

Deteção de degradações de pavimentos rodoviários com recurso à técnica YOLO (You Only Look Once)

Maria Carolina Cortez de Oliveira

Dissertação para obtenção do Grau de Mestre em
Engenharia civil
(mestrado integrado)

Orientadora: Prof. Doutor Bertha Maria Batista dos Santos
Co-orientador: Prof. Doutor Pedro Gabriel de Faria Lapa Barbosa de Almeida

Outubro de 2025

Folha em branco

Declaração de Integridade

Eu, Maria Carolina Cortez de Oliveira, que abaixo assino, estudante com o número de inscrição 43131 do Mestrado Integrado em Engenharia Civil da Faculdade Universidade da Beira Interior, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 13/10/2025

Maria Oliveira

(assinatura conforme Cartão de Cidadão ou preferencialmente assinatura digital no documento original se naquele mesmo formato)

Folha em branco

Agradecimentos

Agradeço primeiramente a Deus, pela oportunidade de estar realizando este sonho e por me guiar em cada passo desta jornada.

Aos meus pais, cujo esforço e dedicação tornaram possível que eu completasse minha graduação fora do meu país de origem, realizando um sonho que sempre foi meu.

Agradeço a cada professor que tive ao longo do caminho, que não apenas me ensinaram engenharia, mas também valiosas lições de vida, e cujo conhecimento levarei sempre comigo.

Em especial, agradeço à professora Bertha Santos, minha orientadora, que esteve sempre presente, guiando e apoiando nosso trabalho com paciência e dedicação.

Ao coorientador Pedro Almeida, agradeço a paciência e pelos conselhos que contribuíram para que este trabalho alcançasse seu melhor resultado.

Agradeço também ao meu companheiro, Luan Menzen, por estar sempre ao meu lado nesta jornada, oferecendo apoio e incentivo.

Por fim, agradeço a cada pessoa que cruzou meu caminho na Universidade da Beira Interior, desde o corpo docente, colegas e amigos feitos durante esses anos, até aqueles que serviram muitas xícaras de café no bar, tornando os dias mais leves e felizes.

A todos, meu sincero obrigado por fazerem parte desta conquista.

Folha em branco

Resumo

A degradação da superfície dos pavimentos rodoviários resulta da ação conjunta de fatores como a intensidade do tráfego, as condições climáticas e eventuais deficiências construtivas. Com o tempo, diferentes tipos de anomalias manifestam-se à superfície dos pavimentos, comprometendo o conforto de circulação e a segurança dos utentes, além de poderem originar danos estruturais significativos.

O objetivo central deste trabalho consistiu em desenvolver e avaliar uma metodologia baseada em deep learning para a deteção automática de degradações superficiais em pavimentos rodoviários. Para tal, foi utilizada a arquitetura YOLO (You Only Look Once), testada em diferentes versões e ambientes de execução, nomeadamente no Roboflow e no YOLOv8 local, com e sem aumento da base de dados. Foram analisadas imagens reais de um troço rodoviário em Portugal, permitindo avaliar o desempenho dos modelos em diferentes cenários de treino.

Os resultados demonstraram que, mesmo com uma base de dados limitada relativa a um troço reduzido, foi possível alcançar níveis consistentes de precisão, recall e mAP, comprovando a viabilidade da aplicação destas técnicas ao contexto rodoviário. Verificou-se que o Roboflow apresentou maior facilidade de utilização e tempos de execução mais reduzidos, enquanto o YOLO local se destacou pela flexibilidade e potencial de personalização.

Conclui-se que a integração desta abordagem em sistemas de gestão de pavimentos pode reduzir significativamente o tempo necessário para a avaliação da sua condição, possibilitando estratégias de conservação mais preventivas e eficazes. Além disso, abre caminho para futuros desenvolvimentos que explorem bases de dados mais extensas, técnicas avançadas de aumento de dados e a adaptação a diferentes condições ambientais, contribuindo para a prolongação da vida útil das infraestruturas viárias.

Palavras-chave

Degradações superficiais de pavimentos rodoviários; Deep Learning; YOLO; Deteção automática.

Folha em branco

Abstract

The deterioration of road pavement surfaces results from the combined action of factors such as traffic intensity, weather conditions, and possible construction deficiencies. Over time, different types of distresses appear on the pavement surface, compromising driving comfort and user safety, in addition to causing significant structural damage.

The main objective of this work was to develop and evaluate a deep learning-based methodology for the automatic detection of surface distresses in road pavements. To this end, the YOLO (You Only Look Once) architecture was used, tested in different versions and execution environments, namely in Roboflow and local YOLOv8, with and without database augmentation. Real images of a road section in Portugal were analysed, allowing the performance of the models to be evaluated in different training scenarios.

The results showed that, even with a limited database relating to a small section, it was possible to achieve consistent levels of accuracy, recall and mAP, proving the feasibility of applying these techniques to the road context. Roboflow proved to be easier to use and had shorter execution times, while local YOLO stood out for its flexibility and customisation potential.

It can be concluded that integrating this approach into pavement management systems can significantly reduce the time needed to assess pavement conditions, enabling more preventive and effective maintenance strategies. Furthermore, it paves the way for future developments that explore more extensive databases, advanced data augmentation techniques, and adaptation to different environmental conditions, contributing to the extension of the service life of road infrastructure.

Keywords

Road pavement distresses; Deep Learning; YOLO; Automatic detection.

Folha em branco

Índice

1.0 Introdução.....	1
1.1 Objetivos do estudo	2
1.2 Estrutura do documento	3
1.3 Revisão da Literatura.....	4
1.3.1 Processo de seleção dos documentos.....	4
1.3.2 Análise de tendências de pesquisa.....	6
2.0 Métodos e Ferramentas para a Detecção Automática de Degradações em Pavimentos	16
2.1 Modelo de <i>deep learning</i>	16
2.1.1 O Algoritmo YOLO (You Only Look Once)	17
2.1.2 A plataforma Roboflow	18
2.2 Métricas e Parâmetros de Avaliação de Desempenho	19
3.0 Metodologia.....	23
3.1 Definição da base de dados	24
3.2 Pré processamento das imagens.....	25
3.3 Anotação e preparação na plataforma Roboflow	27
3.4 Preparação da Base de Dados para o YOLO	28
3.5 Treino do modelo YOLO	28
3.6 Avaliação, Testes Finais e Análise dos resultados.....	29
3.7 Síntese e Conclusão Metodológica.....	30
4.0 Caso de estudo	31
4.1 Localização e aquisição das imagens para a base de dados	32
4.2 Anotações das imagens	34
4.3 Testes preliminares no Roboflow	36
4.3.1 Roboflow – Versão 1	38
4.3.2 Roboflow – Versão 2	42
4.3.3 Roboflow – Versão 3	47
4.3.4 Roboflow – Versão 4	51
4.4 Testes no modelo YOLO	56
4.4.1 YOLO – Versão 1	57
4.4.2 Versão 2 YOLO.....	63
4.5 Comparação dos 6 modelos.....	71
4.5.1 Testes realizados na plataforma Roboflow.....	71
4.5.2 Testes realizados com o modelo YOLO	73
4.5.3 Comparação entre testes YOLO (Roboflow YOLOv11 versus YOLOv8)	75
5.0 Conclusões	78
5.1 Síntese dos principais resultados e tendências	79
5.2 Limitações do caso de estudo	80
5.3 Indicações para trabalhos futuros	81

