

**USO DE INTELIGÊNCIA ARTIFICIAL E
AEROFOTOGRAMETRIA PARA IDENTIFICAÇÃO DE
FISSURAS EM CICLOVIAS E INTEGRAÇÃO COM GIS**

THARLYS HIKARO PINHEIRO SILVA

**DISSERTAÇÃO DE MESTRADO EM ESTRUTURAS E CONSTRUÇÃO CIVIL
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL

**USO DE INTELIGÊNCIA ARTIFICIAL E AEROFOTOGRAMETRIA
PARA IDENTIFICAÇÃO DE FISSURAS EM CICLOVIAS E
INTEGRAÇÃO COM GIS**

THARLYS HIKARO PINHEIRO SILVA

ORIENTADORA: DSc. MICHELE TEREZA MARQUES CARVALHO

**DISSERTAÇÃO DE MESTRADO EM ESTRUTURAS E CONSTRUÇÃO
CIVIL**

BRASÍLIA/DF: JANEIRO – 2026

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL

**USO DE INTELIGÊNCIA ARTIFICIAL E AEROFOTOGRAMETRIA PARA
IDENTIFICAÇÃO DE FISSURAS EM CICLOVIAS E INTEGRAÇÃO COM GIS**

THARLYS HIKARO PINHEIRO SILVA

DISSERTAÇÃO DE MESTRADO
SUBMETIDA AO DEPARTAMENTO DE
ENGENHARIA CIVIL E AMBIENTAL DA
FACULDADE DE TECNOLOGIA DA
UNIVERSIDADE DE BRASÍLIA COMO
PARTE DOS REQUISITOS NECESSÁRIOS
PARA A OBTENÇÃO DO GRAU DE
MESTRE EM ESTRUTURAS E
CONSTRUÇÃO CIVIL.

APROVADO POR:

Prof.^a Michele Tereza Marques Carvalho, DSc. (UnB)
(Orientadora)

Prof. André Luís Brasil Cavalcante, DSc. (UnB)
(Examinador Interno)

Prof.^a Dayana Bastos Costa, DSc. (UFBA)
(Examinador Externo)

BRASÍLIA/DF – JANEIRO/2026.

Dedico este trabalho à *minha esposa e às minhas filhas*, que foram *meu ponto de referência* quando tudo parecia *ruído e incerteza*.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por sua presença constante e por me conceder saúde, força e sabedoria ao longo de toda esta caminhada.

Aos professores que tive a honra de conhecer e com quem pude aprender, deixo meu agradecimento, pois cada um contribuiu, de alguma forma, para o sucesso desta jornada. Em especial, ao Professor Lenildo, pela disponibilização de equipamentos essenciais ao estudo e pelo compartilhamento de conhecimento, e à minha orientadora, Professora Michele, pelo apoio, orientação, compreensão e confiança que tornaram este trabalho possível.

Aos colegas e amigos que encontrei em Brasília e no PECC, meu muito obrigado por tornarem minha estadia na cidade e na pós-graduação mais leve e divertida, mesmo nos momentos difíceis. Em especial, agradeço ao Paulo Santos, pela amizade, pela disponibilidade constante em ajudar e por compartilhar conhecimentos fundamentais ao longo desta jornada e ao Valmor Filho, por me ensinar tanto sobre drones e levantamentos, e por me acompanhar e orientar em todos eles, sob sol ou chuva.

À minha família, pelo incentivo em todos os momentos, em especial à minha mãe, ao meu pai, ao meu irmão e à minha irmã, por todo apoio, compreensão e presença ao longo desta jornada. E, sobretudo, à minha esposa, Hellen Rodrigues, que, com inteligência emocional e uma resistência superior à do concreto, segurou todas as pontas quando eu mais precisei, e às minhas filhas, por serem minha motivação diária.

A todos que contribuíram, direta ou indiretamente, para esta caminhada, registro minha sincera gratidão.

RESUMO

USO DE INTELIGÊNCIA ARTIFICIAL E AEROFOTOGRAMETRIA PARA IDENTIFICAÇÃO DE FISSURAS EM CICLOVIAS E INTEGRAÇÃO COM GIS

Autor: Tharlys Hikaro Pinheiro Silva

Orientadora: Dr^a. Michele Tereza Marques Carvalho

Programa de Pós-graduação em Estruturas e Construção Civil

Brasília, janeiro de 2026.

A inspeção e o monitoramento de ciclovias são essenciais para a segurança e o conforto dos usuários, porém ainda dependem, em grande parte, de vistorias manuais, com limitações de produtividade, rastreabilidade e padronização. Neste trabalho, foi proposto e validado um fluxo integrado para identificação automatizada de fissuras em ciclovias de pavimento rígido, combinando aerofotogrametria por VANT, modelos de aprendizado profundo e integração dos resultados em ambiente GIS, de modo a transformar saídas de detecção/segmentação em informações georreferenciadas para apoio à gestão da manutenção. Como contribuição científica, o estudo (i) constitui um banco de dados anotado específico para detecção e segmentação de fissuras em ciclovias de pavimento rígido, (ii) apresenta evidências de que modelos de alto desempenho são viáveis e robustos mesmo com conjuntos de dados limitados, desde que haja alta especificidade, (iii) sistematiza e valida um fluxo fim a fim VANT–IA–GIS aplicado à infraestrutura cicloviária e (iv) demonstra a conversão das inferências em camadas vetoriais georreferenciadas, discutindo limitações e ajustes necessários na vetorização das máscaras para preservação da morfologia das fissuras. Foi constituído um banco de dados para detecção com 738 imagens recortadas (1024×1024), totalizando 1.757 anotações de fissuras; após aumento de dados, obteve-se um conjunto consolidado com 3.834 imagens. Quatro arquiteturas foram treinadas e comparadas (YOLOv8m, YOLO11m, YOLO11e e RT-DETR-l), alcançando mAP_{50} entre 0,797 e 0,831, com melhor desempenho do YOLO11l ($P=0,887$; $R=0,737$; $mAP_{50}=0,831$; 12,2 ms/imagem). Para segmentação, foi construído um *dataset* com 413 imagens (640×640) e 895 máscaras anotadas, e três modelos foram avaliados (YOLOv8m-seg, YOLO11m-seg e YOLO11l-seg), com destaque para o YOLO11l-seg ($P=0,846$; $R=0,865$; $mAP_{50}=0,872$; 5,6 ms/imagem). A validação externa foi realizada em ortomosaico georreferenciado, com inferência no QGIS via *plugin Deepness* após conversão para ONNX, evidenciando coerência espacial das detecções e a necessidade de ajustes na vetorização das máscaras para preservar a morfologia das fissuras. Como produto, obtiveram-se camadas vetoriais georreferenciadas aptas a visualização, quantificação e análise espacial, consolidando a viabilidade do método para mapeamento automatizado de patologias em infraestrutura cicloviária e fornecendo uma base para integrações futuras com rotinas de priorização e sistemas de apoio à decisão.

Palavras-chave: Inteligência Artificial; VANT; Gestão da Manutenção; GIS; Ciclovias.

ABSTRACT

USE OF ARTIFICIAL INTELLIGENCE AND AEROPHOTOGRAMMETRY FOR CRACK IDENTIFICATION IN BIKE LANES AND INTEGRATION WITH GIS

Author: Tharlys Hikaro Pinheiro Silva

Supervisor: Michele Tereza Marques Carvalho

Postgraduate Program in Structures and Civil Construction

Brasília, January 2026.

Bike lane inspection and monitoring are essential for user safety and comfort; however, they still rely largely on manual surveys, with limitations in productivity, traceability, and standardization. This study proposes and validates an integrated workflow for automated crack identification in rigid concrete bike lanes, combining UAV-based aerophotogrammetry, deep learning models, and the integration of outputs in a GIS environment, thereby transforming detection/segmentation results into georeferenced information to support maintenance management. As a scientific contribution, this work (i) establishes a dedicated annotated dataset for crack detection and segmentation in rigid bike-lane pavements, (ii) provides evidence that high-performance and robust models are feasible even with limited datasets, provided that high specificity is ensured, (iii) systematizes and validates an end-to-end UAV–AI–GIS workflow applied to cycling infrastructure, and (iv) demonstrates the conversion of model inferences into georeferenced vector layers, discussing limitations and the adjustments required in mask vectorization to preserve crack morphology. A detection dataset was built with 738 cropped images (1024×1024), totaling 1,757 crack annotations; after data augmentation, a consolidated set of 3,834 images was obtained. Four architectures were trained and compared (YOLOv8m, YOLO11m, YOLO11l, and RT-DETR-l), achieving mAP_{50} values between 0.797 and 0.831, with the best performance obtained by YOLO11l ($P=0.887$; $R=0.737$; $mAP_{50}=0.831$; 12.2 ms/image). For segmentation, a dataset with 413 images (640×640) and 895 annotated masks was created, and three models were evaluated (YOLOv8m-seg, YOLO11m-seg, and YOLO11l-seg), with YOLO11l-seg standing out ($P=0.846$; $R=0.865$; $mAP_{50}=0.872$; 5.6 ms/image). External validation was conducted on a georeferenced orthomosaic, with inference performed in QGIS using the Deepness plugin after ONNX conversion, evidencing spatial coherence of the detections and the need for adjustments in mask vectorization to preserve crack morphology. As an output, georeferenced vector layers suitable for visualization, quantification, and spatial analysis were produced, confirming the feasibility of the method for automated mapping of pathologies in cycling infrastructure and providing a basis for future integration with prioritization routines and decision-support systems.

Keywords: Artificial Intelligence; UAV; Maintenance Management; GIS; Bike Lanes.

LISTA DE FIGURAS

Figura 1 – Malha Ciclovária do Distrito Federal.....	14
Figura 2 – Metodologia proposta.....	34
Figura 3 – Fluxo metodológico adotado para o desenvolvimento e treinamento dos modelos de detecção de fissuras em ciclovias	34
Figura 4 – Exemplos de aplicação do protocolo de anotação das fissuras em ciclovias de pavimento rígido (detecção): (a) fissura dividida conforme os limites das placas de concreto; (b) fissuras sem conexão física aparente rotuladas individualmente; (c) ramificações conectadas visualmente são agrupadas em uma única <i>bounding box</i>	37
Figura 5 – Exemplos de transformações aplicadas no processo de <i>data augmentation</i> das imagens de ciclovias de pavimento rígido.....	39
Figura 6 – Fluxo metodológico adotado para o desenvolvimento e avaliação dos modelos de segmentação de fissuras em ciclovias.....	44
Figura 7 – Exemplos de aplicação do protocolo de anotação das fissuras em ciclovias de pavimento rígido (segmentação): (a) fissura segmentada conforme os limites das placas de concreto; (b) fissuras sem conexão física aparente rotuladas individualmente; (c) ramificações conectadas visualmente são agrupadas em uma única máscara.	47
Figura 8 – Fluxo metodológico adotado para o desenvolvimento do estudo de caso e integração IA-GIS.....	52
Figura 9 – Processo de extração de ROIs a partir da imagem original.....	58
Figura 10 – Histograma da distribuição de anotações por imagem no conjunto de detecção ..	60
Figura 11 – Curvas de aprendizado de detecção no conjunto de validação: (a) Perda de validação (modelos YOLO), (b) Perda de validação (RT-DETR), (c) Evolução do mAP ₅₀	61
Figura 12 – Detecção de fissuras em ciclovias usando YOLOv8m, YOLO11m, YOLO11l e RT-DETR-l sobre a influência de fatores como desfoque, alta iluminação, sombra e ruído.	68
Figura 13 – Histograma da distribuição de anotações por imagem no conjunto de segmentação	74
Figura 14 – Curvas de aprendizado de segmentação no conjunto de validação: (a) Perda de validação (seg_loss), (b) Evolução do mAP ₅₀	75

Figura 15 – Avaliação qualitativa da segmentação de fissuras em ciclovias de pavimento rígido: (a) imagem de entrada (ROI), (b) máscara de referência (ground truth), (c) máscara predita pelo modelo YOLOv8m-seg, (d) máscara predita pelo modelo YOLO11m-seg e (e) máscara predita pelo modelo YOLO11l-seg.....	81
Figura 16 – Ortomosaico georreferenciado da ciclovia: (a) Visão geral do trecho mapeado; (b) Detalhe da resolução em nível do pavimento	86
Figura 17 – Ambiente de execução das inferências no QGIS por meio do <i>plugin Deepness</i> ..	88
Figura 18 - Visão geral do ortomosaico após inferências.....	88
Figura 19 – Resultado da vetorização inicial da segmentação, evidenciando a simplificação geométrica das fissuras.	89
Figura 20 – Resultado da vetorização após ajuste, com preservação do contorno real das fissuras segmentadas.....	90
Figura 21 – Validação dos resultados espaciais: (a) Trechos do ortomosaico; (b) Camada de detecção (YOLO11l); (c) Camada de segmentação (YOLO11l-seg).....	91
Figura 22 – Resposta dos modelos de detecção e segmentação em superfícies pavimentadas fora do domínio de treinamento.....	94

LISTA DE TABELAS

Tabela 1 – Características das coletas de imagens por VANT utilizadas no estudo	35
Tabela 2 – Modelos utilizados para detecção	40
Tabela 3 – Configuração dos hiperparâmetros adotados no treinamento dos modelos de detecção	41
Tabela 4 – Modelos utilizados para segmentação	48
Tabela 5 – Configuração dos hiperparâmetros adotados no treinamento dos modelos de segmentação	49
Tabela 6 – Especificações do ambiente de <i>software</i> utilizado na integração IA-GIS	54
Tabela 7 – Distribuição de imagens e anotações entre os conjuntos de dados de detecção	59
Tabela 8 – Desempenho dos modelos de detecção	62
Tabela 9 – Comparação dos resultados deste estudo com trabalhos relacionados na detecção automática de fissuras	64
Tabela 10 – Distribuição de imagens e anotações entre os conjuntos de dados de segmentação	73
Tabela 11 – Desempenho dos modelos de segmentação	77

LISTA DE ABREVIATURAS E SIGLAS

AP	Precisão Média (<i>Average Precision</i>)
CNN	Redes Neurais Convolucionais (<i>Convolutional Neural Networks</i>)
DETR	<i>DEtection TRansformer</i>
DF	Distrito Federal
DL	Aprendizado Profundo (<i>Deep Learning</i>)
FN	Falsos Negativos (<i>False Negative</i>)
FP	Falsos Positivos (<i>False Positive</i>)
GIS	Sistemas de Informação Geográfica (<i>Geographic Information System</i>)
GPU	Unidade de Processamento Gráfico (<i>Graphics Processing Unit</i>)
GSD	<i>Ground Sample Distance</i>
IA	Inteligência Artificial
IoU	Interseção sobre a União (<i>Intersection over Union</i>)
mAP	<i>mean Average Precision</i>
ML	Aprendizado de Máquina (<i>Machine Learning</i>)
O&M	Operação e Manutenção
ODS	Objetivos de Desenvolvimento Sustentável
ONNX	<i>Open Neural Network Exchange</i>
PR	<i>Precision-Recall</i>
ROI	Região de Interesse (<i>Region of Interest</i>)
RT-DETR	<i>Real-Time Detection Transformer</i>
RTK	Posicionamento Cinemático em Tempo Real (<i>Real Time Kinematic</i>)
TP	<i>True Positive</i> (Verdadeiros Positivos)
VANT	Veículo Aéreo Não Tripulado
YOLO	<i>You Only Look Once</i>

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	CONTEXTUALIZAÇÃO	13
1.2	MOTIVAÇÃO DA PESQUISA	15
1.3	OBJETIVOS	16
	1.3.1 Objetivo Geral	16
	1.3.2 Objetivos Específicos	17
1.4	ESTRUTURA DO TRABALHO	17
2	REVISÃO DA LITERATURA.....	18
2.1	INFRAESTRUTURA CICLOVIÁRIA, PAVIMENTO RÍGIDO E FISSURAS	18
2.2	AEROFOTOGRAMETRIA E USO DE VANT NA ENGENHARIA CIVIL.....	22
2.3	INTELIGÊNCIA ARTIFICIAL E VISÃO COMPUTACIONAL	24
2.4	INTEGRAÇÃO IA-GIS	30
3	METODOLOGIA.....	33
3.1	SEÇÃO 1 – DESENVOLVIMENTO E TREINAMENTO DE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA DETECÇÃO DE FISSURAS EM CICLOVIAS..	34
	3.1.1 Coleta de dados – Etapa 1	35
	3.1.2 Preparação do conjunto de dados – Etapa 2.....	36
	3.1.3 Implementação e treinamento dos modelos de detecção – Etapa 3	39
	3.1.4 Avaliação de desempenho dos modelos de detecção – Etapa 4	42
3.2	SEÇÃO 2 – DESENVOLVIMENTO E TREINAMENTO DE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA SEGMENTAÇÃO DE FISSURAS	43
	3.2.1 Coleta de dados – Etapa 1	45
	3.2.2 Preparação do conjunto de dados – Etapa 2.....	46
	3.2.3 Implementação e treinamento dos modelos de segmentação – Etapa 3	48
	3.2.4 Avaliação de desempenho dos modelos de segmentação – Etapa 4	49
3.3	SEÇÃO 3 – ESTUDO DE CASO E INTEGRAÇÃO IA-GIS.....	50
	3.3.1 Levantamento aerofotogramétrico e geração do ortomosaico – Etapa 1	52
	3.3.2 Integração dos modelos de inteligência artificial ao ambiente GIS	53
	3.3.3 Procedimento de validação qualitativa e organização dos resultados	55
4	RESULTADOS E DISCUSSÕES.....	57

4.1	DESEMPENHO DOS MODELOS DE DETECÇÃO DE FISSURAS EM CICLOVIAS (SEÇÃO 1)	57
4.1.1	Caracterização do conjunto de dados para detecção	57
4.1.2	Treinamento dos modelos de detecção	60
4.1.3	Avaliação de desempenho dos modelos de detecção	62
4.1.4	Avaliação qualitativa das inferências dos modelos de detecção	67
4.2	DESEMPENHO DOS MODELOS DE SEGMENTAÇÃO DE FISSURAS EM CICLOVIAS (SEÇÃO 2)	72
4.2.1	Caracterização do conjunto de dados para segmentação	73
4.2.2	Treinamento dos modelos de segmentação	74
4.2.3	Avaliação de desempenho dos modelos de segmentação	76
4.2.4	Avaliação qualitativa das inferências dos modelos de segmentação	79
4.3	ESTUDO DE CASO E INTEGRAÇÃO IA-GIS (SEÇÃO 3)	85
4.3.1	Caracterização do estudo de caso e ortomosaico gerado	85
4.3.2	Integração dos modelos de IA ao ambiente GIS	87
4.3.3	Resultados espaciais e validação qualitativa integrada	90
4.3.4	Produto e aplicabilidade em inspeção e manutenção	95
5	CONCLUSÕES.....	96
	REFERÊNCIAS BIBLIOGRÁFICAS.....	101

1 INTRODUÇÃO

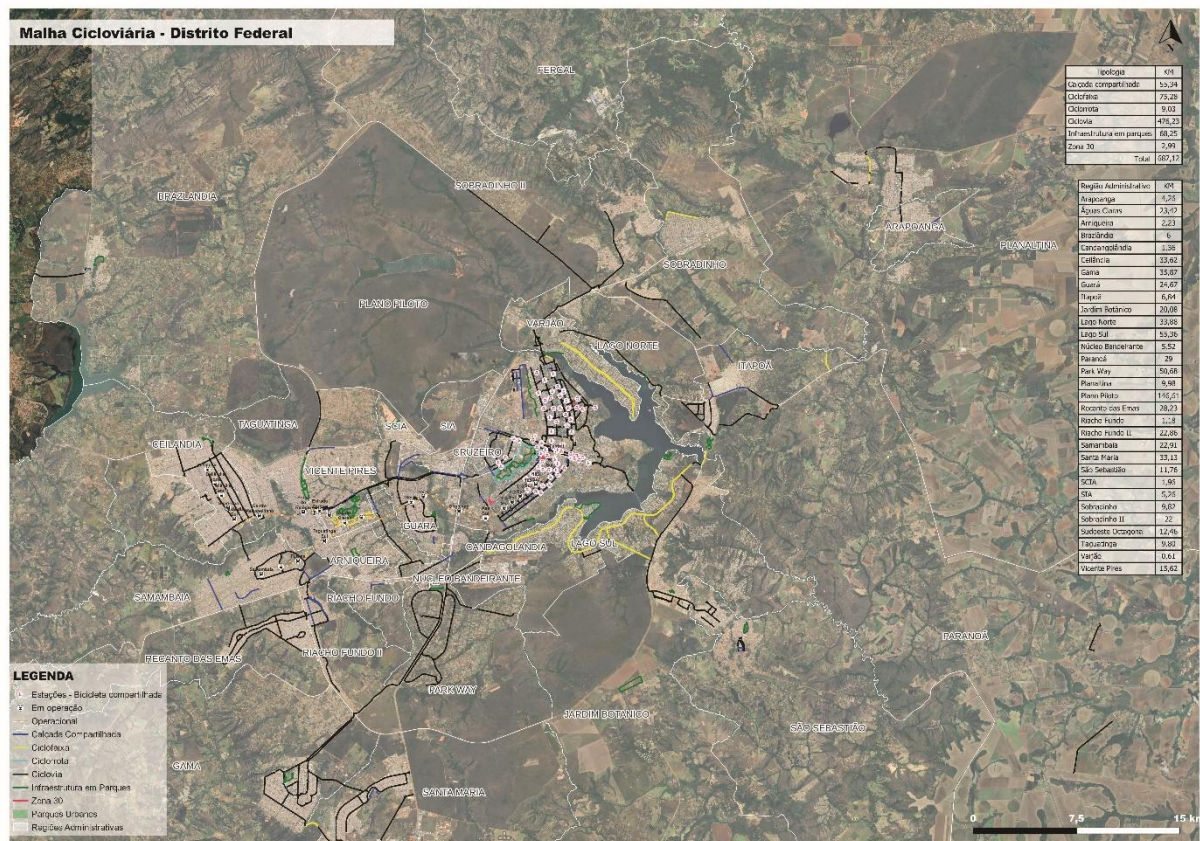
Essa dissertação de mestrado foi desenvolvida no Programa de Pós-Graduação em Estruturas e Construção Civil (PECC) da Universidade de Brasília (UnB) sob a orientação da Professora Dra. Michele Tereza Marques Carvalho, se tratando de uma pesquisa na linha de Gestão e sustentabilidade na construção civil, com área de foco em inteligência artificial e aprendizado de máquina com integração ao Sistema de Informação Geográfica (GIS).

1.1 CONTEXTUALIZAÇÃO

A infraestrutura viária desempenha um papel essencial na mobilidade urbana, sendo a qualidade dos espaços para circulação de bicicletas um elemento fundamental para garantir segurança e conforto aos ciclistas. Essas infraestruturas têm sido amplamente aceitas e utilizadas nas cidades ao redor do mundo devido à sua contribuição significativa em três pilares fundamentais da sociedade: a preservação ambiental, a melhoria da mobilidade urbana e o fortalecimento da saúde pública (Holm, Glümer, Diderichsen, 2012; ITF, 2013; Götschi; Garrard; Giles-Corti, 2016), alinhando-se a metas dos Objetivos de Desenvolvimento Sustentável (ODS), como saúde e bem-estar, cidades e comunidades sustentáveis e ação contra a mudança global do clima (Organização das Nações Unidas, 2015). Além disso, o uso da bicicleta nos deslocamentos complementares consolida-se como uma estratégia eficaz para a redução das emissões de gases de efeito estufa (Motta, 2016).

No cenário nacional, o Distrito Federal (DF) destaca-se por possuir uma das maiores malhas cicloviárias do Brasil. De acordo com a Secretaria de Transporte e Mobilidade (SEMOB, 2024), a rede conta com uma extensão total de 711,4 km, dos quais 494,5 km são compostos por ciclovias (Figura 1). Uma parcela significativa dessa infraestrutura é constituída por pavimentos rígidos de concreto, tipologia que, apesar da elevada durabilidade, está sujeita a degradações específicas ao longo de sua vida útil exigindo uma gestão para sua manutenção.

Figura 1 – Malha Ciclovária do Distrito Federal



Fonte: SEMOB (2024)

Nesse contexto, as fissuras figuram como uma das principais manifestações patológicas em pavimentos de concreto. Embora muitas vezes iniciem como defeitos superficiais, elas representam uma ameaça à integridade estrutural e à segurança. A detecção precoce e precisa dessas anomalias é indispensável para viabilizar intervenções de manutenção preventivas, evitando a propagação de danos que podem levar a falhas ou à intervenções corretivas mais extensas (Oliveira; Correia, 2013; Hsieh; Tsai, 2020; Fang *et al.*, 2020; Ai *et al.*, 2023; Deng *et al.*, 2022).

Entretanto, a gestão da manutenção dessas extensas malhas viárias impõe desafios significativos. Os métodos tradicionais de inspeção, predominantemente visuais e manuais, tendem a ser onerosos, lentos e sujeitos à subjetividade humana (Daulat *et al.*, 2022; Sandu; Varganova; Samii, 2022; West *et al.*, 2024). Diante disso, o uso de veículos aéreos não tripulados (VANT) equipados com câmeras de alta resolução, aliado a técnicas de Inteligência Artificial (IA), surge como uma solução. Estudos recentes indicam que o processamento digital de imagens obtidas por aerofotogrametria permite a identificação automática de patologias, conferindo agilidade e objetividade ao diagnóstico (Omoebamije *et al.*, 2023).

A convergência entre tecnologias emergentes, aerofotogrametria, IA e GIS, oferece, portanto, uma nova abordagem para a gestão da infraestrutura. Enquanto a aerofotogrametria fornece a base de dados visual detalhada e a IA encarrega-se da detecção automatizada das fissuras, o GIS confere o contexto espacial essencial a essas informações, possibilitando organizar as ocorrências em camadas geográficas e bases de dados para consulta, análise e priorização de intervenções. A relevância dessa integração reside na capacidade de não apenas detectar o dano, mas de espacializar seu padrão de ocorrência sobre a malha cicloviária, viabilizando tomadas de decisão baseadas em dados georreferenciados e otimizando o ciclo de vida do ativo.

1.2 MOTIVAÇÃO DA PESQUISA

No contexto das ciclovias, observa-se a carência de diretrizes e metodologias específicas para a identificação, o monitoramento e a manutenção de fissuras em pavimentos rígidos, especialmente no que se refere à aplicação de tecnologias emergentes em escala de rede. Embora a utilização conjunta de inteligência artificial e aerofotogrametria já venha sendo explorada para a inspeção de infraestruturas correlatas, como pavimentos rodoviários (Sierra *et al.*, 2022; Tang *et al.*, 2023) e calçadas (Qiu; Lau, 2023), sua aplicação direcionada a ciclovias permanece incipiente, apesar da crescente expansão desse tipo de infraestrutura em centros urbanos.

Essa lacuna é relevante porque ciclovias apresentam particularidades que impactam diretamente a aquisição e a interpretação de imagens: geometrias mais estreitas, maior proximidade com elementos do ambiente urbano (meio-fio, sinalização horizontal, vegetação, sombras projetadas, mobiliário urbano e tráfego de pedestres), além de interferências recorrentes associadas à operação cotidiana. Somam-se a isso variações de iluminação ao longo do dia, presença de sujeira e manchas superficiais, remendos localizados e heterogeneidade de textura, fatores que elevam a complexidade do problema e podem reduzir o desempenho de modelos treinados em bases genéricas ou obtidas em contextos distintos (Kang *et al.*, 2020; Qiu; Lau, 2023). Assim, torna-se necessário investigar estratégias de aquisição, modelagem e validação que sejam representativas das condições reais de ciclovias e que produzam resultados confiáveis para apoiar decisões para gestão da manutenção.

Além do desafio de identificação em si, observa-se uma lacuna ainda mais abrangente na gestão de infraestruturas urbanas: a integração efetiva entre inspeções automatizadas por IA e Sistemas de Informação Geográfica. Metodologias recentes que combinam VANT e aprendizado

profundo, por exemplo, ainda enfrentam limitações decorrentes de integrações restritas com GIS, o que evidencia como a ausência de fluxos padronizados reduz a contextualização espacial dos defeitos e limita o uso gerencial dos resultados (Hu; Yee; Goff, 2024; Tan *et al.*, 2022). Ainda assim, mesmo quando há boa acurácia na identificação, muitas abordagens não avançam até a etapa em que as ocorrências são estruturadas como informação georreferenciada (camadas vetoriais e banco de dados espacial), permitindo consultas, análises de distribuição, quantificação e priorização de trechos críticos.

Do ponto de vista de operação e manutenção, essa integração é estratégica, pois a gestão de redes cicloviárias exige decisões orientadas por dados: definir onde intervir primeiro, estimar extensão/área de dano, planejar recursos e acompanhar a evolução das manifestações ao longo do tempo. Nesse sentido, torna-se necessária a consolidação de referências científicas e metodológicas capazes de apoiar gestores e órgãos públicos na adoção de práticas mais eficientes de monitoramento e manutenção, com maior rastreabilidade e potencial de padronização

Diante desse cenário, este trabalho propõe uma abordagem integrada para a identificação e o mapeamento de fissuras em ciclovias de pavimento rígido. A metodologia combina aerofotogrametria por VANT, modelos de IA para detecção e segmentação e a integração com ambiente GIS, visando transformar as predições em informação georreferenciada utilizável para análise espacial e apoio à tomada de decisão. Com isso, busca-se contribuir para a sustentabilidade, segurança e durabilidade da infraestrutura cicloviária, oferecendo um fluxo de trabalho aplicável ao contexto de gestão e manutenção de ativos urbanos.

1.3 OBJETIVOS

A seguir serão apresentados o objetivo geral e os específicos desta pesquisa.

1.3.1 Objetivo Geral

Desenvolver e avaliar um fluxo de trabalho baseado em aerofotogrametria por VANT e modelos de inteligência artificial para a identificação automatizada de fissuras em ciclovias de pavimento rígido, integrando os resultados a um ambiente GIS para apoiar o mapeamento georreferenciado e a gestão de operação e manutenção da infraestrutura.

1.3.2 Objetivos Específicos

Os objetivos específicos deste trabalho consistem em:

- Disponibilizar e caracterizar um banco de dados anotado de fissuras em ciclovias de pavimento rígido de concreto, obtido por imagens de VANT e rotulagem manual, voltado a suprir a escassez de dados para essa tipologia e viabilizar o desenvolvimento e a avaliação de modelos de IA.
- Quantificar e comparar o desempenho de arquiteturas de inteligência artificial na detecção e segmentação de fissuras em pavimentos rígidos, identificando as configurações com melhor equilíbrio entre acurácia e eficiência computacional para o cenário estudado.
- Verificar a capacidade de generalização dos modelos selecionados em um estudo de caso independente, com dados adquiridos por VANT com posicionamento cinemático em tempo real (RTK) e ortomosaico georreferenciado, avaliando a robustez do método em um cenário distinto daquele utilizado no treinamento.
- Gerar e avaliar, por meio da integração IA–GIS, produtos geoespaciais derivados das inferências (mapas temáticos e camadas vetoriais georreferenciadas) que viabilizem a quantificação e a análise da distribuição espacial dos danos como suporte à gestão da manutenção e à priorização de trechos críticos.

1.4 ESTRUTURA DO TRABALHO

Este trabalho está organizado em cinco capítulos, estruturados de modo a apresentar sequencialmente a fundamentação, o desenvolvimento metodológico, os resultados obtidos e as conclusões da pesquisa:

- Capítulo 1 – Introdução: Apresenta a contextualização do tema, a problemática relacionada à manutenção de ciclovias de pavimento rígido e a motivação para o uso de tecnologias emergentes. Define também os objetivos geral e específicos que nortearam o estudo.
- Capítulo 2 – Revisão da Literatura: Reúne o embasamento teórico necessário para o desenvolvimento da pesquisa, abordando as características da infraestrutura cicloviária e suas manifestações patológicas, o uso de aerofotogrametria com VANT na engenharia

civil, os fundamentos de inteligência artificial e visão computacional (com foco em identificação de manifestações patológicas) e a integração entre IA e GIS.

- Capítulo 3 – Metodologia: Descreve os procedimentos adotados, estando organizado em três seções principais e interdependentes: (1) desenvolvimento e treinamento dos modelos de detecção de fissuras; (2) desenvolvimento e treinamento dos modelos de segmentação; e (3) realização de um estudo de caso prático para validação externa e integração dos modelos ao ambiente GIS para mapeamento automatizado.
- Capítulo 4 – Resultados e Discussões: Apresenta a análise dos dados obtidos, seguindo a mesma estrutura tripartida da metodologia. Discutem-se primeiramente o desempenho quantitativo e qualitativo dos modelos de detecção, seguidos pela avaliação dos modelos de segmentação. Por fim, são apresentados os resultados espaciais do estudo de caso, demonstrando a aplicação prática do fluxo integrado IA-GIS.
- Capítulo 5 – Conclusões: Sintetiza os principais achados da pesquisa em relação aos objetivos propostos, destaca as contribuições acadêmicas e práticas do trabalho e sugere recomendações para trabalhos futuros.

2 REVISÃO DA LITERATURA

A revisão da literatura apresentada neste capítulo consolida os fundamentos e o estado da arte que sustentam o desenvolvimento desta dissertação. Inicialmente, abordam-se as características da infraestrutura cicloviária e do pavimento rígido, com ênfase nas fissuras como principais manifestações patológicas, nas limitações de abordagens predominantemente manuais e nos desafios de sua manutenção. Na sequência, discute-se o uso da aerofotogrametria com VANT na engenharia civil, destacando seu potencial para inspeções remotas. Posteriormente, são apresentados os fundamentos de Inteligência Artificial e Visão Computacional, focando nas arquiteturas de aprendizado profundo aplicadas à identificação de danos. Por fim, o capítulo examina a integração entre esses modelos e o GIS, contextualizando o mapeamento automatizado como ferramenta de apoio à gestão.

2.1 INFRAESTRUTURA CICLOVIÁRIA, PAVIMENTO RÍGIDO E FISSURAS

A infraestrutura cicloviária vem ganhando relevância no contexto da mobilidade urbana por viabilizar deslocamentos ativos com benefícios ambientais e de saúde pública. Contudo, para que o uso da bicicleta seja seguro e confortável, a qualidade funcional do pavimento torna-se

um requisito central: irregularidades, desníveis e, principalmente, fissuras podem reduzir o conforto de rolamento, aumentar a percepção de risco e elevar a probabilidade de acidentes, sobretudo em ciclovias com tráfego recorrente e exposição contínua às ações do ambiente.

