2.3.3 | Tipos de Ruídos e Artefatos | 23 |
| 3 | Trabalhos Relacionados | 25 |
| 3.1 | Abordagens Baseadas em Degradação Clássica e Simplificada | 28 |
| 3.2 | Integração de Super-Resolução e OCR: Perda Perceptual e Atenção | 31 |
| 3.3 | Modelos de Degradação Avançada e Cenários Naturais | 32 |
| 4 | Metodologia Experimental | 34 |
| 4.1 | Visão Macro do Fluxograma Experimental | 34 |
| 4.2 | Nós 1 e 2: Variação de Protocolos de Degradação | 36 |
| 4.3 | Nós 3 e 4: Incorporação de Backbones de OCR | 37 |
| 4.4 | Nó 5: Avaliação Comparativa | 37 |
| 5 | Estudos Experimentais | 38 |
| 5.1 | Avaliação do Impacto do Método de Degradação na Generalização do Modelo RealESRGAN | 38 |

| | | |
|-----------|---|-----------|
| 5.2 | Avaliação e Comparação do Impacto do Método de Degradação na Generalização entre Modelos de Super-Resolução | 41 |
| 5.2.1 | Impacto Combinado da Metodologia de Degradação e Perda Perceptual | 41 |
| 5.2.1.1 | Desempenho de Modelos Treinados com Degradação Simplificada | 42 |
| 5.2.1.2 | Robustez em Domínio Cruzado com Degradação Multi-Estágio | 43 |
| 5.2.1.3 | Complexidade Ótima de Degradação é Arquitetura-Dependente | 44 |
| 5.2.1.4 | Retornos Decrescentes além de Dois Estágios de Degradação | 44 |
| 5.2.1.5 | Efeitos Arquitetura-Dependentes da Perda Perceptual | 45 |
| 5.2.1.5.1 | Padrão 1: Vantagem de OCR Loss na ESRGAN com um estágio de degradação. | 45 |
| 5.2.1.5.2 | Padrão 2: Colapso das versões com OCR Loss em degradação complexa. | 45 |
| 5.2.1.5.3 | Padrão 3: Dependência de LPR-RSR-EXT em relação à OCR Loss. | 46 |
| 5.2.1.6 | VGG Loss como Escolha Robusta Padrão | 46 |
| 5.2.1.7 | Síntese: Interação entre Degradação, Arquitetura e Função de Perda | 46 |
| 6 | Conclusão | 48 |
| 6.1 | Confirmação das Hipóteses | 48 |
| 6.1.1 | Hipótese 1: Metodologia de Degradação Multi-Estágio | 48 |
| 6.1.2 | Hipótese 2: Perda Perceptual Baseada em OCR | 49 |
| 6.2 | Considerações Finais | 49 |
| | Referências | 51 |

1

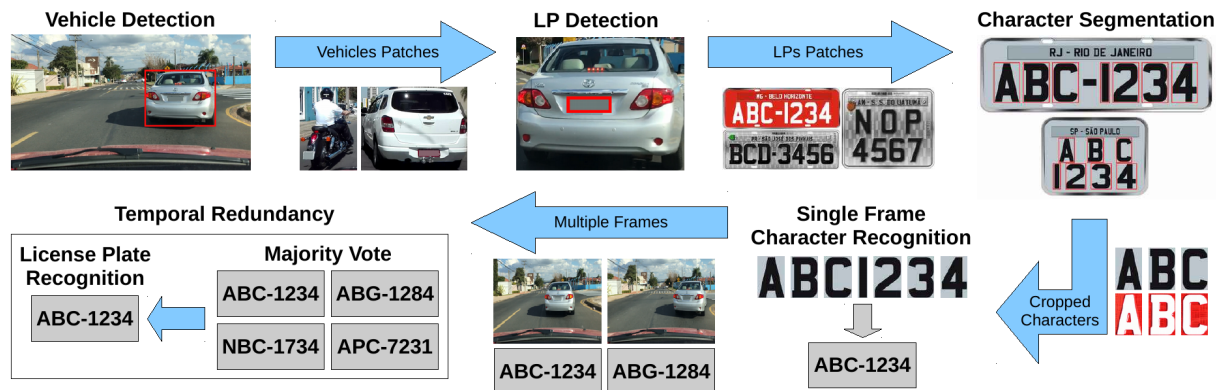
Introdução

Segundo [Gupta et al. \(2021\)](#), o Reconhecimento Automático de Placas Veiculares, denominado *Automatic Number Plate Recognition (ANPR)* ou *Automatic License Plate Recognition (ALPR)* a depender da região, constitui um sistema especializado de Reconhecimento Óptico de Caracteres (OCR – *Optical Character Recognition*). Desenvolvido para atender à demanda crescente de gerenciamento de frota veicular, o ALPR tem como objetivo principal a leitura automática de placas em imagens de veículos, mesmo em condições não ideais ([RAO et al., 2024](#)). Suas aplicações incluem:

- a) Controle inteligente de tráfego urbano;
- b) Automação de sistemas de estacionamento;
- c) Monitoramento de segurança em tempo real;
- d) Integração com redes de transporte inteligente (ITS – *Intelligent Transportation Systems*).

Além desses cenários tradicionais, a tecnologia é adaptável a qualquer contexto que exija identificação automatizada de veículos sem intervenção humana.

Figura 1 – Sequência de módulos que fazem parte de um sistema de ALPR



Fonte: (LAROCA et al., 2018)

Conforme Rao et al. (2024), sistemas ALPR eficientes dependem de técnicas avançadas de visão computacional e aprendizado de máquina, demandando recursos computacionais compatíveis com a aplicação. O emprego de Unidades de Processamento Gráfico (GPUs – *Graphics Processing Units*) é frequente para garantir baixa latência nas inferências, especialmente em aplicações em tempo real.

Estruturalmente, os sistemas de ALPR podem ser projetados com diferentes arquiteturas, dependendo das prioridades de desempenho. Embora algumas implementações adotem *pipelines* complexos para maximizar a taxa de detecção, outras priorizam a velocidade de reconhecimento por meio de estruturas simplificadas (LAROCA et al., 2018; BOBY et al., 2023).

Como ilustrado na Figura 1, um fluxo típico envolve múltiplos estágios de processamento. A abordagem de Larooca et al. (2018) incorpora um detector de veículos como filtro preliminar para reduzir falsos positivos, nos quais objetos não-veiculares são erroneamente classificados como placas. Contudo, como observado por Du et al. (2012), muitas arquiteturas modernas eliminam essa etapa, focando diretamente na localização da placa, segmentação de caracteres e reconhecimento óptico. A segmentação de caracteres apresenta particular desafio em cenários com placas rotacionadas ou anguladas (PAN; CHEN; LUO, 2023), casos em que métodos baseados em detecção semântica demonstram superioridade sobre a análise pixel a pixel.

A arquitetura em cascata apresentada na Figura 1 possui natureza crítica: falhas em qualquer estágio comprometem todo o *pipeline*. Mesmo quando todas as etapas são executadas com sucesso, a conversão final de caracteres visuais para texto pode ser afetada por deformações na imagem, resultando em erros de reconhecimento devido à natureza probabilística do processo.

Os desafios do ALPR são multifatoriais, sendo influenciados por fatores ambientais (iluminação, condições climáticas), configurações de hardware (qualidade da câmera, ângulo de captura) e características intrínsecas das placas (sujeira, danos, tipografia). Qin (2014) classifica

esses desafios em duas categorias:

- a) **Intrínsecos**: relacionados à geometria do objeto (deformação, rotação);
- b) **Extrínsecos**: decorrentes do ambiente (iluminação, movimento da câmera, oclusões).

Complementarmente, [Pourhadi et al. \(2022\)](#) sistematiza os principais fatores de degradação em imagens de ALPR:

- a) **Ruídos e irregularidades luminosas**: variações de iluminação (subexposição, sobreexposição ou distribuição não uniforme) que comprometem a segmentação;
- b) **Artefatos de movimento**: desfoque cinético decorrente do movimento relativo entre câmera e veículo durante a captura;
- c) **Orientação espacial**: ângulo de incidência (θ) entre o plano da placa e o plano focal da câmera, afetando a perspectiva;
- d) **Limitações de resolução**: restrições intrínsecas do sensor ou efeitos da distância câmera-objeto, reduzindo a densidade de pixels por caractere.

Diante desses fatores, a literatura tem proposto abordagens específicas para cada categoria de desafio. Baixa iluminação é tratada por meio de redes de realce de imagem ([SAPUTRA; SUCIATI; FATICHAH, 2024](#)); deformação de perspectiva, por redes de detecção e retificação geométrica ([RAO et al., 2024](#); [SILVA; JUNG, 2018](#)); condições climáticas adversas, por modelos de restauração multi-estágio ([ZAMIR et al., 2021](#)); e desbalanceamento de dados de treinamento, pela combinação de imagens reais e sintéticas ([ASAAD; FAIZABADI; ZAKI, 2023](#)). No contexto brasileiro, a coexistência dos padrões de placa tradicional e Mercosul impõe um desafio adicional, abordado por meio de modelos especializados por padrão ([CABRAL et al., 2021](#)). Os fundamentos dessas técnicas são apresentados no [Capítulo 2](#).

1.1 Motivação



Figura 2 – Imagem de veículo (800×600 pixels)

Fonte: Autor.



Figura 3 – Recorte da placa (65×26 pixels)

Fonte: Autor.

Os sistemas de ALPR enfrentam desafios na transição entre as etapas de localização de placas e reconhecimento de caracteres. Conforme ilustrado nas Figuras 2 e 3, uma placa típica pode ocupar apenas 65×26 pixels em uma imagem de 800×600 pixels, representando menos de 0,3% da área total da captura. Essa limitação acarreta dificuldade na detecção precisa de caracteres por métodos convencionais, alta sensibilidade a ruídos e artefatos de compressão, e deformações geométricas que prejudicam a classificação.

A super-resolução emerge como solução promissora nesse contexto, permitindo a ampliação da resolução da imagem, a restauração de componentes de alta frequência e a correção de artefatos de compressão antes da etapa de OCR (POURHADI et al., 2022). Contudo, a literatura apresenta uma lacuna: a ausência de estudos sistemáticos sobre como a metodologia de degradação sintética utilizada no treinamento afeta a capacidade de generalização dos modelos de SR quando aplicados a conjuntos de dados de ALPR com degradações mais complexas, que simulam imagens naturais e não controladas.

Esta lacuna motivou a investigação apresentada nesta dissertação, cujo foco é a relação entre os protocolos de degradação sintética e o desempenho dos modelos em condições robustas de operação.

1.2 Hipóteses do Trabalho

Este trabalho está fundamentado nas seguintes hipóteses científicas:

- **Metodologia de Degradação Multi-Estágio:** A utilização de protocolos de degradação de múltiplos estágios pode proporcionar modelos de super-resolução com maior capacidade

de generalização para imagens reais de baixa qualidade no cenário de ALPR. A abordagem de degradação de alta ordem, como a implementada no RealESRGAN e detalhada na [seção 2.3](#), pode ser eficaz para simular as degradações complexas encontradas em imagens de placas veiculares capturadas em condições não ideais.

- **Extração de Características para Perda Perceptual:** A incorporação de um *backbone* pré-treinado para reconhecimento óptico de caracteres na função de perda perceptual pode melhorar a capacidade do modelo de super-resolução em reconstruir caracteres de placas veiculares. Ao extrair características tanto da imagem super-resolvida quanto da *ground-truth* utilizando uma rede especializada em OCR, espera-se que o modelo aprenda a preservar e realçar características textuais críticas, resultando em melhor legibilidade e taxa de reconhecimento posterior.

1.3 Objetivos

Este trabalho tem como objetivo principal investigar a capacidade de generalização de modelos de super-resolução aplicados ao ALPR, com ênfase em dois aspectos: a influência do número de estágios de degradação sintética na generalização do modelo, e o impacto da incorporação de função de perda perceptual baseada em redes OCR pré-treinadas. Os objetivos específicos são:

- Realizar um mapeamento sistemático da literatura sobre super-resolução aplicada ao ALPR;
- Treinar e avaliar modelos generativos de super-resolução em cenários de variação do número de estágios de degradação, visando identificar a melhor combinação entre arquitetura, protocolo de treinamento e metodologia de degradação para composição da base de dados;
- Avaliar o reconhecimento de caracteres das placas restauradas por meio de um modelo OCR baseado na arquitetura YOLOv8, pré-treinado em base de dados de placas veiculares.

1.4 Contribuições

Este trabalho oferece três contribuições principais:

1. **Evidência Empírica Sistemática:** Demonstra, por meio de uma matriz de 150 experimentos cruzados (30 modelos \times 5 domínios de teste), que a simulação realista de degradações complexas durante o treinamento é um pré-requisito para a generalização de modelos SR em ALPR, superando o paradigma simplificado ainda prevalente na literatura. Modelos treinados com dois níveis de degradação alcançaram até 18 vezes mais caracteres reconhe-

cidos em testes fora de domínio comparados a modelos treinados apenas com desfoque Gaussiano.¹

2. **Identificação de Relação Não-Monotônica:** Verificação de que a complexidade de degradação apresenta ponto ótimo entre um e dois níveis, além do qual o desempenho se deteriora. Protocolos com três ou quatro estágios reduziram a acurácia média de caracteres reconhecidos em até 59% em relação ao protocolo de dois estágios. Esse resultado contraria a intuição de que maior diversidade de treinamento melhora monotonicamente a generalização, e estabelece limites práticos para a síntese de dados.
3. **Análise de Dependência Arquitetural da Perda Perceptual:** Primeira demonstração sistemática de que a eficácia de OCR Loss versus VGG Loss é contingente à arquitetura SR e à complexidade de degradação. No protocolo de um estágio, ESRGAN com OCR Loss obteve 18,1% de correspondência completa da placa, contra 7,4% com VGG Loss. Em contrapartida, RealESRGAN alcançou média de 3,56 caracteres reconhecidos com VGG Loss contra 1,22 com OCR Loss. Esse resultado tem implicações diretas para o projeto de sistemas ALPR-SR.