Folha em branco

Lista de Figuras

Figura 1 - Fluxograma da metodologia de estudo.	24
Figura 2 - Municípios do Fundão e Covilhã.	32
Figura 3 - Localização do troço rodoviário em estudo	33
Figura 4 - Exemplo de imagem usada no estudo	34
Figura 5 - Métricas para versão 1 Roboflow	38
Figura 6 – Limiar de confiança para versão 1 Roboflow	39
Figura 7 - Curvas de perda da versão 1 Roboflow	41
Figura 8 - Exemplo de deteção das degradações presentes numa imagem com recurso à versão 1 – Roboflow – Roboflow 3.0	42
Figura 9 - Métricas para versão 2 Roboflow	43
Figura 10 – Limiar de confiança para versão 2 Roboflow	44
Figura 11 - Curvas de perdas da versão 2 Roboflow	45
Figura 12 - Exemplo de deteção das degradações presentes numa imagem com recurso à versão 2 – Roboflow – YOLOv11	46
Figura 13 - Métricas para versão 3 Roboflow	47
Figura 14 – Limiar de confiança da versão 3 Roboflow.....	48
Figura 15 - Curvas de perda da versão 3 Roboflow	49
Figura 16 - Exemplo de deteção das degradações presentes numa imagem com recurso à versão 3 – Roboflow – Roboflow 3.0 com base de dados ampliada	51
Figura 17 - Métricas para versão 4 Roboflow	52
Figura 18 – Limiar de confiança da versão 4 Roboflow.....	52
Figura 19 - Curvas de perdas da versão 4 Roboflow	54
Figura 20 - Exemplo de deteção das degradações presentes numa imagem com recurso à versão 4 – Roboflow – YOLOv11 com base de dados ampliada	56
Figura 21 - Resultados da versão 1 YOLO.....	57
Figura 22 - Curva de precisão-confiança para o teste versão 1 YOLO	58
Figura 23 - Curva de Revocação-Confiança para o teste versão 1 YOLO	59
Figura 24 - Curva de F1-Confiança para o teste versão 1 YOLO	60
Figura 25 - Matriz de confusão da versão 1 YOLO	61
Figura 26 - Métricas da versão 1 YOLO	62
Figura 27 - Exemplo de deteção com o modelo versão 1 do YOLO.....	63
Figura 28 - Resultados da versão 2 YOLO.....	64
Figura 29 - Curva de Precisão-Confiança para a versão 2 YOLO.....	65
Figura 30 - Curva de Revocação-Confiança para a versão 2 YOLO.....	66
Figura 31 - Curva de F1-Confiança para a versão 2 YOLO	67
Figura 32 - Matriz de confusão da versão 2 YOLO	68
Figura 33 - Métricas da versão 2 YOLO	69
Figura 34 - Exemplo de deteção com o modelo versão 2 do YOLO.....	70

Folha em branco

Lista de Tabelas

Tabela 1 - Processo de seleção dos documentos.....	5
Tabela 2 - Informações básicas relacionadas aos documentos selecionados.....	8
Tabela 3 - Caracterização dos documentos de acordo com a aplicação do YOLO na deteção de danos no pavimento.	11
Tabela 4 -Métricas e parâmetros de avaliação utilizados no estudo.....	22
Tabela 5 - Classificação das degradações.....	35
Tabela 6 - Matriz de confusão da versão 1 Roboflow	40
Tabela 7 - Matriz de confusão da versão 2 Roboflow	44
Tabela 8 – Matriz de confusão da versão 3 Roboflow	48
Tabela 9 - Matriz de confusão da versão 4 Roboflow	53
Tabela 11 – Síntese dos resultados obtidos para os 6 modelos treinados.....	71
Tabela 12 - Comparação dos resultados obtidos na plataforma Roboflow	72
Tabela 13 - Comparação dos resultados obtidos com o modelo YOLO	74
Tabela 14 - Comparação dos resultados entre os modelos obtidos no Roboflow- YOLOv11 e no modelo YOLO-YOLOv8	76

Folha em branco

Lista de Acrónimos

AI	Artificial intelligence (Inteligência artificial)
CLS	Classification Loss (Perdas de classificação)
CNN	Convolutional Neural Network (Rede neural convolucional)
CPU	Central Processing Unit (Unidade central de processamento)
CVPR	Computer Vision and Pattern Recognition (Visão por computador e reconhecimento de padrões)
DFL	Distribution Focal Loss (Perda focal de distribuição)
DL	Deep Learning (Aprendizado profundo)
FN	False Negative (Falso negativo)
FP	False Positive (Falso positivo)
GPU	Graphics Processing Unit (Unidade de processamento gráfico)
IoU	Intersection Over Union (Interseção sobre a união)
mAP	Mean Average Precision (Precisão média média)
ML	Machine learning (Aprendizados de máquina)
PP	Pontos Percentuais
TN	True Negative (Verdadeiro negativo)
TP	True Positive (Verdadeiro positivo)
VANT	Veículo Aéreo Não Tripulado
XL	Extra Large (Extra grande)
YOLO	You Only Look Once
BD	Base de dados

Folha em branco

Capítulo 1

1.0 Introdução

A relevância de infraestruturas rodoviárias de qualidade tem vindo a ser progressivamente reconhecida, sobretudo com a expansão das redes de transporte e comunicação. Neste contexto, o estado de conservação dos pavimentos assume um papel determinante para o funcionamento eficiente e seguro das estradas. Ao longo do tempo, os pavimentos sofrem desgastes devido à ação combinada do tráfego, das condições climáticas e, por vezes, de deficiências construtivas. Estes fatores conduzem ao aparecimento de diferentes formas de degradação superficial, que podem comprometer o conforto e a segurança dos utilizadores, e colocar em risco a integridade estrutural da infraestrutura.

A manutenção adequada dos pavimentos constitui, assim, uma decisão estratégica e economicamente vantajosa. Intervenções regulares permitem identificar e corrigir problemas em fase inicial, evitando obras de grande envergadura, reduzindo custos de reabilitação e minimizando os impactos negativos para os utilizadores. Neste sentido, torna-se essencial investir em métodos de monitorização mais ágeis e fiáveis, que permitam avaliar a condição dos pavimentos de forma rápida e precisa.

Nos últimos anos, as técnicas de visão computacional e *deep learning* têm demonstrado elevado potencial na deteção automática de degradações em pavimentos rodoviários. Entre os algoritmos de referência destaca-se o YOLO (You Only Look Once), desenvolvido por Joseph Redmon e colaboradores em 2016, no âmbito da Conference on Computer Vision and Pattern Recognition (CVPR) [REDMON et al, 2016]. Este modelo introduziu uma abordagem inovadora para a deteção de objetos em tempo real, combinando rapidez e elevada precisão. Desde então, o YOLO tem vindo a ser continuamente aprimorado, originando diversas versões que refletem os avanços tecnológicos na área do processamento de dados.

A aplicação desta técnica à gestão de pavimentos requer a utilização de bases de dados de imagens representativas da realidade. Estas imagens podem ser obtidas através de câmeras instaladas em veículos terrestres ou aéreos que percorrem os troços rodoviários, ou ainda recorrendo a plataformas como o Google Street View. Posteriormente, as imagens são anotadas e introduzidas no YOLO, de modo a identificar automaticamente o tipo e o nível de gravidade dos danos presentes na superfície do pavimento. Esta automatização permite implementar medidas de intervenção de forma mais ágil e direcionada, evitando a progressão para danos estruturais graves que, numa fase posterior, seriam mais complexos e onerosos de reparar.