Do ponto de vista do projeto e da operação, o pavimento de ciclovias deve atender a requisitos básicos relacionados ao desempenho superficial. Recomenda-se uma superfície regular, impermeável, antiderrapante e, preferencialmente, com aparência agradável, de modo a favorecer o conforto e a segurança do ciclista. Além disso, como a ciclovia frequentemente se desenvolve próxima a áreas de circulação de pedestres, é desejável que haja diferenciação visual (cor, textura ou padrão) entre as faixas, reduzindo conflitos de uso. Em trechos de interferência com acessos veiculares (garagens, estacionamentos e travessias), ainda que ciclovias não sejam dimensionadas para grandes solicitações, é comum a recomendação de reforço local da estrutura (por exemplo, com armação e camada de concreto magro), justamente para mitigar danos associados a carregamentos não previstos em condições normais de operação (BRASIL, 2007).

A escolha do tipo de pavimentação em ciclovias tende a considerar disponibilidade de materiais e mão de obra, custo, rapidez de execução, conforto de rolamento e manutenção. Entre as soluções usuais, destacam-se: pavimentos flexíveis (tipicamente asfálticos), pavimentos rígidos (concreto) e pavimentos em blocos intertravados. Cada alternativa apresenta compromissos distintos. Em geral, pavimentos rígidos oferecem elevada durabilidade e boa estabilidade geométrica; pavimentos flexíveis podem ter menor custo inicial e execução mais rápida; e blocos intertravados favorecem intervenções localizadas e reuso, embora possam demandar maior controle executivo para evitar irregularidades e recalques diferenciais ao longo do tempo (Silva *et al.*, 2023; Petronio; Rodgher; Florian, 2024).

No caso específico do pavimento rígido, trata-se de um sistema caracterizado por elevada rigidez em relação às camadas inferiores, o que permite à placa de concreto distribuir tensões e absorver esforços de maneira distinta do comportamento típico de pavimentos flexíveis (DNIT, 2006). Os pavimentos de concreto se destacam por utilizar concreto como camada de rolamento, podendo ser executados por diferentes técnicas (moldagem *in loco* ou soluções pré-moldadas), com implicações diretas no planejamento, no controle tecnológico, na execução e na manutenção ao longo do ciclo de vida (Balbo, 2009). Para ciclovias, é comum a adoção de concreto simples moldado *in loco*, com argumentos técnicos associados ao conforto de rolamento, à durabilidade e à facilidade de manutenção. Estudos aplicados a ciclovias apontam

que, apesar do custo inicial frequentemente superior ao de soluções flexíveis, o desempenho e a vida útil podem justificar a escolha do pavimento rígido quando se considera o comportamento ao longo do tempo e a necessidade de intervenções (Bastos; Mota, 2013; Pontes Júnior, 2019; Petronio; Rodgher; Florian, 2024).

Sob a ótica de desempenho mecânico e funcional, parâmetros usuais de dosagem e resistência são empregados para garantir integridade e conforto ao ciclista. Em aplicações reportadas para ciclovias, são citados consumos de cimento da ordem de 350 kg/m³ e resistências à compressão em torno de 25 MPa, como valores compatíveis com o desempenho requerido, desde que associados a adequada execução e controle (Pontes Júnior, 2019). Nessa perspectiva, o desempenho do pavimento rígido não depende apenas das propriedades do concreto, mas também do detalhamento e execução de juntas, da condição de suporte (base/sub-base) e do controle de fatores que influenciam retração, fissuração e transferência de carga (Silva Júnior, 2018)

As manifestações patológicas em pavimentos rígidos de ciclovias incluem um conjunto de anomalias: fissuras transversais, longitudinais, diagonais, de canto e por retração plástica, além de quebras localizadas, esborcinamento de juntas, manchas, buracos, alçamento de placas e placas divididas, entre outras. Entre elas, as fissuras são particularmente críticas porque funcionam como “porta de entrada” para água e partículas, favorecendo deterioração progressiva e aumento de custos de manutenção. Além disso, fissuras podem estar associadas tanto a mecanismos estruturais (por exemplo, fadiga por solicitações repetidas e condições de suporte inadequadas) quanto a mecanismos construtivos e de materiais (retração, cura deficiente, variações térmicas, detalhes de juntas e selagem) (Hora; Pereira, 2020). Briceño (2019), por exemplo, discute a relação entre aplicação repetida de cargas e danos por fadiga em pavimentos rígidos, com potencial evolução para fissuração e desintegração da placa.

Evidências em campo reforçam a predominância das fissuras como dano recorrente em ciclovias de concreto e mostram que sua evolução pode ser acelerada por fatores executivos e de uso. No caso de Brasília-DF, a ciclovia da Esplanada dos Ministérios (trecho central e intensamente utilizado) foi investigada em inspeções em diferentes momentos. Um estudo prévio (2017) reportou o trecho como aquele com maior número de danos avaliados, totalizando 799 ocorrências mapeadas, o que motivou seu acompanhamento posterior (Silva Júnior, 2018). Posteriormente, em inspeções realizadas em 2020, foram contabilizadas 1987 manifestações patológicas, com predominância de fissuras de canto (976 incidências), longitudinais (536) e

por retração plástica (252), evidenciando um aumento expressivo em relação ao levantamento anterior, no qual as fissuras de canto e longitudinais mais que quintuplicaram, além do acréscimo de ocorrências por retração plástica e transversais (Hora, 2021).

A interpretação desse crescimento aponta para causas combinadas. Em particular, a execução inadequada das juntas e a ausência/deficiência de selagem aparecem como fatores fortemente associados ao aumento de fissuração, por favorecerem infiltração, movimentações e concentração de tensões. Além disso, o uso inadequado da ciclovia, como circulação/estacionamento de veículos e a fixação de estruturas temporárias para eventos, introduz ações e cargas para as quais o pavimento não foi dimensionado, acelerando o processo de deterioração e comprometendo o desempenho funcional (Hora, 2021). Esses achados reforçam que a durabilidade da infraestrutura cicloviária não depende apenas do tipo de pavimento escolhido, mas também de qualidade construtiva, operação compatível com o projeto e estratégias efetivas de manutenção e fiscalização.

Nesse cenário, a gestão da manutenção se beneficia diretamente de inspeções periódicas e monitoramento contínuo, com foco na identificação precoce de fissuras e no acompanhamento de sua evolução espacial e temporal. Tradicionalmente, inspeções visuais *in loco* são amplamente utilizadas por demandarem baixo investimento inicial e dependerem de procedimentos consolidados. Porém, há limitações relevantes: subjetividade na avaliação, dependência de mão de obra especializada, baixa escalabilidade para grandes extensões e dificuldade em manter dados estruturados e padronizados para análises comparativas ao longo do tempo. Na prática, metodologias de inspeção podem exigir instrumentos simples (ex.: trena e fissurômetro para medições), registros por trecho e organização sistemática das ocorrências, mas ainda assim a coleta é trabalhosa e suscetível a inconsistências quando realizada em redes extensas (Albuquerque, 2020).

Assim, abre-se espaço para abordagens automatizadas e reprodutíveis, capazes de aumentar a frequência de inspeção, reduzir a subjetividade e, principalmente, gerar produtos digitais estruturados que alimentem rotinas de priorização de intervenções. É exatamente nesse ponto que tecnologias como aerofotogrametria com VANT e métodos de inteligência artificial se tornam estratégicos por permitirem mapear fissuras com alta cobertura espacial, registrar a condição do pavimento em bases georreferenciadas e apoiar decisões de manutenção, fundamento que conduz aos próximos tópicos desta revisão.

2.2 AEROFOTOGRAMETRIA E USO DE VANT NA ENGENHARIA CIVIL

A adoção de VANT, popularmente chamados de drones, tem se consolidado como uma alternativa eficiente para aquisição de dados em inspeções de infraestrutura e edificações, sobretudo quando se busca ampliar a cobertura espacial, reduzir o tempo de levantamento e aumentar a frequência de monitoramento. Em comparação com vistorias exclusivamente presenciais, os VANT permitem a coleta sistemática de imagens com repetibilidade e em curto intervalo de tempo, favorecendo estratégias de manutenção mais proativas e contribuindo para a segurança viária (Khan *et al.*, 2024; Wang; Ye, 2022).

Além disso, por operarem remotamente, reduzem a exposição de equipes a situações de risco e eliminam, em muitos casos, a necessidade de interdição de áreas, andaimes, plataformas elevatórias e mobilização de grandes equipes, o que tende a diminuir custos operacionais e aumentar a viabilidade de inspeções periódicas (Nooralishahi *et al.*, 2021; Azari; O'shea; Campbell, 2021; Shafiee *et al.*, 2021).

No contexto da engenharia civil, essa capacidade de aquisição rápida e com alta densidade de informações tem sido explorada em inspeções de pontes, fachadas, taludes, edificações e instalações industriais, possibilitando tanto avaliações visuais remotas quanto análises mais detalhadas apoiadas por produtos derivados do processamento fotogramétrico. A depender do objetivo e do plano de voo, imagens obtidas por VANT podem ser processadas para gerar nuvens de pontos, malhas texturizadas, modelos tridimensionais e, especialmente, ortomosaicos georreferenciados, produtos que representam a superfície de forma ortorretificada, com escala uniforme e compatível com medições e integração em GIS (Dorafshan; Maguire, 2018; Mandirola *et al.*, 2022). Para aplicações em pavimentos, o ortomosaico se torna particularmente relevante por permitir a leitura contínua do trecho e a vinculação direta com coordenadas, favorecendo a espacialização de danos e a comparação temporal entre campanhas.

A aerofotogrametria aplicada a VANT baseia-se, em linhas gerais, na reconstrução geométrica a partir de múltiplas imagens sobrepostas, associando princípios de visão estereoscópica e algoritmos de correspondência de pontos. Em um fluxo típico, realiza-se o planejamento de voo e aquisição das imagens (frequentemente em tomada nadiral para pavimentos, 90° para baixo), o alinhamento por técnicas do tipo *Structure-from-Motion* (SfM), a geração de nuvem de pontos densa por *Multi-View Stereo* (MVS), a produção de modelo de superfície e a ortorretificação e mosaico final. O resultado é um produto cartográfico contínuo, cujo valor prático está na

compatibilidade com análises métricas e com a sobreposição de camadas temáticas em ambiente GIS.

A qualidade do ortomosaico e sua utilidade para inspeção dependem diretamente de fatores de aquisição e processamento, tais como altura de voo (e consequentemente o *Ground Sample Distance* – GSD), sobreposição longitudinal e lateral, estabilidade do voo, iluminação, presença de sombras, textura do pavimento e parâmetros de calibração da câmera. Em pavimentos, há um desafio adicional: fissuras tendem a ser feições finas e de baixo contraste, o que exige GSD suficientemente pequeno para preservar a visibilidade dos danos. Em contrapartida, reduzir a altura de voo aumenta a quantidade de imagens e o custo computacional do processamento, além de exigir maior rigor no controle de desfoque por movimento, exposição e variações radiométricas. Também é importante notar que a etapa de mosaico pode introduzir efeitos de reamostragem e suavização, o que demanda cuidados quando o objetivo é detectar feições muito estreitas; por isso, em aplicações de detecção automatizada, é comum empregar estratégias de inferência por blocos (*tiles*) ou janelas deslizantes no ortomosaico ou mesmo analisar imagens individuais em alta resolução, conforme as necessidades do fluxo de trabalho (Song; Liu; Yuan, 2024).

No contexto de detecção automática de fissuras, o GSD também impacta diretamente a escala aparente dos danos e, consequentemente, o desempenho de modelos de visão computacional. Estudo recente discute o GSD como um parâmetro-chave para a detecção automática, destacando que variações de GSD podem exigir estratégias de adequação entre aquisição e processamento (Boutet; Hallot, 2025).

Além desses aspectos técnicos, a definição da altitude e do plano de voo é frequentemente limitada por restrições legais e de segurança operacional, como limites regulamentares de altura, exigência de autorizações e condicionantes para operação em áreas urbanas e/ou próximas a zonas sensíveis, o que pode reduzir a flexibilidade para atingir GSD muito pequenos em determinados trechos (ANAC, 2017; DECEA, 2020).

A georreferência é outro elemento central quando a finalidade é integrar resultados em ambiente GIS. Em geral, pode ser obtida por diferentes arranjos, como posicionamento embarcado com RTK, uso de pontos de controle em solo ou combinações desses métodos. A escolha influencia a acurácia posicional final e, consequentemente, a confiabilidade de análises espaciais e da localização dos danos ao longo da malha cicloviária. Para aplicações de manutenção, essa consistência espacial é crucial: mapas temáticos, quantificação por trechos e priorização de

intervenções dependem de que os produtos estejam corretamente referenciados e sejam comparáveis entre diferentes campanhas.

Quando associada à inteligência artificial, a aerofotogrametria com VANT amplia significativamente a escalabilidade do processo de inspeção. As imagens capturadas podem alimentar modelos de aprendizado profundo (*deep learning* - *DL*) capazes de identificar automaticamente manifestações patológicas, reduzindo o esforço manual de triagem e aumentando a reprodutibilidade dos diagnósticos. Estudos recentes indicam desempenhos elevados na detecção automatizada de danos a partir de imagens de VANT, com valores de *mean average precision* (mAP) variando amplamente conforme o tipo de dano, qualidade das imagens, método e conjunto de dados utilizados (He *et al.*, 2023; Khan *et al.*, 2024; Silva *et al.*, 2020; Silva *et al.*, 2023). Entretanto, essa mesma literatura aponta um ponto crítico: a generalização dos modelos é fortemente condicionada pela representatividade do banco de dados empregado no treinamento. Em muitos casos, o uso de bases genéricas — com padrões de superfície, condições de iluminação e contextos distintos do cenário real de operação — limita o desempenho quando o método é transferido para ambientes urbanos específicos. Assim, a obtenção de imagens em inspeções reais, contemplando variações de textura, juntas, sinalização, sombras e diferentes estágios de degradação, torna-se fundamental para construir modelos mais robustos e aderentes ao problema aplicado (Arafin; Billah; Issa, 2023).

Dessa forma, a aerofotogrametria com VANT não se restringe à captura de imagens, mas constitui a base de um fluxo de trabalho digital no qual dados georreferenciados de alta resolução podem ser transformados em produtos cartográficos (como ortomosaicos) e, posteriormente, analisados por modelos de visão computacional. Essa integração cria as condições necessárias para a automação da identificação de fissuras e para a organização dos resultados em ambiente GIS, conectando a etapa de aquisição (VANT/fotogrametria) ao núcleo de processamento apresentado no tópico seguinte, sobre inteligência artificial e visão computacional.

2.3 INTELIGÊNCIA ARTIFICIAL E VISÃO COMPUTACIONAL

A Inteligência Artificial pode ser compreendida como o campo científico dedicado ao desenvolvimento de máquinas e sistemas capazes de executar tarefas associadas à inteligência humana, incluindo a interpretação de informações, reconhecimento de padrões, tomada de decisão e aprendizagem a partir de dados. Nesse contexto, a IA tornou-se particularmente

relevante com a expansão da capacidade de processar grandes volumes de dados e construir modelos estatísticos em larga escala, impulsionando avanços em áreas como visão computacional, robótica, veículos autônomos, tradução automática, reconhecimento de voz, diagnóstico médico e aplicações generativas (Baduge *et al.*, 2022). Embora o termo tenha origem nas primeiras décadas do desenvolvimento da computação, o crescimento recente da IA é fortemente associado à consolidação de duas frentes tecnológicas: o aprendizado de máquina (*Machine Learning* – ML) e o DL, viabilizados pela combinação entre maior capacidade computacional (especialmente via GPUs), desenvolvimento de algoritmos mais eficientes e ampla disseminação de bibliotecas e *softwares* que facilitam implementação e reprodução de resultados (Baduge *et al.*, 2022).

O aprendizado de máquina é um subcampo da IA orientado à construção de modelos capazes de aprender padrões a partir de dados e realizar previsões ou classificações sem que todas as regras sejam explicitamente programadas. Em termos operacionais, esse processo envolve ajustar parâmetros internos do algoritmo com base no desempenho observado em exemplos anteriores, promovendo melhora progressiva ao longo das iterações de treinamento (Oliveira, 2018; Baduge *et al.*, 2022). O aprendizado profundo, por sua vez, constitui uma especialização do ML baseada em modelos com múltiplas camadas de processamento, capazes de extrair representações hierárquicas dos dados de entrada, das mais simples às mais abstratas, o que o torna especialmente adequado a dados de alta dimensão, como imagens, vídeos e sinais (Deng; Yu, 2014; Baduge *et al.*, 2022). De forma geral, enquanto métodos tradicionais de ML frequentemente exigem a definição e extração manual das características da imagem, o DL se destaca por aprender automaticamente características discriminantes diretamente dos dados.

No centro de grande parte das aplicações contemporâneas de DL em visão computacional estão as redes neurais artificiais. Essas redes são estruturas computacionais inspiradas no princípio de funcionamento de redes neurais biológicas, organizadas em camadas de nós (neurônios) conectados, em que as saídas de uma camada alimentam as entradas da camada seguinte. Esse arranjo possibilita a propagação do sinal e o aprendizado de padrões por meio do ajuste de pesos sinápticos e limiares (Van Gerven; Bohte, 2018). De modo didático, o funcionamento pode ser descrito em duas etapas: fase de aprendizado, na qual os parâmetros são recalibrados para reforçar respostas corretas e reduzir respostas incorretas, e fase de utilização, quando a rede, com parâmetros estabilizados, passa a ser empregada para resolver o problema-alvo, podendo ser reavaliada e refinada em ciclos posteriores de otimização (Oliveira Jr. *et al.*, 2007).

Dentro desse universo, as Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs) ocupam papel central por serem projetadas para processar dados com estrutura em grade, como imagens. As CNNs operam com camadas convolucionais, *pooling* e camadas densas (totalmente conectadas), formando um pipeline que extrai características visuais por meio de filtros, reduz dimensionalidade preservando informações relevantes e, por fim, realiza tarefas como classificação, regressão ou detecção. A convolução permite capturar padrões locais (bordas, texturas, contrastes) e, em camadas mais profundas, padrões mais complexos (formas e arranjos espaciais), com eficiência computacional superior a arquiteturas totalmente conectadas para dados visuais (Baduge *et al.*, 2022). Uma estrutura típica de CNN aplicada a classificação pode ser representada por um fluxo de camadas convolucionais com ativação (como ReLU), *pooling*, repetição do bloco para aprendizado progressivo de características, seguido de *flatten* e camadas densas culminando em uma camada de decisão (por exemplo, *Softmax*) (Tabian; Fu; Khodaei, 2019). Esse princípio geral sustenta uma ampla variedade de arquiteturas e aplicações, incluindo inspeções automatizadas de superfícies de concreto e pavimentos.

Nos últimos anos, o avanço das CNNs e da visão computacional baseada em DL impulsionou o aumento de pesquisas voltadas à detecção automatizada de fissuras e outros danos em infraestruturas, em função da precisão, robustez e capacidade de operação em diferentes condições de coleta de imagem (Nishikawa *et al.*, 2011; Dung; Anh, 2019; Han; Yang; Yang, 2024). No domínio de pavimentos, em particular, a detecção de fissuras tem relevância por se tratar de uma manifestação frequentemente predominante, com potencial de evolução progressiva e impacto direto na funcionalidade. Diversos trabalhos reportam aplicações de DL para inspeção visual, com uso de CNNs e arquiteturas modernas para reconhecimento e localização de danos em imagens, superando abordagens tradicionais em velocidade, objetividade e consistência de resultados, desde que haja dados representativos para treinamento (Jiang; Pang; Li, 2021; Santos; Carvalho, 2025; Wang *et al.*, 2024).

Do ponto de vista da formulação do problema, a identificação de fissuras por aprendizado profundo costuma ser organizada em duas tarefas principais: detecção de objetos e segmentação (Han; Yang; Yang, 2024). Na detecção, busca-se localizar e classificar regiões de interesse por meio de caixas delimitadoras (*bounding boxes*), o que é adequado quando o objetivo é identificar a presença e a posição aproximada do dano. Na segmentação, busca-se delinear a forma do dano em nível de pixel, possibilitando representações mais precisas e,

consequentemente, extração de métricas geométricas com maior fidelidade, como área degradada e extensão linear de fissuras, fundamentais quando o foco é quantificação e suporte à manutenção.

A detecção de objetos pode ser abordada por arquiteturas de duas etapas (*two-stage*), como a família R-CNN (Girshick *et al.*, 2014), ou de uma etapa (*one-stage*), como SSD (Liu *et al.*, 2016) e YOLO. Na família R-CNN, a rede primeiro propõe regiões candidatas e, em seguida, classifica e refina essas regiões; já em detectores *one-stage* a predição ocorre de forma direta sobre a imagem, priorizando eficiência e rapidez, característica especialmente relevante em aplicações que demandam tempo de resposta reduzido. A série YOLO, introduzida inicialmente como “You Only Look Once”, consolidou-se como referência em detecção em tempo real por unificar o processo de localização e classificação em uma única passagem de inferência, equilibrando desempenho e velocidade (Redmon *et al.*, 2016). Esse equilíbrio se torna particularmente vantajoso em inspeções de pavimentos, em que o volume de dados pode ser elevado e a execução precisa ser escalável.

Em aplicações de fissuras, muitos cenários demandam respostas rápidas e a delimitação aproximada da região danificada é suficiente para triagem e planejamento inicial de inspeção. Nesses casos, métodos baseados em *bounding boxes* tendem a apresentar menor custo computacional do que abordagens em nível de pixel, favorecendo uso prático em grandes áreas, como ortomosaicos (Su *et al.*, 2024). Por essa razão, a série YOLO têm sido amplamente empregada em inspeção de pavimentos e estruturas.

Como exemplos, Wang *et al.* (2023) utilizaram YOLOv5 com ajustes de hiperparâmetros e dados sintéticos/virtuais para aprimorar o reconhecimento de rachaduras e buracos em pavimentos asfálticos, evidenciando o potencial desses modelos em cenários com múltiplas classes de dano. Em uma perspectiva semelhante, Qiu e Lau (2023) compararam diferentes arquiteturas e destacaram a adequação dos modelos ResNet50-based YOLOv2 e YOLOv4-tiny, que demonstraram o melhor equilíbrio entre precisão na identificação de fissuras pequenas e velocidade de processamento necessária para inspeções em tempo real.

A evolução dos modelos YOLO acompanha avanços em *backbones*, *necks*, funções de perda, estratégias de aumento de dados e refinamentos de treinamento, frequentemente incorporados por implementações amplamente disseminadas na comunidade. No contexto desta dissertação, a escolha por versões recentes, disponibilizadas e consolidadas em bibliotecas amplamente adotadas, como a Ultralytics, que oferece documentação, modelos pré-treinados e rotinas

padronizadas de treinamento, validação e exportação, favorece a reprodutibilidade e a integração com diferentes ambientes computacionais, além de permitir comparação direta entre variações arquiteturais e tamanhos de modelo (Jocher *et al.*, 2023; 2024; Su *et al.*, 2024). Também têm sido propostos ajustes e variantes específicas para aprimorar a detecção de fissuras, considerando os desafios típicos desse objeto (feições finas, baixo contraste e alta variabilidade), como discutido em propostas recentes de adaptação de arquiteturas para o domínio de trincas (Su *et al.*, 2024).

Além do domínio tradicional das CNNs, modelos baseados em *Transformers* passaram a ganhar espaço na visão computacional ao incorporar mecanismos de atenção capazes de modelar relações espaciais de longo alcance. Um marco dessa transição foi o DETR (*DEtection TRansformer*), que reformulou a detecção como um problema de predição direta de conjuntos e reduziu dependências de heurísticas clássicas como a supressão de não-máximos (*Non-Maximum Suppression* – NMS) (Carion *et al.*, 2020).

Apesar dos benefícios conceituais, limitações de custo computacional e convergência motivaram variantes mais eficientes, como o Deformable DETR, que introduz atenção deformável para reduzir complexidade e acelerar treinamento (Zhu *et al.*, 2020). Mais recentemente, o RT-DETR (*Real-Time Detection Transformer*) foi proposto com foco em viabilizar desempenho em tempo real, aproximando a eficiência de inferência das CNNs modernas sem perder a capacidade de capturar contexto global, o que pode ser relevante em imagens com ruído visual, variações de textura e múltiplos padrões concorrentes (Zhao *et al.*, 2023).

A literatura recente também aponta que modelos baseados em *Transformers* podem apresentar vantagens em cenários com restrição de dados rotulados, ao explorar estratégias como aprendizado semi-supervisionado e melhor aproveitamento de grandes volumes de dados não rotulados (Singh; Hoskere, 2023; Eltouny; Sajedi; Liang, 2024; Liu *et al.*, 2024). Ainda, arquiteturas híbridas, combinando CNNs e *Transformers*, têm sido investigadas como alternativa para unir eficiência de extração local das convoluções com a capacidade global de atenção dos *Transformers*, com ganhos reportados em alguns cenários de segmentação e detecção (Zhang *et al.*, 2024).

De maneira geral, enquanto as CNNs baseiam sua eficácia na localidade e no compartilhamento de pesos, o que as torna eficientes e rápidas para extração de características de bordas e texturas de fissuras, elas encontram limites na captura de dependências de longo alcance e no

entendimento contextual da cena. Por outro lado, a abordagem baseada em Transformers supera essa limitação através do mecanismo de atenção global, que permite ao modelo correlacionar informações de toda a imagem simultaneamente e estabelecer dependências globais entre pixels distantes, superando o limite de localidade das CNNs convencionais. Contudo, essa flexibilidade arquitetônica impõe um limite de eficiência computacional, exigindo otimizações (como as presentes no RT-DETR) para que o modelo possa operar em tempo real sem demandar recursos excessivos de hardware.

Assim, o contraponto entre CNNs e Transformers representa não apenas uma diferença arquitetural, mas também uma tendência de pesquisa em inspeção automatizada, justificando comparações experimentais entre famílias de modelos quando o objetivo inclui desempenho e viabilidade operacional.

Embora a detecção por caixas seja eficiente, a tarefa de inspeção pode exigir delimitação mais precisa do dano, especialmente quando se pretende estimar quantitativos associados à manutenção (Santos; Carvalho, 2025). A segmentação, ao operar em nível de pixel, permite delinear a geometria real da fissura, tornando possível medir extensão, área e padrões de distribuição com maior fidelidade, algo particularmente importante quando os resultados precisam alimentar bases georreferenciadas e análises espaciais. Nesse sentido, Han, Yang e Yang (2024) apontam que a identificação de fissuras por DL se organiza majoritariamente entre detecção e segmentação, com escolhas condicionadas pelo equilíbrio entre custo computacional, precisão geométrica e finalidade do produto.

Diferentes estratégias têm sido reportadas para segmentação e quantificação. Kang *et al.* (2020), por exemplo, propuseram um método automatizado de detecção, localização e quantificação utilizando um fluxo baseado em CNN com Faster R-CNN para identificação das regiões associadas às fissuras. Já revisões recentes em inspeção automatizada apontam que arquiteturas modernas (tanto CNNs quanto Transformers e híbridos) têm ampliado a capacidade de segmentar e caracterizar danos com maior robustez, fortalecendo o uso de saídas em nível de pixel para apoiar decisões de reparo e manutenção (Santos; Carvalho, 2025; Dang *et al.*, 2023; Zuo *et al.*, 2025).

Do ponto de vista metodológico, o desempenho de modelos de DL depende fortemente da qualidade e representatividade do conjunto de dados, além de escolhas relacionadas a arquitetura, hiperparâmetros e estratégias de treinamento. Em inspeção de fissuras, desafios como desbalanceamento de classe (muitos pixels de fundo e poucos de fissura), variabilidade

de textura, baixa relação sinal-ruído e diversidade de condições de iluminação tornam comum a adoção de técnicas de aumento de dados (*data augmentation*) para aumentar diversidade do conjunto de treinamento. Essa técnica amplia virtualmente a base de dados por meio de transformações geométricas e fotométricas (rotações, redimensionamentos, recortes, alterações de brilho/contraste e cor), contribuindo para melhorar generalização e reduzir sobreajuste (*overfitting*) (Lee; Lee; Kim, 2022; Maharana; Mondal; Nemade, 2022).

Outra estratégia amplamente utilizada é o aprendizado por transferência (*transfer learning*), que consiste em reutilizar conhecimento aprendido por um modelo previamente treinado em uma tarefa ou grande conjunto de dados, adaptando-o a um novo domínio relacionado. Ao partir de pesos já ajustados para extrair características visuais gerais, o modelo pode alcançar melhor desempenho com menos dados rotulados e menor tempo de treinamento, desde que a adaptação seja conduzida adequadamente (Olivas *et al.*, 2009; Goodfellow; Bengio; Courville, 2016). Em aplicações de detecção e segmentação de fissuras, essa abordagem é especialmente útil quando o banco de dados é limitado ou quando se busca robustez em cenários distintos do treinamento original, sendo comum o ajuste fino (*fine-tuning*) e/ou congelamento parcial de camadas para equilibrar estabilidade e adaptação (Chen *et al.*, 2023; Munawar *et al.*, 2021; Wang *et al.*, 2024).

Em síntese, a literatura evidencia que a inspeção automatizada por visão computacional pode ser formulada tanto como detecção (com ganho de eficiência e viabilidade em tempo real) quanto como segmentação (com ganho de precisão geométrica e capacidade de quantificação). A evolução de arquiteturas, de CNNs clássicas e detectores como YOLO à incorporação de Transformers como DETR e RT-DETR, amplia o repertório de soluções disponíveis e justifica abordagens comparativas para selecionar modelos com melhor equilíbrio entre acurácia e custo computacional. Esses fundamentos sustentam a aplicação desenvolvida nesta dissertação, em que imagens obtidas por aerofotogrametria são processadas por modelos de DL para identificar fissuras e produzir resultados compatíveis com integração espacial, tema que se conecta ao estado da arte em detecção de fissuras e às etapas subsequentes de organização dos dados para uso em GIS.

2.4 INTEGRAÇÃO IA-GIS

Projetos de infraestrutura linear, como estradas, ferrovias e ciclovias, apresentam extensão significativa, elevada variabilidade de condições ao longo do traçado e forte dependência do contexto espacial (uso do solo, drenagem, sombreamento, entorno construído). Nesse cenário,

a adoção do GIS tem se consolidado como estratégia para organizar, analisar e comunicar informações georreferenciadas ao longo do ciclo de vida do ativo, do planejamento e projeto à operação e manutenção (O&M) (Ojo *et al.*, 2019; Karimi; Iordanova, 2021; Malinverni *et al.*, 2020). A partir dessa base espacial, a integração com IA, especialmente visão computacional e aprendizado profundo, amplia o potencial de automação da inspeção, permitindo que anomalias detectadas em imagens sejam vinculadas a coordenadas e convertidas em informações acionáveis para gestão.

Em termos de representação, o GIS opera principalmente com dois modelos de dados: *raster* e vetor (Longley *et al.*, 2013). Em aplicações de inspeção com VANT, o *raster* normalmente corresponde a ortomosaicos georreferenciados (frequentemente em *GeoTIFF*), que preservam a continuidade espacial e permitem análises distribuídas ao longo de grandes extensões. Já os resultados da IA, caixas delimitadoras (detecção) e, sobretudo, máscaras (segmentação), podem ser convertidos em feições vetoriais (linhas/polígonos), viabilizando consultas, filtros e cálculos métricos (extensão, área, densidade e distribuição espacial de danos) em camadas temáticas dentro do GIS (Dipankar; Suman, 2023).

A literatura recente mostra que, embora existam muitos estudos de detecção/segmentação de fissuras por visão computacional, a etapa de georreferenciamento e integração efetiva em GIS ainda é frequentemente negligenciada, o que limita análises espaciais e reduz o valor do dado para planejamento e priorização de manutenção. Hu, Yee e Goff (2024) destacam explicitamente essa lacuna e propõem um fluxo para pontes que combina criação de ortomosaicos georreferenciados, segmentação por *deep learning* e visualização em ambiente GIS, visando superar a fragmentação entre “detectar” e “gerir” o dano no contexto espacial do ativo.

Entre os estudos aplicados, observa-se um padrão metodológico recorrente: aquisição de imagens (VANT ou sistemas móveis), geração de produto georreferenciado (ortomosaico), inferência por modelos de IA e mapeamento do dano em camadas GIS. No contexto de pavimentos de concreto em ambiente portuário, Tsaimou *et al.* (2024) apresentam uma metodologia que combina imagens de VANT e aprendizado profundo para identificação de fissuras e construção de mapas 2D georreferenciados dos danos, reforçando o valor da etapa de mapeamento para inspeções em áreas extensas e de difícil controle manual. Em abordagem semelhante voltada a pavimentos, o estudo de Dipankar e Suman (2023) explicitamente enquadra como objetivo a detecção e mapeamento de fissuras, relatando a plotagem dos

resultados no GIS como parte do fluxo de trabalho, o que evidencia a consolidação dessa integração como produto de inspeção.

Além do caso de ortomosaicos, há exemplos em que a georreferência é explorada para vincular defeitos a modelos e elementos do ativo. Yoon *et al.* (2023) propõem um método de inspeção externa de edificações com VANT e CNN em que um sistema de coordenadas geográficas e pontos de controle/coordenação são usados para obter localização e, a partir disso, as detecções (incluindo fissuras) são convertidas em coordenadas vetoriais, permitindo representar a posição dos defeitos e automatizar a modelagem 3D para fins de manutenção. Essa abordagem reforça um ponto central para O&M: detectar o defeito é importante, mas localizá-lo de forma consistente e rastreável é o que viabiliza gestão, histórico e tomada de decisão.

Em termos de implicações para a gestão e a operação, a integração GIS–IA potencializa: (a) automação e produtividade na inspeção; (b) rastreabilidade espacial do dano (onde ocorreu, como evolui, com que densidade); (c) quantificação (extensão/área de fissuras a partir de feições vetoriais) e (d) suporte a ações preventivas e priorização por criticidade, sobretudo em ativos extensos. Ao mesmo tempo, permanecem desafios relevantes para consolidar a integração como prática operacional: dependência da qualidade do dado (iluminação, sombras, obstruções, variabilidade de textura), necessidade de padronização de produtos (formato, sistema de referência, atributos mínimos), interoperabilidade e, frequentemente, validação manual em ciclos de atualização, seja porque o modelo falha em alguns cenários, seja porque parte do processo demanda intervenção quando objetos/defeitos não são reconhecidos (Yoon *et al.*, 2023)

Em síntese, os estudos analisados reforçam que a integração GIS–IA é especialmente promissora para O&M de infraestrutura linear, mas ainda está em fase de consolidação: a literatura indica avanço consistente na detecção/segmentação, porém o fluxo completo ainda aparece em menor número de trabalhos e com distintos níveis de maturidade. A consolidação dessa integração tende a depender de fluxos mais padronizados, com foco em interoperabilidade, enriquecimento semântico dos defeitos (tipo, severidade, métricas) e automação de atualização do banco espacial, de modo que a saída da IA se transforme, de fato, em informação de gestão.