1.5 Estrutura da Dissertação

A organização deste documento segue:

- **Capítulo 2:** Fundamentação teórica acerca da super-resolução e metodologias de degradação de imagens;
- **Capítulo 3:** Revisão sistemática da literatura;
- **Capítulo 4:** Metodologia experimental;
- **Capítulo 5:** Análise de resultados;
- **Capítulo 6:** Conclusões.

¹ <<https://www.kaggle.com/datasets/cristianoec/resultados-dissertacao>>

2

Fundamentação Teórica

Este capítulo apresenta os conceitos fundamentais que sustentam este trabalho. São abordados os desafios de qualidade de imagem no contexto do ALPR, os métodos de super-resolução e sua evolução arquitetural, as métricas de avaliação de qualidade de imagem e a metodologia de degradação sintética adotada no treinamento dos modelos investigados.

Segundo [Hamdi, Chan e Koo \(2021\)](#) e [Abdelaziz, Chan e Koo \(2021\)](#), o desempenho do reconhecimento de caracteres é proporcional à qualidade da imagem fornecida como entrada. A precisão é comprometida quando a imagem possui baixa resolução ou está degradada. No contexto de ALPR, a velocidade do obturador de câmeras convencionais frequentemente não é suficiente para capturar imagens nítidas de veículos em movimento ([NASCIMENTO et al., 2023](#)), resultando em imagens com desfoque e perda de detalhes. Como consequência, o sistema torna-se suscetível à confusão entre caracteres visualmente similares — como “I” e “1” —, o que compromete a acurácia do reconhecimento.

No estudo de [Wang et al. \(2021a\)](#), é descrita a influência negativa das condições ambientais sobre o desempenho do ALPR, especialmente em capturas realizadas durante neblina ou sob dispersão de partículas na atmosfera, condições que introduzem degradações relevantes na qualidade da imagem. O autor aponta que a substituição por equipamentos de melhor qualidade seria a abordagem mais direta para mitigar esses efeitos, mas ressalta que os custos envolvidos tornam essa alternativa inviável na maioria das aplicações práticas.

Em [Pourhadi et al. \(2022\)](#), destaca-se que, além das limitações do sensor da câmera, a distância entre o objeto e o equipamento de captura também é fator determinante para a resolução obtida. No cenário de ALPR, a placa ocupa uma área reduzida na imagem após o recorte, resultando em regiões de interesse de baixa resolução. O estudo também observa que a redução de resolução pode ser intencional em aplicações com restrições de armazenamento ou largura de banda de transmissão, conforme também discutido por [Nascimento et al. \(2022\)](#).

2.1 Super-Resolução de Imagens

A super-resolução (SR) de imagem única (*Single Image Super-Resolution* – SISR) consiste no problema de reconstruir uma imagem de alta resolução $\mathbf{y} \in \mathbb{R}^{rH \times rW \times C}$ a partir de uma imagem de baixa resolução $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$, onde r é o fator de escala, H e W são as dimensões espaciais e C é o número de canais (XIN et al., 2021). Por ser um problema inverso mal-posto — múltiplas imagens HR podem ser consistentes com uma mesma LR —, a reconstrução precisa é um desafio fundamental, em especial quando há necessidade de preservar a identidade de caracteres textuais (NASCIMENTO et al., 2022). Quanto maior o fator de escala da ampliação, mais complexa se torna a tarefa.

No contexto do ALPR, a SISR é particularmente relevante pois permite ampliar a resolução das regiões de placa antes da etapa de OCR, restaurando componentes de alta frequência e reduzindo artefatos de compressão (POURHADI et al., 2022). Quando o conteúdo a ser restaurado é textual, são empregados métodos de Super-Resolução em Imagens de Texto em Cena (*Scene Text Image Super-Resolution* – STISR), que incorporam restrições específicas para a reconstrução fiel de caracteres (NASCIMENTO et al., 2022).

Os métodos de super-resolução organizam-se em três categorias principais:

- a) *Métodos baseados em interpolação* utilizam funções paramétricas ou não paramétricas — como interpolação bilinear ou bicúbica — para estimar os pixels em uma grade de alta resolução (ZHANG et al., 2012). São computacionalmente eficientes, mas tendem a produzir imagens com desfoques e artefatos visíveis em regiões de borda, onde as componentes de alta frequência são críticas;
- b) *Métodos baseados em reconstrução* definem restrições explícitas sobre a imagem reconstruída, como suavidade de bordas e co-ocorrência de padrões. Em Xin et al. (2021), por exemplo, foram construídos dicionários de 10 caracteres em alta resolução para mapear as características dos dígitos de placas, explorando a padronização do domínio. O desempenho pode ser relevante em problemas específicos e bem definidos, mas a capacidade de generalização é limitada (XIN et al., 2021);
- c) *Métodos baseados em aprendizagem* partem do pressuposto de que os detalhes de alta frequência ausentes na imagem LR podem ser aprendidos a partir de pares de imagens HR e LR (ZHANG et al., 2012). Redes neurais profundas são treinadas para modelar um mapeamento de LR para HR mesmo para imagens não vistas durante o treinamento, processo denominado super-resolução às cegas. Atualmente, esse paradigma apresenta os melhores resultados em termos de generalização, especialmente quando combinado com arquiteturas baseadas em Redes Generativas Adversárias (GANs – *Generative Adversarial Networks*) (XIN et al., 2021; HAMDI; CHAN; KOO, 2021).

2.1.1 Evolução das Arquiteturas de Super-Resolução

Os métodos de SR baseados em aprendizagem evoluíram de forma consistente ao longo dos anos, com marcos arquiteturais que ampliaram progressivamente a qualidade das imagens reconstruídas.

A Rede Neural Convolutiva de Super-Resolução (SRCNN – *Super-Resolution Convolutional Neural Network*) (DONG et al., 2015) foi uma das primeiras arquiteturas a demonstrar que redes convolucionais profundas podem superar os métodos baseados em interpolação e reconstrução em tarefas de SR, ao tratar as restrições excessivamente restritivas dos métodos anteriores (ABDELAZIZ; CHAN; KOO, 2021). As abordagens baseadas em redes convolucionais dessa geração utilizavam a função de perda de Erro Quadrático Médio (MSE – *Mean Squared Error*), que, embora produza resultados favoráveis em métricas como PSNR e SSIM, frequentemente gera imagens com aparência borrada e perda de texturas de alta frequência (HAMDI; CHAN; KOO, 2021).

O maior avanço em termos de qualidade visual ocorreu com a introdução das GANs ao problema de SR, exemplificada pela Rede Generativa Adversária de Super-Resolução (SRGAN – *Super-Resolution Generative Adversarial Network*) (LEDIG et al., 2017). Introduzidas por Goodfellow et al. (2014), as GANs operam com duas entidades em competição: um gerador e um discriminador. No contexto da SR, o gerador recebe uma imagem LR e produz uma imagem HR candidata, enquanto o discriminador classifica se a imagem recebida é real (HR original) ou gerada. O objetivo do treinamento é que o gerador produza imagens indistinguíveis das reais sob a perspectiva do discriminador. Além das perdas do gerador e do discriminador, o processo incorpora a função de perda perceptual, calculada com base em características semânticas extraídas de camadas profundas de redes pré-treinadas, em vez de comparações pixel a pixel. Isso melhora a fidelidade visual da imagem restaurada em relação à referência original.

A Rede Generativa Adversária de Super-Resolução Aprimorada (ESRGAN – *Enhanced Super-Resolution Generative Adversarial Network*) (WANG et al., 2018) introduziu dois avanços relevantes sobre a SRGAN. O primeiro foi a substituição dos blocos residuais por Blocos Densos Residuais-em-Residual (RRDB – *Residual-in-Residual Dense Block*), que estabelecem conexões diretas entre camadas iniciais e finais da rede e eliminam a normalização em lote. Essa mudança permite que redes mais profundas sejam treinadas de forma mais estável e resulta em ganhos de PSNR em comparação à SRGAN, como observado por Kabiraj et al. (2023). O segundo avanço foi a reformulação do discriminador: em vez de classificar imagens simplesmente como reais ou geradas, o discriminador da ESRGAN avalia se uma imagem é mais realista do que outra (*relativistic discriminator*), o que contribui para maior nitidez e fidelidade dos detalhes de alta frequência.

A versão mais recente dessa linha arquitetural é o RealESRGAN (WANG et al., 2021b), cujo foco é a generalização para imagens reais. Enquanto SRGAN e ESRGAN assumem

degradações simplificadas no treinamento, o RealESRGAN adota um modelo de degradação de segunda ordem — detalhado na [seção 2.3](#) — que simula combinações mais complexas e realistas de ruído, desfoque, compressão e redimensionamento. Essa abordagem amplia o espaço de degradações aprendido pelo modelo, tornando-o mais robusto a condições não controladas.

No levantamento sistemático realizado neste trabalho, observou-se que a SRGAN e a ESRGAN são as arquiteturas mais prevalentes entre os estudos de SR aplicada ao ALPR, tanto em suas versões originais quanto em variantes modificadas, conforme detalhado no [Capítulo 3](#).

2.2 Métricas de Qualidade de Imagem

Esta seção apresenta as principais métricas utilizadas para avaliar a qualidade de imagens reconstruídas por modelos de SR. É importante destacar que PSNR e SSIM, embora amplamente adotadas, apresentam limitações no contexto de GANs para SR: como ressaltado por [Tian et al. \(2022\)](#) e [Yang et al. \(2024\)](#), essas métricas operam a nível de pixel e não capturam adequadamente a qualidade visual perceptual, que é o objetivo central dos modelos generativos. Alguns estudos, como [Ledig et al. \(2017\)](#), recorrem a avaliações por opinião humana (*Mean Opinion Score*); em contextos que exigem avaliação automatizada, métricas como o Índice Perceptual ([BLAU et al., 2018](#)) foram propostas como alternativas.

As métricas descritas a seguir foram utilizadas como referência em trabalhos relacionados, conforme sistematizado por [Sereethavekul e Ekpanyapong \(2023\)](#).

2.2.1 FID – Fréchet Inception Distance

O FID mede a distância entre a distribuição das imagens geradas e a distribuição das imagens reais de referência, utilizando as ativações das camadas profundas da rede Inception v3. Camadas profundas capturam representações semânticas de alto nível — como formas e objetos — em contraste com as camadas iniciais, que respondem a padrões simples como bordas. Um valor de FID igual a 0 indica que as distribuições são idênticas; valores maiores indicam maior divergência.

$$FID = \|\mu_1 - \mu_2\|^2 + tr \left(\sum_1 + \sum_2 - 2 \left(\sum_1^{\frac{1}{2}} \cdot \sum_1 \cdot \sum_2^{\frac{1}{2}} \right)^{\frac{1}{2}} \right) \quad (2.1)$$

- μ_1 e μ_2 : médias dos vetores de características das imagens reais e geradas, respectivamente;
- \sum_1 e \sum_2 : matrizes de covariância dos vetores de características;
- tr : traço da matriz (soma dos elementos da diagonal principal).

2.2.2 PSNR – Peak Signal-to-Noise Ratio

O PSNR mede a relação entre a potência máxima do sinal e a potência do ruído introduzido pela reconstrução, calculado a partir do MSE entre a imagem reconstruída e a referência. Para imagens de 8 bits, valores típicos situam-se entre 25 e 48 dB, sendo que valores mais altos indicam menor distorção pixel a pixel.

$$PSNR = 10 \cdot \log_{10} \left(\frac{256^2}{MSE} \right) \quad (2.2)$$

- MSE : erro quadrático médio entre os pixels da imagem reconstruída e os da imagem de referência.

2.2.3 SCC – Spatial Correlation Coefficient

O SCC é definido como o coeficiente de concordância espacial para processos estacionários de segunda ordem, com o objetivo de detectar desalinhamentos entre duas imagens. Os valores resultantes variam de 0 a 1.

$$\rho^c(h) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \rho_{xy} R(h, \phi_{xy}) \quad (2.3)$$

- σ_x e σ_y : desvios padrão das variáveis aleatórias X e Y ;
- ρ_{xy} : coeficiente de correlação entre X e Y , com $|\rho^c(h)| \leq |\rho_{xy}(h)| \leq 1$;
- $R(h, \phi_{xy})$: função de correlação com vetor paramétrico ϕ .

2.2.4 SSIM – Structural Similarity Index Measure

O SSIM avalia a similaridade entre duas imagens com base em três componentes: luminância, contraste e estrutura. Ao contrário do PSNR, que opera exclusivamente a nível de pixel, o SSIM incorpora informações estruturais da cena. O valor resultante varia entre 0 e 1, onde valores mais altos indicam maior similaridade estrutural.

$$SSIM(i_1, i_2) = [l(i_1, i_2)]^\alpha \cdot [c(i_1, i_2)]^\beta \cdot [s(i_1, i_2)]^\gamma \quad (2.4)$$

- i_1, i_2 : imagens de referência e reconstruída;
- $l(i_1, i_2), c(i_1, i_2), s(i_1, i_2)$: funções de comparação de luminância, contraste e estrutura;
- $\alpha > 0, \beta > 0, \gamma > 0$: pesos relativos de cada componente.

2.2.5 VIF – Visual Information Fidelity

O VIF quantifica a fidelidade da informação visual por meio de coeficientes de wavelet, comparando a informação que o cérebro humano seria capaz de extrair da imagem de referência com a que pode ser extraída da imagem reconstruída. O valor resultante varia de 0 a 1, onde 1 indica fidelidade perfeita.

$$VIF = \frac{\sum_{j(\text{subbands})} I(\vec{C}^{N_j}; \vec{F}^{N_j} | s^{N_j})}{\sum_{j(\text{subbands})} I(\vec{C}^{N_j}; \vec{E}^{N_j} | s^{N_j})} \quad (2.5)$$

- $I(\vec{C}^N; \vec{F}^N | s^N)$ e $I(\vec{C}^N; \vec{E}^N | s^N)$: informação extraível das imagens de referência e reconstruída em uma sub-banda específica;
- \vec{C}^{N_j} : coeficientes da sub-banda j da imagem de referência.