1.1 Objetivos do estudo

O presente trabalho tem como objetivo central desenvolver e validar uma metodologia para a deteção de degradações em pavimentos rodoviários recorrendo a técnicas de deep learning, com particular destaque para o uso do algoritmo YOLO (You Only Look Once). Pretende-se, desta forma, contribuir para a modernização e automatização dos processos de monitorização de infraestruturas rodoviárias, tradicionalmente dependentes de inspeções visuais realizadas a pé, frequentemente demoradas e propensas à subjetividade inerente à avaliação humana na identificação das degradações.

De modo a concretizar este objetivo geral, foram estabelecidos diversos objetivos específicos que estruturam a investigação. O primeiro consiste em realizar uma revisão da literatura, identificando os principais avanços no domínio da deteção automática de degradações em pavimentos, bem como as limitações ainda existentes. No segundo pretende-se construir uma base de dados representativa, através da recolha, seleção e anotação de imagens da superfície de pavimentos. Para a anotação das imagens, é usada a plataforma Roboflow, que facilita a gestão, organização e anotação das imagens, assegurando um treino consistente e adequado do modelo. Esta plataforma também disponibiliza técnicas para deteção automática de objetos.

O terceiro objetivo específico prende-se com a implementação e parametrização da técnica YOLO, avaliando o seu desempenho na identificação de diferentes tipos de degradações, tais como fissuras, e covas. Este processo inclui a análise comparativa de métricas de precisão, revocação e tempo de processamento, permitindo aferir a viabilidade da aplicação em cenários reais de inspeção rodoviária. Paralelamente, busca-se verificar a aplicabilidade da metodologia em diferentes condições de luminosidade, textura e desgaste, de forma a testar a robustez do modelo.

Neste trabalho é analisado um conjunto de imagens de uma secção de um pavimento rodoviário localizado no Fundão, em Portugal, devidamente anotadas e processadas no modelo YOLO juntamente com modelos realizados no Roboflow, a fim de compará-los, com o objetivo de avaliar a eficácia desta técnica na deteção automática de degradações. Pretende-se, assim, contribuir para a integração de soluções baseadas em deep learning nos sistemas de gestão de pavimentos, potenciando uma monitorização mais eficiente e sustentada, capaz de prolongar a vida útil das infraestruturas e de reforçar a segurança rodoviária.

Por fim, pretende-se discutir os resultados obtidos à luz das necessidades atuais de gestão e manutenção de infraestruturas rodoviárias, destacando o contributo da visão por computador para a tomada de decisão mais eficiente, baseada em dados objetivos e fiáveis. Assim, o trabalho ambiciona não apenas desenvolver um estudo na área da deteção automática de degradações de pavimentos, mas também a geração de conhecimento aplicável à prática, com potencial para apoiar entidades responsáveis na preservação e segurança da rede viária.

1.2 Estrutura do documento

O presente trabalho encontra-se organizado em cinco capítulos, estruturados de forma sequencial e lógica, iniciando pelo enquadramento teórico ao qual se segue a aplicação prática e conclusões finais. Esta organização visa não apenas a clareza expositiva, mas também a articulação entre as diferentes fases da investigação, assegurando a coerência global do estudo.

No Capítulo 1 – Introdução, é realizado o enquadramento geral do tema, abordando-se a importância da monitorização e manutenção dos pavimentos rodoviários e as limitações dos métodos convencionais de inspeção. Integra-se neste capítulo a revisão bibliográfica, onde se analisam estudos prévios e abordagens existentes no domínio da deteção automática de degradações em pavimentos. São também apresentados a motivação do estudo, os objetivos do trabalho e a sua relevância científica e prática.

O Capítulo 2 – Métodos e Ferramentas para a Deteção Automática de Pavimentos, centram-se na descrição das bases técnicas que sustentam a investigação. São abordados os princípios fundamentais da visão por computador, das redes neuronais convolucionais (CNNs) e dos modelos de deteção em tempo real, com destaque para a família YOLO (You Only Look Once). Este capítulo contempla igualmente a descrição de ferramentas de suporte, como a plataforma Roboflow, que se revelou essencial no processo de gestão, anotação e preparação dos dados, e ainda as métricas de avaliação utilizadas na análise de desempenho dos modelos.

O Capítulo 3 – Metodologia e Estrutura de Estudo, descreve em detalhe a estratégia adotada para alcançar os objetivos definidos. É apresentada a metodologia de investigação, explicitando cada uma das suas etapas e respetiva articulação. Este capítulo apresenta os critérios de seleção e anotação de imagens de pavimentos rodoviários, a utilização do Roboflow para normalização e expansão do conjunto de dados, bem como os procedimentos de treino, validação e teste do modelo YOLO. São igualmente apresentados os parâmetros de configuração do algoritmo, as métricas de avaliação escolhidas — como precisão, revocação, mean Average Precision (mAP) e tempo de processamento — e os métodos utilizados para aferir a robustez da abordagem em diferentes cenários e condições.

O Capítulo 4 – Caso de Estudo apresenta de forma detalhada a aplicação prática da investigação, desenvolvida integralmente com base nos dados recolhidos no âmbito da investigação em pavimentos desenvolvida no Departamento de Engenharia e Arquitetura da Universidade da Beira Interior. Neste capítulo é descrito o processo de aquisição das imagens no local de estudo, bem como a sua preparação e anotação. São expostos os procedimentos adotados para a execução do modelo, tanto na plataforma Roboflow como no ambiente de treino com o YOLO, descrevendo-se os parâmetros utilizados em cada configuração. De seguida, são apresentados os resultados quantitativos obtidos, incluindo métricas de desempenho como precisão, revocação, mAP e tempo de deteção. Os resultados são comparados de forma objetiva, de modo a evidenciar as

diferenças entre as execuções realizadas e a destacar os pontos fortes e limitações de cada abordagem experimental.

O Capítulo 5 – Conclusões é dedicado à um breve resumo dos principais resultados, as limitações encontradas no caso de estudo e a sugestões de linha de investigações futuras.

1.3 Revisão da Literatura

A gestão eficiente dos pavimentos rodoviários depende, em grande medida, da capacidade de monitorizar de forma sistemática e rigorosa o seu estado de conservação. Tradicionalmente, este processo tem assentado em inspeções visuais e ensaios complementares, os quais, apesar de fornecerem informação relevante, apresentam limitações evidentes, como a necessidade de deslocação de equipas especializadas, o tempo de execução e a subjetividade inerente à avaliação humana.

O desenvolvimento das técnicas de visão computacional e de *deep learning* veio abrir novas perspectivas para a deteção automática de degradações superficiais em pavimentos, permitindo recolher e processar grandes volumes de dados com rapidez e elevada precisão. Entre os algoritmos mais utilizados, destaca-se o YOLO (You Only Look Once) pela sua capacidade de identificar objetos em tempo real e com elevada eficiência.

O objetivo desta revisão é, assim, identificar as principais tendências, vantagens e limitações do uso do YOLO na deteção automática de degradações de pavimentos. Ao sistematizar o estado da arte desta área emergente, o estudo estabelece uma base sólida para a avaliação do potencial de aplicação desta tecnologia em sistemas de gestão de pavimentos rodoviários, contribuindo para processos de monitorização mais rápidos, objetivos e sustentáveis.

1.3.1 Processo de seleção dos documentos

A base de dados científica escolhida para pesquisar e recolher os dados brutos utilizados na análise é a Scopus, uma base de dados da Elsevier criada em 2004 que contém resumos, artigos e citações com curadoria especializada. Os dados foram recolhidos entre março e julho de 2024.