Este capítulo apresentou os fundamentos e o estado da arte que sustentam a presente pesquisa, articulando três frentes complementares: a aquisição de dados por aerofotogrametria com VANT e a geração de produtos georreferenciados (ortomosaicos) para inspeção de pavimentos;

o uso de modelos de aprendizado profundo, com destaque para detectores e segmentadores baseados em YOLO, para identificação automatizada de fissuras e o papel dos Sistemas de Informação Geográfica na estruturação, visualização e análise espacial de danos em ativos lineares. A literatura evidencia avanços relevantes em desempenho de detecção/segmentação, mas também indica que fatores como resolução/GSD, variações de iluminação e qualidade do dado, além de escolhas de treinamento e estratégias de inferência, influenciam diretamente a robustez dos resultados em cenários reais.

Além disso, a revisão aponta uma lacuna recorrente: mesmo quando modelos alcançam boa acurácia, muitas abordagens não avançam para a etapa de transformar as inferências em informação espacialmente estruturada (camadas vetoriais e banco de dados geográfico), limitando análises de distribuição, quantificação e priorização de intervenções. Assim, o capítulo reforça a relevância de fluxos padronizados que conectem aquisição por VANT, inferência por IA e integração em GIS como base para aplicações de gestão da manutenção. Com esse enquadramento, o próximo capítulo descreve a metodologia adotada para constituir os conjuntos de dados, treinar e avaliar modelos de detecção e segmentação, validar externamente em ortomosaico georreferenciado e operacionalizar a integração dos resultados em ambiente GIS.

3 METODOLOGIA

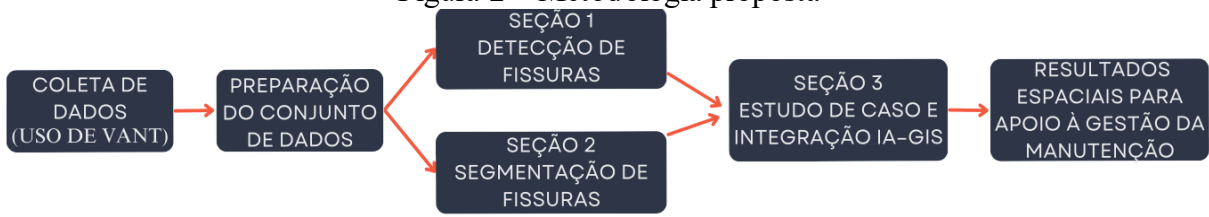
A metodologia adotada nesta pesquisa foi estruturada de modo a atender ao objetivo geral e aos objetivos específicos propostos, organizando o desenvolvimento da pesquisa de forma sistemática, consistente e reproduzível.

O fluxo metodológico foi dividido em três seções principais, sequenciais e interdependentes, descritas ao longo deste capítulo. A visão global dessa estrutura é apresentada na Figura 2.

Em resumo, as três seções que compõem a metodologia desta dissertação compreendem:

- Seção 1 – Desenvolvimento e treinamento de modelos de inteligência artificial para detecção de fissuras em ciclovias;
- Seção 2 – Desenvolvimento e treinamento de modelos de inteligência artificial para segmentação de fissuras;
- Seção 3 – Estudo de caso para validação externa dos modelos de IA e integração em ambiente GIS para mapeamento automatizado.

Figura 2 – Metodologia proposta

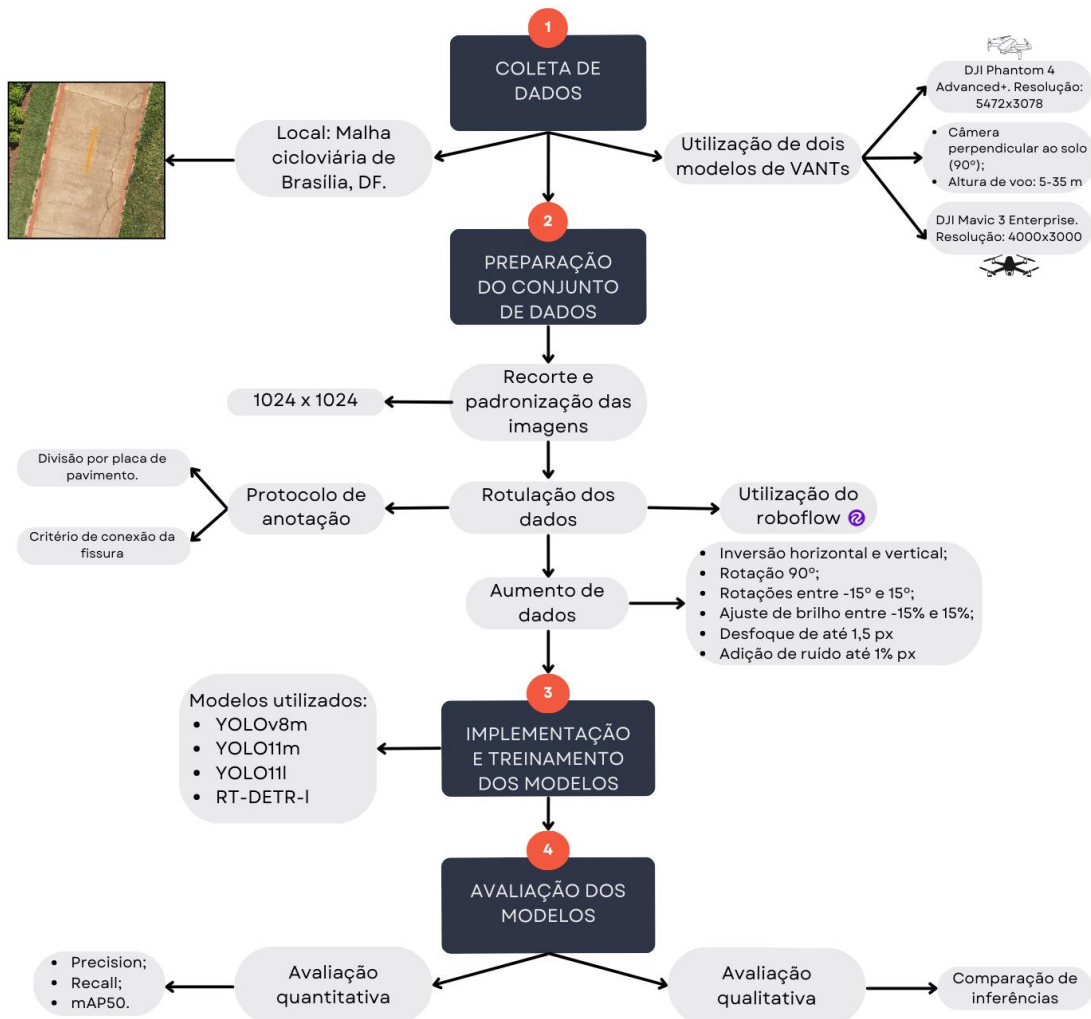


Fonte: O autor.

3.1 SEÇÃO 1 – DESENVOLVIMENTO E TREINAMENTO DE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA DETECÇÃO DE FISSURAS EM CICLOVIAS

Nesta seção, apresenta-se a metodologia adotada para o desenvolvimento e a avaliação de modelos de inteligência artificial aplicados à detecção automática de fissuras em cicloviadas, a partir de imagens obtidas por VANT. Esta seção é composta por quatro etapas, organizadas de forma sequencial e interdependente, conforme ilustrado na Figura 3.

Figura 3 – Fluxo metodológico adotado para o desenvolvimento e treinamento dos modelos de detecção de fissuras em cicloviadas



Fonte: O autor.

Inicialmente, realiza-se a coleta de dados, por meio de levantamento aerofotogramétrico da malha cicloviária selecionada. Em seguida, procede-se à preparação do conjunto de dados, abrangendo o recorte, a padronização das imagens, a rotulação das fissuras e o *data augmentation*. A terceira etapa corresponde à implementação e ao treinamento dos modelos de detecção, utilizando diferentes arquiteturas de aprendizado profundo. Por fim, os modelos treinados são submetidos a uma avaliação de desempenho, contemplando análises quantitativas e qualitativas, com o objetivo de verificar sua robustez e capacidade de generalização frente às condições reais de aquisição das imagens.

3.1.1 Coleta de dados – Etapa 1

O estudo foi conduzido em trechos de ciclovias com pavimento de concreto rígido localizadas em Brasília, DF. As imagens utilizadas foram obtidas a partir de duas coletas distintas, envolvendo diferentes equipamentos e parâmetros operacionais. A Tabela 1 sintetiza os principais parâmetros de e a origem de cada conjunto.

Tabela 1 – Características das coletas de imagens por VANT utilizadas no estudo

Fonte	Drone	Local	Resolução	Altura de voo	Modo de voo
(Santos; Dusi, 2022)	DJI Phantom 4 Advanced+	Asa Sul	5472x3078	5m	Manual
(O autor)	DJI Mavic 3 Enterprise	Eixo Monumental	4000x3000	12m - 35m	Automatizado

Fonte: O autor.

As imagens coletadas por Santos e Dusi (2022) correspondem a trechos localizados na Asa Sul e foram adquiridas com o VANT DJI Phantom 4 Advanced+. Já as imagens obtidas com o DJI Mavic 3 Enterprise foram coletadas pelo autor em trechos de ciclovia do Eixo Monumental. Para garantir uma captura padronizada e ortogonal do pavimento, todos os voos foram executados com a câmera apontada perpendicularmente para o solo (ângulo de 90°).

Com o objetivo de gerar um banco de imagens que representasse condições reais de operação, os sobrevoos foram realizados em alturas variadas, com o VANT posicionado entre 5 metros e 35 metros em relação ao nível do pavimento. A utilização de dois sensores distintos e a variação na altura de voo resultaram em um *dataset* heterogêneo, com ampla variabilidade em termos de resolução espacial, condições de iluminação natural, presença de sombreamento de árvores e edificações, e ruídos diversos (por exemplo folhas, galhos, sujidade e detritos sobre o pavimento). Essa variabilidade é intencional, visando o desenvolvimento de um modelo de

identificação de fissuras que seja robusto, capaz de manter desempenho consistente frente às diferentes condições operacionais observadas em ciclovias urbanas de pavimento rígido.

3.1.2 Preparação do conjunto de dados – Etapa 2

A preparação do conjunto de dados para o treinamento dos modelos envolveu três subetapas sequenciais de pré-processamento: o recorte e padronização das imagens, a rotulação das fissuras e, por fim, o aumento de dados.

Inicialmente, abordou-se o desafio imposto pela alta resolução das imagens brutas obtidas pelos VANT, que inviabilizam o processamento computacional direto para o treinamento de modelos de aprendizado de máquina. O treinamento dos modelos utilizando as imagens em sua resolução nativa demandaria elevado custo computacional, ao passo que o redimensionamento integral dessas imagens para dimensões menores resultaria em perda significativa de detalhes, tornando as fissuras, que são anomalias de pequena escala, praticamente indistinguíveis.

Para contornar essa limitação, foi adotada uma etapa de pré-processamento baseada na extração de regiões de interesse (ROI). Essa estratégia fundamenta-se na abordagem metodológica proposta por Santos e Carvalho (2025), que demonstraram que o isolamento de danos em recortes de menor dimensão, em detrimento do uso da imagem completa redimensionada, aumenta significativamente a capacidade dos modelos de detectar patologias de pequena escala, como fissuras, ao preservar a densidade de pixels original do objeto. As imagens originais foram segmentadas em recortes com dimensões padronizadas de 1024×1024 pixels, preservando a escala e a resolução espacial do pavimento em cada recorte. Dessa forma mantêm-se os detalhes finos das fissuras e, ao mesmo tempo, padroniza-se a dimensão das imagens de entrada para as etapas subsequentes.

Os recortes gerados contemplaram tanto trechos de pavimento com a presença de fissuras quanto trechos íntegros, sem manifestações patológicas visíveis. A inclusão de amostras negativas no conjunto de dados visa auxiliar o modelo a diferenciar adequadamente padrões de fissuração de variações naturais da textura do pavimento, contribuindo para a redução de falsos positivos e para a melhoria do desempenho geral da detecção (Gao *et al.*, 2019).

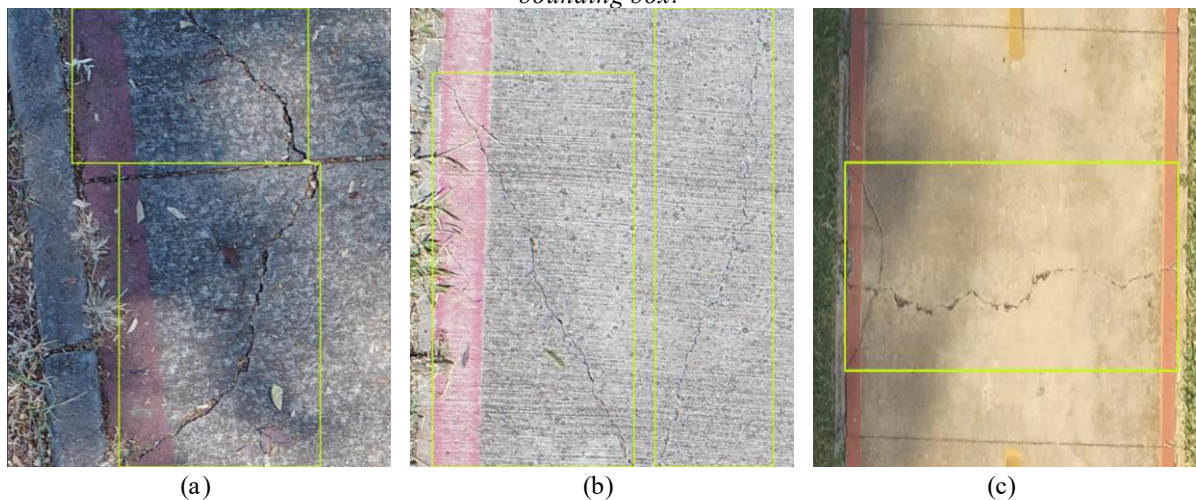
É importante ressaltar que uma única imagem bruta de alta resolução pôde originar múltiplas ROI distintas. Adicionalmente, devido à alta sobreposição inerente aos levantamentos aerofotogramétricos, era comum que ROI quase idênticas fossem extraídas de imagens

consecutivas. Para evitar redundância e garantir a diversidade do *dataset*, realizou-se um processo de verificação manual, mantendo apenas uma versão de cada região de interesse única.

Após a etapa de recorte, procedeu-se à rotulação do conjunto de dados utilizando a plataforma de anotação de imagens *Roboflow* (Dwyer *et al.*, 2024). O método de anotação foi o de caixas delimitadoras, onde cada fissura visível foi demarcada por um retângulo. Para garantir a consistência e a padronização em todo o *dataset*, foi estabelecido um protocolo de anotação com as seguintes diretrizes:

- **Divisão por Placa de Pavimento:** Cada anotação foi contida dentro de uma única placa de concreto. Fissuras que se estendiam por múltiplas placas foram seccionadas na junta de dilatação, recebendo uma caixa delimitadora distinta para cada segmento, conforme exemplificado na Figura 4a. Essa abordagem possibilita, em trabalhos futuros, análises relacionadas à densidade de danos por placa de pavimento.
- **Critério de Conexão da Fissura:** A definição de uma ou múltiplas anotações por placa foi baseada na conexão visual entre as fissuras. Fissuras sem conexão física aparente dentro da mesma placa foram rotuladas individualmente, cada uma com sua própria caixa (Figura 4b). Em contrapartida, uma fissura contínua, mesmo que apresentasse ramificações, foi encapsulada por uma única caixa delimitadora abrangente, como ilustrado na Figura 4c.

Figura 4 – Exemplos de aplicação do protocolo de anotação das fissuras em ciclovias de pavimento rígido (detecção): (a) fissura dividida conforme os limites das placas de concreto; (b) fissuras sem conexão física aparente rotuladas individualmente; (c) ramificações conectadas visualmente são agrupadas em uma única *bounding box*.



Fonte: O autor.

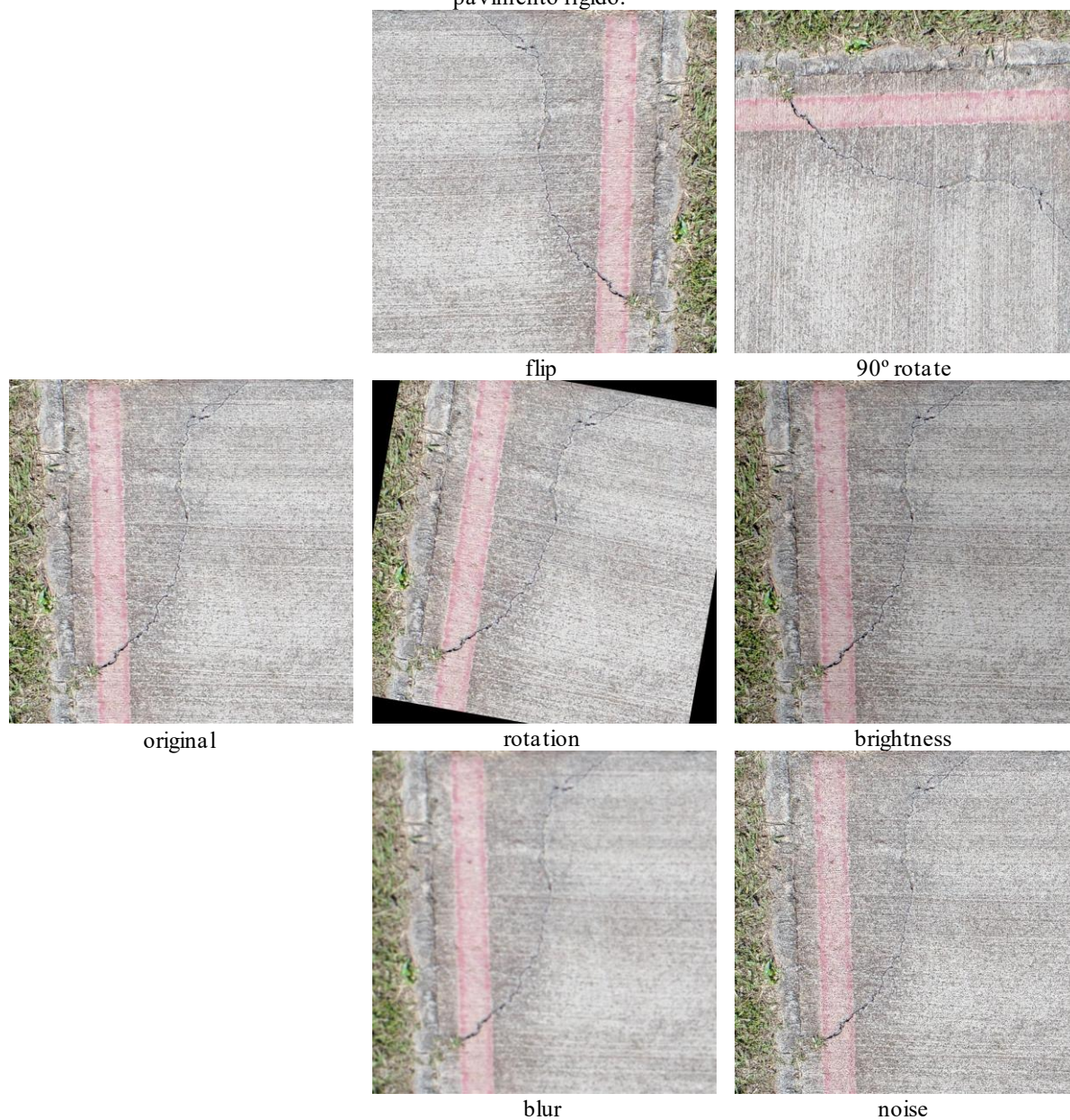
O processo de rotulação foi executado de forma semi-automatizada para otimizar a eficiência e consistência. Inicialmente, as primeiras 100 imagens do conjunto de dados foram rotuladas

manualmente. Este subconjunto inicial foi então utilizado para treinar um modelo preliminar dentro do próprio ambiente *Roboflow*. Subsequentemente, a função "*Label Assist*" da plataforma foi empregada, permitindo que o modelo treinado sugerisse automaticamente caixas delimitadoras nas imagens restantes, que eram então validadas, ajustadas ou descartadas. Esta abordagem híbrida acelerou o processo de rotulação, mantendo a qualidade do *dataset* final.

Por fim, com o conjunto de dados rotulados, aplicou-se a técnica de *data augmentation* utilizando-se a plataforma *Roboflow*. Segundo Mumuni e Mumuni (2022), a técnica de *data augmentation* em visão computacional fundamenta-se na geração de instâncias sintéticas de treinamento, visando expandir a diversidade do conjunto de dados sem a necessidade de novas coletas em campo. O processo emprega transformações que simulam variações físicas e ambientais, englobando alterações geométricas (como rotações e translações), ajustes no espaço de cor (brilho e contraste) e a introdução de ruídos, estratégias para reduzir *overfitting* e aprimorar a robustez dos modelos de aprendizado profundo. O protocolo de aumento de dados foi configurado para gerar 6 novas variações para cada imagem original do conjunto de treinamento, resultando em um fator de expansão total de 7x.

O conjunto de transformações aplicadas aleatoriamente a cada imagem incluem: inversão horizontal e vertical (*flip*), rotações de 90° (*90° rotate*), rotações entre -15° e +15° (*rotation*), ajustes de brilho entre -15% e +15% (*brightness*), aplicação de desfoque (*blur*) de até 1,5 pixels e adição de ruído (*noise*) em até 1% dos pixels. A Figura 5 ilustra visualmente o resultado da aplicação dessas transformações sobre uma imagem amostral do pavimento, destacando a imagem original (à esquerda) em contraste com as diversas variações sintéticas geradas (à direita), demonstrando a variabilidade introduzida no treinamento. Essas transformações foram selecionadas para aumentar a robustez do modelo a variações de ponto de vista, iluminação e ruídos visuais, comuns em levantamentos com VANT.

Figura 5 – Exemplos de transformações aplicadas no processo de *data augmentation* das imagens de ciclovias de pavimento rígido.



Fonte: O autor.

3.1.3 Implementação e treinamento dos modelos de detecção – Etapa 3

Para a tarefa de identificação automática de fissuras em ciclovias, foi adotada uma abordagem baseada em aprendizado profundo para detecção de objetos. O fluxo metodológico desta etapa compreendeu a seleção de arquiteturas representativas do estado da arte, o treinamento dos modelos por meio de *transfer learning* e *fine-tuning* dos parâmetros, de modo a possibilitar uma análise comparativa consistente entre diferentes paradigmas de detecção.

Considerando as duas vertentes principais abordadas no tópico 2.3, que fala sobre as CNN e as arquiteturas baseadas em *transformers*, este trabalho avaliou arquiteturas representativas de cada paradigma, com o objetivo de comparar seu desempenho na detecção automática de fissuras em ciclovias de pavimento rígido. No grupo das redes convolucionais, foram selecionadas as arquiteturas YOLOv8 (Jocher, Ayush, Jing, 2023) e YOLO11 (Jocher, Jing, 2024), reconhecidas pelo equilíbrio entre precisão e eficiência computacional. A YOLOv8 corresponde à iteração consolidada mais recente no momento do estudo, enquanto a YOLO11 foi incorporada por se tratar de uma arquitetura recente, ainda pouco explorada em estudos aplicados à inspeção automatizada de infraestrutura civil, permitindo avaliar seu desempenho em um contexto real de detecção de fissuras em ciclovias de pavimento rígido. Para representar o paradigma baseado em Transformers, foi escolhida a arquitetura RT-DETR (Zhao *et al*, 2023), considerando sua concepção orientada a aplicações em tempo real e a adoção recente em estudos de detecção de objetos.

Considerando essas vertentes, foram selecionados para este estudo os modelos YOLOv8m, YOLO11m, YOLO11l e RT-DETR-l, cujas especificações de porte e complexidade estão detalhadas na Tabela 2. A seleção de diferentes portes (médio e grande) e paradigmas visa fornecer subsídios para definir a configuração mais eficiente para cenários reais de inspeção, possibilitando avaliar o impacto dessas variáveis no processo de treinamento e nos resultados.

Tabela 2 – Modelos utilizados para detecção

Modelo	Porte	Parâmetros (M)
YOLOv8m	Médio	25.9
YOLO11m	Médio	20.1
YOLO11l	Grande	25.3
RT-DETR-l	Grande	32.0

Fonte: O autor.

Definidas as arquiteturas, a etapa de treinamento seguiu as práticas consolidadas em visão computacional. O conjunto de dados anotado foi dividido aleatoriamente na proporção de 70% para treinamento, 20% para validação e 10% para teste. O subconjunto de treinamento foi utilizado para o ajuste dos pesos e aprendizado das características visuais das fissuras, enquanto o de validação foi empregado para monitoramento contínuo do desempenho, prevenção de *overfitting*, ajuste de hiperparâmetros e seleção do modelo com melhor desempenho. Por fim, o subconjunto de teste foi utilizado exclusivamente para a avaliação do desempenho final e da capacidade de generalização em dados não vistos durante o treinamento.

A estratégia de treinamento adotada baseou-se em *transfer learning*. Em vez de treinar os modelos a partir do zero, foram utilizados pesos pré-treinados em grandes bases de dados de referência¹, permitindo que os modelos partissem de representações visuais previamente aprendidas. A partir dessa inicialização, foi realizado um processo de *fine-tuning*, no qual os pesos das redes foram refinados para especializar cada modelo (Tabela 2) na tarefa específica de detecção de fissuras em pavimentos de ciclovias, aproveitando o conhecimento previamente adquirido sobre formas e texturas.

Com o objetivo de assegurar a padronização e a comparabilidade entre os experimentos, todos os modelos foram treinados utilizando os mesmos hiperparâmetros, conforme apresentado na Tabela 3.

Tabela 3 – Configuração dos hiperparâmetros adotados no treinamento dos modelos de detecção

Hiperparâmetro	Código	Valor
Tamanho da imagem de entrada	imgsz	1024
Taxa de aprendizado inicial	lr0	0.001
Tamanho do lote	batch	16
Número máximo de épocas	epochs	300
Paciência (parada antecipada)	patience	30

Fonte: O autor.

O tamanho das imagens de entrada foi fixado em 1024×1024 pixels, compatível com a etapa de recorte adotada na preparação do conjunto de dados. A taxa de aprendizado inicial foi definida como 0,001, o tamanho do lote (*batch size*) como 16, e o número máximo de épocas como 300, sendo empregado um critério de parada antecipada (*early stopping*) com paciência de 30 épocas. A padronização dessas variáveis visa garantir que as diferenças de desempenho observadas sejam decorrentes da arquitetura da rede, e não de variações na configuração do treinamento.

Todos os experimentos computacionais deste estudo, incluindo o treinamento e a avaliação dos modelos, foram implementados em linguagem Python, utilizando a biblioteca PyTorch. O ambiente de *hardware* consistiu em uma estação de trabalho equipada com processador AMD Ryzen 7 5700X3D (8-Core, 3.00 GHz), 32 GB de memória RAM e uma unidade de processamento gráfico (GPU) NVIDIA GeForce RTX 4090, garantindo capacidade

¹ O conjunto de dados utilizado para o pré-treinamento dos modelos foi o COCO (Common Objects in Context). Trata-se de uma base de dados de larga escala amplamente utilizada como *benchmark* em visão computacional, contendo mais de 330 mil imagens e 80 categorias de objetos rotulados, permitindo que a rede inicie o treinamento com capacidade robusta de extração de características (Lin et al., 2014).

computacional adequada para o treinamento de modelos de grande porte e para a execução dos experimentos de forma eficiente e reproduzível.

Os modelos treinados a partir desse procedimento foram, então, submetidos a uma etapa de avaliação quantitativa e qualitativa, descrita a seguir, com o objetivo de comparar seu desempenho e analisar sua adequação ao contexto de inspeção automatizada de ciclovias.

3.1.4 Avaliação de desempenho dos modelos de detecção – Etapa 4

A performance dos modelos de detecção de objetos foi avaliada de forma quantitativa para estabelecer uma comparação direta e objetiva. A análise é fundamentada na curva de *Precision-Recall* (PR), que representa a relação entre a precisão (P) e *recall* (R) do modelo em diferentes limiares de confiança. P mede a acurácia das detecções, indicando a proporção de identificações corretas dentre todas as detecções realizadas pelo modelo, enquanto R mede a capacidade do modelo de encontrar todas as fissuras que realmente existem no conjunto de dados. As formulações matemáticas para ambas as métricas são apresentadas nas Equações 1 e 2, respectivamente.

$$P = \frac{TP}{TP + FP} \quad \text{Equação 1}$$

$$R = \frac{TP}{TP + FN} \quad \text{Equação 2}$$

Onde:

TP representa os Verdadeiros Positivos (detecções corretas);

FP os Falsos Positivos (detecções incorretas);

FN os Falsos Negativos (fissuras não detectadas pelo modelo).

Para a classificação das detecções, adotou-se o critério da Interseção sobre a União (*Intersection over Union* – IoU), que quantifica a sobreposição geométrica entre a caixa delimitadora predita pelo modelo e a caixa de referência (*ground truth*) anotada manualmente. Uma predição foi considerada correta, sendo classificada como Verdadeiro Positivo, quando o valor de IoU entre as caixas fosse superior a 0,50. Predições com valores de IoU inferiores a esse limiar foram consideradas incorretas, contribuindo para a contabilização de falsos positivos ou falsos negativos, conforme o caso.

Para consolidar a curva PR em uma métrica única que quantifique o desempenho global do modelo, utilizou-se a Precisão Média (*Average Precision* - AP), definida como a área sob a curva PR, conforme apresentado na Equação 3. A AP sintetiza a capacidade do modelo de manter alta precisão em diferentes níveis de *recall*, ou seja, valores mais elevados de AP indicam melhor desempenho do modelo, refletindo simultaneamente maior precisão nas detecções e maior capacidade de identificar a maioria das ocorrências reais das manifestações patológicas no conjunto de dados.

$$AP = \int_0^1 P(R) d(R) \quad \text{Equação 3}$$

Como métrica principal de comparação entre os modelos, foi adotada a mAP com limiar de IoU igual a 0,50, denominada mAP₅₀. Essa métrica fornece uma avaliação sintética e abrangente da capacidade geral de detecção dos modelos. Considerando que este estudo se concentra em uma única classe de objeto (“fissura”), o valor de mAP é numericamente equivalente ao valor da AP calculada para essa classe.

Para garantir a consistência e a reprodutibilidade do processo de avaliação, o cálculo das métricas foi realizado por meio de um *script* desenvolvido em Python, que executou os modelos finalizados sobre o conjunto de teste previamente separado do *dataset*.

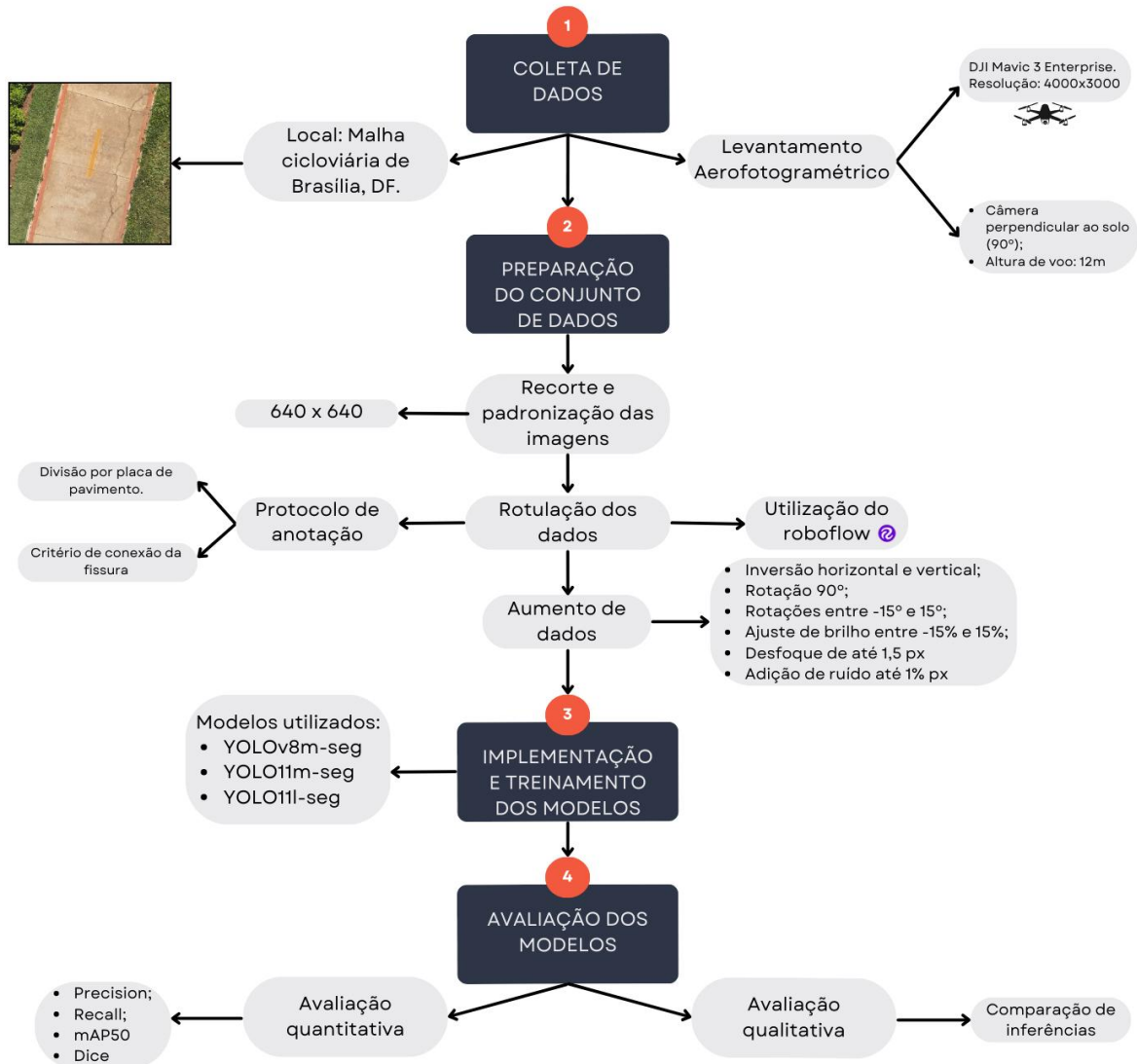
Além da avaliação quantitativa, foi realizada uma análise qualitativa para examinar a robustez das arquiteturas frente a condições adversas comuns em levantamentos aerofotogramétricos. Essa avaliação consistiu na comparação visual das inferências geradas por cada modelo em um conjunto de imagens não utilizadas no treinamento, abrangendo diferentes cenários de iluminação (exposição solar e sombreamento), desfoque e interferências visuais, como sujeira ou marcas no pavimento. As detecções foram comparadas lado a lado para cada cenário, permitindo identificar padrões de acerto e erro, bem como pontos fortes e limitações específicas de cada arquitetura no contexto de aplicação proposto.