2.2.6 LPIPS – Learned Perceptual Image Patch Similarity

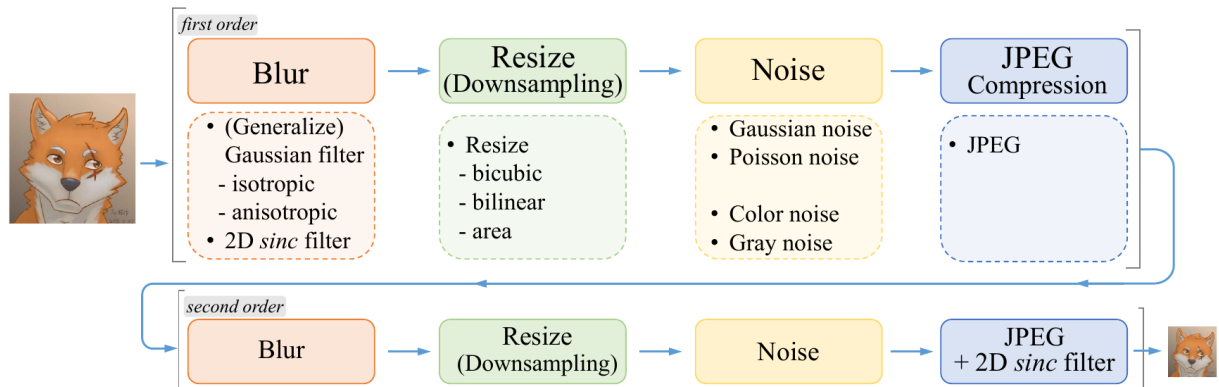
O LPIPS foi proposto por [Zhang et al. \(2018\)](#) com o objetivo de quantificar diferenças visuais percebidas por humanos de forma que métricas como PSNR e SSIM não conseguem capturar. Enquanto estas operam a nível de pixel, o LPIPS utiliza ativações de camadas profundas de redes neurais convolucionais treinadas para classificação para extrair características semânticas das imagens, aproximando-se da avaliação humana de qualidade visual.

Um exemplo ilustrativo é o da compressão de imagens: ao reduzir componentes de frequência média para preservar as altas frequências — perceptualmente mais relevantes —, o PSNR pode indicar perda de qualidade enquanto o LPIPS sinaliza baixa degradação perceptual. Da mesma forma, imagens desfocadas podem apresentar valores satisfatórios de PSNR e SSIM, mas valores elevados de LPIPS, refletindo a perda de legibilidade percebida pelos humanos.

Para o cálculo, as características extraídas das imagens comparadas são combinadas por meio de métricas de distância semântica. Quanto menor o valor de LPIPS, maior a similaridade perceptual entre as imagens. Como indicado pelos próprios autores, nenhuma métrica automática representa integralmente as preferências da percepção humana, mas o LPIPS oferece uma aproximação mais alinhada com a avaliação subjetiva do que as métricas pixel a pixel tradicionais.

2.3 Metodologia de Degradação e Ruídos

Figura 4 – Visão geral da geração de dados sintéticos no RealESRGAN



Fonte: (WANG et al., 2021b)

Wang et al. (2021b) propôs uma metodologia para simular degradações complexas encontradas em imagens do mundo real por meio de um modelo de degradação de alta ordem, ilustrado na Figura 4. Este modelo é central para o treinamento de redes de SR que precisam lidar com degradações variadas e desconhecidas, como as introduzidas por sistemas de captura, edição e transmissão de imagens. A metodologia organiza-se em torno de dois níveis de complexidade: o modelo clássico de degradação e sua extensão para alta ordem.

2.3.1 Modelo Clássico de Degradação

O modelo clássico, descrito na Equação 2.6, sintetiza imagens de baixa resolução (LR) a partir de imagens de alta resolução (HR) por meio de quatro etapas sequenciais:

- **Desfoque:** a imagem HR é convoluída com um *kernel* de desfoque, como filtros Gaussianos isotrópicos ou anisotrópicos, filtros Gaussianos generalizados ou *kernels* com distribuição em platô. A escolha do *kernel* determina o tipo e a intensidade da degradação simulada;
- **Redimensionamento:** a imagem desfocada é subamostrada por um fator de escala r , utilizando algoritmos como interpolação bilinear, bicúbica ou por área. Essa etapa é responsável pela redução de resolução;
- **Adição de ruído:** ruídos como o gaussiano e o de Poisson são adicionados para simular imperfeições de captura ou transmissão;
- **Compressão JPEG:** a imagem é submetida à compressão JPEG, introduzindo artefatos de bloco típicos de imagens transmitidas ou armazenadas em formatos com perdas.

$$\mathbf{x} = \mathcal{D}(\mathbf{y}) = [(\mathbf{y} \otimes \mathbf{k}) \downarrow_r + \mathbf{n}]_{\text{JPEG}}, \quad (2.6)$$

onde \mathbf{y} é a imagem HR, \mathbf{k} é o *kernel* de desfoque, \downarrow_r é a operação de subamostragem por fator r , \mathbf{n} é o ruído adicionado e $[\cdot]_{\text{JPEG}}$ representa a compressão JPEG.

2.3.2 Modelo de Degradação de Alta Ordem

Para simular degradações mais complexas, o RealESRGAN estende o modelo clássico para um processo de n estágios, no qual o mesmo conjunto de operações de degradação é aplicado sequencialmente com parâmetros distintos a cada etapa. Isso permite capturar combinações mais realistas de degradações, como múltiplas compressões JPEG, redimensionamentos consecutivos e adição de ruídos em diferentes estágios do processamento. A Equação 2.7 descreve formalmente esse processo:

$$\mathbf{x} = \mathcal{D}^n(\mathbf{y}) = (\mathcal{D}_n \circ \dots \circ \mathcal{D}_2 \circ \mathcal{D}_1)(\mathbf{y}), \quad (2.7)$$

onde cada \mathcal{D}_i representa uma instância independente do modelo clássico com parâmetros distintos. Na implementação original do RealESRGAN, $n = 2$, o que configura um processo de segunda ordem.

2.3.3 Tipos de Ruídos e Artefatos

O modelo de degradação do RealESRGAN incorpora diferentes tipos de ruído e artefatos para ampliar a diversidade das degradações simuladas:

- **Ruído gaussiano:** ruído aditivo com distribuição gaussiana, parametrizado pelo desvio padrão σ . Pode ser aplicado de forma colorida (independente por canal RGB) ou em tons de cinza (mesmo ruído em todos os canais). Simula imperfeições introduzidas por sensores ou transmissão;
- **Ruído de Poisson:** segue uma distribuição de Poisson, com intensidade proporcional ao valor de pixel. É característico de imagens capturadas em baixa luminosidade, onde o número de fótons incidentes por pixel é reduzido, introduzindo flutuações estatísticas;
- **Artefatos de ringing e overshoot:** resultam de filtros que limitam altas frequências, como o filtro *sinc* (Equação 2.8), ou de operações de realce excessivo. O modelo incorpora tais filtros intencionalmente durante o treinamento, de forma que a rede aprenda a suprimir esses artefatos;
- **Artefatos de compressão JPEG:** blocos visíveis e perda de detalhes em altas frequências, controlados pelo fator de qualidade q . Simulam degradações comuns em imagens compartilhadas em plataformas digitais ou transmitidas por redes com compressão;

- **Redimensionamento com algoritmos variados:** a utilização alternada de interpolação por área, bilinear e bicúbica introduz diferentes padrões de suavização e *overshoot*, ampliando a variedade de degradações sintéticas geradas.

$$k(i, j) = \frac{\omega_c}{2\pi\sqrt{i^2 + j^2}} J_1(\omega_c\sqrt{i^2 + j^2}), \quad (2.8)$$

A combinação do modelo de alta ordem com a variedade de ruídos e artefatos descritos permite gerar pares de treinamento sintéticos que aproximam as degradações encontradas em imagens reais de ALPR, superando as limitações dos protocolos simplificados — como a interpolação bicúbica isolada — que predominam na literatura e que, conforme investigado neste trabalho, restringem a capacidade de generalização dos modelos treinados.

3

Trabalhos Relacionados

Para encontrar artigos relacionados à aplicação da super-resolução em sistemas de ALPR, foram realizadas buscas utilizando a seguinte *string* de busca: **“optical character recognition” AND (“Super Resolution” OR “Image Enhancement” OR “Super Sampling” OR “Super-Sampling” OR “Super-Resolution”) AND (“license plate recognition” OR “number plate recognition”)**. As buscas foram conduzidas nas bases ACM Digital Library, El Compendex, IEEE Digital Library, ISI Web of Science, Science@Direct, Scopus e Springer Link.

Após consolidar os resultados de todas as bases de dados, inicialmente foram identificados 817 itens. No entanto, essa contagem incluía tipos de conteúdo não exclusivamente acadêmicos, como capítulos de livros ou resumos de conferências, que foram removidos, reduzindo o total para 627 itens. Além disso, foram identificados e removidos 190 artigos duplicados, resultando em um total de 397 itens únicos que precisaram ser verificados.

Após essas filtragens, foram considerados os seguintes critérios de inclusão:

- a) Pelo menos uma técnica de super-resolução e uma técnica de reconhecimento óptico de caracteres foram abordadas;
- b) Métricas foram apresentadas para comparar as imagens super-resolvidas com a imagem verdadeira (GT) ou para medir a acurácia alcançada com os resultados do trabalho;
- c) O artigo foi publicado em ou após 2021.

Para exclusão, foram considerados os seguintes critérios:

- a) É um artigo de revisão sistemática, *survey* ou meta-análise, sem proposta metodológica original;
- b) Está fora do contexto de ALPR;
- c) É um livro ou outro tipo de conteúdo não-artigo proveniente da exportação das bases;

- d) Foi publicado antes de 2021;
- e) O artigo não está disponível publicamente ou por meio do acesso CAFe do periódico CAPES;
- f) A técnica de super-resolução não foi relacionada ao reconhecimento óptico de caracteres.

A [Tabela 1](#) apresenta os dados do processo de filtragem para cada uma das bases de busca.

Tabela 1 – Distribuição de conteúdos duplicados, rejeitados e aceitos relacionados à super-resolução para cada base de busca

| Base | Total | Duplicados | Rejeitados | Aceitos |
|----------------------|------------|------------|------------|-----------|
| ACM Digital Library | 12 | 0 | 12 | 0 |
| El Compendex | 275 | 108 | 162 | 5 |
| IEEE Digital Library | 41 | 16 | 23 | 2 |
| ISI Web of Science | 13 | 2 | 10 | 1 |
| Science@Direct | 27 | 1 | 22 | 4 |
| Scopus | 194 | 98 | 89 | 7 |
| Springer Link | 255 | 8 | 246 | 1 |
| Total | 817 | 233 | 564 | 20 |

Fonte: Elaborada pelo autor.

As questões de pesquisa que orientaram a análise dos trabalhos aceitos foram:

- a) Qual abordagem foi utilizada para o reconhecimento dos caracteres?
- b) Quais conjuntos de dados foram utilizados?
- c) Quais técnicas de super-resolução foram empregadas?
- d) Quais métricas de avaliação foram adotadas?
- e) Quais foram os resultados em termos de acurácia ou qualidade da imagem restaurada?
- f) O que foi proposto como contribuição original?

A [Tabela 2](#) apresenta uma síntese cronológica dos 20 trabalhos aceitos, com destaque para a arquitetura SR, o protocolo de síntese de imagens LR, as métricas utilizadas e os principais resultados reportados. Essa visão permite identificar a trajetória de evolução das abordagens ao longo do período analisado (2021–2024).

Tabela 2 – Síntese cronológica dos 20 trabalhos aceitos no mapeamento sistemático

| Referência | Ano | Arq. SR | Protocolo LR | Resultado principal |
|----------------------------------|------|----------------------|-------------------------------|---|
| (HAMDI; CHAN; KOO, 2021) | 2021 | D_GAN_ESR | Bicúbica + CycleGAN | Redução da taxa de erro de OCR |
| (WANG et al., 2021a) | 2021 | ESRGAN | Real + sintético | Acurácia sistema: 99,5% (W-LPR) |
| (SHOMEI; SAMS, 2021) | 2021 | ESRGAN (RRDB) | Bicúbica | Melhora do mAP (LPDB-A) |
| (ABDELAZIZ; CHAN; KOO, 2021) | 2021 | SRGAN-based | Bicúbica 4× + ruídos | Acurácia OCR: 12,8%→62,8% |
| (XIN et al., 2021) | 2021 | Dicionário + CNN | Redimensionamento | PSNR superior ao SRGAN |
| (POURHADI et al., 2022) | 2022 | SRGAN | Bicúbica (4×, 2×) | Acurácia sistema: +18% |
| (BOBY et al., 2022) | 2022 | ESRGAN | Não especificado | +7,46% acurácia (YOLOv3) |
| (NASCIMENTO et al., 2022) | 2022 | MPRNet + PS/PU | Gaussiano iterativo (SSIM) | Acurácia: 69,8% (RodoSol) |
| (HAQUE et al., 2022) | 2022 | ESRGAN (8×) | Filtro bilateral + cinza | Acurácia: 78% (SR) vs 91% (HR) |
| (KABIRAJ et al., 2023) | 2023 | ESRGAN + RRDB | Bicúbica | Acurácia: 15%→85% |
| (PAN; CHEN; LUO, 2023) | 2023 | DBAGAN (ESRGAN) | Bicúbica (SRLP 50×150) | SCSR: +12% acurácia vs sem SR |
| (NASCIMENTO et al., 2023) | 2023 | MPRNet+PS/PU+GP | Bicúbica+Gaussiano (SSIM≤0,1) | 74,2% (≥5 char, RodoSol) |
| (BOBY et al., 2023) | 2023 | ESRGAN | Não avaliado isoladamente | SR integrada sem avaliação separada |
| (PATTANAİK; BALABAN-TARAY, 2023) | 2023 | DCTGAN | Não especificado | Acurácia global: 0,983 |
| (SHI et al., 2023) | 2023 | BCSR (ResNet+Bezier) | TextZoom (real) | +8,4% acurácia (com ajuste fino) |
| (LU; GU; WANG, 2023) | 2023 | GAN + GPP | CycleGAN (tradução de estilo) | Inferior ao RPNNet; superior em desfoque |
| (ZHANG et al., 2023) | 2023 | DeblurGANv2 | Real (CCPD: blur, db, fn) | Levenshtein: 3→2 com desfoque; mAP@0.5: 99,7% |
| (KIM; KIM; PARK, 2024) | 2024 | AFA-Net | Bicúbica + desfoque movimento | +15,28% acurácia; +6,47% sim. sequencial |
| (PAN; TANG; TJAHIJADI, 2024) | 2024 | LPSRGAN (n-RCD) | Multi-estágio (n-RCD) | Melhor taxa de reconhecimento geral |
| (YANG et al., 2024) | 2024 | HDSN (Swin Transf.) | Pragmática (multi-componente) | Superior em todos os níveis (TextZoom) |

Fonte: Elaborada pelo autor.