Os documentos foram selecionados com base em vários critérios, sendo o mais importante o acesso aberto, ou seja, publicações que estão disponíveis gratuitamente na Internet, e as palavras chaves “YOLO, pavement deterioration, pavement inspection, learning, deep learning, and automatic detection”, considerando o título, o resumo e o conteúdo das palavras-chave do artigo. A inclusão de artigos de acesso aberto na revisão garante que o conteúdo publicado está disponível e possa ser visto pelo maior número possível de pessoas.

Outros critérios incluem a seleção de estudos escritos em língua inglesa, em fase final de publicação, pertencentes à área temática da engenharia e tipo de documento artigo ou comunicação de conferência. O processo de seleção teve como objetivo garantir a qualidade e

relevância dos documentos selecionados, refletindo um firme compromisso com elevados padrões de excelência científica.

A Tabela 1 apresenta todas as etapas do processo de seleção dos artigos utilizados nesta revisão da literatura.

Tabela 1 - Processo de seleção dos documentos

Fase do processo	Crítérios de seleção
1	Pesquisa na base de dados Scopus utilizando a expressão “YOLO AND (pavement distress OR pavement inspection) AND (deep learning OR automatic detection)”.
2	Seleção de documentos com acesso aberto, em fase final de publicação, em língua inglesa, área temática “Engenharia” e tipo de documento artigo ou comunicação científica.
3	Confirmação da presença das palavras-chave “YOLO and Pavement distress.
4	Documentos em que o algoritmo YOLO foi efetivamente utilizado para detectar danos em pavimentos rodoviários.

Na fase 1, a informação foi seleccionada através de uma pesquisa avançada utilizando um conjunto de termos relevantes relativos à utilização do algoritmo YOLO para a detecção de degradações de pavimentos rodoviários. Os seguintes termos e operadores lógicos foram considerados como essenciais para a pesquisa documental: “YOLO AND (pavement distress OR pavement inspection) AND (deep learning OR automatic detection)”.

Em seguida, na fase 2, os resultados foram filtrados para incluir apenas publicações de acesso aberto, garantindo que todos os artigos selecionados estivessem disponíveis para consulta. Além disso, a pesquisa foi restrita a documentos escritos em inglês, em fase final de publicação e a artigos e comunicações de conferências.

Em seguida, na fase 3, foi verificada a presença dos termos “YOLO” e “pavement distress” nos documentos.

Finalmente, na fase 4, os artigos foram refinados para incluir apenas aqueles que descreviam pesquisas sobre o uso do algoritmo YOLO para detectar danos no pavimento, incluindo aqueles que usavam modelos híbridos (uso do YOLO junto com outros algoritmos).

Como resultado do processo de seleção, 16 documentos foram considerados (Tabela 2).

1.3.2 Análise de tendências de pesquisa

Esta seção tem como objetivo identificar as principais tendências de investigação relacionadas com a utilização do algoritmo YOLO na detecção de danos em pavimentos, com base na seleção de documentos obtida.

A Tabela 2 apresenta as informações básicas contidas nos documentos selecionados para este estudo.

Os estudos estão ordenados por ano, do mais antigo ao mais recente, e são apresentados os autores, revista científica ou conferência, país e palavras-chave. A maioria dos estudos citados é recente, destacando o avanço novo e rápido das técnicas de inteligência artificial e sua crescente aplicação na gestão de pavimentos. O campo está em expansão, com expectativa de crescimento e inovação contínuos.

A área temática escolhida para a pesquisa de documentos foi a engenharia, mas há muitas outras, como a engenharia da computação, matemática, física e ciência dos materiais, que aparecem simultaneamente nas informações das revistas. A maioria dos documentos citados nesta pesquisa foi publicada em revistas científicas internacionais de alto impacto, enquanto apenas dois dos artigos analisados foram publicados em atas de conferências.

Como a investigação se centrou na aplicação do algoritmo YOLO, a palavra-chave “YOLO” aparece em 8 dos 16 documentos analisados. A palavra “deep learning” ou similar também aparece em 8 documentos como palavra-chave, relacionados com a detecção de danos (aparece 7 vezes) em pavimentos flexíveis, com especial ênfase em fendas (aparece 4 vezes) e covas (aparece 2 vezes).

A Tabela 3 fornece informações relevantes sobre as características dos estudos analisados em cada documento.

Verifica-se uma participação significativa dos países asiáticos, cujo domínio pode ser atribuído às políticas governamentais que incentivam o desenvolvimento tecnológico e priorizam o investimento em inteligência artificial. Este cenário reflete o compromisso dos países asiáticos em liderar a inovação neste campo. Além disso, a Europa e as Américas, especialmente a América do Norte, têm apresentado um crescimento significativo na investigação e desenvolvimento de tecnologias relacionadas com a inteligência artificial, consolidando-se como regiões importantes para a investigação e aplicação dessas inovações.

A coluna “Tipo de pavimento”, tal como indicado anteriormente na análise das palavras-chave, mostra que o tipo de pavimento estudado é principalmente o flexível (betuminoso), com apenas um estudo de caso desenvolvido sobre um pavimento rígido (betão). No que diz respeito ao ambiente rodoviário, a maioria dos estudos foram realizados em áreas urbanas. Apenas um estudo foi realizado em meio rural, refletindo a maior atenção dada os pavimentos urbanos, que

sofrem mais desgaste devido ao uso intensivo e a sistemas de gestão de pavimentos menos desenvolvidos.

No que diz respeito à versão do algoritmo YOLO utilizada, as publicações mais recentes utilizam as versões mais recentes do algoritmo, refletindo os desenvolvimentos tecnológicos que melhoram a detecção de danos no pavimento. A maioria dos documentos analisados também compara os resultados obtidos utilizando diferentes versões do algoritmo, destacando a superioridade das versões mais recentes.

Tabela 2 - Informações básicas relacionadas aos documentos selecionados. (Parte 1/3)

Ano	Título	Autores (Citação)	Jornais/Atas	Palavras-chave	Doc N.º
2019	Asphalt pavement pothole detection using deep learning method based on YOLO neural network	Ukhwah, E.; Yuniarno, E.; Suprpto, Y. (UKHWAH et al., 2019)	International seminar on intelligent technology and its application (ISITIA)	Pothole detection, YOLO, Computer vision, Detection, Distress detection	1
2020	<i>Deep machine learning approach to develop a new asphalt pavement condition index.</i>	Majidifard, H.; Adu-Gyamf, Y.; Buttlar, W. (MAJIDIFARD et al.,2020)	Construction and Building materials. ScienceDirect	Pavement monitoring, Pavement distresses detection, Deep learning, Google API, Machine learning, Pavement condition prediction, YOLO, Image processing	2
2020	Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses	Majidifard, H.; Adu-Gyamf, Y.; Buttlar, W. (MAJIDIFARD et al.,2020)	TRR Journal of the transportation research board	-	3
2020	Pavement distresses detection and classification based on YOLO network	Du, Y.; Pan, N.; Xu, Z.; Deng, F.; Shen, Y.; Kang, H. (DU et al., 2020)	International journal of pavement engineering	Pavement distress, Object detection, Image classification, YOLO network	4
2022	Detection of pothole for repair works of asphalt flexible pavement optimization using YOLO	Halim, M.; Ibrahim, A.; Osman, M.; Kader, M.; Termizi, M.; Abu, A. (HALIM et al., 2022)	AIP conference proceedings	-	5
2022	Computers Vision Based pothole detection under challenging conditions	Bucko, B.; Lieskovska, E.; Zábovska, K.; Zábovsky, M. (BUCKO et al., 2022)	Sensors	Pothole detection, Pavement distress, YOLO v3	6