3.2 SEÇÃO 2 – DESENVOLVIMENTO E TREINAMENTO DE MODELOS DE INTELIGÊNCIA ARTIFICIAL PARA SEGMENTAÇÃO DE FISSURAS

Nesta seção é apresentada a metodologia adotada para o desenvolvimento e a avaliação dos modelos de segmentação de fissuras em ciclovias de pavimento rígido. A abordagem evolui da detecção de objetos, cujo objetivo é identificar a presença e a localização aproximada das

fissuras por meio de caixas delimitadoras, para a segmentação de instância, que visa a delimitação das fissuras em nível de pixel, gerando máscaras que permitam, futuramente, a extração de propriedades físicas como extensão e largura. Esse procedimento demanda maior detalhamento espacial e maior rigor no processo de anotação dos dados. O fluxo metodológico adaptado para esta tarefa está ilustrado na Figura 6.

Figura 6 – Fluxo metodológico adotado para o desenvolvimento e avaliação dos modelos de segmentação de fissuras em ciclovias



Fonte: O autor.

Em função das especificidades inerentes à tarefa de segmentação, algumas adaptações metodológicas foram necessárias em relação ao fluxo apresentado na Seção 3.1. Essas adaptações envolvem, principalmente, a seleção e a aquisição do conjunto de dados, a estratégia de recorte e padronização das imagens, o tipo de anotação empregado e a configuração dos modelos utilizados. Ainda que a estrutura geral do fluxo metodológico seja mantida, as

modificações introduzidas refletem as demandas computacionais mais elevadas e a necessidade de maior precisão geométrica associadas à segmentação de fissuras.

O fluxo metodológico adotado nesta etapa é organizado em quatro etapas principais: (1) coleta de dados, (2) preparação do conjunto de dados, (3) implementação e treinamento dos modelos de segmentação e (4) avaliação dos modelos.

3.2.1 Coleta de dados – Etapa 1

A coleta de dados para a etapa de segmentação de fissuras foi realizada em trechos de ciclovias com pavimento de concreto rígido localizadas em Brasília, DF, por meio de levantamento aerofotogramétrico com veículo aéreo não tripulado (VANT). Diferentemente da etapa de detecção, nesta fase foram utilizadas exclusivamente imagens obtidas com o VANT DJI Mavic 3 Enterprise, em voos automatizados realizados a uma altura aproximada de 12 m em relação ao nível do pavimento.

A adoção dessa altura de voo está diretamente associada às exigências da tarefa de segmentação, que demanda maior nível de detalhamento espacial para a delimitação precisa das fissuras em nível de pixel. Em alturas mais elevadas, a resolução espacial resultante pode comprometer a identificação das bordas e a continuidade das fissuras, dificultando a geração de máscaras acuradas. Nesse sentido, o voo a 12 m representa a menor altura operacional viável em levantamentos automatizados com VANT de uso profissional, permitindo maximizar a resolução das imagens sem recorrer a voos manuais, que seriam menos representativos de cenários práticos de inspeção em larga escala.

Além do critério técnico relacionado à resolução espacial, a seleção desse subconjunto de dados buscou aproximar a metodologia das condições reais de aplicação, nas quais um operador tende a empregar voos automatizados para o levantamento de extensos trechos de ciclovias ou outras infraestruturas urbanas. Dessa forma, a estratégia adotada prioriza a viabilidade operacional do método proposto, alinhando o processo de aquisição de dados às práticas mais prováveis em aplicações de engenharia. Adicionalmente, essa restrição amostral permite avaliar o desempenho dos modelos em um cenário controlado de menor variabilidade de sensores, isolando a capacidade intrínseca de segmentação da influência de resoluções mistas.

O conjunto de imagens utilizado nesta etapa corresponde, portanto, a um recorte específico do banco de dados empregado na etapa de detecção, oriundo das coletas sintetizadas na Tabela 1,

selecionado de forma a atender simultaneamente às demandas computacionais da segmentação e aos objetivos do estudo. Essa seleção possibilita investigar o desempenho dos modelos de segmentação em um contexto controlado e específico, mantendo coerência com a abordagem metodológica adotada e permitindo avaliar, nos resultados, a eficácia da segmentação mesmo a partir de um conjunto de dados menos extenso, porém tecnicamente adequado à tarefa.

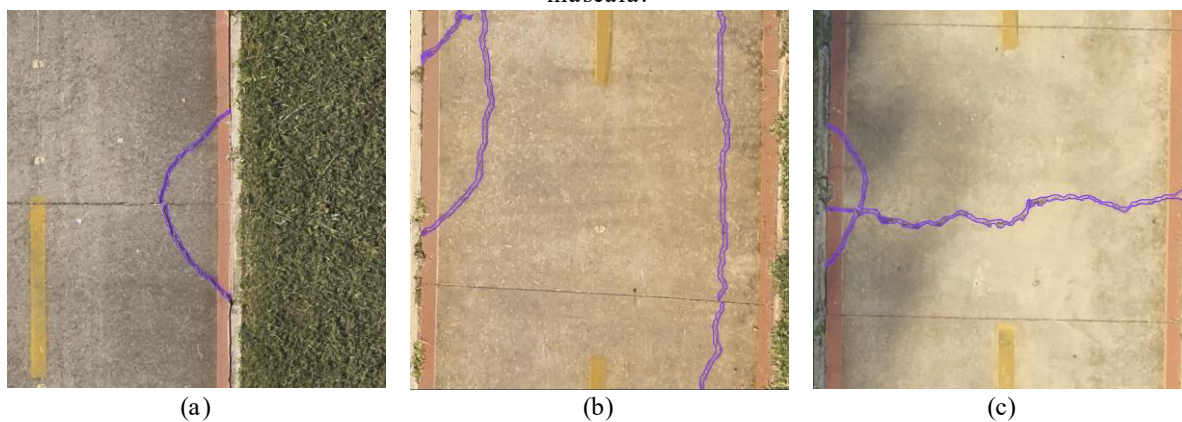
3.2.2 Preparação do conjunto de dados – Etapa 2

Uma vez selecionado o subconjunto de imagens de alta resolução, procedeu-se à etapa de preparação do conjunto de dados, que seguiu princípios metodológicos semelhantes aos adotados na etapa de detecção, com adaptações específicas decorrentes das particularidades da segmentação. Essas adaptações envolveram a estratégia de recorte e padronização das imagens, o tipo de anotação empregado e o rigor do processo de validação manual.

Diferentemente da abordagem adotada na detecção, o recorte das imagens em ROI foi padronizado em 640×640 pixels. Essa redução dimensional, em comparação aos 1024 pixels utilizados anteriormente, justifica-se pela maior complexidade das redes de segmentação, que demandam mais memória da GPU para processar as máscaras de pixels. A escolha dessa resolução também se alinha à entrada nativa dos modelos empregados, garantindo um equilíbrio adequado entre custo computacional e precisão. Dessa forma, foi possível reduzir o esforço de processamento associado ao treinamento, sem comprometer o nível de detalhe necessário para a delimitação das fissuras.

Na sequência, a etapa de rotulação das fissuras foi realizada na plataforma Roboflow, introduzindo outra alteração metodológica desta seção: a migração da anotação por caixas delimitadoras para a demarcação por polígonos. Nesse processo, cada fissura visível foi delimitada manualmente por um polígono ajustado à sua geometria real, permitindo a geração de máscaras representativas das áreas fissuradas. Para garantir a consistência entre as etapas de detecção e segmentação, manteve-se o protocolo de anotação previamente estabelecido, incluindo a divisão das anotações por placa de pavimento e o critério de conexão visual das fissuras. Fissuras contínuas dentro de uma mesma placa foram representadas por um único polígono, enquanto fissuras desconectadas foram anotadas separadamente. Exemplos desse processo de anotação são apresentados na Figura 7.

Figura 7 – Exemplos de aplicação do protocolo de anotação das fissuras em ciclovias de pavimento rígido (segmentação): (a) fissura segmentada conforme os limites das placas de concreto; (b) fissuras sem conexão física aparente rotuladas individualmente; (c) ramificações conecta das visualmente são agrupadas em uma única máscara.



Fonte: O autor.

Com o objetivo de otimizar o processo de anotação e reduzir o esforço manual, adotou-se uma abordagem semi-automatizada semelhante à empregada na etapa de detecção. Inicialmente, um subconjunto das imagens foi anotado manualmente, sendo utilizado para o treinamento de um modelo preliminar de segmentação dentro do ambiente Roboflow. Em seguida, a funcionalidade *Label Assist* da plataforma foi empregada para gerar sugestões automáticas de polígonos nas imagens restantes, que foram posteriormente avaliadas.

Diferentemente da etapa de detecção, o processo de validação manual nesta fase foi conduzido de forma mais rigorosa, uma vez que pequenas imprecisões na delimitação das bordas das fissuras podem impactar significativamente o desempenho dos modelos de segmentação. Todas as máscaras geradas foram revisadas, com ajustes vértice a vértice, de modo a assegurar que os polígonos acompanhassem fielmente a geometria irregular das fissuras observadas nas imagens, excluindo áreas de pavimento íntegro. Essa precisão é fundamental para que o modelo aprenda a segmentar a patologia em si e não apenas a sua vizinhança.

Por fim, após a conclusão da anotação e validação das máscaras, aplicou-se o processo de aumento de *data augmentation*, replicando as mesmas técnicas e parâmetros definidos na etapa de detecção, incluindo rotações, espelhamentos, ajustes de brilho, desfoque e ruído. A adoção das mesmas estratégias de aumento de dados visou ampliar a variabilidade do conjunto de treinamento e aumentar a robustez dos modelos, mantendo consistência metodológica entre as diferentes etapas do estudo.

3.2.3 Implementação e treinamento dos modelos de segmentação – Etapa 3

A implementação e o treinamento dos modelos de segmentação de fissuras seguiram uma abordagem baseada em aprendizado profundo, utilizando arquiteturas da família YOLO adaptadas para a tarefa de segmentação de instância. A escolha por essas arquiteturas visou manter coerência metodológica com a etapa de detecção, além de possibilitar a comparação indireta entre os desempenhos das abordagens de detecção e segmentação.

Foram avaliados três modelos de segmentação derivados das arquiteturas YOLOv8 (Jocher, Ayush, Jing, 2023) e YOLO11 (Jocher, Jing, 2024): YOLOv8m-seg, YOLO11m-seg e YOLO11l-seg. Esses modelos compartilham a mesma base arquitetural e os mesmos princípios gerais de extração de características e geração de previsões, diferindo principalmente quanto ao porte e ao número de parâmetros, conforme apresentado na Tabela 4. Essa variação permite analisar o impacto da complexidade do modelo na qualidade das máscaras geradas e no custo computacional associado, possibilitando investigar o equilíbrio entre precisão da segmentação e eficiência computacional, aspecto relevante para aplicações práticas de inspeção automatizada.

Tabela 4 – Modelos utilizados para segmentação

Modelo	Porte	Parâmetros (M)
YOLOv8m-seg	Médio	27.3
YOLO11m-seg	Médio	22.4
YOLO11l-seg	Grande	27.6

Fonte: O autor.

O conjunto de dados anotado foi dividido aleatoriamente em três subconjuntos distintos, mantendo a proporção adotada na etapa de detecção: 70% das imagens foram destinadas ao treinamento, 20% à validação e 10% ao teste.

A estratégia de treinamento adotada baseou-se em *transfer learning*, utilizando pesos pré-treinados disponibilizados pelos desenvolvedores das arquiteturas YOLOv8 e YOLO11. A partir desses pesos iniciais, realizou-se o *fine-tuning* dos modelos para a tarefa específica de segmentação de fissuras em ciclovias, permitindo acelerar a convergência do treinamento e melhorar a capacidade de generalização, especialmente considerando o volume reduzido do conjunto de dados.

Para assegurar a comparabilidade entre os experimentos, manteve-se o ambiente computacional, descrito em 3.1.3, e a maioria dos hiperparâmetros. A única alteração foi a

redução da dimensão de entrada (*imgsz*) para 640 pixels, conforme justificado anteriormente pela demanda de memória da tarefa de segmentação. Os parâmetros específicos adotados no treinamento estão detalhados na Tabela 5.

Tabela 5 – Configuração dos hiperparâmetros adotados no treinamento dos modelos de segmentação

Hiperparâmetro	Código	Valor
Tamanho da imagem de entrada	<i>imgsz</i>	640
Taxa de aprendizado inicial	<i>lr0</i>	0.001
Tamanho do lote	<i>batch</i>	16
Número máximo de épocas	<i>epochs</i>	300
Paciência (parada antecipada)	<i>patience</i>	30

Fonte: O autor.

3.2.4 Avaliação de desempenho dos modelos de segmentação – Etapa 4

A avaliação do desempenho dos modelos de segmentação foi realizada de forma quantitativa, utilizando o conjunto de teste previamente definido e o *script* computacional descrito na seção anterior. Nesta etapa, buscou-se mensurar a capacidade dos modelos em delimitar corretamente as fissuras em nível de pixel, a partir da comparação entre as máscaras segmentadas automaticamente e as máscaras de referência anotadas manualmente.

Para garantir consistência metodológica com a etapa de detecção, adotaram-se as mesmas métricas globais de desempenho empregadas na Seção anterior. Embora os conceitos matemáticos de *precision*, *recall* e *mAP* permaneçam os mesmos apresentados nas Equações 1, 2 e 3 (Seção 3.1.4), a base de cálculo para a classificação dos acertos sofreu uma alteração conceitual com a substituição do *IoU* de caixas pelo *IoU* de Máscaras (*Mask IoU*).

Diferentemente da detecção, onde se avalia a sobreposição de retângulos, o *Mask IoU* calcula a interseção pixel a pixel entre a máscara predita pelo modelo e a máscara de referência (polígono anotado manualmente). Essa métrica é significativamente mais rigorosa, pois penaliza predições que, embora localizem a fissura, falham em delinear sua morfologia exata. Dessa forma, a métrica principal adotada para a comparação dos modelos foi a *mAP*₅₀, essa métrica sintetiza o desempenho global do modelo ao considerar diferentes limiares de confiança, sendo calculada a partir da área sob a curva PR, considerando um limiar mínimo de *IoU* igual a 0,50 entre as máscaras segmentadas e as máscaras de referência.

No entanto, considerando as particularidades das métricas de segmentação e a necessidade de comparabilidade com o estado da arte, foram incorporados dois indicadores adicionais: o F1-

Score e o Coeficiente Dice. Em tarefas de segmentação binária, o Coeficiente Dice é matematicamente equivalente ao F1-*Score* (média harmônica entre precisão e *recall*). Portanto, ambos foram derivados diretamente dos valores de *Precision* e *Recall* obtidos no conjunto de teste, conforme a Equação 4:

$$\text{Dice} = \text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Equação 4}$$

Essa métrica quantifica a similaridade espacial entre a predição e o gabarito, sendo amplamente utilizada para validar a fidelidade geométrica da segmentação e permitindo a comparação direta com trabalhos que reportam o Dice ou o *Fscore*.

Dessa forma, a métrica principal adotada para a hierarquização dos modelos foi a mAP₅₀, enquanto o F1-*Score* e o Dice serviram como indicadores complementares para validação cruzada com outros trabalhos acadêmicos. Todas as métricas foram computadas exclusivamente sobre o conjunto de teste, composto por imagens não utilizadas durante o treinamento, assegurando uma avaliação imparcial da capacidade de generalização dos modelos. A análise comparativa entre os diferentes modelos de segmentação considerou não apenas os valores absolutos das métricas obtidas, mas também a influência do porte do modelo e do número de parâmetros no desempenho alcançado, permitindo avaliar o compromisso entre a qualidade das máscaras geradas e o custo computacional associado.

Além da avaliação quantitativa, foi conduzida uma análise qualitativa para examinar a fidelidade geométrica das máscaras geradas e a capacidade dos modelos em delinear corretamente a morfologia das fissuras. Essa avaliação consistiu na comparação visual direta entre as máscaras preditas e o *ground truth* em imagens do conjunto de teste, abrangendo diferentes padrões de fissuração (lineares, ramificadas e curvilíneas), variações de contraste e a presença de elementos de interferência, como sinalização horizontal. As segmentações foram analisadas quanto à sua aderência aos contornos reais, preservação da continuidade e espessura do traçado, permitindo identificar a precisão dos modelos em nível de pixel e sua robustez frente à complexidade visual do pavimento.

3.3 SEÇÃO 3 – ESTUDO DE CASO E INTEGRAÇÃO IA-GIS

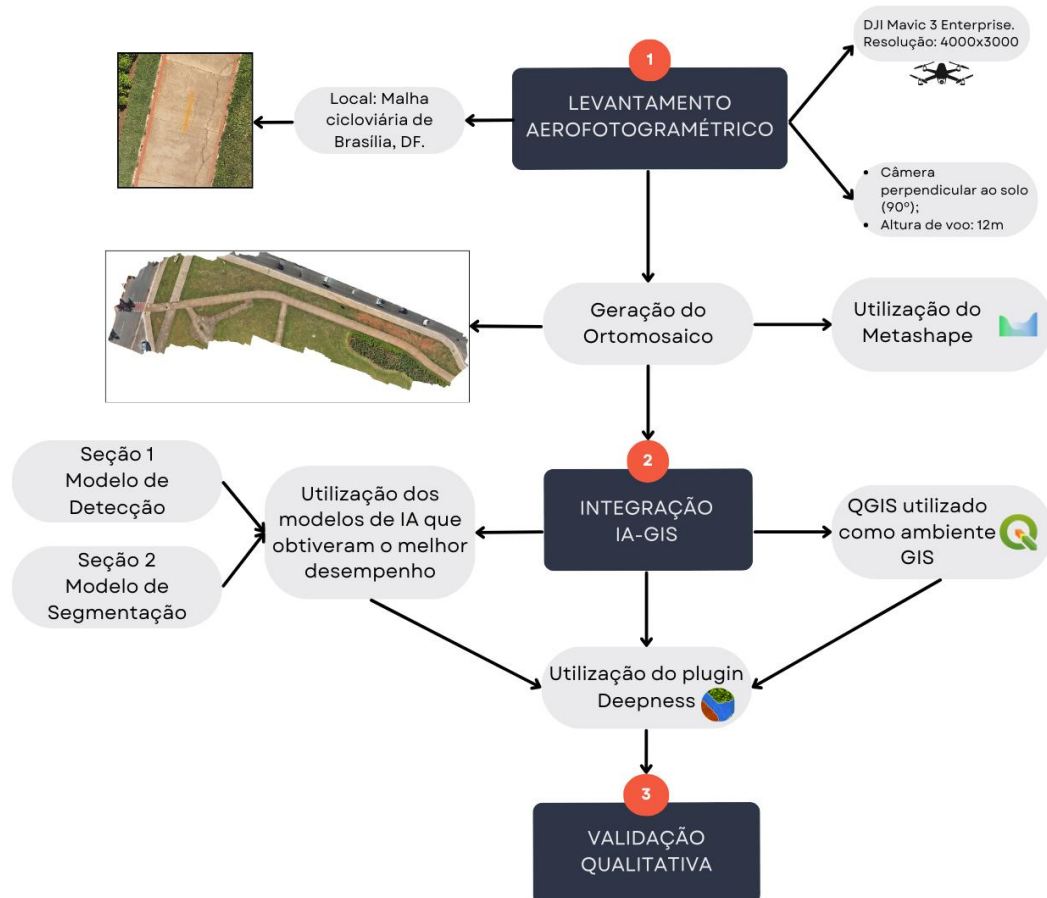
Nesta seção é apresentado o estudo de caso desenvolvido para a validação externa e a aplicação prática dos modelos de detecção e segmentação de fissuras treinados nas etapas anteriores, bem como sua integração em ambiente GIS. Diferentemente das seções anteriores, voltadas ao

desenvolvimento e à avaliação quantitativa dos modelos de IA, esta etapa tem como objetivo validar a capacidade de generalização dos modelos em dados inéditos e consolidar um fluxo de integração para o mapeamento automatizado de patologias.

Para esse fim, foi realizado um novo levantamento aerofotogramétrico de um trecho de ciclovia localizado em Brasília, DF, seguido da geração de um ortomosaico georreferenciado. Os modelos previamente treinados foram então integrados ao ambiente GIS por meio de uma abordagem baseada em ferramentas de código aberto (*open source*), utilizando o *software* QGIS (QGIS, 2025) e o complemento (*plugin*) *Deepness* (Aszkowski *et al.*, 2023). Essa integração possibilitou a execução dos modelos diretamente sobre o ortomosaico georreferenciado, estabelecendo um fluxo metodológico que conecta a detecção e a segmentação automatizada de fissuras à representação geográfica das manifestações patológicas, com potencial aplicação em atividades de inspeção, mapeamento e apoio à manutenção de infraestruturas urbanas.

O fluxo metodológico adotado para a integração entre inteligência artificial e GIS, bem como as principais etapas envolvidas desde a aquisição dos dados até a validação dos resultados, é apresentado na Figura 8. A validação nesta etapa é conduzida de forma predominantemente qualitativa e espacial, considerando a coerência geométrica das predições, a continuidade das fissuras identificadas e a aderência dos resultados ao contexto real de inspeção, com potencial aplicação em atividades de mapeamento, inspeção e apoio à manutenção de infraestruturas urbanas.

Figura 8 – Fluxo metodológico adotado para o desenvolvimento do estudo de caso e integração IA-GIS



Fonte: O autor.

3.3.1 Levantamento aerofotogramétrico e geração do ortomosaico – Etapa 1

Com o objetivo de validar externamente os modelos de detecção e segmentação de fissuras desenvolvidos nas seções anteriores e avaliar sua aplicação em um cenário real de inspeção, foi realizado um novo levantamento aerofotogramétrico em um trecho de ciclovia de pavimento rígido localizado em Brasília, DF. Este trecho não foi utilizado em nenhuma etapa de treinamento, validação ou teste dos modelos, garantindo a independência dos dados empregados nesta fase do estudo.

A aquisição das imagens foi realizada por meio de um voo aerofotogramétrico utilizando um VANT DJI Mavic 3 Enterprise, equipado com módulo RTK, visando assegurar maior precisão no georreferenciamento dos produtos gerados. Durante o levantamento, a câmera foi mantida posicionada perpendicularmente ao pavimento (ângulo de 90°), de modo a minimizar distorções geométricas e garantir a captura ortogonal da superfície da ciclovia.

O levantamento foi realizado a uma altura nominal de voo de 12 m em relação ao nível do pavimento, adotando-se a mesma configuração operacional utilizada na etapa de segmentação (3.2.1). Essa escolha visa assegurar a compatibilidade entre as condições de aquisição dos dados e aquelas empregadas no desenvolvimento dos modelos, permitindo avaliar seu desempenho em um cenário coerente com o domínio de aplicação considerado.

As imagens adquiridas foram posteriormente processadas no *software* Agisoft Metashape (AGISOFT, 2025), seguindo o fluxo padrão de processamento fotogramétrico, incluindo o alinhamento das imagens, a reconstrução da nuvem de pontos, a geração do modelo de superfície e, por fim, a criação do ortomosaico georreferenciado. O produto foi exportado no formato *raster GeoTIFF* (arquivo .tif), preservando as informações espaciais necessárias para sua posterior integração em ambiente GIS.

Quanto à qualidade do georreferenciamento, a utilização de um VANT equipado com módulo RTK contribuiu para reduzir as incertezas associadas ao posicionamento por satélite do equipamento em modo autônomo, resultando em maior consistência posicional dos produtos gerados. Além disso, a aquisição com câmera nadiral (90°) e a altura controlada favorecem a ortorretificação do pavimento, reduzindo distorções geométricas. Como verificação operacional, o ortomosaico foi sobreposto a camadas de referência no QGIS, observando-se coerência no alinhamento de feições do pavimento (bordas e elementos lineares), o que indica consistência espacial suficiente para as análises propostas. Assim, o produto GeoTIFF georreferenciado é considerado adequado ao objetivo do estudo de espacializar as inferências no ambiente GIS.

O ortomosaico gerado constitui a base espacial para a aplicação dos modelos de inteligência artificial desenvolvidos neste trabalho, permitindo a execução das inferências diretamente sobre um produto georreferenciado contínuo e representativo do pavimento da ciclovia, de modo que as detecções e segmentações preservem a referência espacial do levantamento.

3.3.2 Integração dos modelos de inteligência artificial ao ambiente GIS

A integração entre os modelos de IA desenvolvidos e o ambiente GIS foi realizada por meio de uma abordagem baseada em ferramentas *open source*, adotando o *software* QGIS como plataforma central de processamento e análise espacial. Essa etapa teve como objetivo viabilizar a aplicação prática dos modelos de detecção e segmentação de fissuras diretamente sobre o

ortomosaico georreferenciado, permitindo a avaliação do desempenho dos modelos em um contexto espacial real e operacional.

Inicialmente, o ortomosaico gerado foi importado para o ambiente QGIS no formato *GeoTIFF*, preservando suas informações de referência espacial. Para a execução dos modelos de inteligência artificial sobre esse dado *raster*, utilizou-se o *plugin Deepness* (Aszkowski *et al.*, 2023), que possibilita a integração de modelos de aprendizado profundo ao fluxo de trabalho do QGIS, permitindo a realização de inferências diretamente sobre imagens georreferenciadas, convertendo as predições de detecção e segmentação para camadas vetoriais.

O processamento de ortomosaicos de grandes dimensões pelo *plugin* é realizado por meio de uma estratégia de inferência fatiada. Neste processo, a imagem georreferenciada é particionada automaticamente em *tiles* com dimensões definidas pela resolução de entrada nativa de cada modelo carregado (1024x1024 pixels para os modelos de detecção e 640x640 pixels para os de segmentação), sendo o resultado a recomposição vetorial das predições individuais de cada bloco.

Para garantir a reprodutibilidade dos experimentos e a compatibilidade entre as bibliotecas de processamento, foi estabelecido um ambiente computacional controlado. As especificações detalhadas das versões de *software*, interpretadores e complementos utilizados na etapa de integração estão listadas na Tabela 6.

Tabela 6 – Especificações do ambiente de *software* utilizado na integração IA-GIS

<i>Software</i> / Componente	Versão / Especificação
<i>Software</i> GIS	QGIS 3.40.6 'Bratislava'
Linguagem de Script	Python 3.12.10 (Embarcado no QGIS)
Complemento (Plugin)	Deepness 0.6.5

Fonte: O autor.

Para garantir a interoperabilidade entre o ambiente de treinamento (PyTorch) e o ambiente de produção (QGIS), foi necessária a conversão dos modelos treinados para o formato padrão aberto ONNX (*Open Neural Network Exchange*), compatível com o mecanismo de execução utilizado pelo *plugin*. Esse procedimento, executado via *script* em Python, garantiu a preservação dos pesos treinados e da arquitetura original dos modelos, bem como a

reprodutibilidade do processo, dissociando a execução da dependência das bibliotecas originais de treinamento.

Durante os testes preliminares de validação da integração, no entanto, identificou-se uma limitação técnica no algoritmo de vetorização nativo do complemento Deepness. Embora a geração de caixas delimitadoras (detecção) apresentasse resultados satisfatórios, o processo de segmentação aplicava uma simplificação baseada em envoltória convexa (*convex hull*) aos polígonos gerados. Esse comportamento, embora eficiente para feições regulares e compactas (como edificações ou lotes), mostrou-se inadequado para a representação de fissuras, cuja morfologia é irregular e não convexa. A aplicação desse algoritmo resultava na eliminação das concavidades e ramificações da patologia, comprometendo a fidelidade do mapeamento das anomalias.

Diante dessa limitação, foram realizadas adaptações no código-fonte do complemento, com o objetivo de ajustar o processo de vetorização das máscaras de segmentação. O algoritmo de pós-processamento foi modificado para utilizar o contorno exato da predição, substituindo a aproximação convexa pela preservação integral dos vértices gerados pelo modelo de segmentação. Essa adaptação metodológica garantiu que a camada vetorial final (*shapefile*) passasse a apresentar contornos mais precisos e compatíveis com a geometria real das manifestações patológicas observadas no pavimento.

Essa integração permitiu não apenas a execução dos modelos em ambiente GIS, mas também a geração de produtos espaciais passíveis de análise, como camadas vetoriais de fissuras associadas a coordenadas geográficas reais. Dessa forma, a etapa de integração IA–GIS estabelece um elo entre os resultados obtidos em nível de imagem e sua aplicação em contextos de inspeção, monitoramento e suporte à tomada de decisão em infraestrutura viária.

3.3.3 Procedimento de validação qualitativa e organização dos resultados

Após a integração dos modelos de detecção e segmentação ao ambiente GIS e a execução das inferências sobre o ortomosaico georreferenciado, foi definido um procedimento sistemático para a validação qualitativa dos resultados obtidos. Esta etapa teve como objetivo avaliar, de forma visual e geométrica, a adequação das saídas dos modelos ao contexto real de aplicação, complementando a avaliação quantitativa realizada nas seções anteriores. A opção por uma análise qualitativa nesta fase justifica-se pela inexistência de rótulos anotados (*ground truth*) para o ortomosaico da área de estudo de caso, o que inviabiliza o cálculo de métricas estatísticas

tradicionais. Além disso, buscou-se priorizar a validação da capacidade de generalização do modelo em um cenário de uso prático, focando na utilidade dos dados georreferenciados para a gestão de ativos.

Para o estudo de caso, foram selecionados os modelos que apresentaram melhor desempenho nas avaliações quantitativas das etapas de detecção e segmentação, conforme descrito nas Seções 3.1.4 e 3.2.4, respectivamente. Essa seleção permitiu concentrar a análise qualitativa nas configurações mais promissoras, reduzindo a influência de modelos com desempenho inferior e garantindo coerência entre as etapas metodológicas.

A validação qualitativa foi conduzida no ambiente QGIS por meio da sobreposição das saídas dos modelos, caixas delimitadoras e máscaras segmentadas vetorizadas, ao ortomosaico georreferenciado. Essa abordagem possibilitou a inspeção visual assistida dos resultados, considerando a correspondência entre as fissuras identificadas automaticamente e as feições visíveis no pavimento. A análise foi realizada de forma exploratória e sistemática ao longo de todo o trecho da ciclovia.

Como critérios para a validação qualitativa, foram considerados aspectos relacionados à aderência geométrica das detecções e segmentações ao pavimento real, incluindo: a coerência espacial das fissuras identificadas, verificando se o posicionamento das caixas e das máscaras corresponde ao traçado visível no ortomosaico; a integridade geométrica, avaliada pela continuidade e extensão das feições, a fim de observar se o modelo preservou o comprimento total da patologia sem fragmentações excessivas; e a confiabilidade da classificação, analisando tanto a ocorrência de falsos positivos (detecções incorretas em áreas sãs) quanto de falsos negativos (omissão de fissuras visíveis no pavimento).

Os resultados da inferência foram organizados em camadas vetoriais distintas no ambiente GIS, separando-se as saídas provenientes dos modelos de detecção e de segmentação. Essa organização permitiu a visualização individual ou conjunta das diferentes representações das fissuras, bem como sua sobreposição ao ortomosaico de referência. A estruturação dos dados nesse formato visa facilitar análises posteriores, possibilitando sua utilização em procedimentos de mapeamento, quantificação e apoio à tomada de decisão em atividades de monitoramento e manutenção de ciclovias.

As análises e as discussões dos resultados obtidos a partir desse procedimento de validação qualitativa são apresentadas no capítulo seguinte, no qual se avalia o desempenho dos modelos

no contexto do estudo de caso e se discutem as implicações práticas da integração entre inteligência artificial e sistemas de informação geográfica para a inspeção automatizada de infraestruturas ciclovárias.

4 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados e discutidos os resultados obtidos a partir do desenvolvimento, treinamento e aplicação dos modelos de inteligência artificial propostos nesta pesquisa. A análise está organizada de forma a refletir a estrutura metodológica adotada no Capítulo 3, permitindo uma leitura contínua e coerente entre as etapas de método e os respectivos resultados alcançados.

A discussão inicia-se pela análise do desempenho dos modelos de detecção de fissuras, seguida pela avaliação quantitativa dos modelos de segmentação. Por fim, apresenta-se a validação externa da metodologia através do estudo de caso, onde se discutem os resultados da integração dos modelos ao ambiente GIS e sua aplicabilidade prática na inspeção de infraestruturas.

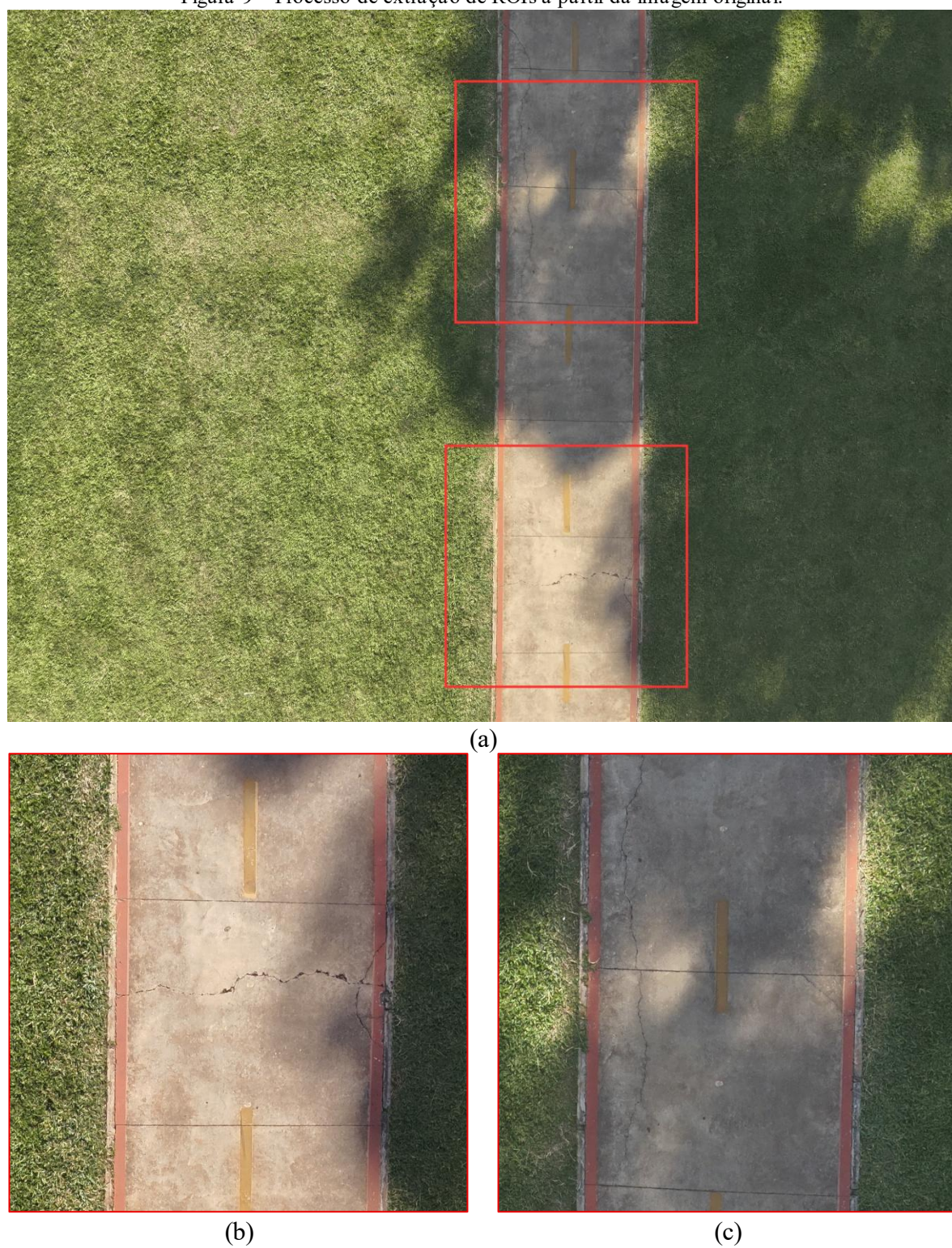
4.1 DESEMPENHO DOS MODELOS DE DETECÇÃO DE FISSURAS EM CICLOVIAS (SEÇÃO 1)

Esta seção apresenta e discute os resultados obtidos para os modelos de detecção automática de fissuras em ciclovias de pavimento rígido. A análise concentra-se no desempenho quantitativo dos modelos avaliados, bem como na interpretação qualitativa das inferências geradas, considerando as particularidades das imagens aerofotogramétricas utilizadas, as potencialidades e limitações das arquiteturas analisadas no contexto da aplicação proposta.