A análise cronológica da [Tabela 2](#) revela uma evolução em três fases distintas. Em 2021, todos os cinco trabalhos adotaram protocolos de degradação simplificados — predominantemente

interpolação bicúbica — e arquiteturas baseadas em SRGAN ou ESRGAN, com avaliação centrada em métricas de qualidade de imagem (PSNR, SSIM). Em 2022, os quatro trabalhos do período mantiveram a dependência do protocolo bicúbico, mas avançaram na integração da SR com pipelines de ALPR mais complexos e na adoção de funções de perda perceptual que incorporam informações de OCR, como em [Nascimento et al. \(2022\)](#). Em 2023, o conjunto de sete trabalhos evidencia uma bifurcação: a maioria ainda emprega degradação simplificada, mas surgem propostas com protocolos mais realistas (como o TextZoom real em [Shi et al. \(2023\)](#)) e com integração fim a fim entre SR e reconhecimento (como em [Kim, Kim e Park \(2024\)](#), publicado em 2024 mas com desenvolvimento desse período). Por fim, em 2024, os três trabalhos analisados rompem diretamente com o paradigma de degradação simplificada, propondo modelos de degradação multi-estágio ([Pan, Tang e Tjahjadi \(2024\)](#)) e pragmática ([Yang et al. \(2024\)](#)) como elementos centrais da contribuição. As seções a seguir detalham essa evolução de forma temática.

3.1 Abordagens Baseadas em Degradação Clássica e Simplificada

Nesta seção, são agrupados os trabalhos que, embora tenham proposto avanços nas arquiteturas de rede, basearam a síntese de imagens de baixa resolução (LR) em métodos simplificados, como interpolação bicúbica ou ruídos básicos. Esses trabalhos concentram-se no período de 2021 a 2023.

[Hamdi, Chan e Koo \(2021\)](#) propôs a *Double Generative Adversarial Network for Image Enhancement and Super Resolution* (D_GAN_ESR), composta por duas GANs em cascata: a primeira voltada à nitidez da imagem e a segunda ao aumento de resolução por fator 4×. As imagens LR foram sintetizadas por redimensionamento bicúbico seguido de tradução de estilo via CycleGAN para introduzir ruídos característicos de câmeras analógicas governamentais. A avaliação foi realizada com PSNR, SSIM e PSNR-F, além de ferramentas de OCR (Tesseract e EasyOCR). No conjunto de dados 1, a D_GAN_ESR atingiu PSNR de 31,29 dB e SSIM de 0,781, com redução observada na taxa de erro de reconhecimento.

Ainda em 2021, [Wang et al. \(2021a\)](#) propôs um sistema de ALPR para placas coreanas voltado a condições adversas, como neblina, baixa iluminação e baixa resolução. O sistema utiliza ESRGAN para a etapa de SR, precedida por remoção de neblina (algoritmo *dark channel prior*) e melhoria de iluminação (CLAHE). Foram introduzidos dois modelos de reconhecimento: W-LPR, treinado com imagens completas de placa, e SC-LPR, baseado em segmentação de caracteres. O W-LPR alcançou 99,5% de acurácia com tempo total de 0,062 s por imagem, contra 98,9% e 0,126 s do SC-LPR.

Em [Shomee e Sams \(2021\)](#), foi proposto um sistema de ALPR para Bangladesh, cujas placas apresentam mais de 100 classes de caracteres, incluindo números, texto e variações

de cor por tipo de veículo. Um conjunto de dados específico (LPDB-A) foi construído com 1.928 imagens de veículos e 2.662 imagens de placas, das quais 720 sintéticas. A ESRGAN, treinada com imagens LR obtidas por interpolação bicúbica a partir de um modelo pré-treinado PSNR+RRDB, integrou o pipeline ao lado de YOLOv4. A inclusão da ESRGAN e de dados sintéticos elevou o mAP de teste, com YOLOv4 alcançando 98,35% de mAP para detecção.

Abdelaziz, Chan e Koo (2021) propôs uma arquitetura CNN+GAN inspirada na SRGAN que, além da super-resolução, trata desfoques e ruídos específicos de ALPR. O conjunto de dados foi composto por 1.000 imagens com versões LR obtidas por redimensionamento bicúbico 4× acrescido de desfoques e ruídos aleatórios. A abordagem aumentou o PSNR médio das imagens desfocadas de 30,18 para 31,69 dB e elevou a acurácia do reconhecimento via Tesseract 4 de 12,8% para 62,8%.

Xin et al. (2021) abordou o problema de super-resolução de placas por um método baseado em dicionário de aprendizagem, sem uso de GANs. Dez dicionários foram construídos a partir de imagens HR de cada dígito, cujos coeficientes de representação esparsa são compartilhados com as imagens LR. Uma rede convolucional de sub-pixel processa e combina as informações dos dicionários com as imagens super-resolvidas. Em experimentos com 100 imagens (80 treino, 20 validação), a abordagem obteve PSNR de 33,31 dB e SSIM de 0,856 na imagem 1, valores superiores aos do SRGAN (PSNR 33,29 dB, SSIM 0,843) e do VDSR (PSNR 33,06 dB, SSIM 0,849). O método destaca-se em imagens desfocadas de placas, mas apresenta limitações de generalização para padrões não previstos nos dicionários.

Em 2022, Pourhadi et al. (2022) integrou a SRGAN a um sistema de ALPR combinado com YOLOv5 para detecção e reconhecimento. Um mecanismo de seleção baseado em altura de caractere (limiar de 32 pixels) determina quando a SR deve ser acionada, controlando o custo computacional. As imagens LR foram sintetizadas por redimensionamento bicúbico. Com 40 blocos residuais, a configuração escolhida obteve PSNR de 22,38 dB, SSIM de 0,791 e MSE de 474,61. A acurácia do sistema integrado com SR atingiu 92,08%, ante 74,63% sem SR — um ganho de 18 pontos percentuais no conjunto de teste de 400 imagens.

Boby et al. (2022) avaliou o impacto da ESRGAN em pipelines com YOLOv3 e WPOD-Net para detecção de placas. No experimento com 50 imagens de teste, a inclusão da ESRGAN elevou a acurácia de 58,44% para 65,90% na configuração com YOLOv3 (+7,46 pontos percentuais) e de 65,30% para 66,62% com a WPOD-Net (+1,32 pontos percentuais). Os autores concluíram que, dado o custo adicional de complexidade, a ESRGAN não se justifica no pipeline com WPOD-Net, cujo ganho marginal é de apenas 1,32 pontos percentuais.

Haque et al. (2022) propôs um sistema de ALPR para Bangladesh com ESRGAN em fator de escala 8×, restaurando imagens de placas de 32×24 para 256×192 pixels. O treinamento utilizou 2.211 imagens convertidas para escala de cinza e filtradas bilateralmente. O detector YOLOv4 atingiu 96,2% de acurácia. Para a ESRGAN, foram obtidos MSE de 0,055 e PSNR de 12,587 dB. A acurácia de reconhecimento de caracteres com as imagens SR foi de 78%, contra

91% com as imagens HR originais.

Em 2023, [Kabiraj et al. \(2023\)](#) acoplou a ESRGAN a um módulo de OCR para placas de baixa qualidade, substituindo as camadas de normalização em lote por RRDB. O conjunto de dados utilizou 1.000 imagens HR com 182 imagens reais adicionais para teste, e as versões LR foram obtidas por transformação bicúbica. A inclusão da ESRGAN elevou a acurácia de 15% para 85%, superando a abordagem PROSR (abaixo de 70%).

[Pan, Chen e Luo \(2023\)](#) propôs a DBAGAN, baseada na ESRGAN com mecanismos de atenção espacial e de canais, para o problema de Super-Resolução de Caractere Único (SCSR – *Single Character Super-Resolution*). O conjunto de dados SRLP foi construído com placas redimensionadas para 50×150 pixels. A DBAGAN obteve PSNR de 26,27 dB, SSIM de 0,824 e acurácia de 91,4%, superando a ESRGAN (89,3%) e a SRGAN (86,0%). A abordagem SCSR apresentou ganho de 12 pontos percentuais de acurácia em relação à configuração sem super-resolução.

[Pattanaik e Balabantaray \(2023\)](#) propôs um sistema de ALPR baseado em DCTGAN para super-resolução e desfoque, combinado com o modelo Xception para reconhecimento de caracteres. A DCTGAN opera no domínio da frequência por meio da Transformada Discreta do Cosseno (DCT), reduzindo a complexidade computacional ao converter operações de multiplicação em operações de adição. O sistema utilizou um localizador de placa HaarCascade e limiarização adaptativa de Bernsen para segmentação. Os resultados reportados foram: taxa de detecção de placa de 0,993, taxa de reconhecimento de caracteres de 0,990 e acurácia global de 0,983.

[Boby et al. \(2023\)](#) propôs um sistema integrado de ALPR em que a ESRGAN é aplicada após a detecção da placa para melhorar a qualidade visual antes do reconhecimento de caracteres. O detector de veículos foi treinado com o Stanford Cars e o detector de placas com o Croatian License Plate. Embora a ESRGAN esteja presente no pipeline, o trabalho não realizou avaliação isolada da etapa de SR nem comparou os resultados com e sem SR, o que limita a caracterização quantitativa do impacto da super-resolução no sistema.

[Zhang et al. \(2023\)](#) propôs um sistema de pipeline para detecção e reconhecimento de placas em tempo real em cenários de tráfego complexo, integrando DeblurGANv2 para remoção de desfoque, YOLOv7 para detecção de placa e LPRNet para reconhecimento de caracteres. A motivação central é que ambientes reais apresentam imagens com desfoque por movimento, iluminação excessiva ou insuficiente, e placas capturadas a distâncias extremas, condições não cobertas adequadamente pelos sistemas convencionais. Diferentemente dos demais trabalhos desta seção, as imagens utilizadas para treinamento e avaliação não foram sintetizadas por protocolos de degradação, mas extraídas diretamente dos subconjuntos CCPD_blur (20.611 imagens), CCPD_db (10.132 imagens com luminosidade extrema) e CCPD_fn (20.967 imagens com distância extrema), combinados em um conjunto de 47.652 imagens (proporção 7:1:2 para treino, validação e teste). Para o módulo de reconhecimento (LPRNet), foi utilizado o CCPD

em conjunto com placas sintéticas geradas conforme o padrão GA36-2007 e GA36-2014, para compensar o desbalanceamento de classes entre províncias chinesas. O detector YOLOv7 obteve mAP@0.5 de 99,7%, com mAP@0.5:0.95 adicional após a introdução do módulo de remoção de desfoque. A avaliação do sistema completo foi realizada com a distância de Levenshtein entre a sequência predita e o rótulo verdadeiro, com a inclusão do módulo DeblurGANv2 reduzindo a distância de 3 para 2 no exemplo reportado, evidenciando o impacto da etapa de pré-processamento sobre a acurácia de reconhecimento.

Embora os trabalhos desta seção demonstrem progressos nas arquiteturas de SR e na integração com pipelines de ALPR, a dependência de degradações bicúbicas ou gaussianas simples restringe a capacidade de generalização para os artefatos encontrados em câmeras de vigilância reais, como compressão JPEG, ruído de sensor correlacionado e múltiplos estágios de transmissão.

3.2 Integração de Super-Resolução e OCR: Perda Perceptual e Atenção

Este grupo de trabalhos reconhece que métricas tradicionais como PSNR e SSIM são insuficientes para guiar a super-resolução no contexto de ALPR, propondo a integração de mecanismos de atenção e funções de perda baseadas em reconhecimento de caracteres (OCR Loss). As propostas deste grupo surgem a partir de 2022 e avançam até 2024.

[Nascimento et al. \(2022\)](#) propôs o uso da camada *Pixel Shuffle* (PS) com módulos de atenção estendidos da MPRNet para super-resolução de placas. A função de perda perceptual foi adaptada para incorporar a taxa de reconhecimento de caracteres via algoritmo de Levenshtein, em conjunto com o SSIM para qualidade visual. As imagens LR foram geradas por aplicação iterativa de ruído gaussiano aleatório até atingir o nível de degradação desejado, medido por SSIM. O conjunto de dados RodoSol-ALPR foi utilizado, com 8.000 imagens para treino e 4.000 para validação. A abordagem proposta obteve 69,8% de acurácia no reconhecimento de todos os caracteres, contra 61,4% da abordagem sem SR e 48,2% do MPRNet, com PSNR de 26,4 dB e SSIM de 0,89.