Tabela 2 - Informações básicas relacionadas aos documentos selecionados. (Parte 2/3)

Ano	Título	Autores (Citações)	Jornais/Atas	Palavras-chave	Doc N.º
2022	Pothole and patch detection on asphalt pavement using deep convolutional neural network	Espínola, A.; Freitas, G.; Júnior, E. (ESPÍNOLA et al., 2022)	CILAMCE-PANACM-2021	Pavement defect, Convolutional Neural Network, YOLO	7
2023	Automated distress detection, classification and measurement for asphalt urban pavements using YOLO	Conti, P.; Lird, A.; Cid, H. (CONTI et al., 2023)	Engineering proceeding	Asphalt pavement, Urban pavements, Distress detection, Deep learning, Convolutional Neural Network	8
2023	<i>Automated pavement crack detection with deep learning methods: What are the main factors and how to improve the performance?</i>	Gong, H.; Tesic, J.; Tao, J.; Luo, X.; Wang, F. (GONG et al., 2023)	TRR Journal of the transportation research board	Data and data science, Deep learning, Pavement condition evaluation, Cracking, Detection	9
2023	Road deterioration detection: A machine learning based system for automated paved pavement crack identifications and analysis	Ganeshan, D.; Sharif, M.; Apegyei, A. (GANESHAN et al., 2023)	International conference on innovation and intelligence for information, computing and technologies (3ICT)	Object detection, YOLO, Data, Pavement distress identification, Automated	10
2023	A deep learning network model for road pavement damage detection from urban street view imagery	Ren, M.; Zhang, X.; Chen, X.; Zhou, B.; Feng, Z. (REN et al., 2023)	Internacional journal of Applied Earth observation and geoinformation	Pavement damage, YOLO v5, Street-view images, Deep learning	11

Tabela 2 - Informações básicas relacionadas aos documentos selecionados. (Parte 3/3)

Ano	Título	Autores (Citações)	Jornais/Atas	Palavras-chave	Doc N.º
2024	Crack YOLO: Rural pavement distress detection model with complex scenarios	Sun, Y.; Song, W.; Zhang, J.; Teng, Q. (SUN et al., 2024)	Electronics	Crack detection, Deep learning, Rural roads, Hybrid attention mechanism	12
2024	Automatic detection of urban pavement distress and dropped objects with a compressive dataset collected via smartphone	Xu, L.; Fu, K.; Ma, T.; Tang, F.; Fan; J. (XU et al., 2024)	Buildings	Pavement distress, Dropped objects, Semantic segmentation	13
2024	Automated pavement cracks detection and classification using deep learning	Naffa, S.; Ashour, K.; Mohamed, R.; Essam, H., Emad, D.; Elhenawy, M., Ashqar, H., Hassan, A.; Alhadidi, T. (NAFFA et al., 2024)	International Conference on Computing and Machine Intelligence (ICMI)	Deep learning, Management, Signs, Detection, Pavement crack detection	14
2024	Real time pavement crack detection based on artificial intelligence	Ya'acob, N.; Zuraimi, M.; Rahman, A.; Yusof, A.; Ali, D. (YA'ACOB et al., 2024)	Journal of advanced research in applied sciences and engineering technology	Crack, Pavement, Road, Damages, Machine Learning	15
2024	Research and optimization of YOLO based method for automatic pavement detect detection	Yao, H.; Fan, Y.; Wei, X.; Liu, Y.; Cao, D.; You, Z. (YAO et al., 2024)	Electronic research archive (AIMS)	Pavement engineering, Pavement distress, Object detection, Optimization strategy, YOLO algorithm	16

Tabela 3 - Caracterização dos documentos de acordo com a aplicação do YOLO na detecção de danos no pavimento. (Parte 1/3)

Doc N.	País	Tipo de pavimento	Ambiente	Versão YOLO	Tipo de degradações detetadas	Fonte ou método utilizado para capturar imagens	Nº de imagens	Posicionamento	Tamanho da imagem (Pixel)
1	Indonésia	Flexível	Urbano	YOLO-V3	Covas	Pavement view camera	448	Não especificado	1280x960, 1624x1234, 2048x1536 e 512x384
2	Estados Unidos	Flexível	Urbano	YOLO-V2	Fendas por reflexão, transversais, longitudinais e em bloco, pele de crocodilo, fendas por reflexão e covas.	Google Street View	7237	Coordenadas GPS das imagens do Google Street View	Não especificado
3	Estados Unidos	Flexível	Urbano	YOLO-V2	Fendas transversais, longitudinais, em bloco e transversal selada, pele de crocodilo e covas.	Google Street View	7237	Coordenadas GPS das imagens do Google Street View	Não especificado
4	China	Flexível	Urbano	YOLO-V3	Fendas transversais, e longitudinais, pele de crocodilo, covas, reparações e tampas de saneamento.	Não especificado	45788	Não especificado	1920x1080
5	Malásia	Flexível	Urbano	YOLO-V4	Covas	Iphone 7	130	Não especificado	416x416
6	Eslováquia	Flexível	Urbano	YOLO-V3	Covas	Não especificado	2099	Não especificado	1920x1080
7	Brasil	Flexível	Urbano	YOLOV4, YOLOV5 e YOLOV7	Covas e reparações	Garmin VIRB ultra 30 action camera	360	Coordenadas GPS das imagens do Google Street View	224x224, 320x320 e 416x416.

Tabela 3 - Caracterização do documento de acordo com a aplicação do YOLO na detecção de danos no pavimento. (Parte 2/3)

Doc N.	País	Tipo de pavimento	Ambiente	Versão YOLO	Tipo de degradações	Fonte ou método utilizado para capturar imagens	Nº de imagens	Posicionamento	Tamanho da imagem (Pixel)
8	Chile	Flexível	Urbano	YOLOV5 e YOLOV7	Fendas transversais, longitudinais e de fadiga, covas e remendos.	GoPro hero 8 camera	50000	Não especificado	400x400
9	Estados Unidos	Flexível	Urbano	YOLOV5	Fendas transversais e longitudinais, pele de crocodilo e covas.	Não especificado	5789	Coordenadas GPS das imagens do Google Street View	64x64
10	Inglaterra	Flexível	Urbano	YOLOV5 e YOLOV8	Fendas transversais e longitudinais, reparações, pele de crocodilo e tampas de saneamento.	Não especificado	201	Não especificado	Não especificado
11	China	Flexível	Urbano	YOLOV5	Fendas transversais e longitudinais, covas, reparações, pele de crocodilo e tampas de saneamento.	Google Street View	156304	Coordenadas GPS das imagens do Google Street View	1024x1024
12	China	Flexível	Rural	YOLOV5 e YOLOV7	Fissura transversal, fissura longitudinal e fissura reticular.	Não especificado	9801	Não especificado	1200x800
13	China	Flexível	Urbano	YOLOV5, YOLOV7 e YOLO-V8	Fendas transversais e longitudinais, covas, pele de crocodilo e fenda oblíqua.	Smartphone	2000	Não especificado	512x512
14	Egito	Flexível	Urbano	YOLO-V5 e YOLO-V8	Fendas transversais e longitudinais, pele de crocodilo, covas, danos em travessias pedonais, deformações localizadas e tampas de saneamento.	Smartphone e Google Street View	74420	Coordenadas GPS das imagens do Google Street View	600x600, 600x420 e 600x300