4.1.1 Caracterização do conjunto de dados para detecção

O conjunto de dados inicial, obtido por meio do levantamento aerofotogramétrico, foi composto por imagens brutas com resoluções nativas de 5472x3078 e 4000x3000 pixels. Após as etapas de preparação metodológica, que incluiu o recorte de regiões de interesse, o *dataset* consistiu em 738 imagens, todas com a dimensão padronizada de 1024x1024 pixels. A Figura 9 ilustra o processo de extração das ROIs, no qual a imagem original (a) apresenta áreas destacadas para recorte, enquanto as imagens (b) e (c) mostram exemplos de fissuras sob diferentes condições de iluminação, incidência solar direta e sombreamento, respectivamente.

Figura 9 – Processo de extração de ROIs a partir da imagem original.



Fonte: O autor.

O processo de rotulação deste conjunto de imagens resultou em um total de 1.757 anotações de caixas delimitadoras, distribuídas na classe única "fissura". Em média, o banco de dados possui 2,4 anotações por imagem, indicando que múltiplas fissuras por imagem são um cenário

comum. A Tabela 7 detalha a distribuição completa do *dataset*, apresentando a divisão tanto das imagens quanto das anotações entre os conjuntos de treinamento, validação e teste.

Tabela 7 – Distribuição de imagens e anotações entre os conjuntos de dados de detecção

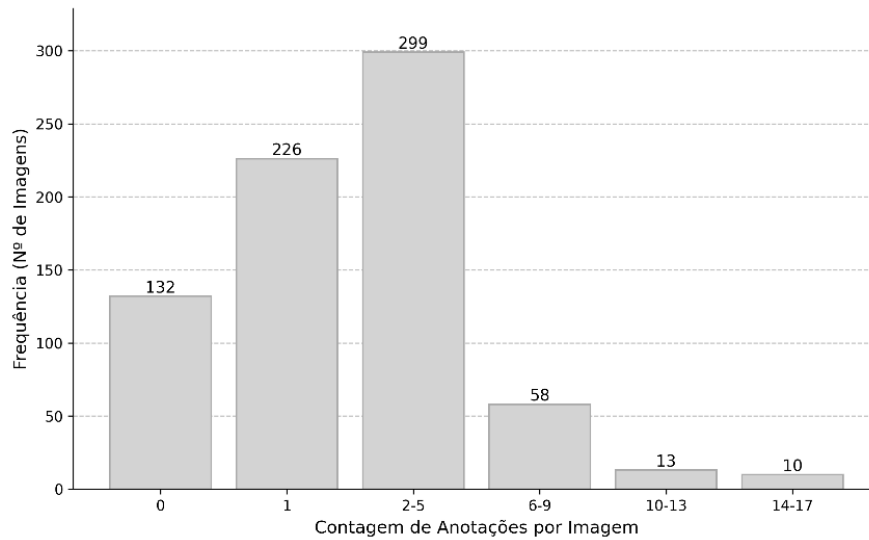
Conjunto	Nº de Imagens	Proporção de Imagens (%)	Nº de Anotações	Proporção de Anotações (%)
Treinamento	516	70	1157	66
Validação	148	20	408	23
Teste	74	10	192	11
Total	738	100	1757	100

Fonte: O autor.

Ao analisar a Tabela 7, nota-se uma alta correspondência entre a proporção de imagens e a de anotações em cada conjunto. A distribuição das anotações acompanha de perto a divisão de 70-20-10 das imagens, com diferenças percentuais mínimas em cada subconjunto. Essa proximidade geral na densidade de fissuras entre os conjuntos valida o desafio proposto e garante que a avaliação de desempenho dos modelos seja justa e fidedigna.

Para analisar a densidade de fissuras por imagem, a Figura 10 apresenta o histograma da contagem de anotações. Observa-se que a maioria das imagens possui entre 1 e 5 anotações, e que uma porção significativa do *dataset* (132 imagens) corresponde a amostras negativas (pavimento íntegro), essenciais para treinar o modelo a reduzir falsos positivos. Adicionalmente, o histograma revela a existência de menos imagens com altíssima densidade de fissuras (acima de 10 anotações), garantindo que o modelo seja exposto a cenários de dano mais severo. Observa-se que essa maior concentração de anotações por imagem tende a ocorrer nos recortes provenientes de voos em maiores altitudes. Nesses casos, uma área de 1024x1024 pixels representa uma seção maior do pavimento, aumentando a probabilidade de abranger múltiplas fissuras em uma única amostra.

Figura 10 – Histograma da distribuição de anotações por imagem no conjunto de detecção



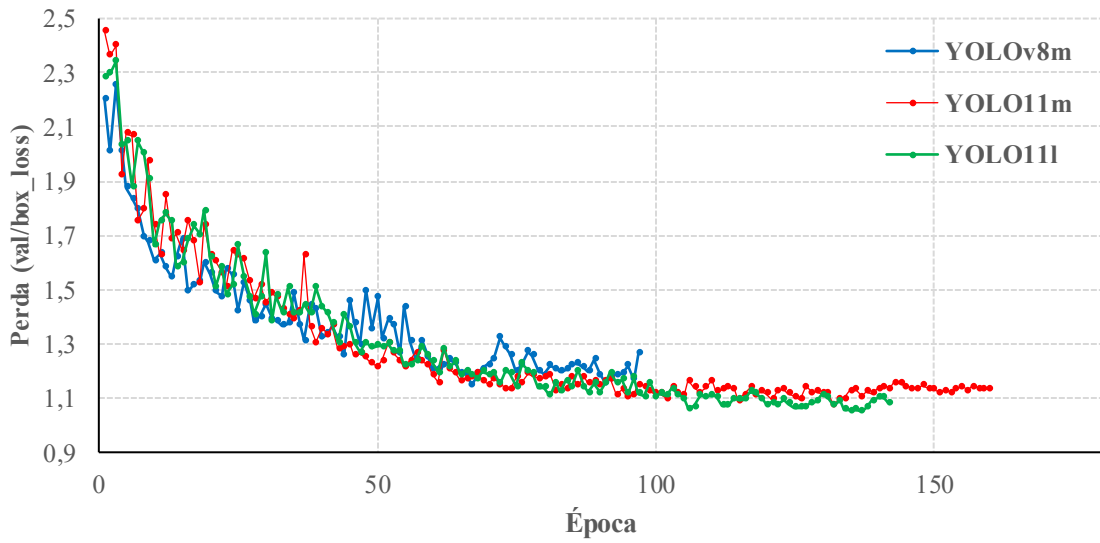
Fonte: O autor.

Desta forma, partindo das 738 imagens únicas, a aplicação das técnicas de aumento de dados (conforme detalhado na Seção 3.2.3) sobre o conjunto de treinamento resultou em um total de 3.612 amostras para esta etapa. Somando-se as 148 imagens de validação e as 74 de teste, o conjunto de dados completo utilizado nos experimentos compreende 3.834 imagens. Este *dataset* final e consolidado constituiu a base para o processo de treinamento e a subsequente avaliação comparativa dos modelos, cujos resultados de performance são detalhados a seguir.

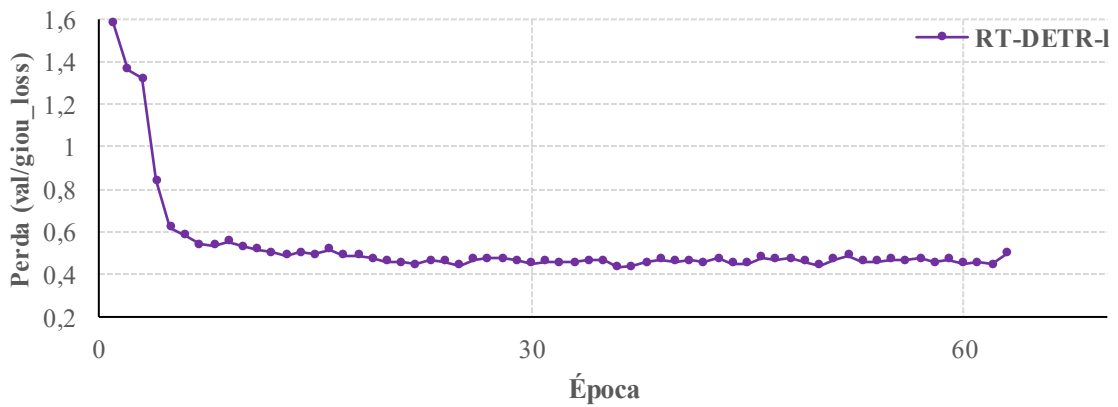
4.1.2 Treinamento dos modelos de detecção

O comportamento de cada um dos quatro modelos foi monitorado durante a fase de treinamento para analisar a estabilidade da convergência e a evolução da performance no conjunto de validação. A Figura 11 apresenta as curvas de aprendizado comparativas, exibindo a perda de validação e a métrica mAP_{50} .

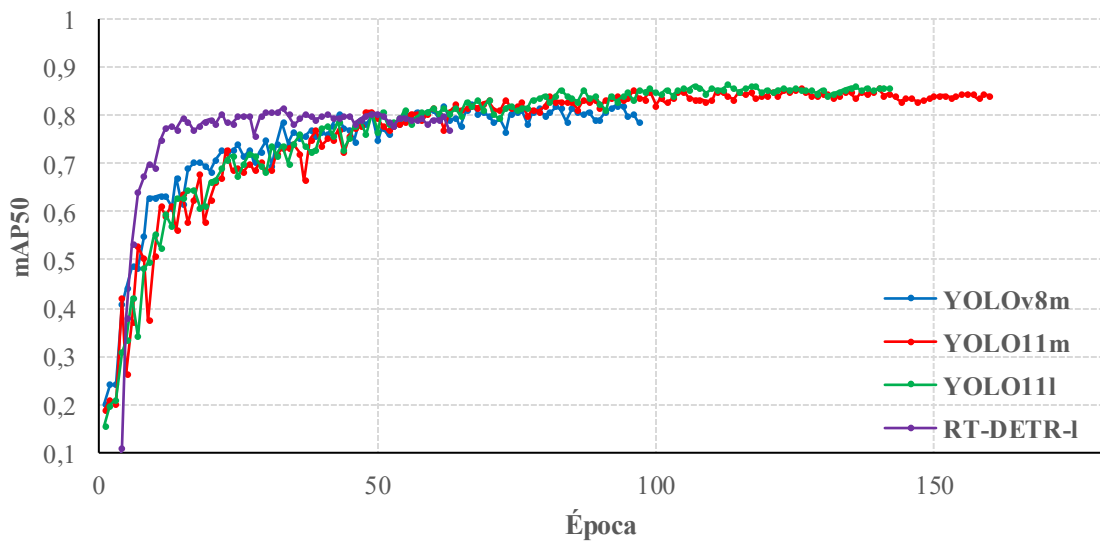
Figura 11 – Curvas de aprendizado de detecção no conjunto de validação: (a) Perda de validação (modelos YOLO), (b) Perda de validação (RT-DETR), (c) Evolução do mAP_{50} .



(a)



(b)



(c)

Fonte: O autor.

A análise das curvas de perda no conjunto de validação, Figura 11 (a) e (b), demonstra que todas as quatro arquiteturas apresentaram uma convergência estável. Conforme observado o valor da perda para cada modelo diminuiu consistentemente ao longo das épocas e atingiu um platô, indicando que o processo de treinamento foi bem-sucedido, sem sinais de instabilidade.

A Figura 11(c) apresenta a evolução da métrica mAP_{50} ao longo do treinamento, revelando as diferentes características de aprendizado de cada arquitetura. Todos os modelos exibiram incremento acentuado nas primeiras épocas, indicando aprendizado eficaz dos padrões de fissuras. Observa-se que o modelo RT-DETR-l apresentou a convergência mais rápida, atingindo seu platô de performance em aproximadamente 30 épocas. Em contrapartida, os modelos da família YOLO, embora tenham necessitado de mais iterações até a estabilização completa, alcançaram picos da métrica mAP_{50} ligeiramente superiores. Esses resultados destacam que, embora o RT-DETR-l tenha convergido em menor número de ciclos, todos os modelos atingiram patamares finais de precisão muito próximos.

Com base nesses resultados, procedeu-se à etapa final de avaliação, na qual a performance dos modelos foi sistematicamente comparada a partir de métricas quantitativas e exemplos qualitativos de detecções no conjunto de teste. Essa análise buscou verificar a capacidade efetiva de generalização das redes e identificar eventuais limitações práticas na identificação automática de fissuras.

4.1.3 Avaliação de desempenho dos modelos de detecção

Nesta seção, é apresentado o desempenho final dos modelos, avaliados sobre o conjunto de teste, que não foi utilizado durante nenhuma etapa do treinamento. Esta avaliação representa a avaliação da capacidade de generalização de cada arquitetura para detectar fissuras em dados não vistos. Os resultados de performance dos modelos estão consolidados na Tabela 8.

Tabela 8 – Desempenho dos modelos de detecção

Modelo	P	R	mAP_{50}	Velocidade de Inferência (ms)
YOLOv8m	0,823	0,766	0,823	8,8
YOLO11m	0,893	0,714	0,823	9,8
YOLO11l	0,887	0,737	0,831	12,2
RT-DETR-l	0,847	0,745	0,797	13,6

Fonte: O autor.

A Tabela 8 apresenta os principais indicadores de desempenho dos modelos no conjunto de teste, abrangendo métricas quantitativas de precisão, *recall* e mAP_{50} , além do tempo médio de inferência por imagem. Esses resultados permitem comparar de forma abrangente o comportamento de cada arquitetura em termos de acurácia e eficiência computacional.

Observa-se que o modelo YOLO11l obteve o maior valor de mAP_{50} (83,1%), indicando desempenho ligeiramente superior na detecção correta de fissuras em diferentes condições de imagem. O YOLOv8m e o YOLO11m apresentaram mAP_{50} muito próximos (82,3%), o que demonstra que mesmo em versões mais compactas da arquitetura YOLO podem alcançar níveis elevados de acurácia quando bem ajustadas. Por outro lado, o RT-DETR-l, embora tenha registrado o menor mAP_{50} (79,7%), ainda assim manteve desempenho competitivo considerando a complexidade do cenário.

A análise das métricas de P e R revela nuances relevantes. O YOLO11m obteve o maior índice de precisão (89,3%), indicando menor incidência de falsos positivos. Entretanto, esse modelo apresentou também o menor *recall* (71,4%), sugerindo maior seletividade na detecção e eventual perda de fissuras menos evidentes. O YOLOv8m destacou-se por apresentar *recall* mais alto (76,6%) com precisão equilibrada (82,3%), configurando-se como uma opção de maior sensibilidade. Já o RT-DETR-l apresentou um balanço intermediário entre precisão (84,7%) e *recall* (74,5%), o que pode contribuir para resultados mais consistentes em cenários com alta variabilidade visual.

Do ponto de vista operacional, o tempo médio de inferência variou significativamente entre os modelos. O YOLOv8m apresentou a maior velocidade (8,8 ms por imagem), seguido pelo YOLO11m (9,8 ms) e YOLO11l (12,2 ms). O RT-DETR-l, embora tenha demonstrado capacidade de convergência rápida durante o treinamento, apresentou o maior tempo de processamento por imagem (13,6 ms), permanecendo, ainda assim, em patamar compatível com aplicações em tempo real.

Em relação ao risco de overfitting, não foram observados indícios consistentes de memorização do conjunto de treinamento. Esse diagnóstico é sustentado por dois aspectos: o comportamento das métricas monitoradas durante o treinamento, nas quais a evolução no conjunto de validação acompanha a redução das perdas, sem degradação sistemática do desempenho em validação nas épocas finais e a proximidade entre o desempenho obtido em validação e o desempenho reportado no conjunto de teste, composto por imagens não utilizadas no treinamento. Adicionalmente, a estratégia adotada (uso de pesos pré-treinados, aumento de variabilidade via

aumento de dados e seleção do melhor ponto de treinamento com base na validação) contribui para reduzir a tendência de sobreajuste, favorecendo a generalização para imagens não vistas.

De maneira geral, os resultados indicam que as arquiteturas YOLO, especialmente na configuração 111, proporcionam melhor equilíbrio entre alta acurácia e tempos de inferência compatíveis com aplicações em larga escala. O RT-DETR-1 mostrou desempenho competitivo e rapidez de convergência, mas apresentou limitações em mAP₅₀ e maior demanda computacional. Esses achados reforçam a importância de considerar não apenas a precisão final, mas também aspectos práticos de desempenho operacional na escolha de modelos para inspeção automatizada de fissuras.

A comparação dos resultados obtidos neste trabalho com pesquisas recentes que aplicaram técnicas de aprendizado profundo para a detecção de fissuras é essencial para contextualizar o desempenho das arquiteturas avaliadas. A Tabela 9 apresenta uma síntese de trabalhos relacionados, destacando o número de imagens utilizadas (*dataset*), a resolução empregada no treinamento e os valores de mAP₅₀ obtidos por diferentes modelos. Essa análise contribui para evidenciar avanços metodológicos, potenciais limitações e as especificidades de cada configuração experimental.

Tabela 9 – Comparação dos resultados deste estudo com trabalhos relacionados na detecção automática de fissuras

Referência	Modelo	Total de imagens	Tamanho da Imagem	mAP ₅₀
Jiang, Pang, Li, 2021	YOLOV3	541	1024x1024	0,524
	Fast-YOLO	689		0,559
	SSD	726		0,475
	MobileNet-SSD	679		0,433
	FasterR-CNN	525		0,563
Liu <i>et al.</i> , 2024	YOLOv5	3213 (7788 instâncias de fissuras)	640x480	0,713
	YOLOv3-SPP			0,705
	YOLOv3			0,688
	Faster R-CNN			0,573
Zuo <i>et al.</i> , 2025	Pavement-DETR	2440	640×640	0,871
	RT-DETR-R18			0,794
	YOLOv8s			0,793
	YOLOv8m			0,791
	YOLOv9s			0,740

Referência	Modelo	Total de imagens	Tamanho da Imagem	mAP ₅₀
	YOLOv9m			0,811
	YOLOv10s			0,825
	YOLOv10m			0,730
	Deformable-DETR			0,776
	Dab-DETR			0,767
	Conditional-DETR			0,766
	DINO-DETR			0,797
Zhao <i>et al.</i> , 2025	RT-DETR-baseline	7397*	640x640	0,821
	RT-DETR-improved			0,841
Este estudo	YOLOv8m	738	1024x1024	0,823
	YOLO11m			0,823
	YOLO11l			0,831
	RT-DETR-l			0,797

(*) O número exato de imagens não é informado; o estudo utilizou dados do subconjunto chinês do RDD2022.

Fonte: O autor.

No estudo de Jiang *et al.* (2021), diferentes arquiteturas foram empregadas para detectar danos em concreto, incluindo fissuras, manchas e exposição de armaduras. Para permitir uma comparação mais adequada, foram considerados apenas os resultados relativos à classe “crack” (fissuras), que apresentou mAP₅₀ variando entre 0,433 (MobileNet-SSD) e 0,563 (Faster R-CNN). Apesar do uso de um *dataset* similar em resolução e tamanho ao deste trabalho, observa-se que o desempenho foi inferior ao alcançado neste trabalho, em que os modelos YOLOv8m, YOLO11m, YOLO11l e RT-DETR-l apresentaram mAP₅₀ superiores a 0,79, evidenciando a evolução das arquiteturas de detecção de objetos mais recentes em alcançar maior precisão mesmo em bases de dados moderadas.

Liu *et al.* (2024) propuseram um sistema de detecção baseado em YOLOv5, YOLOv3-SPP, YOLOv3 e Faster R-CNN para identificar fissuras e buracos (*potholes*) em pavimentos de concreto asfáltico, utilizando imagens capturadas por câmeras visíveis e térmicas. Para efeito comparativo, foram consideradas apenas as métricas associadas às imagens visíveis e às classes de fissuras. Os valores de mAP₅₀ apresentados não correspondem diretamente ao mAP global reportado no estudo, mas foram recalculados especificamente para esta análise por meio da média dos AP₅₀ obtidos individualmente para cada classe de fissura em cada modelo.

Esse procedimento permitiu isolar o desempenho relativo exclusivamente às fissuras, resultando em mAP_{50} que variaram entre 0,573 (Faster R-CNN) e 0,713 (YOLOv5). Embora o trabalho tenha empregado um conjunto de dados mais extenso, composto por 3.213 imagens no total, não foi possível determinar quantas dessas imagens continham exclusivamente fissuras. Entretanto, o número de instâncias de fissuras anotadas fora substancialmente superior ao deste estudo (7.788 instâncias no trabalho de Liu *et al.*, em comparação com 1.757 instâncias utilizadas neste trabalho). Além disso, as imagens foram redimensionadas para 640×480 px durante o pré-processamento, ao passo que o presente estudo empregou imagens de entrada com resolução de 1024×1024 px. Essa diferença de abordagem na definição das classes, associada à escolha do modelo e à resolução superior, pode ter contribuído para os resultados mais elevados observados nos modelos do estudo.

O trabalho de Zuo *et al.* (2025) avaliou o desempenho de diferentes arquiteturas de detecção, incluindo o Pavement-DETR — um modelo baseado em RT-DETR com aprimoramentos específicos, como mecanismos de atenção espacial e uma nova função de perda combinada — além de variantes do YOLOv8, YOLOv9, YOLOv10, Deformable-DETR, Dab-DETR, Conditional-DETR e DINO-DETR. Nesse estudo, que envolveu a detecção simultânea de seis categorias de defeitos em 2.440 imagens UAV de pavimentos, o Pavement-DETR apresentou o maior desempenho, com mAP_{50} de 0,871. Entre os demais modelos, os valores de mAP_{50} variaram entre 0,730 (YOLOv10m) e 0,825 (YOLOv10s), demonstrando que mesmo arquiteturas recentes apresentam diferenças significativas na acurácia dependendo da configuração e da estratégia de treinamento.

No presente trabalho, utilizando um *dataset* consideravelmente menor (738 imagens únicas) e restrito à detecção exclusiva de fissuras, obtiveram-se valores de mAP_{50} variando de 0,797 (RT-DETR-l) a 0,831 (YOLO11l). Esses resultados superam a maioria dos modelos avaliados por Zuo *et al.* (2025), ficando abaixo apenas do Pavement-DETR (0,871) e próximos do YOLOv10s (0,825). Além do volume reduzido de dados, destaca-se que as imagens foram processadas em resolução de 1024×1024 pixels, superior à utilizada no estudo de referência (640×640 pixels), o que preserva mais detalhes das fissuras e pode ter contribuído para o desempenho alcançado, ainda que com maior custo computacional. Esse cenário reforça que a especialização da tarefa e o uso de *transfer learning* podem mitigar a dependência de grandes volumes de dados anotados em aplicações de detecção de fissuras, e que a padronização do pré-processamento, o balanceamento do conjunto e a preservação de detalhes na entrada, seja por maior resolução ou

por análise em blocos, influenciam de forma relevante o desempenho e a estabilidade do modelo (Katsigiannis *et al.*, 2023; Lee *et al.*, 2023; Fan *et al.*, 2025).

O estudo de Zhao *et al.* (2025) propôs uma versão aprimorada do RT-DETR, incorporando um *backbone* especializado (PCAM), mecanismos de atenção CGA e uma nova função de perda denominada Foclar-MPDIoU. Os experimentos foram conduzidos com imagens do subconjunto chinês do *dataset* RDD2022, resultando em um mAP_{50} de 0,841 com precisão e *recall* de 86,4% e 80,1%, respectivamente. Embora o número exato de imagens utilizadas não seja especificado, o *dataset* RDD2022 é notoriamente mais amplo que o empregado neste trabalho, o que pode favorecer o desempenho. Ainda assim, os resultados obtidos pelos modelos YOLO11l e RT-DETR-l neste estudo apresentaram métricas próximas, demonstrando que o uso de arquiteturas de base atualizadas pode atingir níveis competitivos de acurácia. Isso ocorre porque a evolução contínua dessas famílias de modelos já incorpora nativamente otimizações de *backbone* e mecanismos de atenção que dispensam customizações manuais complexas para a maioria das aplicações práticas (Terven; Cordova-Esparza, 2023; Zuo *et al.*, 2025).

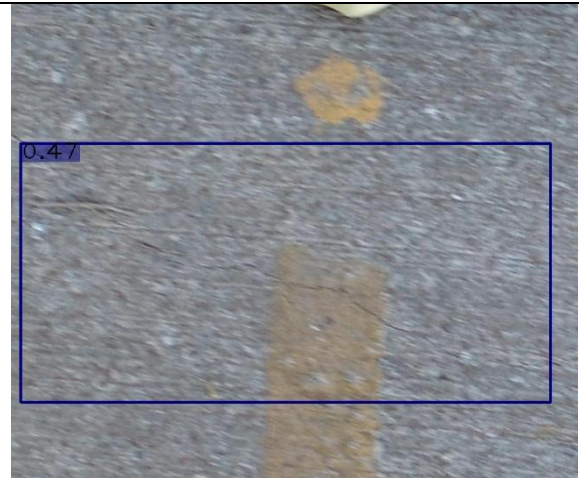
De modo geral, esta análise comparativa evidencia que a utilização de imagens de maior resolução, associada a estratégias de ajuste fino e ao uso de modelos de última geração, contribuíram para que o presente estudo atingisse métricas competitivas em relação à trabalhos da área. Esses resultados reforçam o potencial das abordagens baseadas em YOLO e RT-DETR para aplicações de inspeção automatizada de pavimentos de concreto, especialmente quando o foco está na detecção específica de fissuras em contextos urbanos específicos.

4.1.4 Avaliação qualitativa das inferências dos modelos de detecção

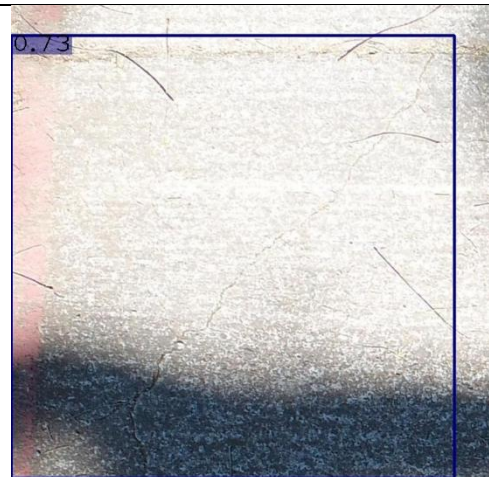
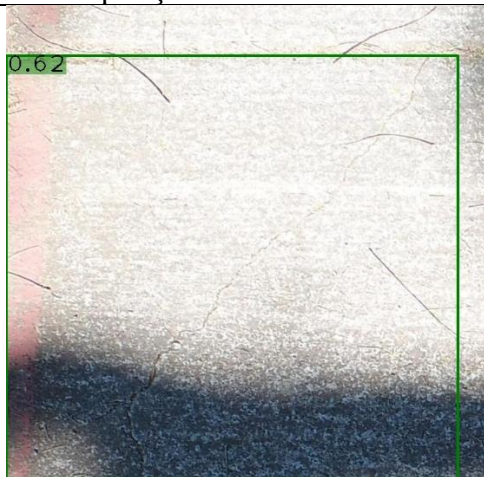
Este tópico apresenta exemplos de inferência realizados com os modelos treinados, incluindo situações específicas como imagens com desfoque (*blur*), alta exposição solar, áreas de sombra pronunciada e imagens contendo ruído visual e obstrução parcial devido a galhos e folhas. As inferências foram realizadas considerando um limiar mínimo de confiança de 0,30, de modo a preservar todas as detecções potencialmente relevantes para análise qualitativa, incluindo aquelas com pontuações mais baixas. Os exemplos apresentados na Figura 12 ilustram a robustez e as limitações observadas nos detectores em diferentes condições de iluminação, nitidez e interferências visuais.

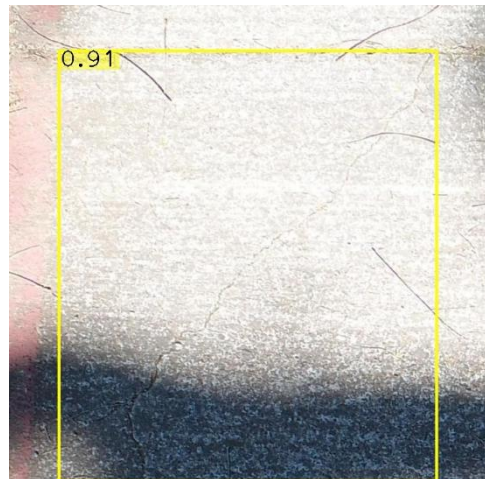
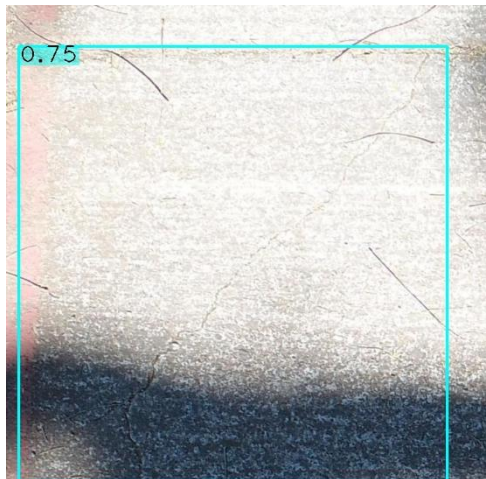
Figura 12 – Detecção de fissuras em ciclovias usando YOLOv8m, YOLO11m, YOLO11l e RT-DETR-l sobre a influência de fatores como desfoque, alta iluminação, sombra e ruído.

Desfoque (blur)

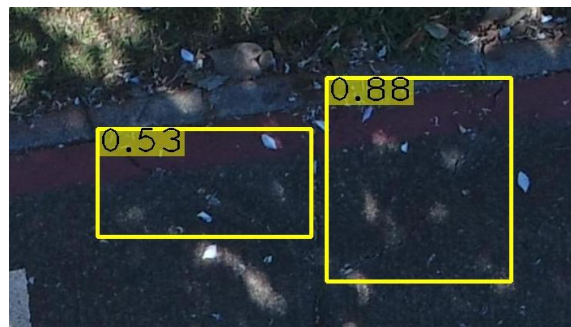
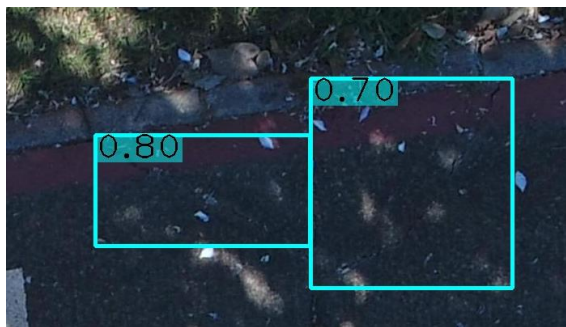
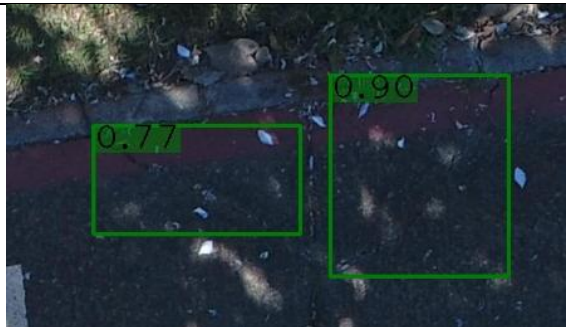


Alta exposição solar

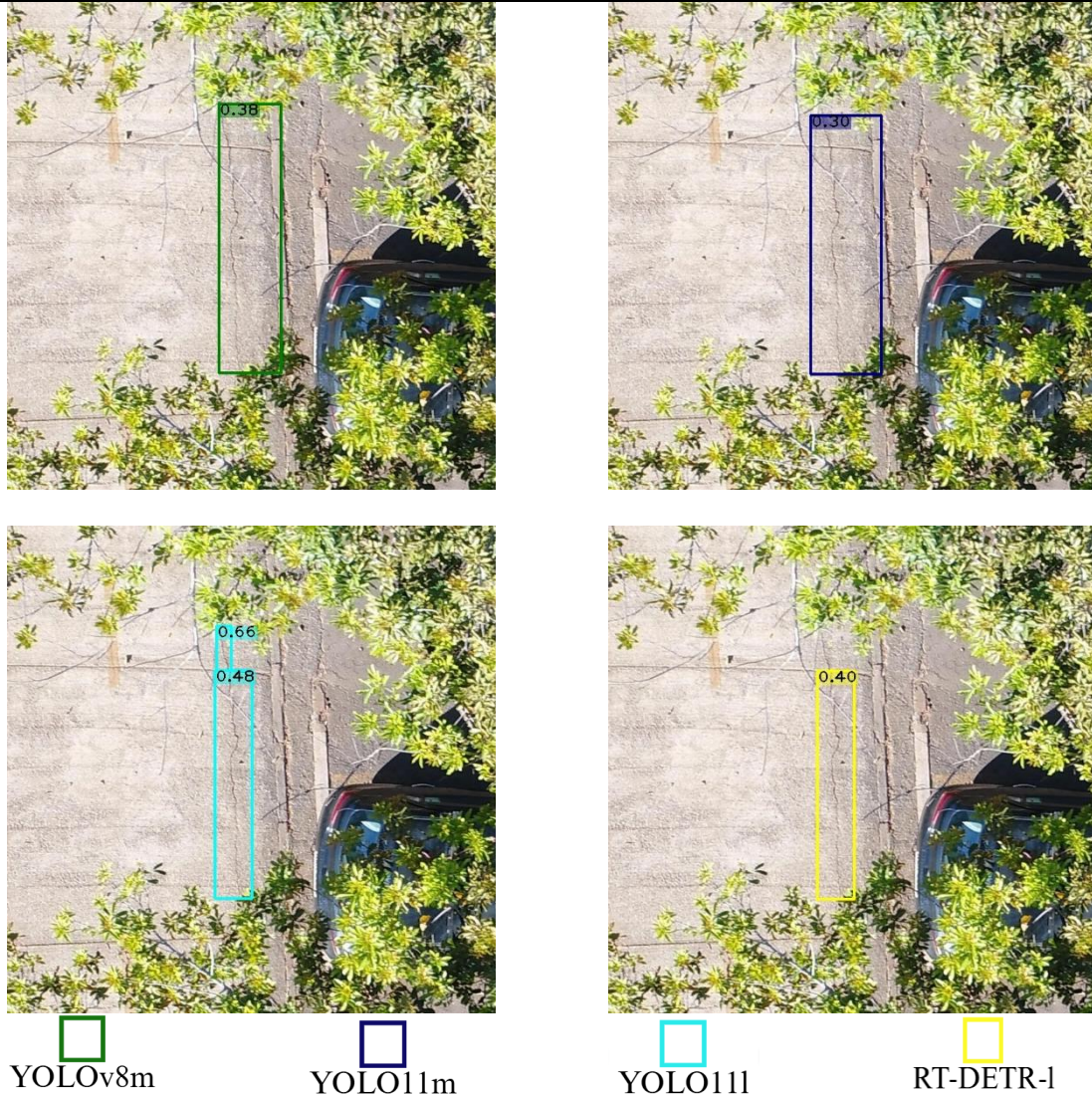




Sombra



Ruído Visual



Fonte: O autor.