Em [Nascimento et al. \(2023\)](#), foram incorporadas melhorias à abordagem anterior: um módulo de atenção trifold (PLTFAM) que analisa relações intercanais, posicionais e geométricas (Geometrical Perception Unit – GP), e uma nova função de perda perceptual baseada em rede pré-treinada para reconhecimento de placas. O limite máximo de degradação foi ajustado para SSIM de 0,1. No conjunto RodoSol, a nova proposta alcançou 74,2% de acurácia para identificação de pelo menos 5 dos 7 caracteres, contra 42,2% do trabalho anterior. No conjunto PKU, a acurácia foi de 97,3%, contra 82,5% anterior.

[Shi et al. \(2023\)](#) propôs o método BCSR (*Bezier Curve-based Super-Resolution*) para

reconstrução de imagens textuais espacialmente deformadas. Um módulo de geração de prioridade textual (TPG – *Text Prior Generation*) utiliza curvas de Bezier cúbicas para delimitar regiões de texto curvadas, evitando a inclusão de plano de fundo. A função de perda TP Loss mede a similaridade entre as características textuais geradas pelo TPG e a imagem HR de referência. O modelo foi avaliado no conjunto TextZoom, com 17.367 imagens de treino e o restante para teste. Sem ajuste fino, a acurácia de reconhecimento aumentou 5,1%; com ajuste fino, o ganho foi de 8,4 pontos percentuais.

Lu, Gu e Wang (2023) propôs um sistema de reconhecimento de placas em condições naturais (*in wild*), composto por remoção de degradação via CycleGAN, super-resolução com GPP (*Gradient Profile Prior*) e reconhecimento por CNN-LSTM com função de perda CTC. O GPP é incorporado para gerar bordas nítidas no processo de SISR. O sistema foi avaliado no conjunto CCPD (300 mil imagens). Em subconjuntos de imagens com desfoque, o método proposto superou o RPNNet; nos demais subconjuntos, o RPNNet apresentou maior acurácia, com diferença atribuída pelos autores à correção de inclinação realizada pelo RPNNet.

Em 2024, Kim, Kim e Park (2024) introduziu o problema Joint-IRLPRNet, que une restauração de imagem e reconhecimento de placas em treinamento fim a fim. A rede AFA-Net (*Adaptive Feature Attention Network*) combina MPRNet para remoção de desfoque de movimento e DeFiAN para aumento de resolução 4×, com composição de características por módulos de atenção multi-nível. O conjunto de dados LBLP foi construído com 2.779 imagens de placas capturadas por câmera de bordo, em versões com resolução original (256×96) e com desfoque e redução 4× (64×24). Os resultados mostraram ganhos de 15,28 pontos percentuais na acurácia de reconhecimento, 6,47 pontos percentuais na similaridade sequencial e 3,89 pontos percentuais na similaridade de caracteres em relação ao baseline. O esquema Joint-IRLPRNet apresentou resultados superiores ao treinamento isolado da rede de SR.

A convergência desses trabalhos indica que a integração da função de perda com informações semânticas de OCR é uma direção promissora para a SR em ALPR. Contudo, como observado nos experimentos de Nascimento et al. (2022) e Nascimento et al. (2023), a eficácia dessa estratégia depende da qualidade das características extraídas pelo OCR, que por sua vez é sensível à qualidade das imagens LR de treinamento.

3.3 Modelos de Degradação Avançada e Cenários Naturais

Os trabalhos mais recentes, publicados em 2024, abordam diretamente a limitação dos modelos de degradação simplificados, propondo métodos mais robustos para simular as condições reais de captura de placas veiculares.

Pan, Tang e Tjahjadi (2024) propôs o modelo de degradação *N-stage Random Combination Degradation* (n-RCD), que estende os modelos tradicionais ao combinar aleatoriamente múltiplos componentes de degradação — limitações de sensor, transmissão de rede, compressão JPEG e

ruído de câmera — em n estágios sequenciais. A motivação central é que os modelos tradicionais, ao limitarem o espaço de degradação a três formas básicas (desfoque, redução dimensional e ruído), geram queda de desempenho quando os modelos são testados em condições reais. A arquitetura LPSRGAN foi desenvolvida para operar sobre esse espaço de degradação mais amplo, incorporando blocos RRDB+ com módulos de agregação de características e discriminador baseado em U-Net3+. Uma função de perda perceptual OCR Loss foi proposta, calculada via CTC entre a sequência de caracteres predita pelo OCR e a sequência rotulada. O conjunto LicensePlateDataset10k foi construído com 10.000 imagens de treino. Nos experimentos, foi observado que até três estágios de degradação o desempenho melhora, sendo esse o número de estágios adotado. O modelo n-RCD individualmente contribuiu com ganho de 0,61 na acurácia e 2,01 dB de PSNR; a OCR Loss contribuiu com ganho adicional de 0,83 na acurácia e 0,85 dB de PSNR em relação ao modelo sem esse componente.

Yang et al. (2024) adotou a arquitetura HDSN, baseada em Swin Transformer, para super-resolução de imagens textuais em cena. O protocolo de degradação denominado “pragmático” incorpora desfoque gaussiano generalizado, ruído aleatório, redução dimensional por bicúbica e variações de foco e movimentação. O modelo foi pré-treinado no conjunto MJSynth (9 milhões de imagens) e ajustado com o TextZoom. Nos experimentos em sete conjuntos de dados de cenas textuais (IC03, IC13, IC15, IIIT5K, CUTE80, SVT e SVTP), o HDSN superou os demais modelos comparados em retenção de detalhes e acurácia de reconhecimento em todos os níveis de dificuldade avaliados.

A maioria dos trabalhos analisados no período 2021–2023 ainda depende de modelos de degradação simplificados (bicúbico/gaussiano), conforme sistematizado na Tabela 2. Apenas as propostas de 2024 tratam a modelagem realista da degradação como contribuição central, o que está diretamente alinhado à lacuna investigada nesta dissertação e valida a relevância da investigação sobre protocolos de degradação progressivamente complexos.

4

Metodologia Experimental

Este capítulo apresenta a metodologia experimental desenvolvida para validar as hipóteses levantadas nesta dissertação. A estrutura foi projetada para isolar e quantificar o impacto de duas variáveis principais na generalização de modelos de ALPR: (1) a complexidade do protocolo de degradação utilizado no treinamento e (2) a especialização da função de perda perceptual por meio de *backbones* de OCR.

A apresentação segue a lógica de um fluxograma de pesquisa, onde cada componente é examinado em três níveis: a questão de pesquisa associada, o método proposto para abordagem e os resultados esperados.

4.1 Visão Macro do Fluxograma Experimental

Com base na revisão bibliográfica apresentada no [Capítulo 3](#), que identificou a lacuna na validação de modelos em cenários robustos decorrente da predominância de degradações simplificadas, o fluxograma experimental foi estruturado em três fases principais:

1. **Preparação de Dados e Protocolos de Degradação** (Nós 1 e 2): Adaptação dos conjuntos de treinamento e testes com diferentes níveis de complexidade de degradação (Simplificada vs. Robusta);
2. **Adaptações Arquiteturais e Treinamento** (Nós 3 e 4): Configuração dos modelos de super-resolução, com foco na substituição de extratores de características genéricos por *backbones* especializados em OCR;
3. **Avaliação Comparativa e Generalização** (Nó 5): Comparação sistemática entre as abordagens nos domínios de degradação, do mais simplificado ao mais complexo.

A Tabela 3 apresenta o mapeamento entre as etapas do experimento e as seções deste capítulo.

Tabela 3 – Mapeamento entre fases do experimento e seções analíticas

| Nó | Seção | Foco |
|-----------|--------------|---|
| 1 e 2 | 4.2 | Variação de protocolos de degradação |
| 3 e 4 | 4.3 | Adaptações arquiteturais (OCR Backbone) |
| 5 | 4.4 | Avaliação sistemática |

Fonte: Elaborada pelo autor.

A Figura 5 apresenta a visão unificada da metodologia. O fluxo principal conduz da preparação dos dados, passando pelo treinamento sob diferentes paradigmas de degradação, até a avaliação final de generalização.

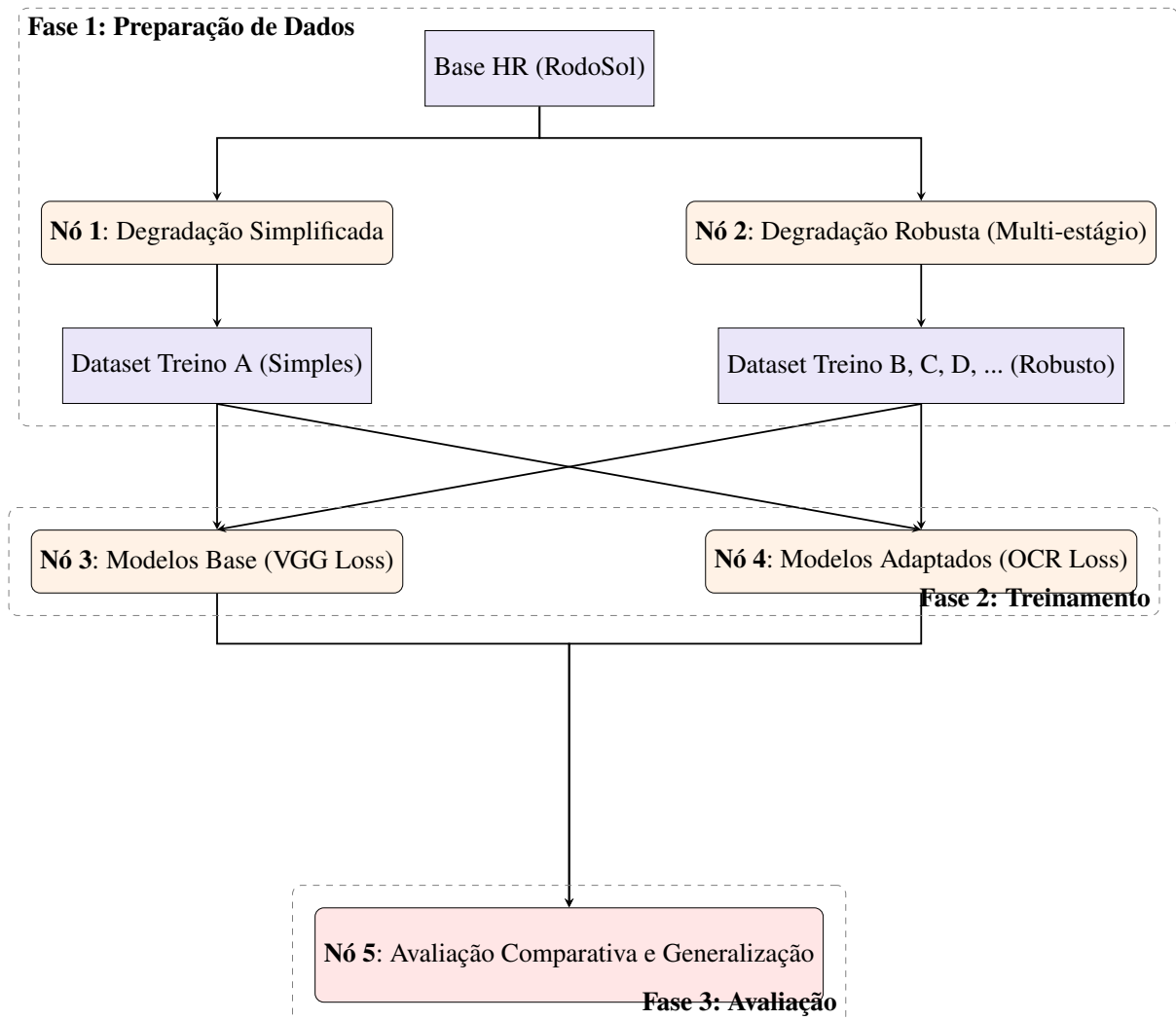


Figura 5 – Diagrama do fluxo experimental integrado, destacando a preparação dos dados (Nós 1 e 2), o treinamento com adaptações arquiteturais (Nós 3 e 4) e a avaliação cruzada (Nó 5).

4.2 Nós 1 e 2: Variação de Protocolos de Degradação

Questão de Pesquisa: O uso de metodologias de degradação de alta ordem (robusta) durante o treinamento resulta em generalização superior em comparação aos métodos simplificados tradicionais?

Metodologia: Para isolar o impacto do método de degradação, foram geradas bases de dados de treinamento que diferem na quantidade de estágios pelos quais as imagens foram submetidas no *pipeline* de degradação, sendo todas as versões derivadas de uma mesma base de imagens de alta resolução (Rodosol-ALPR):

- **Degradação Simplificada (Nó 1):** Síntese de imagens LR utilizando apenas interpolação bicúbica e *gaussian blur*, replicando a abordagem predominante na literatura (ABDELAZIZ;

CHAN; KOO, 2021);

- **Degradação Robusta** (Nó 2): Utilização do *pipeline* de degradação de alta ordem do RealESRGAN, com variação no número de estágios para identificar o ponto de equilíbrio entre complexidade e preservação de características.

Resultados Esperados:

- Modelos treinados com degradação simplificada devem apresentar desempenho elevado em testes sintéticos do mesmo domínio, mas desempenho reduzido em imagens com composição de ruídos mais complexa;
- A degradação robusta deve ampliar o espaço de representação aprendido pelo modelo, permitindo a recuperação de placas em cenários não vistos durante o treinamento.

4.3 Nós 3 e 4: Incorporação de Backbones de OCR

Questão de Pesquisa: A substituição de extratores de características genéricos (como VGG19) por *backbones* especializados em reconhecimento de caracteres melhora a preservação semântica dos textos nas placas?

Metodologia: Esta etapa concentra-se na modificação da função de perda perceptual (*Perceptual Loss*). Em vez de penalizar o modelo apenas por diferenças visuais genéricas, propõe-se guiar o gradiente com base na legibilidade dos caracteres.