Tabela 3 - Caracterização do documento de acordo com a aplicação do YOLO na detecção de danos no pavimento. (Parte 3/3)

Doc N.	País	Tipo de pavimento	Ambiente	Versão YOLO	Tipo de degradações	Fonte ou método utilizado para capturar imagens	Nº de imagens	Posicionamento	Tamanho da imagem (Pixel)
15	Malásia	Flexível	Urbano	YOLO-V5	Fendas transversais e longitudinais, e covas.	Smartphone Android	9053	Coordenadas GPS das imagens do Google Street View	Não especificado
16	China	Rígido	Urbano	YOLO-V4, YOLO-V5 e YOLO-V7	Fendas transversais e longitudinais, pele de crocodilo e covas.	Não especificado	7710	Não especificado	Não especificado

Nos últimos anos, tem-se verificado um aumento expressivo de estudos que aplicam o YOLO à detecção de anomalias em pavimentos rodoviários. Para esta revisão bibliográfica foram analisados 16 artigos científicos de acesso aberto, indexados na base de dados Scopus, publicados entre 2019 e 2024. Estes trabalhos foram desenvolvidos em diferentes contextos geográficos, com particular destaque para países asiáticos (sobretudo a China, que contribuiu com 5 publicações), os Estados Unidos (3 artigos), bem como contribuições da Europa e da América Latina. A predominância de estudos em países asiáticos reflete políticas públicas orientadas para o investimento em inteligência artificial, enquanto as contribuições da Europa e da América reforçam a tendência de internacionalização da investigação nesta área.

As publicações selecionadas abordam diferentes tipos de degradação, incluindo fendas longitudinais e transversais, fendas pele de crocodilo, covas e reparações. Os estudos analisados recorreram a diversas versões do YOLO (do YOLOv2 ao YOLOv8), refletindo a rápida evolução desta arquitetura, e utilizaram diferentes fontes de imagens, desde câmeras de alta definição montadas em veículos, *smartphones*, câmeras instaladas em veículos aéreos não tripulados (VANTs), até imagens provenientes do Google Street View. A utilização de imagens captadas com recurso a veículos aéreos não tripulados tem sido cada vez mais frequentes na área de detecção de degradações de pavimentos rodoviários e aeroportuários, como pode ser observado no artigo de Feitosa et al.

Verificou-se ainda uma grande variabilidade no número de imagens utilizadas, que variou entre pouco mais de uma centena e mais de 150 mil imagens, influenciando diretamente a precisão dos resultados obtidos.

O tipo de danos no pavimento abordado em cada estudo é outro tema muito relevante. Todos os artigos mencionam o estudo da detecção de fendas ou covas, sendo os danos mais comuns analisados as fendas longitudinais e transversais, a pele de crocodilo e as covas. Alguns estudos ainda referem as reparações e a identificação de tampas de saneamento.

O número de imagens utilizadas para treinar o algoritmo varia significativamente entre os estudos selecionados, apresentando valores entre 130 e 156304 imagens. À medida que o número de imagens anotadas aumenta, também aumenta a qualidade dos resultados do YOLO, ou seja, a taxa de identificação correta dos danos aumenta.

Informações adicionais, como a forma como essas imagens foram obtidas, o tamanho da imagem, o posicionamento e a fonte ou tipo de câmera utilizada para capturar as imagens, também são detalhadas na Tabela 3.

Os estudos que fornecem informações sobre a fonte das imagens referem-se à utilização de imagens do Google Street View (4 estudos) ou capturadas por câmeras (3 estudos) ou

smartphones (4 estudos), mostrando uma tendência para a utilização de sistemas simplificados de captura de imagens ou dados disponíveis.

Apenas 7 estudos relataram o uso de sistemas de posicionamento, e todos eles se referiram ao GPS.

Por fim, em relação ao tamanho da imagem, também aqui há uma variação significativa, provavelmente associada ao equipamento utilizado para capturá-las. No entanto, predominam os formatos com dimensões em torno de 400x400 pixels.

Após analisar os principais aspectos dos 16 documentos selecionados, conclui-se que o uso da inteligência artificial para avaliar o estado dos pavimentos rodoviários tem perspectivas promissoras. Fatores como custo-benefício, economia de tempo, qualidade dos dados obtidos e avanços no desenvolvimento tecnológico em comparação com os sistemas existentes são elementos-chave na decisão de investir neste tipo de abordagens.

Além disso, os resultados indicam que todos os pontos positivos estão a convergir para a automatização total do processo de avaliação do estado do pavimento num futuro próximo.

Nos estudos que utilizam o YOLO para análise de pavimentos, observa-se uma tendência clara de ampliação do escopo das detecções, abrangendo uma variedade cada vez maior de tipos de degradações. Essa expansão busca tornar os modelos mais versáteis e aplicáveis a diferentes contextos rodoviários. Além disso, muitos pesquisadores têm recorrido a bases de dados de acesso aberto, como o Google Street View, pela facilidade de obtenção das imagens e pela ampla cobertura geográfica. Embora essas imagens nem sempre apresentem alta qualidade ou condições ideais de iluminação e enquadramento, elas representam uma alternativa prática e econômica para o desenvolvimento e o treino de modelos de detecção automática baseados em deep learning.

Em resumo, os estudos analisados sugerem que os sistemas de monitorização do estado do pavimento baseados em YOLO podem melhorar o tratamento dos dados, reduzir o tempo de processamento e melhorar a precisão dos dados.

Capítulo 2

2.0 Métodos e Ferramentas para a Detecção Automática de Degradações em Pavimentos

A detecção automática de degradações superficiais em pavimentos rodoviários é um domínio de investigação em rápido crescimento, impulsionado pela evolução das técnicas de visão computacional e de *deep learning*. Neste capítulo, é apresentado o modelo de *deep learning*, os principais programas e arquiteturas de detecção utilizados no âmbito da engenharia rodoviária, com destaque particular para o algoritmo YOLO (You Only Look Once) e para a plataforma Roboflow. Esta última tem vindo a ganhar relevância pela sua versatilidade na preparação de dados, no treino de modelos e na implementação de soluções baseadas em visão computacional. Além da análise destas abordagens, são descritas as métricas de avaliação de desempenho mais comuns, que são fundamentais para a comparação objetiva entre modelos e para a validação da sua eficácia em cenários reais.

2.1 Modelo de *deep learning*

A detecção automática de objetos através de técnicas de *deep learning* baseia-se, de forma geral, na utilização de redes neuronais convolucionais (CNN) [VALAPPIL, N et al, 2019] capazes de extrair características relevantes das imagens e associá-las a classes previamente definidas. No caso dos pavimentos rodoviários, estas classes correspondem a diferentes tipos de degradações superficiais, como fendas, covas e reparações [REDMON J et al, 2015].

O desenvolvimento de um modelo de aprendizagem profunda segue, habitualmente, quatro etapas principais:

- Preparação do conjunto de dados.
- Definição da arquitetura.
- Treino e validação do modelo.
- Teste e avaliação de desempenho.

- **Preparação do Conjunto de Dados**

A qualidade do modelo depende em grande medida da representatividade e da organização do conjunto de dados. Nesta fase são realizadas a recolha de imagens, a anotação manual dos objetos de interesse, a normalização de formatos e, por vezes, a aplicação de técnicas de aumento de dados (*data augmentation*), que visam diversificar artificialmente o conjunto de treino e melhorar a capacidade de generalização do algoritmo.