Na condição de desfoque, todos os modelos conseguiram identificar a fissura principal, ainda que com redução perceptível nas pontuações de confiança. O RT-DETR-l apresentou a detecção com maior confiabilidade e caixa mais bem ajustada ao contorno da fissura, evidenciando boa capacidade de preservação das características relevantes mesmo em imagens com baixa nitidez. No entanto, esse modelo também gerou uma detecção adicional em uma região sem trinca aparente, configurando um falso positivo induzido pelo ruído visual. As variantes YOLO11l, YOLO11m e YOLOv8m mantiveram desempenho satisfatório, com detecções consistentes da fissura principal e menor propensão a falsos positivos, embora com *scores* ligeiramente inferiores e delimitações menos precisas. Esse comportamento reforça a necessidade de considerar o equilíbrio entre a capacidade de identificar todas as fissuras presentes na imagem (*recall*) e a tendência de produzir detecções incorretas (falsos positivos). Tal balanço é

determinante para selecionar o modelo mais adequado, especialmente em cenários sujeitos à degradação da qualidade visual, onde o ruído ou desfoque pode ser incorretamente interpretado como patologia (Cao; Liu; He, 2020; Gao; Cao; Yi, 2025; Lecca; Bianco, 2025; Gao; Cao; Yi, 2025).

Nas imagens com alta exposição solar, todos os modelos foram capazes de identificar corretamente a fissura principal, evidenciando boa estabilidade frente ao aumento de luminosidade e redução de contraste. O RT-DETR-l apresentou a detecção mais confiável, com pontuação de 0,91 e caixa bem ajustada à extensão da fissura, indicando menor sensibilidade ao brilho intenso. O YOLO11l também manteve desempenho consistente, com *score* elevado e boa delimitação. As variantes YOLO11m e YOLOv8m exibiram detecções corretas, ainda que com pontuações ligeiramente inferiores e caixas um pouco mais amplas que a extensão real da fissura, indicando menor precisão no ajuste dos contornos. Esses resultados demonstram que a exposição solar intensa, embora provoque redução moderada na confiança, não inviabilizou a detecção pelas diferentes arquiteturas.

Nas imagens com sombra pronunciada, todos os modelos conseguiram identificar corretamente as duas fissuras presentes, demonstrando boa capacidade de detecção mesmo em regiões com contraste reduzido. O YOLO11m apresentou o desempenho mais consistente, com as maiores pontuações de confiança e caixas bem ajustadas aos contornos das fissuras. O YOLOv8m também exibiu resultados satisfatórios, com *scores* elevados e delimitações adequadas. As variantes YOLO11l e RT-DETR-l mantiveram detecções corretas, embora com valores de confiança inferiores e caixas relativamente bem delimitadas, porém com ligeira redução na precisão em comparação ao YOLOv8m. Esses achados indicam que, apesar da interferência visual causada pela sombra, todas as arquiteturas foram capazes de identificar as fissuras, com variações na confiança e no detalhamento das predições.

Na situação de ruído visual e obstrução parcial causada por vegetação, observou-se maior variabilidade na forma como os modelos interpretaram o cenário em relação ao protocolo de anotação adotado neste estudo. O YOLO11l foi a única arquitetura que produziu duas caixas delimitadoras separadas, em conformidade com a regra de divisão por placa de pavimento, apresentando *scores* moderados e bom alinhamento aos contornos. O YOLO11m e o YOLOv8m identificaram corretamente a presença das duas fissuras, porém encapsularam ambas em uma única *bounding box*, contrariando a estratégia de anotação definida. O RT-DETR-l, por sua vez, detectou apenas uma das fissuras, deixando de englobar a segunda trinca

presente na outra placa, ainda que a caixa gerada tenha apresentado bom ajuste sobre a região identificada. Esse comportamento evidencia que, em situações com alta interferência visual, os modelos podem divergir na forma de delimitar múltiplas ocorrências próximas.

De modo geral, a avaliação qualitativa reforçou a observação de que todos os modelos possuem elevada capacidade de detecção em condições favoráveis e mantêm desempenho razoavelmente estável em cenários de variação de luminosidade e presença de desfoque. Entretanto, situações de ruído visual intenso e obstrução parcial demonstraram maior variabilidade nas respostas, tanto na forma de delimitação das fissuras quanto na confiança atribuída às predições. Entre as arquiteturas avaliadas, o YOLO11l destacou-se pelo desempenho mais equilibrado, com resultados que corroboram as métricas quantitativas obtidas anteriormente e evidenciam maior consistência em diferentes condições. Por outro lado, embora o RT-DETR-l tenha apresentado menor desempenho geral nas métricas de validação, demonstrou comportamento competitivo na avaliação qualitativa, principalmente em cenários de alta luminosidade e imagens com desfoque, sugerindo potencial promissor para aplicações futuras em ambientes complexos. Esses achados ratificam que a especialização dos modelos garantiu resultados satisfatórios e competitivos mesmo com um *dataset* reduzido. Contudo, indicam também que a inclusão incremental de exemplos focados nessas situações de interferência visual extrema constitui um caminho natural para elevar ainda mais a robustez e a generalização do sistema em trabalhos futuros.

Embora os modelos de detecção tenham apresentado desempenho satisfatório, a delimitação geométrica das fissuras permanece limitada à representação por caixas delimitadoras, motivando a análise dos modelos de segmentação apresentada na seção seguinte.

4.2 DESEMPENHO DOS MODELOS DE SEGMENTAÇÃO DE FISSURAS EM CICLOVIAS (SEÇÃO 2)

Esta seção apresenta e discute os resultados obtidos para os modelos de segmentação de fissuras em ciclovias de pavimento rígido, considerando tanto o desempenho quantitativo quanto a qualidade geométrica das máscaras geradas. A análise considera o subconjunto de dados específico para esta etapa, discutindo a capacidade dos modelos em delimitar a geometria das fissuras.

4.2.1 Caracterização do conjunto de dados para segmentação

Diferentemente da etapa de detecção, o conjunto de dados utilizado para a etapa de segmentação foi composto apenas por imagens capturadas pelo drone DJI Mavic 3 Enterprise, a uma altura de voo de 12m, originalmente com resolução de 4000×3000 pixels. A partir dessas imagens, foram extraídas regiões de interesse correspondentes às áreas de ciclovias, as quais foram recortadas e padronizadas para a dimensão de 640×640 pixels, resultando em um total de 413 imagens utilizadas nos experimentos de segmentação.

Essa redução na dimensão de entrada, em comparação aos 1024x1024 pixels utilizados na detecção, é uma estratégia comum em tarefas de segmentação de instâncias para viabilizar o custo computacional, dado que a predição de máscaras pixel a pixel demanda significativamente mais memória da GPU (Bakhtiarnia; Zhang; Iosifidis, 2022; Minaee et al., 2022).

O processo de rotulação dessas imagens, realizado por meio da demarcação de polígonos representativos das fissuras, resultou em um total de 895 anotações. A distribuição dessas amostras entre os conjuntos de treinamento, validação e teste é apresentada na Tabela 10.

Tabela 10 – Distribuição de imagens e anotações entre os conjuntos de dados de segmentação

Conjunto	Nº de Imagens	Proporção de Imagens (%)	Nº de Anotações	Proporção de Anotações (%)
Treinamento	290	70	643	72
Validação	81	20	165	18
Teste	42	10	87	10
Total	413	100	895	100

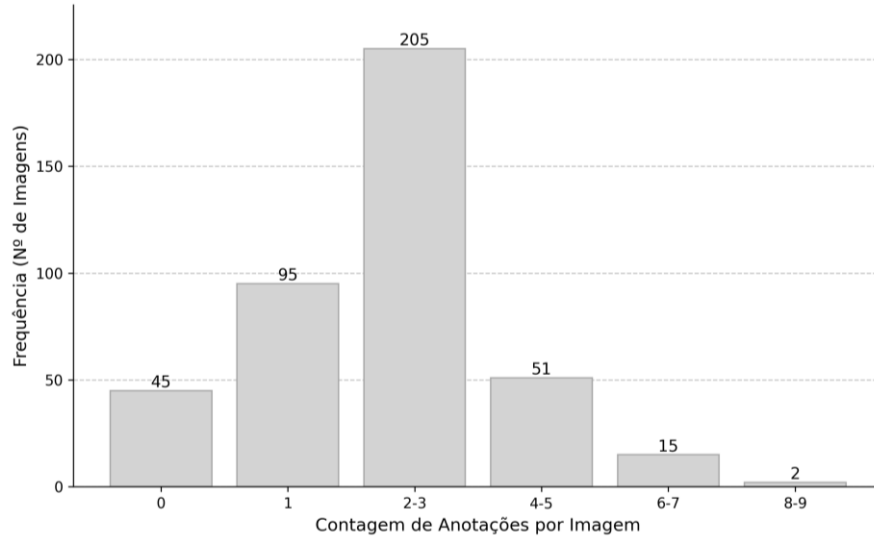
Fonte: O autor.

A análise da Tabela 10 demonstra que a proporção de anotações (72% no treino, 18% na validação e 10% no teste) segue a distribuição das imagens (70-20-10), indicando que a densidade de fissuras é uniforme entre os subconjuntos, o que previne vieses no treinamento. Esse balanceamento é particularmente relevante para tarefas de segmentação, nas quais a diversidade geométrica das fissuras influencia diretamente o aprendizado das máscaras.

Em relação à densidade de anomalias, o conjunto de dados apresenta uma média de 2,2 anotações por imagem. Este valor é muito próximo à média observada no *dataset* de detecção (2,4), sugerindo que, mesmo com a redução da área da imagem para 640 x 640, a complexidade do cenário se mantém. Isso indica que as fissuras tendem a aparecer agrupadas ou ramificadas,

desafiando a capacidade do modelo de distinguir instâncias individuais adjacentes. Para detalhar essa característica, a Figura 13 apresenta o histograma da distribuição de anotações por imagem.

Figura 13 – Histograma da distribuição de anotações por imagem no conjunto de segmentação



Fonte: O autor.

O histograma revela um pico acentuado na faixa de 2 a 3 anotações por imagem (representando 205 imagens, ou cerca de 50% do total). Essa concentração favorece a avaliação da segmentação de instâncias, pois inclui um número recorrente de imagens com múltiplas ocorrências por cena, permitindo verificar se o modelo consegue delimitar e separar máscaras distintas mesmo quando as fissuras estão próximas.

Além disso, observa-se a presença de 45 imagens sem anotações, constituindo as amostras negativas. A inclusão dessas imagens de pavimento íntegro é fundamental para o treinamento de segmentação, pois ensina a rede neural a distinguir texturas de asfalto saudável, contribuindo para a redução de falsos positivos na avaliação final.

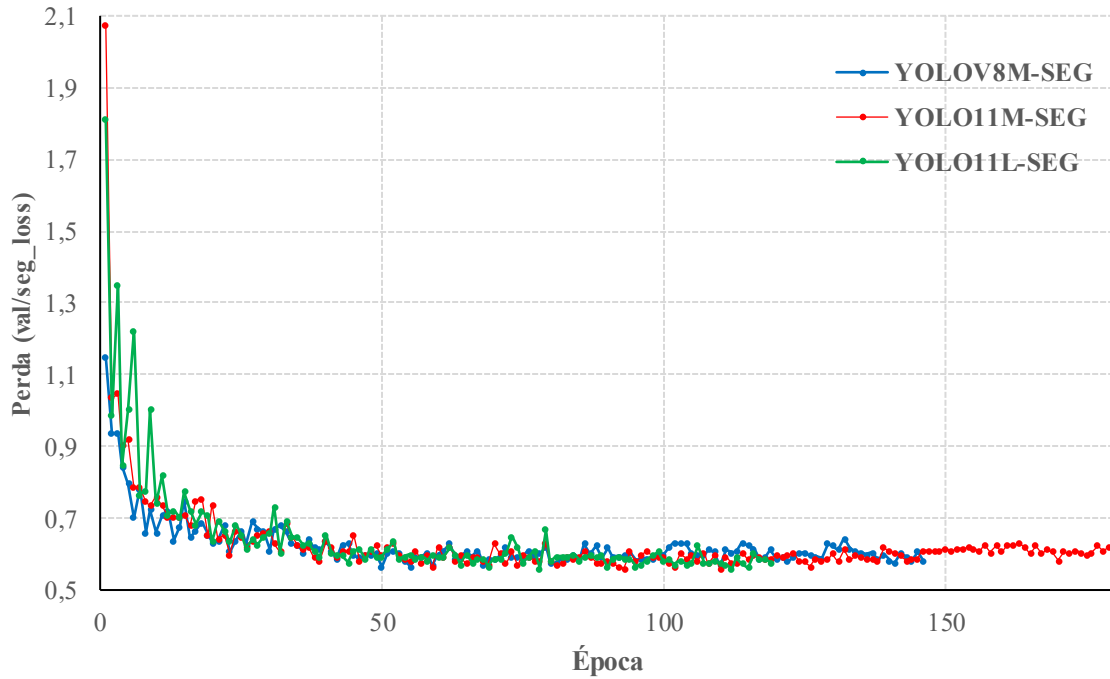
De modo geral, o conjunto de dados preparado para a segmentação apresenta características compatíveis com os objetivos do estudo, combinando precisão geométrica das anotações, diversidade de cenários visuais e uma quantidade de amostras suficiente para avaliar o desempenho dos modelos em condições controladas.

4.2.2 Treinamento dos modelos de segmentação

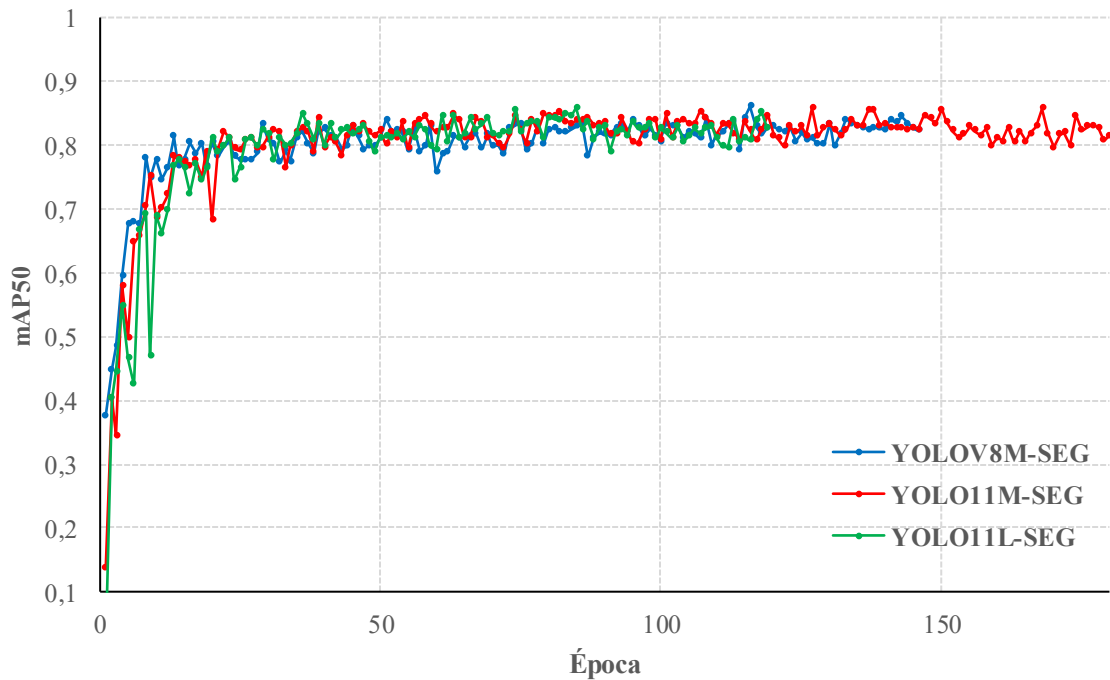
Assim como na etapa de detecção, o comportamento de cada arquitetura foi monitorado durante a fase de treinamento para analisar a estabilidade da convergência e a evolução da precisão das

máscaras geradas a partir da evolução das curvas de perda (*seg_loss*) e da métrica mAP₅₀ no conjunto de validação ao longo das épocas de treinamento, conforme observa-se na Figura 14.

Figura 14 – Curvas de aprendizado de segmentação no conjunto de validação: (a) Perda de validação (*seg_loss*), (b) Evolução do mAP₅₀



(a)



(b)

Fonte: O autor.

A análise quantitativa das curvas de perda, Figura 14(a), confirma uma convergência acentuada nas épocas iniciais. Observa-se que o modelo YOLOv8m apresentou a adaptação mais rápida, com uma redução maior do erro logo nos primeiros ciclos. Contudo, essa diferença foi rapidamente suprimida. A partir da 50ª época, todas as arquiteturas atingiram um platô de estabilidade técnica com valores de perda oscilando minimamente entre 0,56 e 0,60, indicando que os modelos atingiram o limite de aprendizado possível com o conjunto de dados disponível, sem evidências de *overfitting* ou divergência.

A evolução da métrica mAP_{50} , Figura 14(b), corrobora esse desempenho consistente. De forma coerente com o comportamento observado nas curvas de perda, todos os modelos atingem rapidamente valores elevados de mAP_{50} , ultrapassando 0,75 ainda nas primeiras dezenas de épocas. A partir desse ponto, observa-se um crescimento mais gradual até a estabilização, com valores finais próximos a 0,86, indicando bom desempenho na segmentação das fissuras. A variação na duração do treinamento (encerrando-se entre 120 e 197 épocas sob o mesmo critério de parada automática) não resultou em diferenças significativas na acurácia final. Esses resultados demonstram que, para a tarefa de segmentação de fissuras, do presente estudo, neste nível de resolução (640 x 640), as otimizações arquiteturais do YOLO11 entregam performance equivalente à versão v8 estabelecida, com todas as redes demonstrando alta capacidade de generalização.

De modo geral, os resultados indicam que todas as arquiteturas foram capazes de aprender de forma eficaz os padrões espaciais das fissuras em nível de pixel. Além disso, a similaridade entre os desempenhos finais sugere que, para cenários de segmentação de fissuras com imagens de alta resolução e conjuntos de dados moderados, modelos de porte intermediário podem oferecer um equilíbrio mais adequado entre desempenho e custo computacional, sem a necessidade de arquiteturas excessivamente complexas.

4.2.3 Avaliação de desempenho dos modelos de segmentação

A avaliação final do desempenho dos modelos de segmentação foi conduzida sobre o conjunto de teste, composto por imagens não utilizadas durante as etapas de treinamento e validação, com o objetivo de analisar a capacidade de generalização das arquiteturas avaliadas na delimitação de fissuras em nível de pixel. A análise considerou as métricas de precisão (P), *recall* (R) e *mean Average Precision* com limiar de sobreposição IoU igual a 0,50 (mAP_{50}), além do tempo médio de inferência por imagem, conforme apresentado na Tabela 11.

Tabela 11 – Desempenho dos modelos de segmentação

Modelo	P	R	mAP ₅₀	Dice ou <i>Fscore</i>	Velocidade de Inferência (ms)
YOLOv8m-seg	0,883	0,787	0,839	0,832	5,2
YOLO11m-seg	0,853	0,854	0,858	0,854	5,3
YOLO11l-seg	0,846	0,865	0,872	0,855	5,6

Fonte: O autor.

Os resultados indicam que todas as arquiteturas avaliadas apresentaram desempenho satisfatório na tarefa de segmentação, com valores de mAP₅₀ superiores a 0,83, evidenciando elevada capacidade de identificação e delimitação das fissuras, mas denotando uma superioridade da arquitetura YOLO11.

O modelo YOLO11l-seg apresentou o maior valor de mAP₅₀ (0,872), associado também ao maior índice de *recall* (0,865), indicando maior sensibilidade na identificação das regiões fissuradas. Esse comportamento é coerente com o maior porte do modelo e com o maior número de parâmetros, o que amplia sua capacidade de representação espacial das fissuras.

O modelo YOLO11m-seg, por sua vez, apresentou um desempenho muito próximo ao da versão de maior porte, com mAP₅₀ de 0,858 e *recall* de 0,854, destacando-se pelo equilíbrio entre precisão e sensibilidade. Ao cruzar os dados de desempenho com o porte dos modelos, observa-se que o YOLO11m-seg se destaca pela eficiência arquitetural. Com apenas 22,4 milhões de parâmetros, significativamente menor que os 27,3 milhões do YOLOv8m-seg, ele foi capaz de superar o antecessor em quase 2 pontos percentuais de mAP₅₀. Além disso, o equilíbrio entre precisão e *recall*, observado principalmente nas arquiteturas YOLO11, assegura a detecção da maioria das fissuras existentes sem a geração excessiva de falsos positivos, comportamento consistente com os bons resultados de convergência previamente observados na Seção 4.2.2.

O modelo YOLOv8m-seg apresentou o maior valor de precisão (0,883), indicando menor incidência de falsos positivos, porém com *recall* inferior (0,787), o que sugere um comportamento mais conservador na segmentação das fissuras: o modelo erra pouco quando faz uma predição, mas tende a omitir fissuras (falsos negativos), o que é indesejável em inspeções de segurança onde a detecção de patologias é crítica.

Quanto ao risco de overfitting, a verificação seguiu os mesmos critérios adotados na etapa de detecção (Seção 4.1.3), com análise do comportamento treino–validação ao longo das épocas e comparação entre validação e teste. Não se observaram divergências sistemáticas que indicassem perda de generalização no conjunto de teste.

Por fim, em relação à viabilidade operacional, a análise do tempo médio de inferência revelou diferenças pouco significativas entre os modelos, variando de 5,2 ms a 5,6 ms por imagem. Essa proximidade indica que, para a resolução de entrada utilizada (640 x 640) e o *hardware* empregado, o aumento de complexidade do modelo YOLO11 (Grande) não penaliza o tempo de resposta de forma significativa em comparação aos modelos médios. Todos os modelos operam confortavelmente em tempo real, validando sua aplicabilidade em sistemas de monitoramento contínuo. Assim, a escolha entre os modelos pode ser orientada predominantemente pelas métricas de segmentação, sem comprometer de forma relevante o desempenho computacional.

Ao comparar os resultados obtidos com estudos recentes de segmentação de fissuras com imagens aéreas e integração a GIS, observa-se coerência quanto à faixa de desempenho alcançada e aos fatores que influenciam diretamente as métricas. Hu, Yee e Goff (2024), por exemplo, reportam Dice = 88,1%, precisão = 77,5% e *recall* = 76,5% para segmentação de fissuras aplicada a inspeção com VANT, indicando bom desempenho global na delimitação pixel a pixel. No presente trabalho, os melhores resultados foram obtidos pelos modelos da família YOLO11, com destaque para o YOLO11l-seg, que atingiu Dice = 85,5%, precisão = 84,6% e *recall* = 86,5%, e para o YOLO11m-seg, com Dice = 85,4%, precisão = 85,3% e *recall* = 85,4%. Nota-se, portanto, que, embora o Dice reportado por Hu, Yee e Goff (2024) seja ligeiramente superior, os modelos deste estudo apresentam precisão e *recall* mais elevados, sugerindo maior capacidade de recuperar a maior parte das regiões fissuradas (sensibilidade) sem aumentar proporcionalmente os falsos positivos. Essa diferença de comportamento pode estar associada a particularidades do cenário e do conjunto de dados (padrões de fissuração, contraste fissura–fundo, presença de ruídos e critérios de avaliação), já que o Dice é diretamente determinado pelo balanço entre precisão e *recall* e pode variar conforme o limiar de decisão adotado e o rigor na sobreposição das máscaras, especialmente em fissuras finas e descontínuas.

Complementarmente, Tsaimou *et al.* (2024) analisaram ortofotos de VANT para mapeamento de fissuras em pavimentos de concreto com integração em GIS, por meio de uma abordagem automatizada baseada em técnicas de visão computacional. Em condições consideradas mais favoráveis, os autores reportam *F-score* médio entre 76,78% e 87,26%; contudo, é importante notar que esse valor médio é influenciado pelo desempenho muito alto da classe “não fissura” (F1 entre 96,13% e 98,75%), o que pode mascarar parcialmente as dificuldades na classe de interesse. Quando se observa especificamente o desempenho para a classe “área fissurada”, o

F1-score varia entre 74,04% e 76,12% nos cenários mais estáveis, mas cai para 57,44% no cenário com ruído intenso, associado principalmente a marcas de pneus e sombras, que geraram confusões com padrões lineares não patológicos. Os autores atribuem explicitamente essa queda ao aumento de falsos positivos em regiões com ruído escuro (marcas de pneu), além de discutirem que altitude de voo e tamanho de pixel (GSD) afetam a detectabilidade, sobretudo de fissuras mais finas.

Embora envolvam métodos e condições experimentais diferentes, os melhores modelos deste trabalho apresentaram Dice/F-score $\approx 0,854$ – $0,855$, com precisão $\approx 0,846$ – $0,853$ e recall $\approx 0,854$ – $0,865$, indicando um equilíbrio mais consistente entre omissões e comissões na segmentação em nível de pixel. Esse contraste com Tsaimou et al. (2024) reforça dois pontos: (1) o desempenho em fissuras tende a degradar significativamente quando há padrões lineares concorrentes (marcações escuras no pavimento e sombras), e (2) alcançar *recall* alto sem aumento desproporcional de falsos positivos é determinante para uso em manutenção, pois reduz o risco de subestimar trechos críticos. Assim, a evidência da literatura corrobora a interpretação de que o desempenho quantitativo observado aqui é competitivo e, ao mesmo tempo, que a robustez em cenários com ruído depende fortemente das condições de aquisição e da composição do *dataset*.

Em síntese, os resultados demonstram a superioridade da arquitetura YOLO11 para a inspeção automatizada de pavimentos. A capacidade desta família de modelos em aliar alta sensibilidade (*Recall*) a um delineamento preciso das fissuras (mAP₅₀), sem incorrer em custos computacionais proibitivos, supera as limitações observadas na geração anterior (YOLOv8). Diante da insignificante diferença no tempo de inferência, o modelo YOLO11-seg consolida-se como a solução mais robusta para a aplicação proposta, garantindo o máximo detalhamento da patologia necessário para a tomada de decisão em manutenção, enquanto mantém a viabilidade de processamento em tempo real.

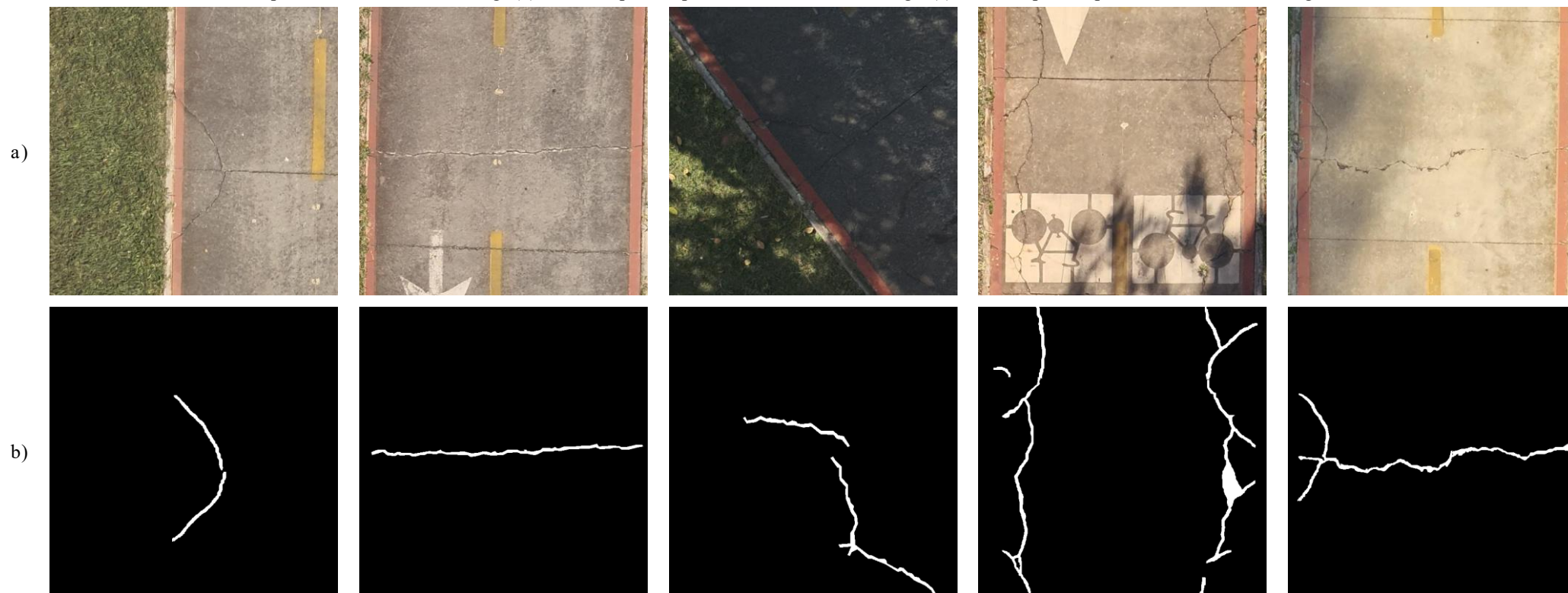
4.2.4 Avaliação qualitativa das inferências dos modelos de segmentação

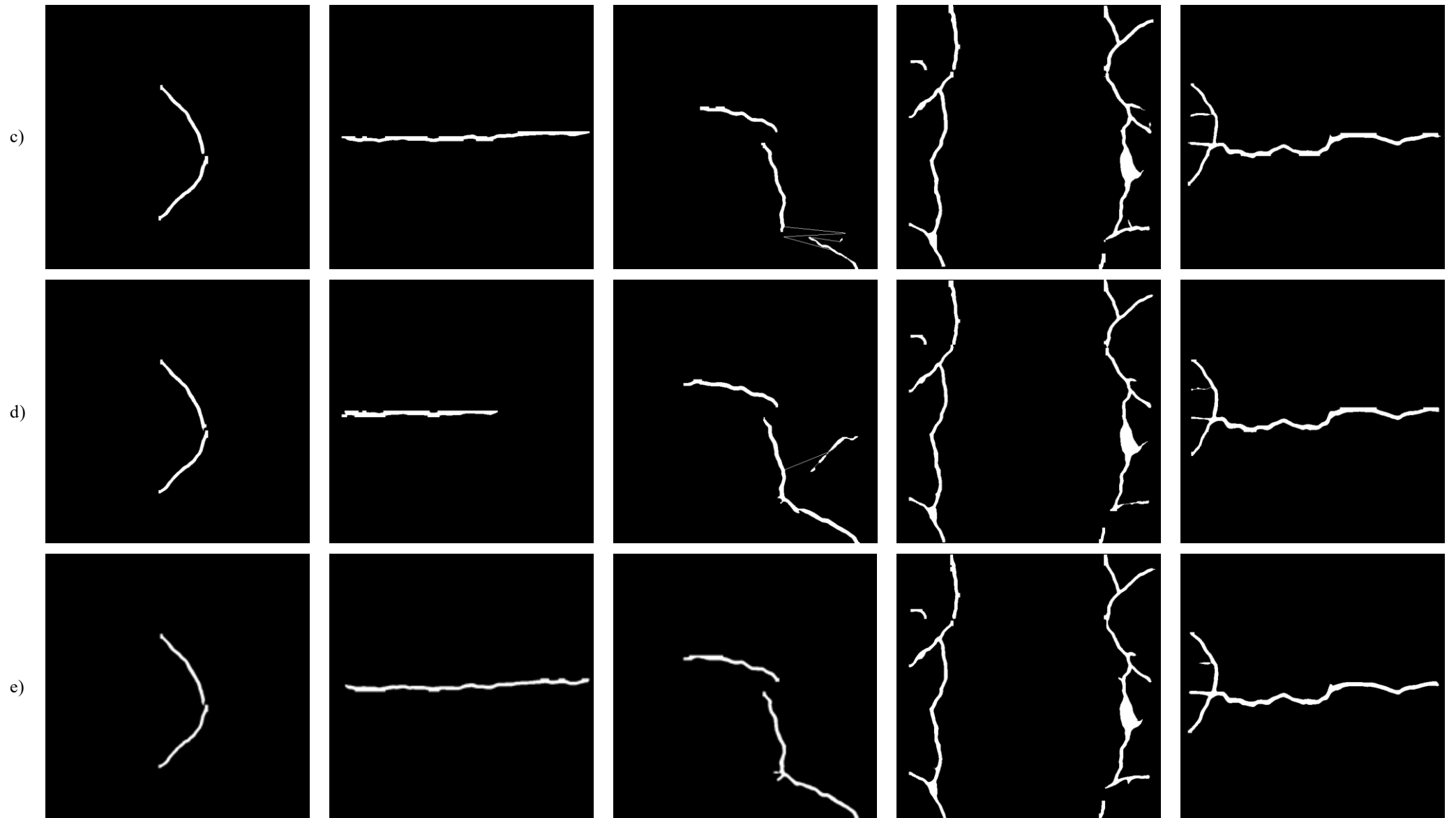
A avaliação qualitativa das inferências teve como objetivo analisar, de forma visual e comparativa, a capacidade dos modelos de segmentação em representar adequadamente a geometria real das fissuras, complementando os resultados quantitativos apresentados nas seções anteriores. Para isso, foram selecionadas amostras representativas do conjunto de teste, contemplando diferentes padrões geométricos de fissuras, níveis de contraste e condições de

iluminação. A Figura 15 apresenta exemplos dessas inferências, organizadas de modo a permitir a comparação direta entre a imagem de entrada (ROI), a máscara de referência (*ground truth*) e as máscaras preditas pelos modelos YOLOv8m-seg, YOLO11m-seg e YOLO11l-seg.