1. Treinamento de modelos base (RealESRGAN/ESRGAN/LPR-RSR-EXT) utilizando a perda perceptual padrão baseada em VGG;
2. Treinamento de variantes substituindo o extrator VGG pelo extrator do modelo OCR utilizado no trabalho da LPR-RSR-EXT, e vice-versa;
3. Comparação direta da taxa de acerto de caracteres entre as versões.

Resultados Esperados: A hipótese é que o uso de *backbones* de OCR reduzirá a ocorrência de reconstruções incorretas de caracteres (como transformar um “8” em “B”), priorizando a estrutura morfológica dos glifos em detrimento de texturas de fundo irrelevantes.

4.4 Nó 5: Avaliação Comparativa

A fase final (Nó 5) consolidará os resultados por meio de uma matriz de avaliação cruzada, confrontando os modelos Simples vs. Robusto e Com OCR vs. Sem OCR. O desempenho foi medido primariamente pela métrica funcional de reconhecimento de placas, alinhando a avaliação com o objetivo final de um sistema ALPR.

5

Estudos Experimentais

Este capítulo apresenta e analisa os experimentos realizados com base nos objetivos e hipóteses formulados nesta dissertação. A [seção 5.1](#) consolida os resultados de um estudo previamente publicado na conferência WorldCist 2025¹, o qual investigou o impacto do método de degradação utilizado na síntese de imagens de baixa resolução sobre a capacidade de generalização da arquitetura RealESRGAN no contexto de ALPR.

A partir dessas evidências iniciais — que indicaram melhora na generalização dos modelos quando treinados com conjuntos de dados degradados de forma mais robusta — este capítulo aprofunda a investigação com o objetivo de analisar com maior granularidade o comportamento dos modelos em diferentes domínios de teste e investigar como as características visuais das placas (fonte, formato e contraste) são afetadas qualitativamente pela super-resolução sob diferentes protocolos de degradação.

Os experimentos aqui descritos vão além dos resultados apresentados na [seção 5.1](#), trazendo contribuições originais que sustentam as propostas centrais desta dissertação. A metodologia empregada em cada etapa será detalhada nas seções subsequentes, acompanhada da análise crítica dos resultados obtidos.

5.1 Avaliação do Impacto do Método de Degradação na Generalização do Modelo RealESRGAN

Com base nos objetivos desta dissertação, foi conduzido um experimento cujo propósito central foi investigar como diferentes metodologias de degradação afetam a capacidade de generalização de modelos de super-resolução baseados em GANs no contexto de ALPR. Este experimento foi originalmente publicado como artigo na conferência WorldCist 2025, sob o

¹ <https://link.springer.com/book/10.1007/978-3-031-97119-8>; Capítulo 40

título “*Study on the impact of the degradation method on the generalization of Super-Resolution models for ALPR*”.

O foco da investigação foi a arquitetura RealESRGAN, em função de sua proposta de operar em um espaço de degradação mais amplo e realista. Foram utilizados dois conjuntos de dados construídos a partir do RodoSol-SR, ambos com imagens de alta resolução (HR) de tamanho 128×128 pixels, derivadas de recortes e retificações das placas. Para a construção das versões de baixa resolução (LR), dois métodos de degradação distintos foram empregados:

- **Degradação simplificada:** interpolação bicúbica com adição de desfoque gaussiano, correspondendo ao método predominante em trabalhos anteriores da literatura;
- **Degradação robusta:** baseada na metodologia original do RealESRGAN, com uma sequência de transformações que simula o processo de degradação de imagens capturadas por sistemas de vigilância, incluindo ruído, compressão JPEG, redimensionamentos e efeitos de sensores.

Dois modelos distintos da arquitetura RealESRGAN foram treinados:

- **Modelo especializado:** treinado com imagens degradadas pelo método simplificado;
- **Modelo generalista:** treinado com imagens degradadas pelo método robusto.

Ambos os modelos foram treinados com os mesmos hiperparâmetros e estrutura (gerador RRDBNet e discriminador UNet), utilizando função de perda combinando L1, perda perceptual com VGG19 e perda adversarial (GAN).

Para avaliar a generalização, foram utilizadas 4.000 imagens de teste sob cada tipo de degradação, avaliadas por métricas tradicionais de qualidade (PSNR, SSIM, MSE, LPIPS) e por desempenho no reconhecimento de caracteres por meio de um modelo OCR baseado em YOLOv8.

Os resultados evidenciam que o modelo generalista apresenta desempenho mais consistente entre os dois conjuntos de teste, enquanto o modelo especializado apresentou desempenho elevado apenas no conjunto cujo domínio de degradação era compatível com o de seu treinamento. A [Tabela 4](#) resume os valores médios das métricas de qualidade das imagens super-resolvidas.

Tabela 4 – Resultados médios de PSNR, SSIM, MSE, LPIPS e tempo de execução (ms) por modelo e conjunto de teste.

| Modelo / Conjunto | PSNR | SSIM | MSE | LPIPS | Tempo (ms) |
|--------------------------|-------------|-------------|------------|--------------|-------------------|
| Generalista / robusto | 18,000 | 0,672 | 1342,885 | 0,243 | 94,203 |
| Generalista / simples | 18,776 | 0,707 | 1235,146 | 0,237 | 82,376 |
| Especializado / robusto | 13,196 | 0,223 | 3560,135 | 0,537 | 80,353 |
| Especializado / simples | 17,250 | 0,712 | 1698,232 | 0,262 | 75,136 |

Fonte: Elaborada pelo autor.

A [Tabela 5](#) mostra os resultados de OCR para as imagens super-resolvidas, considerando o número de placas com 5, 6 ou todos os 7 caracteres corretamente reconhecidos.

Tabela 5 – Desempenho do OCR em função do modelo de super-resolução e do método de degradação.

| Modelo / Conjunto | 7 caracteres | ≥ 6 caracteres | ≥ 5 caracteres |
|--------------------------|---------------------|-----------------------|-----------------------|
| Generalista / robusto | 69 | 129 | 207 |
| Generalista / simples | 113 | 222 | 378 |
| Especializado / robusto | 22 | 46 | 76 |
| Especializado / simples | 315 | 397 | 522 |

Fonte: Elaborada pelo autor.

Os resultados evidenciam que o modelo generalista mantém desempenho aceitável em domínios não vistos durante o treinamento, ao contrário do modelo especializado, cujo desempenho cai de forma acentuada ao ser testado com o protocolo robusto (315 placas com 7 caracteres no domínio simples contra apenas 22 no domínio robusto). Isso indica que a adoção de uma metodologia de degradação mais abrangente na síntese dos dados de treinamento contribui para ampliar a capacidade de generalização da arquitetura RealESRGAN.

Tal observação é relevante no contexto de ALPR, uma vez que os sistemas operam em condições variáveis de iluminação, ângulo, ruído e qualidade ótica. O uso de modelos treinados com maior diversidade de degradações tende a produzir desempenho mais estável em cenários reais.

Adicionalmente, observou-se uma diferença no tempo médio de inferência, com o modelo generalista apresentando tempo inferior ao modelo especializado em um dos cenários — constatação que levanta a hipótese de influência indireta da diversidade do domínio de treinamento sobre o comportamento temporal do modelo, tema a ser investigado em trabalhos futuros.

5.2 Avaliação e Comparação do Impacto do Método de Degradação na Generalização entre Modelos de Super-Resolução

Nesta seção, são detalhados os experimentos realizados para comparar o impacto da metodologia de degradação na capacidade de generalização das arquiteturas RealESRGAN, ESRGAN e LPR-RSR-EXT. Cada modelo foi treinado com cinco conjuntos de dados distintos, com exceção do LPR-RSR-EXT, que foi avaliado também com a base de dados original indicada no artigo. Todos os conjuntos são derivados da mesma base de imagens HR do estudo apresentado na [seção 5.1](#). Os protocolos de degradação aplicados foram:

- **Conjunto 1:** Aplicação exclusiva de *gaussian blur*
- **Conjunto 2:** Metodologia simplificada com 1 estágio de degradação
- **Conjunto 3:** Metodologia RealESRGAN com 2 estágios de degradação
- **Conjunto 4:** Metodologia estendida com 3 estágios de degradação
- **Conjunto 5:** Metodologia estendida com 4 estágios de degradação

Os treinamentos foram conduzidos em um sistema equipado com processador AMD Ryzen 7 5700G, 16 GB de memória RAM DDR4 a 3200 MHz e placa gráfica NVIDIA RTX 3060 com 12 GB de VRAM. As versões baseadas na arquitetura LPR-RSR-EXT, que demandam maior capacidade computacional, foram treinadas em um servidor com processador Intel Core i9-10900K, 64 GB de memória RAM DDR4 e duas placas gráficas NVIDIA RTX 4090 com 24 GB de VRAM cada.

A avaliação considerou duas métricas principais: (1) taxa de reconhecimento de caracteres, medida por um modelo OCR aplicado às imagens super-resolvidas, e (2) métricas tradicionais de qualidade de imagem (PSNR e SSIM). Essa abordagem permite comparar objetivamente a capacidade de cada modelo em reconstruir imagens legíveis para sistemas ALPR sob diferentes protocolos de degradação e funções de perda perceptual.

5.2.1 Impacto Combinado da Metodologia de Degradação e Perda Perceptual

Esta seção investiga simultaneamente as duas hipóteses centrais deste trabalho: (1) o impacto da complexidade do protocolo de degradação na generalização e (2) a eficácia de funções de perda perceptual baseadas em OCR em comparação ao VGG-19 tradicional. A análise integrada dessas variáveis revela interações que não seriam observadas em estudos isolados.

5.2.1.1 Desempenho de Modelos Treinados com Degradação Simplificada

A Figura 6 apresenta as taxas de reconhecimento completo de placas (percentual de placas com todos os 7 caracteres corretamente reconhecidos) para cada combinação de degradação de treinamento e teste, organizada em formato de mapa de calor para facilitar a identificação visual dos padrões de generalização. Modelos treinados exclusivamente com degradação Gaussian apresentam queda acentuada de desempenho quando confrontados com degradações mais complexas.

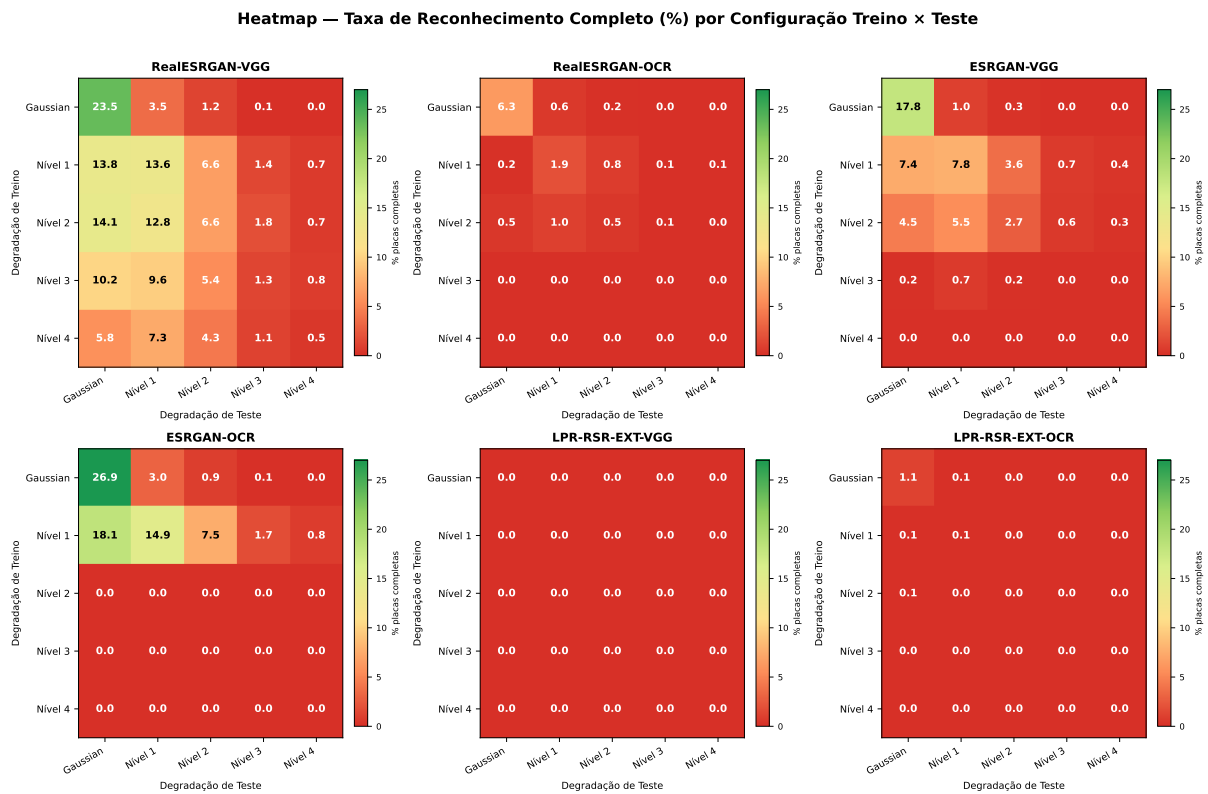


Figura 6 – Mapa de calor — Taxa de Reconhecimento Completo (%) por configuração treino × teste. Eixo vertical: protocolo de treinamento; eixo horizontal: protocolo de teste. Verde indica maior taxa de reconhecimento; vermelho indica menor.

Para RealESRGAN-VGG, o modelo treinado com Gaussian alcança 23,5% de correspondência completa nos testes *in-domain* (Gaussian), mas cai para 1,2% quando testado com imagens que passaram por dois estágios de degradação, e para 0,0% com quatro estágios. Esse padrão de desempenho concentrado na diagonal principal — elevado apenas quando o protocolo de treino e teste são idênticos — replica-se nas demais arquiteturas, indicando que protocolos simplificados restringem a capacidade de generalização.