- **Arquitetura do Modelo**

A escolha da arquitetura constitui uma etapa crítica. No caso do YOLO, a rede é estruturada para processar a imagem numa única passagem, prevendo simultaneamente as classes e as localizações dos objetos na imagem. A evolução das versões do YOLO incorporou melhorias ao nível das camadas convolucionais, das funções de ativação e dos mecanismos de regularização, otimizando o equilíbrio entre velocidade e precisão.

- **Treino e Validação**

O processo de treino consiste em ajustar os parâmetros internos da rede, minimizando as funções de perda associadas à tarefa. Entre estas incluem-se a *Box Loss*, que avalia a precisão das caixas delimitadoras, a *Class Loss*, que mede a correção da classificação, e a *DFL Loss (Distribution Focal Loss)*, introduzida em versões mais recentes do YOLO para melhorar a regressão espacial. A validação é realizada em paralelo, utilizando um subconjunto distinto do conjunto de dados, de modo a acompanhar a evolução do modelo e a prevenir fenómenos de sobreajuste (*overfitting*).

- **Teste e Avaliação**

Após o treino, o modelo é testado com imagens não utilizadas na fase de treino e validação, assegurando uma avaliação objetiva da sua capacidade de generalização. Nesta fase, recorrem-se às métricas como a Precisão, a Revocação, o F1-Score e o mAP@0.5, bem como a parâmetros relacionados com o desempenho computacional, tais como o tempo de inferência, o consumo de memória da GPU e o tamanho das imagens processadas.

- **Ferramentas de Implementação**

A construção de modelos de deteção automática de objetos requer ferramentas que permitam gerir os dados, treinar o algoritmo e avaliar o seu desempenho. Neste contexto, destacam-se duas abordagens complementares: o YOLO, enquanto algoritmo de referência em tarefas de deteção em tempo real, e a plataforma Roboflow, que disponibiliza um ambiente integrado para anotação de imagens, organização de bases de dados e realização de testes em diferentes configurações do modelo. A utilização destas ferramentas garante não apenas a padronização do processo, mas também a possibilidade de comparar resultados obtidos em ambientes distintos, assegurando maior robustez às análises.

2.1.1 O Algoritmo YOLO (You Only Look Once)

O YOLO (You Only Look Once) é um dos algoritmos de referência na área da deteção automática de objetos em imagens e vídeos. Foi apresentado pela primeira vez em 2016, por Joseph Redmon e colaboradores, na Conference on Computer Vision and Pattern Recognition (CVPR) [REDMON et al, 2016], tendo introduzido uma abordagem inovadora face aos métodos existentes até então.

Enquanto algoritmos anteriores baseados em redes convolucionais (como o R-CNN e variantes) dividiam o processo em várias etapas, o modelo de rede neural convolucional (CNN) YOLO propõe um modelo de detecção em tempo real, tratando a tarefa como um problema de regressão única que, numa única passagem é capaz de identificar múltiplos objetos, a sua localização na imagem e a respetiva classe.

Desde a sua primeira versão, o YOLO tem vindo a evoluir continuamente, resultando em sucessivas melhorias de desempenho em termos de velocidade, precisão e capacidade de generalização. Atualmente existem várias versões do algoritmo (YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv7, YOLOv8, entre outras) [TERVEN, J et al, 2023], as quais incorporam avanços significativos em arquiteturas de redes neurais convolucionais, funções de perda e técnicas de treino.

Essa evolução consolidou o YOLO como um dos algoritmos mais utilizados em diferentes áreas, nomeadamente na saúde, vigilância, robótica e, mais recentemente, na detecção de degradações em pavimentos rodoviários, onde a rapidez e a precisão da análise são fatores determinantes.

A principal vantagem do YOLO reside na sua capacidade de combinar elevada precisão com um tempo de processamento reduzido (acima de 30 fotogramas por segundo), o que permite a sua aplicação em cenários que exigem análise em tempo real [HUSSAIN, M et al, 2023]. Além disso, a ampla disponibilidade de versões de código aberto e a vasta comunidade de utilizadores tornam o algoritmo altamente acessível e adaptável a diferentes contextos de investigação e aplicação prática.

2.1.2 A plataforma Roboflow

Desenvolvida com o objetivo de facilitar a criação, gestão e implementação de projetos de visão computacional, o Roboflow reduz as barreiras técnicas para investigadores e profissionais de diversas áreas. Fundada em 2020 nos Estados Unidos, esta plataforma integrada foi concebida para ajudar a superar os desafios comuns no desenvolvimento de modelos de *deep learning*, nomeadamente a preparação de conjuntos de dados, a anotação de imagens, o processamento preliminar, a geração de variações sintéticas, como a aumento de dados, e o treino de modelos [ROBOFLOW, 2025].

Uma das suas principais vantagens é disponibilizar um ambiente intuitivo que permite integrar diferentes arquiteturas de detecção de objetos, incluindo o YOLO, sem ser necessário conhecimentos avançados de programação. O Roboflow oferece também acesso a uma vasta galeria de conjuntos de dados públicos de várias áreas, bem como a ferramentas de partilha colaborativa, que aceleram a investigação e a aplicação prática de modelos de visão computacional.

No contexto da engenharia rodoviária, o Roboflow destaca-se como uma ferramenta particularmente útil, uma vez que simplifica o processo de anotação e gestão de imagens de pavimentos e permite o treino e o teste de modelos de forma centralizada. A integração direta com diferentes versões do YOLO faz da plataforma uma alternativa versátil para o desenvolvimento e comparação de abordagens de deteção automática de degradações, permitindo ao investigador avaliar o desempenho em diferentes configurações de forma mais eficiente.

A complementaridade entre o YOLO, um algoritmo de elevada eficiência, e o Roboflow, uma plataforma de suporte à criação e treino de modelos, é, por conseguinte, uma mais-valia significativa para o desenvolvimento de soluções aplicadas à monitorização e gestão de pavimentos rodoviários.

2.2 Métricas e Parâmetros de Avaliação de Desempenho

A análise do desempenho dos modelos de deteção automática não se limita às métricas clássicas de Precisão (*Precision*), Revocação (*Recall*), Pontuação F1 (*F1-Score*), Interseção sobre União (*Intersection over Union* – IoU) e Precisão Média Média (*Mean Average Precision* – mAP). Em contextos de *deep learning*, como o presente estudo, é igualmente necessário considerar outros indicadores relacionados com o processo de treino e validação, nomeadamente o número de épocas (epochs), o consumo de memória GPU (GPU memory), as funções de perda (Losses) e parâmetros associados ao conjunto de dados, como o número de instâncias e o tamanho das imagens em píxeis [SPRITAN S, 2023].

De seguida são apresentadas uma breve definição das métricas e parâmetros, assim como orientações para a sua interpretação.