De maneira geral, observa-se que todos os modelos foram capazes de identificar corretamente a presença das fissuras e segmentar suas regiões principais, demonstrando boa aderência espacial às máscaras de referência. As predições preservam, em grande parte, a continuidade das fissuras, mesmo em casos de traçados curvilíneos ou irregulares, evidenciando a capacidade das arquiteturas em capturar padrões espaciais complexos em nível de pixel. Adicionalmente, todos os modelos demonstraram robustez frente a ruídos visuais, não apresentando falsos positivos significativos em áreas com sinalização horizontal (pinturas de solo), o que indica adequada distinção entre fissuras e elementos de marcação viária.

Figura 15 – Avaliação qualitativa da segmentação de fissuras em ciclovias de pavimento rígido: (a) imagem de entrada (ROI), (b) máscara de referência (ground truth), (c) máscara predita pelo modelo YOLOv8m-seg, (d) máscara predita pelo modelo YOLO11m-seg e (e) máscara predita pelo modelo YOLO11l-seg.





Fonte: O autor.

Ao analisar as inferências geradas pelo modelo YOLOv8m-seg (Figura 15c), observa-se um desempenho consistente na identificação das fissuras principais, com boa preservação da continuidade ao longo de seu traçado. Entretanto, em alguns exemplos, nota-se a geração de máscaras ligeiramente mais espessas do que o observado no *ground truth*, além da inclusão de pequenos segmentos adicionais em regiões adjacentes à fissura principal, isso é particularmente notável na terceira imagem de entrada, onde a ciclovia está sob forte sombreamento causado pela vegetação. Esse comportamento indica uma tendência à supersegmentação local, possivelmente associada à sensibilidade do modelo a feições lineares ou variações texturais do pavimento. Ainda assim, o modelo mantém adequada coerência geométrica global, sem fragmentação significativa das fissuras, o que é compatível com os valores de *recall* obtidos na avaliação quantitativa.

O modelo YOLO11m-seg (Figura 15d) apresentou desempenho visualmente consistente na segmentação das fissuras, com boa aderência geométrica ao traçado real nas regiões corretamente identificadas. Observa-se que as máscaras geradas acompanham de forma precisa a largura e a orientação da fissura principal, indicando adequada capacidade de delimitação em nível de pixel. Entretanto, em um dos exemplos, nota-se a interrupção da máscara antes da extremidade final da fissura, caracterizando uma subsegmentação parcial em trechos de menor contraste ou continuidade visual. Ainda assim, o modelo mantém bom alinhamento espacial e reduzida incorporação de regiões não relacionadas à fissura, evidenciando um equilíbrio entre precisão geométrica e sensibilidade. Esse comportamento está em consonância com os resultados quantitativos apresentados na Seção 4.2.3, nos quais o YOLO11m-seg apresentou valores elevados de mAP_{50} e *recall*, embora ligeiramente inferiores aos obtidos pela versão de maior porte.

Em relação ao YOLO11l-seg (Figura 15e), observa-se o desempenho qualitativo mais consistente entre as arquiteturas avaliadas. O modelo foi capaz de reproduzir a geometria das fissuras com elevada aderência ao *ground truth*, preservando a continuidade, a curvatura e a extensão das trincas ao longo de sua trajetória. Nos exemplos analisados, o YOLO11l-seg demonstrou maior capacidade de acompanhar fissuras longas e sinuosas, reduzindo a ocorrência de segmentações fragmentadas e mantendo uma representação mais próxima da espessura real das fissuras anotadas.

Além disso, o modelo apresentou melhor desempenho em cenários com fissuras múltiplas ou ramificadas, conseguindo delimitar adequadamente as diferentes ocorrências sem comprometer

sua integridade geométrica. Esse comportamento é coerente com o maior porte do modelo e com o número ampliado de parâmetros, que favorecem uma representação espacial mais rica em nível de pixel, e está alinhado aos resultados quantitativos apresentados na Seção 4.2.3, nos quais o YOLO11l-seg obteve os maiores valores de mAP50 e *recall*, indicando maior sensibilidade e consistência na segmentação das fissuras.

Essa superioridade qualitativa do YOLO11l-seg na preservação da topologia das fissuras corrobora avanços recentes na arquitetura de redes neurais para detecção de patologias. Huang *et al.* (2025) apontam que a nova estrutura de extração de características do YOLO11 permite uma representação mais fiel de objetos finos e irregulares em comparação às versões anteriores, como o YOLOv8, que tendem a apresentar maior fragmentação das máscaras em cenários de baixo contraste ou ruído de fundo. A dificuldade observada nos modelos menores (como o YOLOv8m-seg) em manter a continuidade da fissura em áreas de sombra alinha-se às limitações reportadas por Yuan *et al.* (2025), que identificaram que arquiteturas sem mecanismos de atenção aprimorados frequentemente produzem "falsos positivos" ou segmentações desconexas em pavimentos complexos. Portanto, a análise visual confirma que a escolha da versão "Large" (l) do YOLO11 não apenas melhora as métricas numéricas, mas é decisiva para garantir a fidelidade geométrica, superando os artefatos de supersegmentação comuns em modelos mais leves.

A análise conjunta dos exemplos evidencia que as diferenças entre os modelos estão menos associadas à capacidade de identificar a presença da fissura e mais relacionadas ao nível de refinamento geométrico das máscaras geradas (aderência do contorno, preservação de continuidade e supressão de ruídos). No conjunto avaliado, o YOLOv8m-seg apresentou segmentações consistentes, enquanto as variantes da YOLO11 tenderam a produzir máscaras com maior fidelidade geométrica, sobretudo no equilíbrio entre continuidade, largura aparente da fissura e redução de artefatos. Esse comportamento é coerente com as melhorias reportadas para o YOLO11 em termos de extração de características (*backbone/neck*) e suporte a segmentação de instâncias, especialmente para pequenos objetos (Ultralytics, 2024).

De modo geral, os resultados qualitativos corroboram as análises quantitativas apresentadas nas Seções 4.2.2 e 4.2.3, indicando que as arquiteturas avaliadas são capazes de segmentar fissuras de forma confiável em imagens de ciclovias obtidas por VANT. A superioridade visual observada nos modelos YOLO11, em especial na versão l-seg, reforça sua adequação para aplicações práticas de inspeção automatizada, nas quais a precisão geométrica das máscaras é

fundamental para futuras análises métricas, como estimativa de extensão, orientação e evolução temporal das fissuras.

4.3 ESTUDO DE CASO E INTEGRAÇÃO IA-GIS (SEÇÃO 3)

Nesta seção são apresentados e discutidos os resultados obtidos a partir da aplicação dos modelos de detecção e segmentação de fissuras em um estudo de caso real, bem como sua integração ao ambiente de Sistemas de Informação Geográfica. Diferentemente das análises conduzidas nas seções anteriores, fundamentadas em conjuntos de imagens individuais, esta etapa considera a aplicação dos modelos sobre um produto espacial contínuo e georreferenciado, permitindo avaliar seu comportamento em um cenário real de inspeção de ciclovias.

Para garantir a maior confiabilidade dos produtos cartográficos, foram selecionados os modelos YOLO11 tanto para a tarefa de detecção quanto para a segmentação, visto que esta arquitetura apresentou superioridade métrica consistente (maior mAP e equilíbrio entre *precision e recall*) e qualitativa em ambas as etapas avaliativas anteriores. Esses modelos foram aplicados a um ortomosaico georreferenciado obtido a partir de um novo levantamento aerofotogramétrico, independente dos dados utilizados no treinamento e na validação, possibilitando uma análise externa do desempenho dos modelos no contexto proposto.

Esta seção apresenta os resultados da integração entre os modelos de Aprendizado Profundo e o GIS, demonstrando a eficácia do fluxo proposto para a automatização do mapeamento das fissuras. A partir dessa integração, as saídas de detecção e segmentação passaram a ser analisadas sob uma perspectiva espacial, considerando sua representação sobre o ortomosaico e sua organização em camadas georreferenciadas. Dessa forma, os tópicos subsequentes discutem os procedimentos técnicos adotados, bem como os resultados espaciais obtidos e suas implicações para a análise e o monitoramento de fissuras em ciclovias.

4.3.1 Caracterização do estudo de caso e ortomosaico gerado

O levantamento aerofotogramétrico, realizado conforme 3.3.1, resultou na geração de um ortomosaico georreferenciado de alta resolução espacial, atingindo um GSD de 4,23 mm/px, conforme apresentado na Figura 16. Este valor indica que cada *pixel* da imagem representa fisicamente cerca de 4,2 mm do terreno, conferindo um nível de detalhamento subcentimétrico que garante a visualização clara de fissuras, juntas, remendos e elementos de sinalização,

permitindo a identificação de manifestações patológicas lineares com diferentes extensões e larguras.

Tal resolução demonstra-se compatível com a escala de análise proposta e encontra respaldo na literatura recente, como demonstrado por Boutet e Hallot (2025), que identificam o GSD como um parâmetro crítico para o desempenho de modelos de inteligência artificial. Segundo os autores, embora GSDs extremamente reduzidos permitam a detecção de microfissuras, eles tendem a elevar drasticamente a taxa de falsos positivos devido à confusão com texturas do material. Portanto, o valor obtido de 4,23 mm/px estabelece um equilíbrio operacional favorável: suficiente para a identificação de fissuras estruturais pelos modelos de IA, ao mesmo tempo em que mitiga o ruído excessivo que resoluções mais finas poderiam introduzir.

Figura 16 – Ortomosaico georreferenciado da ciclovia: (a) Visão geral do trecho mapeado; (b) Detalhe da resolução em nível do pavimento



(a)



(b)

Fonte: O autor.

Visualmente, o ortomosaico (Figura 16) evidencia a continuidade espacial da ciclovia ao longo do trecho analisado, bem como a heterogeneidade do entorno imediato, composto por áreas gramadas, canteiros e caminhos adjacentes. A presença de marcações pintadas no pavimento e de variações cromáticas naturais do concreto constitui um aspecto relevante do estudo de caso, uma vez que esses elementos representam potenciais fontes de ruído para os modelos de inteligência artificial e contribuem para a avaliação de sua robustez em ambiente real (Tsaimou *et al.*, 2024; Yuan *et al.*, 2025).

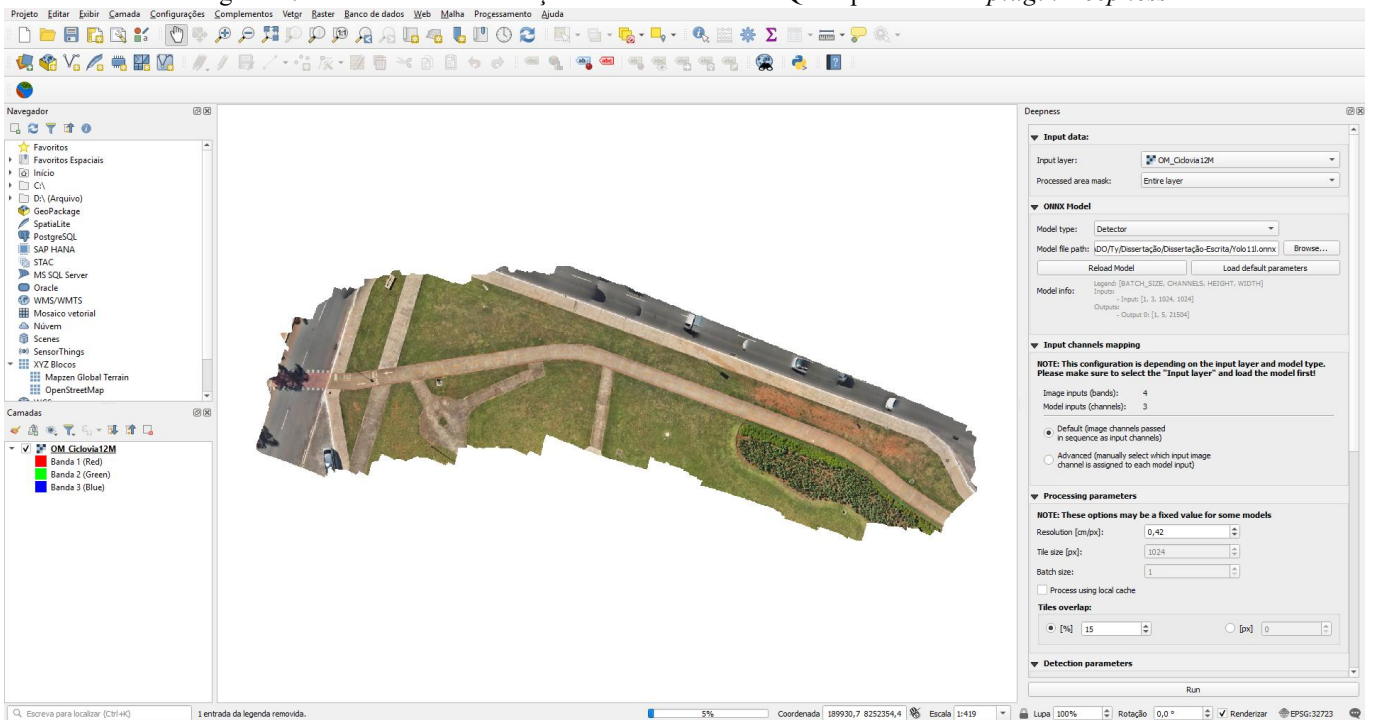
Esse ortomosaico georreferenciado constitui, portanto, a base espacial para a aplicação dos modelos selecionados e para a geração das feições vetoriais analisadas nas seções subsequentes. A partir dele, torna-se possível avaliar não apenas a capacidade dos modelos em identificar fissuras, mas também a coerência espacial e a representatividade dos resultados no contexto do mapeamento de patologias em ciclovias.

4.3.2 Integração dos modelos de IA ao ambiente GIS

Para esta etapa, aplicaram-se diretamente sobre o ortomosaico georreferenciado do estudo de caso os modelos YOLO11l de detecção e de segmentação, selecionados por apresentarem o melhor desempenho nas avaliações discutidas nas seções anteriores deste capítulo, de modo que a análise refletisse a configuração de maior desempenho do método proposto.

O ortomosaico foi inicialmente importado no ambiente QGIS, preservando suas informações de referência espacial, e submetido à inferência por meio do *plugin* Deepness, o que demandou a conversão prévia dos pesos treinados para o formato de interoperabilidade ONNX (*Open Neural Network Exchange*). Essa abordagem possibilitou a execução dos modelos diretamente sobre um produto espacial contínuo, permitindo que as saídas de detecção e segmentação fossem associadas às coordenadas geográficas do levantamento. A Figura 17 apresenta uma visão do ambiente de execução durante a aplicação de um dos modelos sobre o ortomosaico georreferenciado e a Figura 18 apresenta uma visão geral do ortomosaico após a aplicação dos modelos de detecção e segmentação.

Figura 17 – Ambiente de execução das inferências no QGIS por meio do *plugin Deepness*



Fonte: O autor.

Figura 18 - Visão geral do ortomosaico após inferências

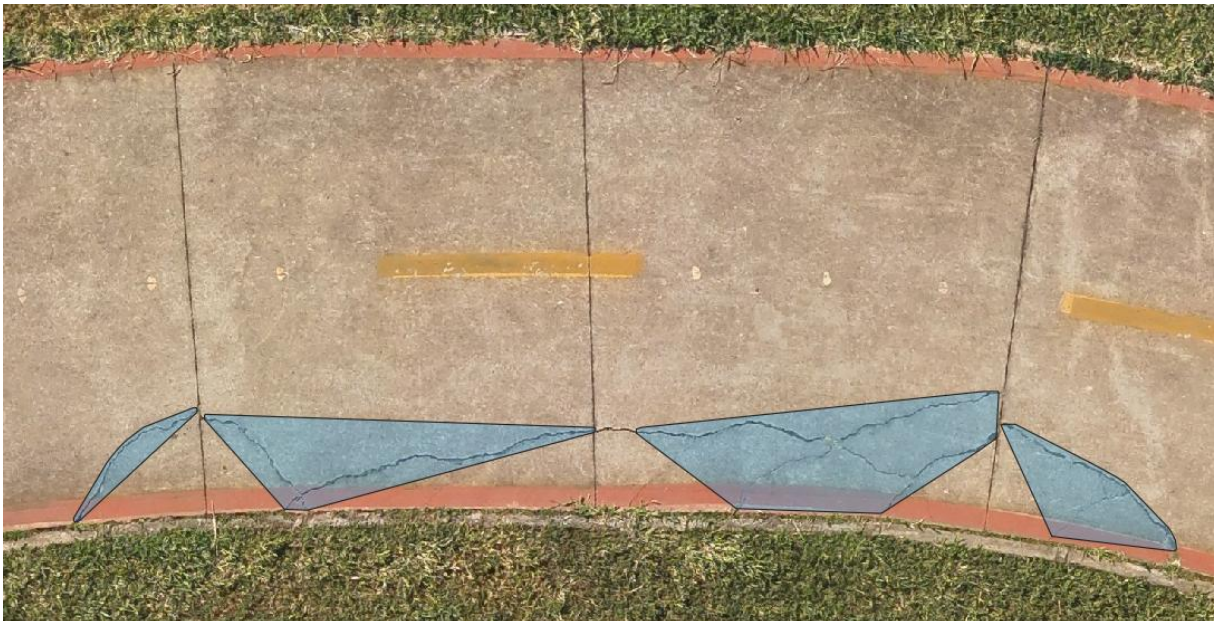


Fonte: O autor.

No caso do modelo de detecção, as caixas delimitadoras geradas apresentaram posicionamento espacial coerente com as regiões do pavimento que continham fissuras visíveis, permitindo a identificação das áreas afetadas ao longo do trecho analisado. Essas saídas demonstraram comportamento estável quando associadas ao sistema de referência espacial do ortomosaico, viabilizando sua representação como informação georreferenciada no ambiente GIS.

Para o modelo de segmentação, a análise inicial dos resultados evidenciou que a vetorização automática das máscaras geradas produzia representações geométricas excessivamente simplificadas. Em particular, a aplicação de uma envoltória convexa resultava na supressão de concavidades e ramificações características das fissuras, comprometendo a correspondência entre a geometria vetorial obtida e o traçado real observado no pavimento. Esse efeito pode ser observado na Figura 19, na qual a feição resultante não representa adequadamente a morfologia da patologia segmentada.

Figura 19 – Resultado da vetorização inicial da segmentação, evidenciando a simplificação geométrica das fissuras.



Fonte: O autor.

Esse comportamento mostrou-se inadequado para a análise espacial das fissuras, uma vez que a simplificação geométrica inviabilizava a avaliação da continuidade e da extensão das manifestações patológicas. A representação obtida, embora compatível com feições compactas e regulares, não preservava as características morfológicas essenciais das fissuras, limitando a interpretação dos resultados no contexto do mapeamento de patologias (Dipankar; Suman, 2023; Han; Yang; Yang, 2024).

Para superar essa limitação e garantir a fidelidade dos dados vetoriais, foi implementada a adaptação algorítmica descrita na metodologia, forçando a preservação dos vértices exatos do contorno da predição. A Figura 20 demonstra o resultado após este ajuste: a camada vetorial passou a delinear com maior precisão a morfologia da fissura, preservando suas ramificações e espessura. Essa correção metodológica foi determinante para viabilizar a geração de

mapeamentos técnicos mais precisos, assegurando que o produto vetorial refletisse a acurácia visual obtida pelo modelo YOLO11l-seg.

Figura 20 – Resultado da vetorização após ajuste, com preservação do contorno real das fissuras segmentadas.



Fonte: O autor.

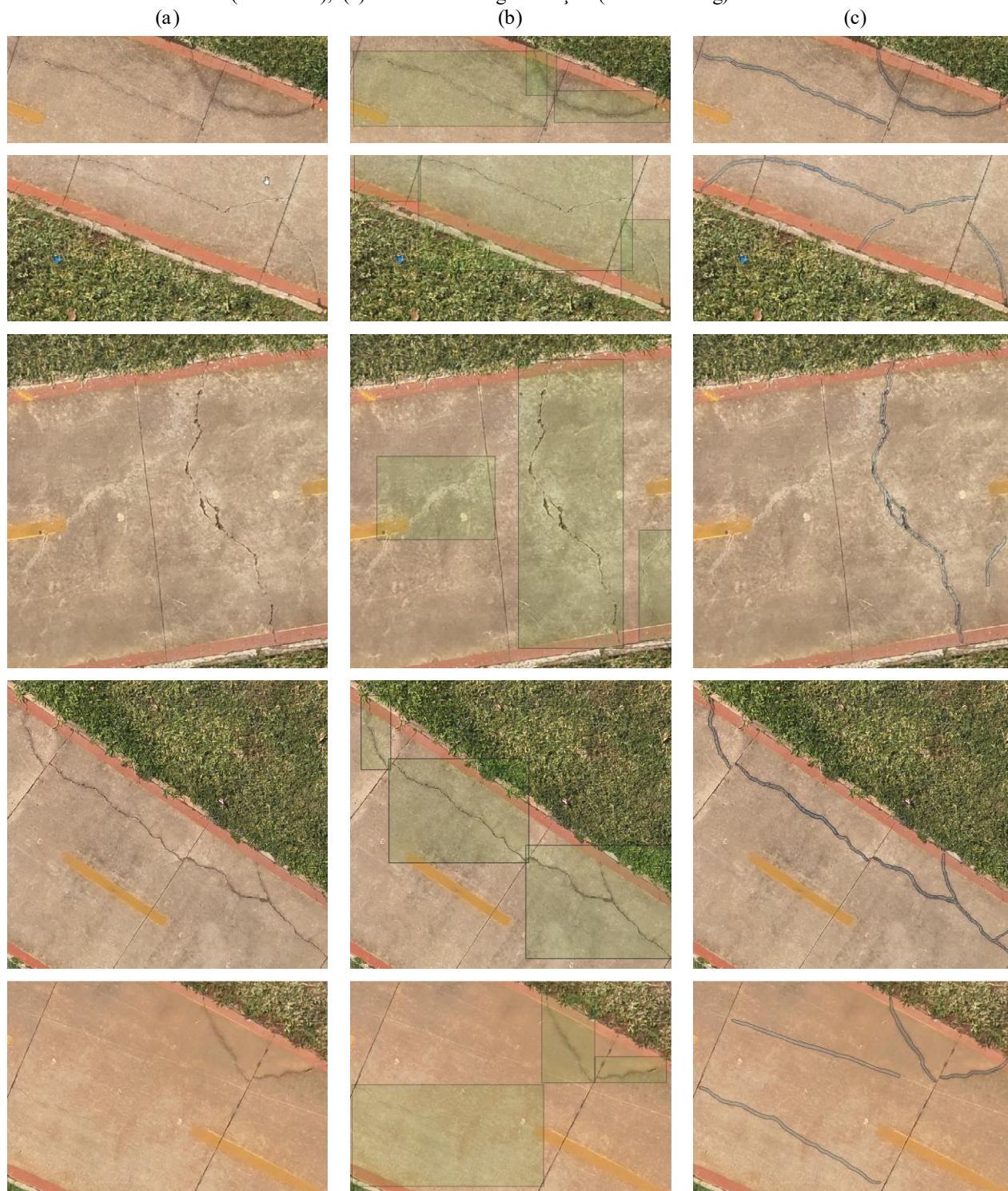
Com isso, a integração entre os modelos de inteligência artificial e o ambiente GIS foi consolidada, possibilitando a geração de resultados espaciais consistentes tanto para a detecção quanto para a segmentação. Essa etapa evidenciou a viabilidade da integração entre modelos de IA e GIS para o mapeamento automatizado de fissuras em infraestruturas cicloviárias, ao transformar saídas de segmentação em dados georreferenciados passíveis de análise, visualização e interpretação espacial. Além do contexto específico das ciclovias, os resultados indicam que o fluxo integrado IA–GIS proposto apresenta potencial de aplicação em outras infraestruturas urbanas, desde que observadas as particularidades de cada sistema. Dessa forma, consolidou-se um fluxo integrado IA–GIS capaz de apoiar a inspeção e o mapeamento sistematizado de fissuras, estabelecendo as bases para a avaliação qualitativa e a discussão dos resultados apresentada na seção seguinte.

4.3.3 Resultados espaciais e validação qualitativa integrada

A análise dos resultados espaciais foi conduzida a partir da sobreposição das saídas dos modelos de detecção e segmentação ao ortomosaico georreferenciado, permitindo uma avaliação qualitativa integrada do desempenho das arquiteturas no contexto real da ciclovia. A Figura 21 apresenta trechos representativos do ortomosaico (coluna a), as respectivas detecções por caixas

delimitadoras (coluna b) e as segmentações vetorizadas das fissuras (coluna c), sintetizando o comportamento observado ao longo de todo o trecho analisado.

Figura 21 – Validação dos resultados espaciais: (a) Trechos do ortomosaico; (b) Camada de detecção (YOLO11); (c) Camada de segmentação (YOLO11-seg)



Fonte: O autor.

De maneira geral, o modelo de detecção apresentou desempenho qualitativo satisfatório, sendo capaz de identificar corretamente as regiões da ciclovia que continham fissuras ao longo de todo o trecho avaliado. Não foram observadas ocorrências relevantes de falsos positivos ou falsos negativos, e todas as fissuras visíveis no pavimento foram adequadamente localizadas pelas caixas delimitadoras, conforme ilustrado na Figura 21 (b). Esse comportamento indica elevada robustez do modelo de detecção frente às variações geométricas e superficiais do pavimento, bem como frente à presença de padrões visuais concorrentes no ambiente urbano.

Em relação ao modelo de segmentação, os resultados indicaram boa capacidade de mapeamento da morfologia das fissuras, com preservação do traçado e da continuidade das manifestações patológicas na maior parte dos casos. Conforme observado na Figura 21 (c), praticamente todas as fissuras presentes no interior da ciclovia foram segmentadas, seja de forma contínua ou em segmentos sucessivos. No entanto, foram identificadas algumas ocorrências pontuais de falsos positivos e falsos negativos, especialmente em fissuras de menor contraste ou em regiões próximas às bordas do pavimento.

As limitações observadas, de caráter pontual, podem ser atribuídas a dois fatores técnicos distintos. A ocorrência de leves descontinuidades na geometria de algumas fissuras longas associa-se ao processo de inferência fatiada (*tiled inference*) realizado pelo *plugin*, que processa o ortomosaico em blocos de 640 x 640 pixels (resolução nativa de treinamento). Essa abordagem pode introduzir interrupções artificiais nas predições justamente nas zonas de transição entre os blocos. Já a incidência de falsos positivos e negativos relaciona-se à natureza mais restrita do conjunto de dados de segmentação em comparação ao de detecção. Enquanto o modelo de detecção beneficiou-se de um banco de dados com maior variabilidade de sensores e alturas de voo, o treinamento da segmentação restringiu-se a imagens obtidas a 12 m de altura (DJI Mavic 3), visando o detalhamento do pixel. Essa especificidade amostral resulta em uma menor capacidade de generalização frente a variações sutis de textura e iluminação que divergem do padrão estrito aprendido, ocasionando ruídos pontuais não observados na etapa de detecção. Diante dessas limitações pontuais, uma estratégia prática para aumentar a confiabilidade do produto é utilizar os modelos de detecção e segmentação de forma complementar, com validação cruzada no ambiente GIS.

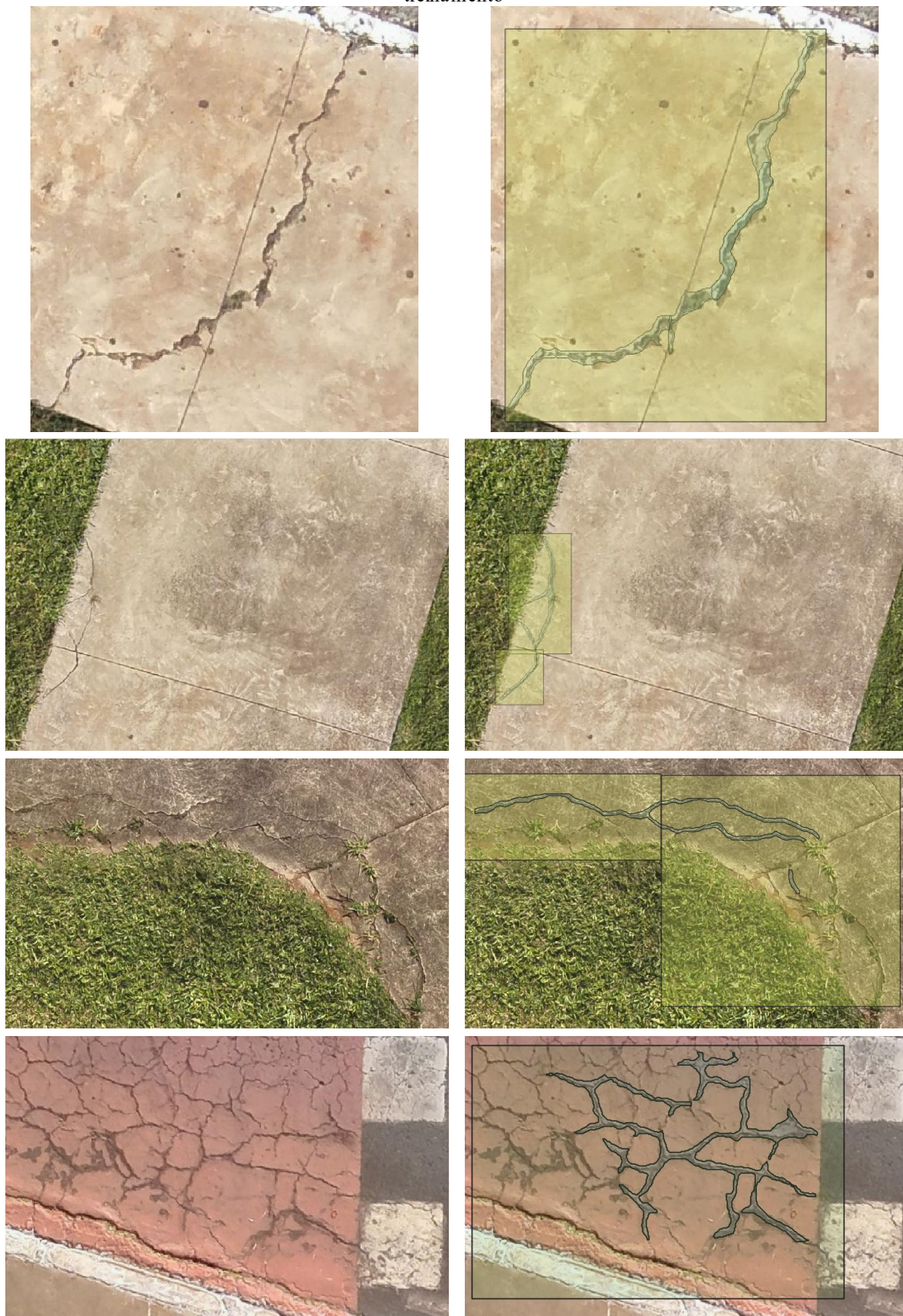
Considerando que o modelo de detecção apresentou elevada robustez na localização das regiões fissuradas no ortomosaico, enquanto a segmentação apresentou ocorrências pontuais de falsos positivos e falsos negativos, a utilização conjunta dos dois modelos se mostra vantajosa como

estratégia de validação cruzada e refinamento do produto espacial. Na prática, a camada de detecção pode funcionar como uma região de interesse para filtrar a camada segmentada: fissuras segmentadas que não intersectam as caixas delimitadoras (ou que intersectam apenas marginalmente) podem ser automaticamente sinalizadas como candidatas a falso positivo e removidas ou direcionadas para revisão. De modo complementar, regiões detectadas que apresentem segmentação ausente/fragmentada podem ser marcadas como “necessita inspeção”, reduzindo o risco de subestimação em áreas de baixo contraste. Essa combinação é particularmente útil em trechos próximos às bordas do pavimento e em fissuras sutis, onde a segmentação tende a ser mais sensível a ruídos e variações de textura. No ambiente QGIS, essa lógica pode ser implementada diretamente via operações espaciais entre camadas (por exemplo, interseção, seleção por localização e filtragem por atributos), aproveitando que as saídas de ambos os modelos são sobrepostas ao ortomosaico durante a validação qualitativa integrada.

Apesar dessas particularidades, um ponto de convergência fundamental foi a aderência de ambos os modelos ao protocolo de anotação estabelecido na metodologia, garantindo coerência entre o padrão aprendido pelo modelo e os resultados obtidos.

Além da análise no domínio específico da ciclovia, avaliou-se o comportamento dos modelos em trechos do ortomosaico correspondentes a superfícies pavimentadas não pertencentes ao foco principal do estudo, como calçadas, passeios e trechos de via asfáltica, os quais foram incorporados ao ortomosaico em função da altura de voo adotada. A Figura 22 apresenta exemplos representativos desses casos, nos quais se observa que, mesmo em superfícies com coloração, textura e acabamento distintos do pavimento da ciclovia, os modelos foram capazes de detectar e segmentar fissuras.

Figura 22 – Resposta dos modelos de detecção e segmentação em superfícies pavimentadas fora do domínio de treinamento



Fonte: O autor.

Nessas superfícies externas ao domínio de treinamento, o desempenho dos modelos mostrou-se inferior ao observado no contexto específico da ciclovia, com maior ocorrência de imprecisões geométricas e respostas menos consistentes. Ainda assim, a presença de detecções e segmentações não aleatórias indica que os modelos aprenderam padrões morfológicos associados às fissuras, como linearidade, ramificação e descontinuidade superficial, e não apenas características específicas do pavimento da ciclovia. Esse comportamento sugere uma capacidade de generalização parcial, ainda que limitada, para contextos pavimentados distintos daqueles utilizados no treinamento.

A análise conjunta das Figuras 21 e 22 evidencia a complementaridade entre os modelos de detecção e segmentação e reforça a robustez do fluxo IA–GIS proposto. Enquanto a detecção se mostrou eficiente para a identificação global das regiões afetadas por fissuras, a segmentação forneceu uma representação mais detalhada da geometria das manifestações patológicas. Em conjunto, essas abordagens permitiram o mapeamento abrangente das fissuras ao longo da ciclovia, com muitos acertos e poucos erros, além de indicar o potencial de aplicação do método em outros contextos urbanos pavimentados, desde que observadas as limitações associadas ao domínio de treinamento. De forma geral, os resultados espaciais observados confirmam que a integração entre os modelos de inteligência artificial e o ambiente GIS possibilita o mapeamento automatizado de fissuras em ciclovias de forma consistente e representativa, validando o fluxo metodológico proposto neste trabalho no contexto do estudo de caso analisado.