O padrão é ainda mais pronunciado na ESRGAN: o modelo treinado com Gaussian (17,8% *in-domain*) cai para 0,3% em imagens com dois estágios de degradação e 0,0% com três ou mais estágios, tanto com VGG quanto com OCR Loss. Essa queda não é gradual — trata-se

de incapacidade do modelo em reconstruir qualquer estrutura reconhecível de caracteres quando a degradação de teste diverge substancialmente do domínio de treinamento.

5.2.1.2 Robustez em Domínio Cruzado com Degradação Multi-Estágio

Em contraste, modelos treinados com um ou dois estágios de degradação demonstram generalização superior. A Figura 7 apresenta a acurácia média de caracteres (escala 0–7), fornecendo granularidade adicional além da métrica binária de correspondência completa.

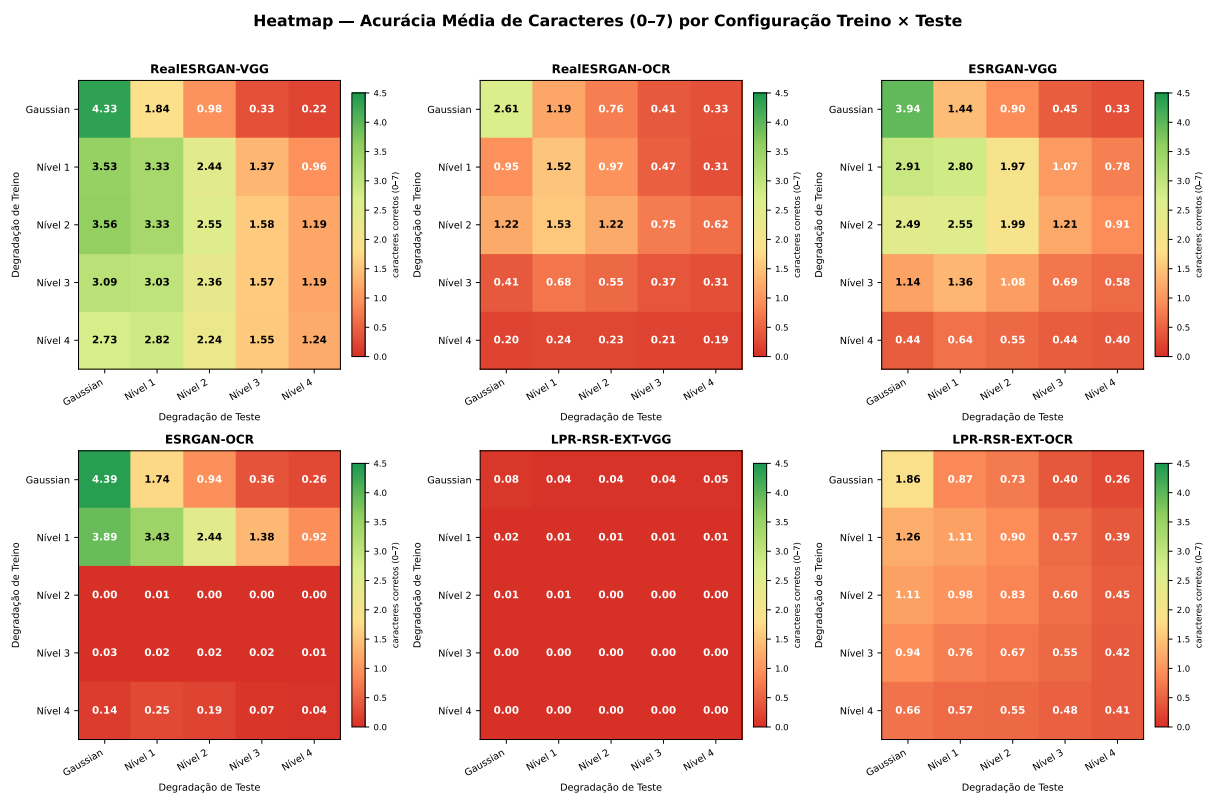


Figura 7 – Mapa de calor — Acurácia Média de Caracteres (escala 0–7) por configuração treino × teste. Eixo vertical: protocolo de treinamento; eixo horizontal: protocolo de teste. Verde indica maior número médio de caracteres reconhecidos; vermelho indica menor.

Para RealESRGAN-VGG, o modelo treinado com dois estágios de degradação alcança 3,56, 3,33, 2,55, 1,58 e 1,19 caracteres corretamente reconhecidos (de 7 no total) nos cinco domínios de teste, respectivamente. Comparado ao modelo treinado apenas com Gaussian, a versão de dois estágios reconhece 1,58 vs. 0,33 caracteres em testes com três estágios de degradação — uma diferença de 1,25 caracteres (razão de 4,8×) — e 1,19 vs. 0,22 caracteres em testes com quatro estágios — diferença de 0,97 caracteres (razão de 5,4×) —, apesar de nunca ter visto tais condições extremas durante o treinamento.

Esse resultado indica que a degradação multi-estágio atua como regularizador, expandindo o espaço de representação aprendido pela rede. O modelo não apenas memoriza a transformação

bicúbica→HR, mas aprende invariâncias mais fundamentais nas estruturas de placas que se mantêm através de diferentes modalidades de corrupção.

5.2.1.3 Complexidade Ótima de Degradação é Arquitetura-Dependente

Analisando ambos os mapas de calor, emerge um padrão consistente: o nível ótimo de degradação de treinamento varia por arquitetura, mas situa-se na faixa intermediária (nível 1 a 2 de degradação), e não nos extremos.

RealESRGAN-VGG: O modelo treinado com dois estágios de degradação apresenta o melhor equilíbrio entre desempenho *in-domain* e generalização. Embora sacrifique acurácia *in-domain* (14,1% vs. 23,5% do modelo Gaussian nos mesmos testes), mantém robustez em testes complexos: 1,8% em três estágios vs. 0,1% para o modelo Gaussian — diferença de 1,7 pontos percentuais (razão de 18×). A média dos cinco domínios de teste posiciona a versão de dois estágios como a configuração de equilíbrio ótimo.

ESRGAN-VGG: O modelo treinado com 1 estágio de degradação fornece o melhor balanço, com 7,8% de correspondência completa em testes com um estágio e 3,6% em testes com dois estágios. A arquitetura ESRGAN apresenta menor tolerância à complexidade de degradação em comparação à RealESRGAN, com melhor desempenho no protocolo de um estágio.

LPR-RSR-EXT: Apresenta desempenho absoluto inferior ao RealESRGAN e ao ESRGAN em todas as configurações, incluindo o melhor caso (apenas Gaussian Blur). Essa discrepância aponta para limitações arquiteturais ou necessidade de ajuste de hiperparâmetros além do escopo deste estudo. Um fator identificado durante a construção do algoritmo de testes é que o modelo OCR utilizado no artigo original da arquitetura demonstrou bom reconhecimento em placas com legibilidade reduzida. Como o presente trabalho utilizou um modelo OCR baseado na arquitetura YOLOv8n treinado na base UFS-ALPR — diferente da Rodosol-ALPR — essa diferença de domínio pode ter influenciado os resultados.

5.2.1.4 Retornos Decrescentes além de Dois Estágios de Degradação

Os modelos treinados com três ou mais estágios de degradação apresentam queda progressiva de desempenho em todos os domínios de teste. Para RealESRGAN-VGG, a versão com quatro estágios alcança 5,8% de correspondência completa em testes Gaussian (comparado a 14,1% para a versão de dois estágios) e 0,5% em testes com quatro estágios (vs. 0,7% para a versão de dois estágios). Em acurácia média de caracteres: 2,73 vs. 3,56 em testes Gaussian — uma diferença de 0,83 caracteres.

A inspeção visual das imagens com três ou mais estágios de degradação revela que as fronteiras de caracteres tornam-se ambíguas mesmo para anotadores humanos. Quando a degradação destrói as estruturas semânticas que o modelo deve aprender a reconstruir, o mapeamento LR→HR torna-se fundamentalmente mal-posto: não há informação suficiente nas

entradas para inferir saídas corretas de forma determinística.

Esse resultado estabelece uma relação **não-monotônica** entre complexidade de degradação e generalização, contrariando a intuição de que maior diversidade de degradação sempre melhora a robustez. Existe um ponto de equilíbrio além do qual estágios adicionais de degradação não expandem o espaço de representação, mas comprometem o sinal de treinamento com exemplos degenerados.

5.2.1.5 Efeitos Arquitetura-Dependentes da Perda Perceptual

A escolha entre VGG Loss e OCR Loss produz efeitos dependentes da arquitetura, revelando três padrões distintos:

5.2.1.5.1 Padrão 1: Vantagem de OCR Loss na ESRGAN com um estágio de degradação.

ESRGAN com OCR Loss treinada com um estágio de degradação alcança o **melhor desempenho absoluto** entre as 30 configurações testadas: 18,1% de correspondência completa em testes Gaussian (Figura 6), contra 7,4% da ESRGAN-VGG na mesma configuração — uma diferença de 10,7 pontos percentuais. Em acurácia média de caracteres (Figura 7): 3,89 vs. 2,91 em testes Gaussian, e 3,43 vs. 2,80 em testes com um estágio — diferenças de 0,98 e 0,63 caracteres, respectivamente.

Esse resultado valida a hipótese de que a perda perceptual guiada por OCR pode melhorar a reconstrução de caracteres, *mas apenas sob condições específicas*. A vantagem existe quando:

1. A arquitetura permite captura de características de texto (ESRGAN com 16 blocos RRDB);
2. O protocolo de degradação é suficientemente simples para preservar estruturas de caracteres reconhecíveis pelo *backbone* OCR.

5.2.1.5.2 Padrão 2: Colapso das versões com OCR Loss em degradação complexa.

A vantagem de OCR Loss **desaparece ou se reverte em estágios superiores de degradação**. ESRGAN-OCR treinada com dois ou mais estágios de degradação exhibe colapso total, com 0,0% de acurácia em todos os domínios de teste (Figura 6). Não se trata de declínio gradual, mas de falha de convergência.

RealESRGAN-OCR apresenta padrão semelhante, embora menos extremo: a versão com um estágio alcança 1,9% de correspondência completa, mas com dois estágios cai para 1,0%, e com três ou mais estágios chega a 0,0%. Em acurácia média de caracteres (Figura 7), RealESRGAN-OCR com dois estágios produz 1,22 caracteres, contra 3,56 da versão VGG na mesma configuração — uma diferença de 2,34 caracteres.

Esse fenômeno aponta para um **limiar crítico de degradação** para OCR Loss, situado entre os estágios 1 e 2. Quando a degradação destrói as estruturas refinadas de caracteres, o *backbone* OCR extrai características ambíguas, gerando gradientes de perda perceptual de alta variância que desestabilizam a otimização do gerador.

5.2.1.5.3 Padrão 3: Dependência de LPR-RSR-EXT em relação à OCR Loss.

LPR-RSR-EXT-VGG apresenta **falha completa** em todas as 25 combinações treino-teste, com aproximadamente 0,0% em ambas as métricas ([Figura 6](#) e [Figura 7](#)). Em contraste, LPR-RSR-EXT-OCR habilita desempenho não-zero: a versão treinada apenas com Gaussian produz 1,1% de correspondência completa e 1,86 caracteres médios em testes Gaussian.

Isso indica que arquiteturas com unidades de percepção geométrica e atenção espacial para texto requerem funções de perda conscientes de caracteres. Características genéricas VGG, projetadas para reconhecimento de objetos gerais, não fornecem gradientes adequados para os componentes especializados em texto, resultando em não-convergência.

Ainda assim, mesmo com OCR Loss, LPR-RSR-EXT alcança no máximo 1,86 caracteres vs. 3,89 para ESRGAN-OCR e 3,56 para RealESRGAN-VGG, indicando que, além da dependência de OCR Loss para convergência, limitações arquiteturais adicionais ou desalinhamento de hiperparâmetros limitam sua competitividade no escopo dos testes realizados.

5.2.1.6 VGG Loss como Escolha Robusta Padrão

A perda perceptual baseada em VGG mantém treinamento estável em todos os estágios de degradação. RealESRGAN-VGG apresenta desempenho consistente, e ESRGAN-VGG treina com sucesso nos níveis 1 e 2 de degradação sem colapso. Quando a degradação corrompe um caractere, as características VGG ainda capturam padrões geométricos (linhas verticais/horizontais, regiões de contraste), enquanto as características OCR tornam-se ambíguas.

Comparando as médias nos cinco domínios de teste para modelos com dois estágios de degradação, VGG Loss supera OCR Loss na RealESRGAN em 2,3× (2,44 vs. 1,07 caracteres) e na ESRGAN de forma absoluta (1,83 vs. 0,00 caracteres). Somente no LPR-RSR-EXT, onde VGG causa falha completa, OCR Loss justifica-se.

5.2.1.7 Síntese: Interação entre Degradação, Arquitetura e Função de Perda

Os resultados estabelecem que a otimização de sistemas ALPR-SR não pode tratar degradação, arquitetura e função de perda como variáveis independentes. A [Tabela 6](#) sintetiza as configurações ótimas identificadas.

Tabela 6 – Configurações ótimas por arquitetura — síntese baseada em desempenho *cross-domain*.
Desempenho reportado como acurácia média de caracteres nos 5 domínios de teste.

| Arquitetura | Degradação Ótima | Perda Ótima | Desempenho (caracteres) |
|--------------------|-------------------------|--------------------|--------------------------------|
| RealESRGAN | 2 estágios | VGG Loss | 2,44 (média) |
| ESRGAN | 1 estágio | OCR Loss | 2,61 (média) |
| ESRGAN | 1 estágio | VGG Loss | 2,04 (média) |
| LPR-RSR-EXT | 2 estágios | OCR Loss | 0,79 (média) |

Escolha padrão: RealESRGAN + 2 estágios de degradação + VGG Loss
Máximo desempenho in-domain: ESRGAN + 1 estágio de degradação + OCR Loss

Fonte: Elaborada pelo autor.

6

Conclusão

Esta dissertação investigou sistematicamente o impacto da metodologia de degradação sintética e da função de perda perceptual na capacidade de generalização de modelos de super-resolução (SR) aplicados ao Reconhecimento Automático de Placas Veiculares (ALPR). Os resultados obtidos validam parcialmente as hipóteses centrais do trabalho e oferecem contribuições práticas e metodológicas para a área, conforme descrito na [seção 1.4](#) e reforçado nas discussões das seções a seguir.

6.1 Confirmação das Hipóteses

6.1.1 Hipótese 1: Metodologia de Degradação Multi-Estágio

Os experimentos confirmaram que a **diversidade e complexidade do protocolo de degradação** utilizado durante o treinamento são fatores determinantes para a generalização dos modelos. Modelos treinados com degradação simplificada (interpolação bicúbica com desfoque gaussiano) apresentaram desempenho elevado apenas em seu domínio específico, com queda acentuada em cenários mais complexos. Modelos treinados com **degradação robusta de múltiplos estágios** (baseada no RealESRGAN) demonstraram capacidade de generalização cruzada superior.

Em termos quantitativos, modelos treinados com dois estágios de degradação (RealESRGAN-VGG) ou um estágio (ESRGAN-VGG/OCR) reconheceram, em testes com três estágios de degradação, até 1,58 caracteres médios contra 0,33 do modelo treinado apenas com Gaussian — uma diferença de 1,25 caracteres (razão de 4,8×). A incorporação de ruído de Poisson, compressão JPEG e desfoques anisotrópicos atuou como regularizador, expandindo o espaço de representação aprendido pela rede e reduzindo o deslocamento de domínio.

Contudo, os resultados evidenciam uma relação **não-monotônica** entre complexidade

de degradação e generalização: estágios de degradação 3 e 4 causaram queda de desempenho em todas as arquiteturas. Para a RealESRGAN-VGG, o modelo com 4 estágios atingiu 2,12 caracteres médios (média dos 5 domínios de teste) contra 2,44 do modelo com 2 estágios, e 0,5% de correspondência completa contra 0,7% — conforme visível nos mapas de calor apresentados no [Capítulo 5](#). Esse fenômeno indica que degradação excessiva destrói as estruturas semânticas que o modelo deve aprender a reconstruir, tornando o mapeamento LR→HR fundamentalmente mal-posto.

6.1.2 Hipótese 2: Perda Perceptual Baseada em OCR

A investigação sobre a substituição do extrator de características VGG-19 por um *backbone* especializado em reconhecimento óptico de caracteres (OCR) na função de perda perceptual revelou um *trade-off* dependente da arquitetura. Os resultados indicam que a eficácia dessa estratégia não é universal, mas contingente a três fatores:

1. Arquitetura do modelo SR:

- **ESRGAN**: OCR Loss proporcionou ganho de 10,7 pontos percentuais (18,1% vs. 7,4% de correspondência completa) quando treinada com um estágio de degradação, correspondendo a uma diferença de 0,98 caracteres médios (3,89 vs. 2,91) nos testes Gaussian;
- **RealESRGAN**: VGG Loss superou OCR Loss com dois estágios de degradação, produzindo 3,56 vs. 1,22 caracteres médios — diferença de 2,34 caracteres —, indicando que a arquitetura já captura características de alta frequência de forma implícita;
- **LPR-RSR-EXT**: a arquitetura especializada em texto apresentou falha completa (0,0%) com VGG Loss, mas convergiu com OCR Loss, atingindo 1,86 caracteres médios no melhor caso, demonstrando dependência fundamental de uma perda consciente de caracteres.

2. **Complexidade da degradação**: OCR Loss apresenta um **limiar crítico** entre o nível 1 e o nível 2 de degradação. ESRGAN-OCR alcançou 18,1% de correspondência completa com um estágio de degradação, mas caiu para 0,0% com dois ou mais estágios. Esse colapso indica que estruturas de caracteres corrompidas além desse limiar geram gradientes ambíguos que desestabilizam o treinamento.

3. **Interação degradação × arquitetura × perda**: os três fatores acima não são independentes. A configuração ótima varia: para ESRGAN, OCR Loss com um estágio é a melhor combinação; para RealESRGAN, VGG Loss com dois estágios é o equilíbrio superior; para LPR-RSR-EXT, OCR Loss é condição necessária para convergência, independentemente do protocolo de degradação.

A hipótese foi **parcialmente validada**: OCR Loss não é universalmente superior, mas oferece vantagens em configurações específicas (ESRGAN com um estágio de degradação: +10,7 pontos percentuais de correspondência completa), ao custo de instabilidade em degradações mais complexas.

6.2 Considerações Finais

Esta dissertação contribui para o campo ao deslocar o foco do design de arquiteturas de SR para a **qualidade e realismo dos dados de treinamento sintético** como variável primária para a generalização em ALPR. Os resultados indicam que, para que a super-resolução cumpra efetivamente seu papel em sistemas de visão computacional aplicados a ambientes reais, os modelos devem ser treinados com protocolos de degradação que reflitam a complexidade das condições de captura encontradas na prática.

Como trabalho futuro, será investigado o comportamento dos melhores modelos de super-resolução identificados neste estudo considerando como entrada imagens reais de baixa resolução submetidas a um processo de degradação real — como captura por câmera de baixa qualidade, compressão realizada pelo software da câmera na conversão de imagem bruta para JPEG, e transmissão via aplicativos de mensagem.

Referências

ABDELAZIZ, A. H.; CHAN, Y. K.; KOO, V. C. Enhancement for license plate recognition using image super resolution technique. In: IEEE. *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. [S.l.], 2021. p. 1–4. Citado 5 vezes nas páginas 16, 18, 27, 29 e 37.

ASAAD, A.; FAIZABADI, A. R.; ZAKI, H. F. M. Synthetic license plate generation: A novel approach for effective license plate recognition in malaysia. In: IEEE. *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. [S.l.], 2023. p. 1–8. Citado na página 12.

BLAU, Y. et al. The 2018 pirm challenge on perceptual image super-resolution. In: *Proceedings of the European conference on computer vision (ECCV) workshops*. [S.l.: s.n.], 2018. p. 0–0. Citado na página 19.

BOBY, A. et al. Investigating the effects of image correction through affine transformations on licence plate recognition. In: IEEE. *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. [S.l.], 2022. p. 1–6. Citado 2 vezes nas páginas 27 e 29.

BOBY, A. et al. Enabling vehicle search through robust licence plate detection. In: IEEE. *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. [S.l.], 2023. p. 1–7. Citado 3 vezes nas páginas 11, 27 e 30.

CABRAL, J. P. et al. An efficient cnn-based approach for automatic brazilian license plate recognition. In: IEEE. *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. [S.l.], 2021. p. 1891–1894. Citado na página 12.

DONG, C. et al. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 38, n. 2, p. 295–307, 2015. Citado na página 18.

DU, S. et al. Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Transactions on circuits and systems for video technology*, IEEE, v. 23, n. 2, p. 311–325, 2012. Citado na página 11.

GOODFELLOW, I. et al. Generative adversarial nets. *Advances in neural information processing systems*, v. 27, 2014. Citado na página 18.

GUPTA, M. et al. Automatic number plate detection for law enforcement using optical recognition system. *Fusion: Practice and Applications*, p. 50–56, 01 2021. Citado na página 10.

HAMDI, A.; CHAN, Y. K.; KOO, V. C. A new image enhancement and super resolution technique for license plate recognition. *Heliyon*, Elsevier, v. 7, n. 11, 2021. Citado 5 vezes nas páginas 16, 17, 18, 27 e 28.

HAQUE, N. et al. Automatic bangla license plate recognition system for low-resolution images. In: IEEE. *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*. [S.l.], 2022. p. 1–6. Citado 2 vezes nas páginas 27 e 29.

- KABIRAJ, A. et al. Number plate recognition from enhanced super-resolution using generative adversarial network. *Multimedia Tools and Applications*, Springer, v. 82, n. 9, p. 13837–13853, 2023. Citado 3 vezes nas páginas 18, 27 e 30.
- KIM, D.; KIM, J.; PARK, E. Afa-net: Adaptive feature attention network in image deblurring and super-resolution for improving license plate recognition. *Computer Vision and Image Understanding*, Elsevier, v. 238, p. 103879, 2024. Citado 3 vezes nas páginas 27, 28 e 32.
- LAROCA, R. et al. A robust real-time automatic license plate recognition based on the yolo detector. In: IEEE. *2018 international joint conference on neural networks (ijcnn)*. [S.l.], 2018. p. 1–10. Citado na página 11.
- LEDIG, C. et al. Photo-realistic single image super-resolution using a generative adversarial network [c]. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. v. 2017, p. 4681–4690. Citado 2 vezes nas páginas 18 e 19.
- LU, Y.; GU, Y.; WANG, B. License plate recognition in wild with super-resolution. In: IEEE. *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE)*. [S.l.], 2023. p. 523–526. Citado 2 vezes nas páginas 27 e 32.
- NASCIMENTO, V. et al. Combining attention module and pixel shuffle for license plate super-resolution. In: IEEE. *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2022. v. 1, p. 228–233. Citado 6 vezes nas páginas 16, 17, 27, 28, 31 e 32.
- NASCIMENTO, V. et al. Super-resolution of license plate images using attention modules and sub-pixel convolution layers. *Computers & Graphics*, Elsevier, v. 113, p. 69–76, 2023. Citado 4 vezes nas páginas 16, 27, 31 e 32.
- PAN, S.; CHEN, S.-B.; LUO, B. A super-resolution-based license plate recognition method for remote surveillance. *Journal of Visual Communication and Image Representation*, Elsevier, v. 94, p. 103844, 2023. Citado 3 vezes nas páginas 11, 27 e 30.
- PAN, Y.; TANG, J.; TIAHJADI, T. Lpsrgan: Generative adversarial networks for super-resolution of license plate image. *Neurocomputing*, Elsevier, p. 127426, 2024. Citado 3 vezes nas páginas 27, 28 e 32.
- PATTANAIK, A.; BALABANTARAY, R. C. Enhancement of license plate recognition performance using xception with mish activation function. *Multimedia tools and applications*, Springer, v. 82, n. 11, p. 16793–16815, 2023. Citado 2 vezes nas páginas 27 e 30.
- POURHADI, N. et al. Yolov5-based alpr improvement using selective sr-gan. In: IEEE. *2022 2nd International Conference on Computing and Machine Intelligence (ICMI)*. [S.l.], 2022. p. 1–6. Citado 6 vezes nas páginas 12, 13, 16, 17, 27 e 29.
- QIN, L. *Online machine learning methods for visual tracking*. Tese (Doutorado) — Troyes, 2014. Citado na página 11.
- RAO, Z. et al. License plate recognition system in unconstrained scenes via a new image correction scheme and improved crnn. *Expert Systems with Applications*, Elsevier, v. 243, p. 122878, 2024. Citado 3 vezes nas páginas 10, 11 e 12.
- SAPUTRA, V. W.; SUCIATI, N.; FATICHAH, C. Low light image enhancement in license plate recognition using uretinex-net and trba. *Procedia Computer Science*, Elsevier, v. 234, p. 404–411, 2024. Citado na página 12.

- SEREETHAVEKUL, W.; EKPANYAPONG, M. Adaptive lightweight license plate image recovery using deep learning based on generative adversarial network. *IEEE Access*, IEEE, v. 11, p. 26667–26685, 2023. Citado na página 19.
- SHI, M. et al. Bcsr: toward arbitrarily oriented text image super-resolution via adaptive bezier curve network. *EURASIP Journal on Advances in Signal Processing*, Springer, v. 2023, n. 1, p. 82, 2023. Citado 3 vezes nas páginas 27, 28 e 31.
- SHOMEE, H. H.; SAMS, A. License plate detection and recognition system for all types of bangladeshi vehicles using multi-step deep learning model. In: IEEE. *2021 Digital Image Computing: Techniques and Applications (DICTA)*. [S.l.], 2021. p. 01–07. Citado 2 vezes nas páginas 27 e 28.
- SILVA, S. M.; JUNG, C. R. License plate detection and recognition in unconstrained scenarios. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 580–596. Citado na página 12.
- TIAN, C. et al. Generative adversarial networks for image super-resolution: A survey. *arXiv preprint arXiv:2204.13620*, 2022. Citado na página 19.
- WANG, H. et al. Robust korean license plate recognition based on deep neural networks. *Sensors*, MDPI, v. 21, n. 12, p. 4140, 2021. Citado 3 vezes nas páginas 16, 27 e 28.
- WANG, X. et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 1905–1914. Citado 2 vezes nas páginas 18 e 22.
- WANG, X. et al. Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European conference on computer vision (ECCV) workshops*. [S.l.: s.n.], 2018. p. 0–0. Citado na página 18.
- XIN, Z. et al. A license plate super-resolution algorithm based on dictionary learning and neural network. In: IEEE. *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*. [S.l.], 2021. p. 151–154. Citado 3 vezes nas páginas 17, 27 e 29.
- YANG, S. et al. Pragmatic degradation learning for scene text image super-resolution with data-training strategy. *Knowledge-Based Systems*, Elsevier, v. 285, p. 111349, 2024. Citado 4 vezes nas páginas 19, 27, 28 e 33.
- ZAMIR, S. W. et al. Multi-stage progressive image restoration. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2021. p. 14821–14831. Citado na página 12.
- ZHANG, K. et al. Single image super-resolution with non-local means and steering kernel regression. *IEEE Transactions on Image Processing*, IEEE, v. 21, n. 11, p. 4544–4556, 2012. Citado na página 17.
- ZHANG, R. et al. A deep learning-based pipeline system for license plate detection and recognition using yolov7 in complex traffic scenes. In: SPIE. *2022 2nd Conference on High Performance Computing and Communication Engineering (HPCCE 2022)*. [S.l.], 2023. v. 12605, p. 18–22. Citado 2 vezes nas páginas 27 e 30.

ZHANG, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 586–595. Citado na página [21](#).