4.3.4 Produto e aplicabilidade em inspeção e manutenção

O produto consolidado deste fluxo de trabalho consiste em um banco de dados geográfico estruturado, contendo as camadas vetoriais das patologias devidamente georreferenciadas. Diferentemente de relatórios de inspeção tradicionais, baseados em planilhas manuais ou registros fotográficos estáticos, o resultado obtido permite a manipulação dinâmica e a consulta espacial dos dados, alinhando-se a abordagens que tratam a etapa de mapeamento em GIS como componente essencial para transformar a detecção em informação útil à gestão do ativo (Hu; Yee; Goff, 2024; Tsaimou *et al.*, 2024).

A conversão das inferências (caixas e máscaras) em vetores (shapefiles ou GeoPackage) viabiliza a extração automatizada de quantitativos físicos essenciais para a gestão da infraestrutura. A partir da camada de segmentação, por exemplo, torna-se possível calcular

indicadores como área fissurada/degradada e métricas lineares associadas às fissuras, a exemplo do que é reportado em fluxos que integram ortomosaicos georreferenciados, segmentação por deep learning e cálculo de métricas (área total e comprimento característico de fissuras) para subsidiar decisões de manutenção (Hu; Yee; Goff, 2024). Em abordagens voltadas a pavimentos, a vetorização e a organização dos resultados em camadas GIS também são exploradas para análises quantitativas e temáticas (com vistas ao planejamento de intervenções (Dipankar; Suman, 2023).

Além da quantificação, a componente espacial do produto permite análises de distribuição de danos. A visualização integrada no ambiente GIS facilita a identificação de zonas críticas de concentração de patologias, permitindo aos gestores priorizarem intervenções em trechos específicos da malha cicloviária baseando-se em critérios objetivos de severidade, em linha com trabalhos que enfatizam o GIS como suporte para comparar regiões fissuradas, localizar áreas-problema e apoiar a priorização de necessidades de manutenção (Dipankar; Suman, 2023). Além disso, ao estruturar as saídas como dados geoespaciais, o fluxo permite acompanhar metadados que variam ao longo do tempo, como a propagação de fissuras, viabilizando análises comparativas entre levantamentos e monitoramento evolutivo do dano (Tsamou *et al.*, 2024)

Dessa forma, os resultados apresentados neste capítulo comprovam que a integração entre a visão computacional e o geoprocessamento é capaz de entregar uma ferramenta de suporte à decisão robusta, preenchendo a lacuna de integração identificada na literatura. O fluxo proposto não apenas automatiza a detecção, mas sistematiza o mapeamento técnico consolidando a contribuição científica do trabalho ao oferecer uma alternativa mais rápida, rastreável e auditável em comparação aos métodos de inspeção visual humana, superando limitações tradicionais de eficiência e acessibilidade aos dados (Chun; Yamane; Tsuzuki, 2021).

5 CONCLUSÕES

Este trabalho investigou e validou um fluxo metodológico integrado para identificação automatizada de fissuras em ciclovias de pavimento rígido, combinando aerofotogrametria por VANT, modelos de inteligência artificial para detecção e segmentação e integração com um ambiente GIS voltado à gestão de inspeção, operação e manutenção. O estudo partiu do entendimento de que a inspeção tradicional de infraestruturas urbanas, frequentemente baseada em vistorias de campo e registros fotográficos pouco estruturados, limita a rastreabilidade

espacial, dificulta a quantificação dos danos e reduz a capacidade de priorização objetiva de intervenções. Nesse contexto, o trabalho foi direcionado não apenas a obter bons indicadores de desempenho em IA, mas principalmente a consolidar um produto georreferenciado e operacionalizável no GIS, transformando inferências (caixas e máscaras) em camadas espaciais úteis para planejamento de manutenção.

Os objetivos propostos foram cumpridos de forma integrada. Primeiramente, foi constituído um banco de dados anotado para a tarefa de detecção, com imagens em resolução adequada à identificação de fissuras e com aumento do conjunto por *data augmentation*; em seguida, foram treinados e avaliados modelos representativos, comparando desempenho por métricas consolidadas (precisão, *recall*, mAP_{50}) e tempo de inferência, de modo a equilibrar qualidade e viabilidade computacional. Em paralelo, foi estruturado um banco de dados específico para segmentação (nível de pixel), e os modelos foram avaliados de forma comparativa, novamente considerando métricas e custo computacional, com destaque à estabilidade do treinamento e à convergência dos modelos para patamares semelhantes ao longo das épocas.

Por fim, foi conduzido um estudo de caso independente com aquisição de dados por VANT e geração de ortomosaico georreferenciado, seguido da execução da inferência em ambiente GIS, utilizando o QGIS e o *plugin Deepness*, com processamento em blocos (*tiling*) e posterior geração de camadas de detecção e segmentação. Essa etapa foi decisiva para demonstrar que o trabalho não se encerra no desempenho do modelo, mas evolui até a entrega de informação espacial acionável para a gestão.

O conjunto utilizado para detecção foi estruturado com 738 imagens, resolução 1024×1024 , com um total de 1757 instâncias anotadas e divisão 70%/20%/10% para treinamento, validação e teste, respectivamente. Essa configuração permitiu avaliar o comportamento de modelos de detecção em um cenário realista de fissuras finas e de alta variabilidade visual.

Entre os modelos testados, observou-se que o desempenho de detecção atingiu patamares consistentes, com destaque para o YOLO11l, que apresentou $mAP_{50} = 0,831$, precisão = 0,887, *recall* = 0,737 e tempo médio de inferência de 12,2 ms por imagem, configurando um bom compromisso entre qualidade e eficiência computacional. Outros modelos também apresentaram desempenho competitivo, como o RT-DETR-l, com $mAP_{50} = 0,797$ e inferência média de 13,6 ms. Em termos práticos, esses resultados indicam que a detecção é efetiva para triagem e identificação global de regiões fissuradas, especialmente quando o objetivo é localizar rapidamente trechos críticos ao longo da ciclovias.

Um ponto metodológico relevante foi a compatibilização entre resolução, escala do dano e altura de voo: o estudo explicitou que a etapa de detecção pode ser conduzida com imagens obtidas em faixas de altura mais amplas (aprox. 5 m a 35 m), desde que a qualidade do produto (nitidez e GSD) mantenha a fissura representada com informação suficiente para o modelo. Isso reforça a aplicabilidade do método em levantamentos operacionais, nos quais restrições de voo e produtividade exigem flexibilidade.

Para segmentação, o conjunto foi estruturado com 413 imagens, resolução 640×640, 895 anotações e divisão 70%/20%/10% para treinamento, validação e teste, respectivamente, reforçando um cenário de dados moderados, porém suficientemente representativos para aprendizagem do padrão espacial das fissuras em nível de pixel. A evolução do treinamento indicou estabilização do desempenho e ausência de evidências de divergência, com mAP₅₀ final próximo de 0,86 para as arquiteturas avaliadas, sugerindo que os modelos atingiram um limite coerente com o conjunto disponível.

Na avaliação final em teste, o melhor desempenho foi obtido pelo YOLO11l-seg, com mAP₅₀ = 0,872, *recall* = 0,865, precisão = 0,846 e tempo médio de inferência de 5,6 ms, seguido de perto por outras arquiteturas (por exemplo, YOLO11m-seg com mAP₅₀ = 0,858). O resultado é relevante pois mostra que a segmentação fornece uma representação mais detalhada e fiel da geometria das fissuras, condição essencial para quantificação e análise espacial (comprimentos, áreas afetadas e padrões de concentração), especialmente quando o objetivo é apoiar a gestão do ativo baseada em evidências.

Apesar do tamanho relativamente reduzido do conjunto anotado, tanto para detecção como para segmentação, os resultados obtidos indicam que, em uma tarefa especializada e com condições de aquisição e pré-processamento padronizadas, é possível alcançar desempenho competitivo na detecção de fissuras finas quando se empregam estratégias como *transfer learning* e *data augmentation*, o que constitui um diferencial do trabalho e reforça a viabilidade de aplicações em contextos com disponibilidade limitada de dados.

Um aspecto técnico destacado pelo trabalho foi a sensibilidade da segmentação à escala: para essa tarefa, a aquisição em altura menor (12 m) foi considerada adequada para preservar a informação necessária à delimitação em nível de pixel, reforçando que o planejamento do voo deve considerar não apenas cobertura e produtividade, mas também a exigência de detalhe do algoritmo e do produto desejado.

A validação por estudo de caso independente teve papel central para demonstrar transferência do método para um cenário distinto daquele utilizado no treinamento e, principalmente, para mostrar que o fluxo é executável no GIS com resultados espacialmente consistentes. A inferência foi realizada diretamente no ortomosaico em ambiente QGIS, com divisão em blocos para contornar limitações de memória e permitir processamento de grandes dimensões, empregando o *plugin Deepness*. Esse procedimento consolidou a etapa-chave de operacionalização: em vez de resultados “soltos” em imagens, o processo gera camadas espaciais sobre o ortomosaico, viabilizando o uso em rotinas de gestão.

A análise conjunta evidenciou a complementaridade entre detecção e segmentação: enquanto a detecção se mostrou mais apropriada para o reconhecimento global e rápido de regiões afetadas, a segmentação forneceu maior detalhamento geométrico. Em conjunto, as duas abordagens permitiram o mapeamento abrangente das fissuras ao longo da ciclovia, com potencial de aplicação em outros contextos pavimentados, desde que respeitadas as limitações do domínio de treinamento. Essa conclusão reforça que a robustez da metodologia não depende de um único tipo de modelo, mas do uso articulado das saídas (caixas e máscaras) em um ambiente espacial.

O produto consolidado do fluxo foi descrito como um banco de dados geográfico estruturado, contendo camadas vetoriais georreferenciadas, o que representa avanço relevante em relação a relatórios tradicionais. Ao converter inferências em vetores (*Shapefile/GeoPackage*), torna-se possível extrair quantitativos físicos, como área degradada e extensão linear das fissuras, além de executar análises espaciais de distribuição (zonas críticas e priorização de trechos). Portanto, o ganho principal do trabalho não é apenas identificar fissuras, mas transformar fissuras em informação de gestão, integrável à lógica de manutenção e planejamento.

Apesar dos resultados promissores, o trabalho evidenciou limitações. A primeira limitação refere-se ao domínio de treinamento. No estudo de caso, a detecção apresentou capacidade de generalização parcial, porém limitada, para superfícies pavimentadas distintas daquelas usadas na fase de treinamento. Isso indica que, embora o modelo possa reconhecer padrões fissurados em outros contextos, a robustez plena depende de diversidade e representatividade do banco de dados, sobretudo quando há variações de textura, material, iluminação, nível de desgaste e interferências visuais.

A segunda limitação foi operacional e ligada à etapa GIS: observou-se que o processo de vetorização inicial das máscaras pode aplicar uma simplificação geométrica baseada em *convex hull*, o que tende a “suavizar” indevidamente fissuras irregulares e reduzir a fidelidade do

contorno real. Essa constatação é particularmente relevante porque o valor da segmentação para manutenção está justamente em preservar geometria e extensão. Portanto, para fins de quantificação rigorosa, a etapa de vetorização deve ser cuidadosamente ajustada ou substituída por rotinas que preservem o contorno, incluindo simplificação controlada, filtragem por ruído e, quando necessário, pós-processamento (por exemplo, esqueletonização para cálculo de comprimento).

Diante desses achados, as contribuições desta dissertação se consolidam em múltiplas frentes. A primeira contribuição é a própria consolidação de um fluxo metodológico completo e replicável, conectando aquisição por VANT, preparação de dados, treinamento e avaliação de modelos, inferência em produtos georreferenciados e integração em GIS, com decisões e parâmetros explicitados de maneira a favorecer reuso e adaptação em outros estudos. A segunda contribuição está na evidência experimental de que, em contextos controlados e bem caracterizados, com parâmetros de aquisição compatíveis com o tamanho do dano e um conjunto de dados representativo daquele cenário, é possível alcançar desempenho robusto e tecnicamente consistente, sustentado tanto por métricas quantitativas quanto pela qualidade visual das predições, sem inviabilizar o uso operacional devido a custo computacional. A terceira contribuição é a integração efetiva com GIS como ponte para a gestão, deslocando o foco do simples reconhecimento automático para a geração de informação espacial acionável (camadas, mapas temáticos e base para quantificação e priorização de intervenções).

A quarta contribuição é o desenvolvimento do banco de dados anotado, construído especificamente para a realidade de ciclovias e fissuras em pavimento rígido, com potencial de apoiar novas investigações, comparações metodológicas e evolução da área; esse banco pode ser disponibilizado mediante solicitação, o que favorece reprodutibilidade e colaboração, respeitando as limitações de direitos e compartilhamento associadas aos dados de aquisição.

Como desdobramentos futuros, recomenda-se ampliar e diversificar o banco de dados, incorporando diferentes ciclovias, materiais, estados de conservação e condições de iluminação e umidade, para reduzir viés de domínio e aumentar a generalização. Também se mostra promissor investigar estratégias de adaptação de domínio e *fine-tuning* em novos trechos, diminuindo perdas quando o cenário de aplicação diverge do treinamento. No eixo GIS, é recomendável aprimorar o pós-processamento de vetorização para preservar a geometria real das fissuras e, quando o objetivo for quantificação, incorporar rotinas que possibilitem medidas mais fiéis (por exemplo, métodos que estimem extensão e conectividade a partir de máscaras

sem deformação excessiva). Por fim, há espaço para consolidar a integração com rotinas de gestão de ativos, aproximando o método de um sistema de apoio à decisão. Nesse sentido, sugere-se investigar a interoperabilidade com plataformas BIM (*Building Information Modeling*), centralizando as informações em um ambiente robusto que possibilite o rastreamento histórico da evolução das patologias e a documentação das intervenções realizadas.

Em síntese, os resultados confirmam que a integração entre IA, aerofotogrametria e GIS possibilita o mapeamento automatizado de fissuras em ciclovias de forma consistente e representativa, validando o fluxo metodológico proposto no estudo de caso e demonstrando seu potencial como base tecnológica para uma gestão da manutenção mais objetiva, rastreável e orientadas por dados.

REFERÊNCIAS BIBLIOGRÁFICAS

ANAC. Agência Nacional de Aviação Civil. **Regulamento Brasileiro de Aviação Civil Especial: RBAC-E nº 94**: Requisitos Gerais para Aeronaves não Tripuladas de Uso Civil. Brasília: ANAC, 2017.

AGISOFT. **Metashape Professional**. Versão 2.1. St. Petersburg: Agisoft LLC, 2025. Software. Disponível em: <https://www.agisoft.com>. Acesso em: 2 dez. 2025.

AI, D. *et al.* Computer vision framework for crack detection of civil infrastructure—A review. **Engineering Applications of Artificial Intelligence**, v. 117, p. 105478, jan. 2023.

ALBUQUERQUE, T. S. **Identificação de risco para manutenção de ciclovias no distrito federal**. Monografia (Graduação) – Faculdade de Tecnologia, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, 2020.

ARAFIN, P.; BILLAH, M.; ISSA, A. Deep learning-based concrete defects classification and detection using semantic segmentation. **Structural Health Monitoring**, v.23, p. 383-409, mai 2023.

ASZKOWSKI, P. *et al.* Deepness: Deep neural remote sensing plugin for QGIS. **SoftwareX**, v. 23, p. 101495, 2023.

AZARI, H.; O'SHEA, D.; CAMPBELL, J. Application of Unmanned Aerial Systems for Bridge Inspection. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2676, n. 1, p. 401–407, ago. 2021.

BADUGE, S. K. *et al.* Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. **Automation in Construction**, v. 141, p. 104440, 2022.

BAKHTIARNIA, Arian; ZHANG, Qi; IOSIFIDIS, Alexandros. Efficient High-Resolution Deep Learning: A Survey. **arXiv:2207.13050**, 2022.

BALBO, J. T. **Pavimentos de Concreto**. São Paulo: Oficina de Textos, 2009.

BASTOS, C; MOTA, E. **Pavimentação de ciclovias**. Rodovias & Vias – Infraestrutura e Desenvolvimento, Ano 14 – Edição 74/2013.

BOUTET, Simon; HALLOT, Pierre. Ground Sampling Distance as a Key Parameter for Automatic Crack Detection in Built Heritage: A Practical Framework With YOLOv5. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XLVIII-M-9-2025, p. 179–186, 2025.

BRASIL. Ministério das Cidades. Secretaria Nacional de Transporte e da Mobilidade Urbana. **Programa Bicicleta Brasil**: caderno de referência para elaboração de plano de mobilidade por bicicleta nas cidades. Brasília, 2007.

BRICEÑO, J. *et al.* Comparación del desprendimiento de material en pavimentos rígidos reforzados con malla electrosoldada o fibras. **Ciencia e Ingeniería**, v. 40, n. 3, p. 331–338, 2019.

CARION, Nicolas *et al.* End-to-end object detection with transformers. In: **European Conference on Computer Vision**. Cham: Springer, 2020. p. 213-229.

CAO, W.; LIU, Q.; HE, Z. Review of Pavement Defect Detection Methods. **IEEE Access**, v. 8, p. 14531–14544, 2020.

CHEN, Y. *et al.* Building Surface Crack Detection Using Deep Learning Technology. **Buildings**, v. 13, n. 7, p. 1814, jul. 2023.

CHUN, P.; YAMANE, T.; TSUZUKI, Y. Automatic Detection of Cracks in Asphalt Pavement Using Deep Learning to Overcome Weaknesses in Images and GIS Visualization. **Applied Sciences**, v. 11, n. 3, p. 892, 20 jan. 2021.

DANG, M. *et al.* CDD-TR: Automated concrete defect investigation using an improved deformable transformers. **Journal of building engineering**, v. 75, p. 106976, set. 2023.

DAULAT, S. *et al.* Challenges of integrated multi-infrastructure asset management: a review of pavement, sewer, and water distribution networks. **Structure and Infrastructure Engineering**, p. 1–20, 9 set. 2022.

DECEA. Departamento de Controle do Espaço Aéreo. **Instrução do Comando da Aeronáutica: ICA 100-40**: Aeronaves não Tripuladas e o Acesso ao Espaço Aéreo Brasileiro. Rio de Janeiro: DECEA, 2020.

DENG, J. *et al.* Review on computer vision-based crack detection and quantification methodologies for civil structures. **Construction and Building Materials**, v. 356, p. 129238, nov. 2022.

DENG, L.; YU, D. Deep Learning: Methods and Applications. **Foundations and Trends® in Signal Processing**, v. 7, n. 3-4, p. 197–387, 2014.

DIPANKAR, A.; SUMAN, S. K. Pavement crack detection based on a deep learning approach and visualisation by using GIS. **International Journal of Pavement Engineering**, v. 24, n. 1, fev. 2023.

DNIT - Departamento Nacional de Infraestrutura de Transportes. **Manual de estudos de tráfego**. Rio de Janeiro: Ministério dos Transportes, 2006.

DORAFSHAN, S.; MAGUIRE, M. Bridge inspection: human performance, unmanned aerial systems and automation. **Journal of Civil Structural Health Monitoring**, v. 8, n. 3, p. 443–476, mai. 2018.

DUNG, C. V.; ANH, L. D. Autonomous concrete crack detection using deep fully convolutional neural network. **Automation in Construction**, v. 99, p. 52–58, mar. 2019.

Dwyer, B., Nelson, J., Hansen, T., *et al.*, "Roboflow," (2024). Disponível em: <<https://roboflow.com>>.

ELTOUNY, K.; SAJEDI, S.; LIANG, X. Dmg2Former-AR: Vision Transformers with Adaptive Rescaling for High-Resolution Structural Visual Inspection. **Sensors**, v. 24, n. 18, p. 6007, 2024.

FAN, Luxin *et al.* Impact of Image Preprocessing and Crack Type Distribution on YOLOv8-Based Road Crack Detection. **Sensors**, v. 25, n. 7, art. 2180, 2025.

FANG, F. *et al.* A novel hybrid approach for crack detection. **Pattern Recognition**, v. 107, p. 107474, nov. 2020.

GAO, L. *et al.* Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. **Sensors**, v. 19, n. 3, p. 684, 7 fev. 2019.

GAO, Xiaohu; CAO, Chunmei; YI, Xiaojing. Using the improved YOLOv11 model to enhance computer vision applications for building crack detection algorithms. **Scientific Reports**, v. 15, art. 38843, 2025.

GIRSHICK, R. *et al.* Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 **IEEE Conference on Computer Vision and Pattern Recognition**, p. 580–587, jun. 2014.

GOODFELLOW, IAN; BENGIO, YOSHUA; COURVILLE, AARON. **Deep Learning**. Cambridge: MIT Press, 2016.

GÖTSCHI, T.; GARRARD, J.; GILES-CORTI, B. Cycling as a Part of Daily Life: A Review of Health Perspectives. **Transport Reviews**, v. 36, n. 1, p. 45-71, 2016.

HAN, C.; YANG, H.; YANG, Y. Enhancing pixel-level crack segmentation with visual mamba and convolutional networks. **Automation in Construction**, v. 168, p. 105770–105770, set. 2024.

HE, X. *et al.* UAV-based road crack object-detection algorithm. **Automation in Construction**, v. 154, p. 105014, out. 2023.

HOLM, A. L.; GLÜMER, C.; DIDERICHSEN, F. Health Impact Assessment of increased cycling to place of work or education in Copenhagen. **BMJ Open**, v. 2, n. 4, p. e001135, 2012.

HORA, K. M. da. **Metodologia Para Avaliação Do Ciclo De Vida De Ciclovias De Pavimento Rígido**. Dissertação (Mestrado) – Programa de Pós-Graduação em Estruturas e Construção Civil, Universidade de Brasília, Brasília, 2021.

HORA, K. M.; PEREIRA, C. H. A. F. Avaliação da Evolução de Danos em Pavimento Rígido de Ciclovia. In: Construção 2020, 2020, Brasília. **ANAIS - Congresso de Construção Civil - CONSTRUÇÃO 2020**, 2020. v. 1.

HSIEH, Y.-A.; TSAI, Y. J. Machine Learning for Crack Detection: Review and Model Performance Comparison. **Journal of Computing in Civil Engineering**, v. 34, n. 5, p. 04020038, set. 2020.

HU, D.; YEE, T.; GOFF, D. Automated crack detection and mapping of bridge decks using deep learning and drones. **Journal of Civil Structural Health Monitoring**, v. 14, n. 3, p. 729–743, jan. 2024.

HUANG, S. et al. A Real-time Concrete Crack Detection and Segmentation Model Based on YOLOv11. **arXiv:2508.11517**, 2025.

ITF. INTERNATIONAL TRANSPORT FORUM. **Cycling, health and safety**. ITF Research Reports. Paris: OECD Publishing, 2013.

JIANG, Y.; PANG, D.; LI, C. A deep learning approach for fast detection and classification of concrete damage. **Automation in Construction**, v. 128, p. 103785, ago. 2021.

JOCHER, Glenn; CHAURASIA, Ayush; QIU, Jing. **Ultralytics YOLOv8**: versão 8.0.0. 2023. Disponível em <<https://github.com/ultralytics/ultralytics>>. Acesso em: 6 jul. 2025.

JOCHER, Glenn; QIU, Jing. **Ultralytics YOLO11**: versão 11.0.0. 2024. Disponível em <<https://github.com/ultralytics/ultralytics>>. Acesso em: 6 jul. 2025.

KANG, D. *et al.* Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. **Automation in Construction**, v. 118, p. 103291, out. 2020.

KARIMI, S.; IORDANOVA, I. Integration of BIM and GIS for Construction Automation, a Systematic Literature Review (SLR) Combining Bibliometric and Qualitative Analysis. **Archives of Computational Methods in Engineering**, v. 28, jan. 2021.

KHAN, M. W. *et al.* Real-Time Road Damage Detection and Infrastructure Evaluation Leveraging Unmanned Aerial Vehicles and Tiny Machine Learning. **IEEE internet of things journal**, v. 11, n. 12, p. 21347–21358, 15 jun. 2024.

KATSIGIANNIS, Stamos *et al.* Deep learning for crack detection on masonry façades using limited data and transfer learning. **Journal of Building Engineering**, v. 76, art. 107105, 2023.

LECCA, Michela; BIANCO, Simone. Common issues and human intervention in object detection from handcrafted features to deep learning: discussion. **Journal of the Optical Society of America A**, v. 42, n. 12, p. 1977–1987, 2025.

LEE, K.; LEE, S.; KIM, H. Y. Bounding-box object augmentation with random transformations for automated defect detection in residential building façades. **Automation in Construction**, v. 135, p. 104138, mar. 2022.

LEE, K.; LEE, S.; KIM, H. Y. Deep Learning-Based Defect Detection Framework for Ultra High Resolution Images of Tunnels. **Sustainability**, v. 15, n. 2, art. 1292, 2023.

LIN, T.-Y. et al. Microsoft COCO: Common Objects in Context. In: **EUROPEAN CONFERENCE ON COMPUTER VISION**. Zurich: Springer, 2014. p. 740–755.

LIU, J. *et al.* Structural damage detection and localization via an unsupervised anomaly detection method. **Reliability Engineering & System Safety**, v. 252, p. 110465, dez. 2024.

LIU, W. *et al.* SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*, v. 9905, p. 21–37, 2016.

LONGLEY, P. A. *et al.* **Sistemas e Ciência da Informação Geográfica**. 3. ed. Porto Alegre: Bookman, 2013.

MAHARANA, K.; MONDAL, S.; NEMADE, B. A Review: Data Pre-Processing and Data Augmentation Techniques. **Global Transitions Proceedings**, v. 3, n. 1, p. 91–99, abr. 2022.

MALINVERNI, E. A. *et al.* A semantic graph database for the interoperability of 3D GIS data. **Applied Geomatics**, v. 14, n. S1, p. 53–66, ago. 2020.

MANDIROLA, M. *et al.* Use of UAS for damage inspection and assessment of bridge infrastructures. **International Journal of Disaster Risk Reduction**, v. 72, p. 102824, abr. 2022.

MINAEE, S. et al. Image Segmentation Using Deep Learning: A Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 7, p. 3523–3542, 2022.

MOTTA, R.A. **Método para a determinação da sustentabilidade de ciclovias**. Tese (Doutorado) – Programa de Pós-Graduação em Transportes, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, 2016.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array**, v. 16, p. 100258, 2022. DOI: 10.1016/j.array.2022.100258.

MUNAWAR, H. S. *et al.* Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages. **Drones**, v. 6, n. 1, p. 5, dez. 2021.

NISHIKAWA, T. *et al.* Concrete Crack Detection by Multiple Sequential Image Filtering. **Computer-Aided Civil and Infrastructure Engineering**, v. 27, n. 1, p. 29–47, mar. 2011.

NOORALISHAHI, P. *et al.* Drone-Based Non-Destructive Inspection of Industrial Sites: A Review and Case Studies. **Drones**, v. 5, n. 4, p. 106, set. 2021.

OJO, S. A. *et al.* Assessment of road infrastructure using remote sensing and GIS methodology for monitoring the condition of paved and unpaved roads. **IOP Conference Series: Materials Science and Engineering**, v. 640, p. 012099, nov. 2019.

OLIVAS, E. S. *et al.* **Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques**. Hershey: Information Science Reference, 2009.

OLIVEIRA JR, H. A.; *et al.* **Inteligência Computacional**. São Paulo: Thomson Learning, 2007.

OLIVEIRA, H.; CORREIA, P. L. Automatic Road Crack Detection and Characterization. **IEEE Transactions on Intelligent Transportation Systems**, v. 14, n. 1, p. 155–168, mar. 2013.

OLIVEIRA, R. F. de. **Inteligência artificial**. Londrina: Editora e Distribuidora Educacional S.A., 2018.

OMOEBAMIJE, O. *et al.* An improved deep learning convolutional neural network for crack detection based on UAV images. **Innovative Infrastructure Solutions**, v. 8, n. 9, ago. 2023.

OPEN GEOSPATIAL CONSORTIUM (OGC). *OGC Indexed 3D Scene Layer (I3S) and Scene Layer Package (.slpk) Format Community Standard Version 1.2*. Arlington, TX, USA: 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Objetivos de Desenvolvimento Sustentável. Brasília, DF: ONU Brasil, 2015. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 10 jul. 2025.

PETRONIO, M. C. DA S.; RODGHER, S. F.; FLORIAN, F. Comparação da viabilidade técnica e econômica entre pavimentações rodoviárias rígidas e flexíveis. **RECIMA21 - Revista Científica Multidisciplinar - ISSN 2675-6218**, v. 5, n. 11, p. e5115877, 7 nov. 2024.

PONTES JÚNIOR, J. A. de M. **Análise de custos, características e desempenho dos sistemas de pavimentação urbana para ciclovias e ciclofaixas**. Dissertação (Mestrado) – Programa de Mestrado Profissional Em Ciências Da Cidade, Universidade de Fortaleza, Fortaleza, 2019.

QGIS. **QGIS Geographic Information System**. QGIS Association, 2025. Disponível em: <http://www.qgis.org>. Acesso em: 12 dez. 2025.

QIU, Q.; LAU, D. Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to unmanned aerial vehicle (UAV) images. **Automation in Construction**, v. 147, p. 104745, mar. 2023.

REDMON, Joseph *et al.* You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2016. p. 779-788.

SANDU, G.; VARGANOVA, O.; SAMII, B. Managing physical assets: a systematic review and a sustainable perspective. **International Journal of Production Research**, p. 1–23, 3 out. 2022.

SANTOS, Hugo Seiji Fujino dos; DUSI, Renzo de Freitas. Relatório PIBIC. Brasília: Universidade de Brasília, Departamento de Engenharia Civil e Ambiental, 2022.

SANTOS, P. A. S.; CARVALHO, M. T. M. Multi-class segmentation of structural damage and pathological manifestations using YOLOv8 and Segment Anything Model. **Automation in Construction**, v. 172, p. 106037, 2025.

SECRETARIA DE TRANSPORTE E MOBILIDADE - SEMOB. **Ciclomobilidade**. Semob, Brasília, dez. de 2024. Disponível em <<http://semob.df.gov.br/bicicletas-compartilhadas/>>. Acesso em 02 de Janeiro de 2025.

SHAFIEE, M. *et al.* Unmanned Aerial Drones for Inspection of Offshore Wind Turbines: A Mission-Critical Failure Analysis. **Robotics**, v. 10, n. 1, p. 26, fev. 2021.

SIERRA, C. *et al.* Development of a Cognitive Digital Twin for Pavement Infrastructure Health Monitoring. **Infrastructures**, v. 7, n. 9, p. 113, ago. 2022.

SILVA JÚNIOR, A. C. **Proposta de metodologia para avaliação de danos de pavimentos rígidos de ciclovias**. Dissertação (Mestrado) – Programa de Pós-Graduação em Estruturas e Construção Civil, Universidade de Brasília, Brasília, 2018.

SILVA, L. A. *et al.* An Architectural Multi-Agent System for a Pavement Monitoring System with Pothole Recognition in UAV Images. **Sensors**, v. 20, n. 21, p. 6205, 30 out. 2020.

SILVA, L. A. *et al.* Automated Road Damage Detection Using UAV Images and Deep Learning Techniques. **IEEE Access**, v. 11, p. 62918–62931, 2023.

SILVA, W. *et al.* Assessment of Interlocking Concrete Block Pavement with By-Products and Comparison with an Asphalt Pavement: A Review. **Applied Sciences**, v. 13, n. 10, p. 5846, 2023.

SINGH, D. K.; HOSKERE, V. Post Disaster Damage Assessment Using Ultra-High-Resolution Aerial Imagery with Semi-Supervised Transformers. **Sensors**, v. 23, n. 19, p. 8235, 2023.

SONG, F.; LIU, B.; YUAN, G. Pixel-Level Crack Identification for Bridge Concrete Structures Using Unmanned Aerial Vehicle Photography and Deep Learning. **Structural Control and Health Monitoring**, v. 2024, n. 1, jan. 2024.

SU, P. *et al.* MOD-YOLO: Rethinking the YOLO architecture at the level of feature information and applying it to crack detection. **Expert Systems with Applications**, v. 237, p. 121346–121346, 2024.

TABIAN, I.; FU, H.; KHODAEI, Z. S. A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures. **Sensors**, v. 19, n. 22, p. 4933, nov. 2019.

TAN, Y. *et al.* Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. **Automation in Construction**, v. 139, p. 104284, jul. 2022.

TANG, R. *et al.* A Knowledge-Guided Fusion Visualisation Method of Digital Twin Scenes for Mountain Highways. **ISPRS International Journal of Geo-Information**, v. 12, n. 10, p. 424–424, out. 2023.

TSAIMOU, C. N. *et al.* Mapping cracks on port concrete pavements by analyzing structural health monitoring metadata with computer vision-based techniques. **Complex Engineering Systems**, v. 4, n. 4, nov. 2024.

ULTRALYTICS. **YOLOv8 vs. YOLO11**: a comparative analysis. 2024. Disponível em: <https://docs.ultralytics.com/compare/yolov8-vs-yolo11/>. Acesso em: 15 dez. 2026.

VAN GERVEN, M.; BOHTE, S. **Artificial Neural Networks as Models of Neural Information Processing**. Lausanne: Frontiers Media, 2018.

WANG, J. *et al.* Building damage inspection method using UAV-based data acquisition and deep learning-based crack detection. **Journal of Civil Structural Health Monitoring**, jul. 2024.

WANG, W. *et al.* Fine-Grained Detection of Pavement Distress Based on Integrated Data Using Digital Twin. **Applied sciences**, v. 13, n. 7, p. 4549–4549, abr. 2023.

WANG, Y.; YE, T. Applications of Artificial Intelligence Enhanced Drones in Distress Pavement, Pothole Detection, and Healthcare Monitoring with Service Delivery. **Journal of Engineering**, v. 2022, out. 2022.

WEST, J. *et al.* Asset maintenance in Australian commercial buildings. **Frontiers in Built Environment**, v. 10, 14 jun. 2024.

YOON, J. *et al.* CNN- and UAV-Based Automatic 3D Modeling Methods for Building Exterior Inspection. **Buildings**, v. 14, n. 1, p. 5–5, 19 dez. 2023.

YUAN, H. *et al.* Research on a Road Crack Detection Method Based on YOLO11-MBC. **Sensors**, v. 24, n. 24, p. 7435, 2025.

ZHANG, C. *et al.* Cracks segmentation of engineering structures in complex backgrounds using a concatenation of Transformer and CNN models driven by scene understanding information. **Structures**, v. 65, p. 106685, jul. 2024.

ZHAO, G. *et al.* Road crack detection based on improved RT-DETR. **Signal Image and Video Processing**, v. 19, n. 7, mai. 2025.

ZHAO, Yian, *et al.* DETRs beat YOLOs on real-time object detection. **arXiv:2304.08069**, 2023.

ZHU, Xizhou *et al.* Deformable detr: Deformable transformers for end-to-end object detection. **arXiv:2010.04159**, 2020.

ZUO, C. *et al.* Pavement-DETR: A High-Precision Real-Time Detection Transformer for Pavement Defect Detection. **Sensors**, v. 25, n. 8, p. 2426, 2025.