- **Precisão (*Precision*)**

Indica a proporção de deteções corretas relativamente ao total de deteções efetuadas. Um valor elevado de precisão significa que o modelo identifica poucos falsos positivos.

$$Precision = \frac{Verdadeiros\ positivos}{Verdadeiros\ positivos + Falsos\ positivos}$$

- **Revocação (*Recall*)**

Mede a capacidade do modelo em identificar corretamente todos os objetos presentes no conjunto de dados. Valores baixos de *Recall* indicam que o modelo deixa escapar um número significativo de ocorrências reais (falsos negativos).

$$Recall = \frac{Verdadeiros\ positivos}{Verdadeiros\ positivos + Falsos\ negativos}$$

- **Pontuação F1 (*F1-Score*)**

Corresponde à média harmónica entre a Precisão e a Revocação, equilibrando os dois indicadores. É particularmente útil quando existe desequilíbrio de classes ou quando se pretende avaliar o compromisso entre a minimização de falsos positivos e falsos negativos.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- **Precisão Média Média (*Mean Average Precision - mAP*)**

O mAP é considerado a métrica de referência em tarefas de detecção de objetos. No presente estudo, adotou-se o mAP@0.5, que corresponde à média da *Average Precision* (AP) de todas as classes para um limiar de IoU $\geq 0,5$. Este valor fornece uma medida global da capacidade do modelo em detetar corretamente as instâncias anotadas, ponderando tanto a precisão como a revocação.

- **Interseção sobre União (*Intersection over Union – IoU*)**

A métrica Intersection over Union (IoU) é usada para avaliar a precisão de modelos de detecção de objetos, como o YOLO. Ela mede o quanto a área da previsão do modelo coincide com a área real do objeto. O valor da IoU varia de 0 a 1, quanto mais próximo de 1, maior é a sobreposição e, portanto, melhor é a detecção. Valores baixos indicam que a previsão está distante ou mal ajustada em relação ao objeto real.

Épocas (Epochs)

Uma época corresponde a uma passagem completa por todo o conjunto de treino. O número de épocas influencia diretamente a aprendizagem do modelo: poucas épocas podem resultar em subajuste (*underfitting*), enquanto um número excessivo pode conduzir a sobreajuste (*overfitting*). Normalmente o número de épocas mais utilizadas nestes tipos de aplicações variam de 100 a 200.

- **Memória GPU (*GPU Memory*)**

A memória GPU corresponde à quantidade de memória da placa gráfica utilizada durante o treino. Trata-se de um fator determinante para a eficiência computacional. Modelos com maior número de parâmetros ou imagens de maior dimensão requerem mais memória, podendo limitar a escalabilidade em dispositivos com menor capacidade.

- **Funções de Perda (Losses)**

Durante o treino, o modelo otimiza diferentes funções de perda que representam o erro em distintas dimensões da tarefa de detecção:

- **Box Loss** – mede o erro na previsão das coordenadas das caixas limitadoras (*bounding box*) relativamente à anotação real. Valores mais baixos indicam maior precisão na localização espacial.
- **Class Loss** – avalia a capacidade do modelo em classificar corretamente cada instância detetada, penalizando erros de atribuição da classe.
- **DFL Loss (Distribution Focal Loss)** – introduzida em versões mais recentes do YOLO, esta perda melhora a regressão das caixas delimitadoras ao modelar a distribuição dos limites, aumentando a precisão espacial.

- **Número de Instâncias (Instances)**

Refere-se ao total de objetos anotados no conjunto de treino e validação. Quanto maior o número de instâncias, mais robusto tende a ser o treino, desde que exista equilíbrio entre as diferentes classes.

- **Tamanho das Imagens (Size in Pixels)**

O tamanho das imagens utilizadas no treino influencia tanto a precisão do modelo como o consumo de recursos computacionais. Imagens de maior resolução permitem maior detalhe na detecção, mas exigem mais memória e tempo de processamento. Nos testes realizados, este parâmetro foi definido de acordo com a configuração padrão do YOLO/Roboflow, sendo ajustado para equilibrar desempenho e eficiência.

- **Limiar de confiança (Confidence threshold)**

O limiar de confiança define a probabilidade mínima para que uma detecção seja aceite como válida. Valores baixos aumentam o número de objetos detetados, mas com maior risco de falsos positivos; já valores altos reduzem erros, mas podem descartar detecções corretas.

A Tabela 4 apresenta o resumo de todas as métricas e parâmetros utilizados ao longo deste estudo.

Tabela 4 - Métricas e parâmetros de avaliação utilizados no estudo

Métrica/Parâmetro	Definição	Interpretação / Relevância
Precisão (<i>Precision</i>)	Proporção de detecções corretas entre todas as positivas.	Mede a capacidade do modelo em evitar falsos positivos. Quanto maior, mais confiável é a classificação.
Revocação (<i>Recall</i>)	Proporção de ocorrências reais identificadas pelo modelo.	Mede a capacidade de encontrar todos os objetos da classe de interesse. Valores baixos indicam perda de instâncias.
Pontuação F1 (<i>F1-Score</i>)	Média harmônica entre Precisão e Revocação.	Equilibra os dois indicadores, útil em bases de dados desequilibradas.
mAP@0.5	Média da Precisão Média de todas as classes para IoU \geq 0,5.	Métrica global mais utilizada em detecção de objetos; avalia simultaneamente a precisão e a revocação.
IoU (<i>Intersection over Union</i>)	Grau de sobreposição entre a caixa predita e a anotada.	Valores mais próximos de 1 indicam maior correspondência espacial entre previsão e realidade.
Épocas (<i>Epochs</i>)	Número de passagens completas pelo conjunto de treino.	Poucas épocas implicam subajuste; excesso de épocas implica sobreajuste.
Memória da GPU (<i>GPU Memory</i>)	Quantidade de memória gráfica usada no treino.	Reflete a eficiência computacional; aumenta com imagens de maior resolução e modelos mais complexos.
<i>Box Loss</i>	Erro na previsão das coordenadas das caixas limitadoras.	Quanto menor, mais precisa é a localização espacial das degradações.
Class Loss	Erro na atribuição da classe do objeto.	Avalia a capacidade de classificar corretamente cada instância.
DFL Loss (<i>Distribution Focal Loss</i>)	Função de perda para regressão de caixas delimitadoras.	Melhora a precisão espacial da detecção em versões recentes do YOLO.
Instâncias (<i>Instances</i>)	Número total de objetos anotados no conjunto de treino/validação.	Bases mais ricas tendem a gerar modelos mais robustos.
Tamanho das imagens (<i>Size in Pixels</i>)	Resolução das imagens usadas no treino (ex.: 416x416, 640x640).	Maior resolução implica maior detalhe, mas maior custo computacional.
Limiar de confiança (<i>Confidence threshold</i>)	Probabilidade mínima para aceitar uma detecção como válida.	Ajusta o equilíbrio entre detecções corretas e falsos positivos.

Capítulo 3

3.0 Metodologia

A metodologia constitui um dos pilares fundamentais de qualquer investigação científica, uma vez que define o caminho seguido para alcançar os objetivos propostos e assegura a validade e a reprodutibilidade dos resultados. No presente trabalho, a metodologia adotada é orientada para a aplicação de técnicas de visão computacional, com destaque para a utilização do algoritmo You Only Look Once (YOLO), na deteção automática de degradações superficiais em pavimentos rodoviários.

O desenho metodológico desenvolvido procura traduzir, de forma clara e estruturada, as diferentes etapas que conduzem este estudo, desde a recolha e preparação dos dados, até a obtenção e análise de resultados. A opção por uma abordagem sistemática e faseada justifica-se pela necessidade de garantir consistência em todas as fases do processo, permitindo não apenas a formação de um modelo confiável, mas também a sua adequada avaliação em termos de desempenho.

De modo a assegurar uma visão global da estratégia metodológica, recorreu-se à construção de um fluxograma que sintetiza as etapas principais do estudo e a sua sequência. Estas etapas incluem a definição da base de dados de imagens, o seu pré-processamento e anotação, a preparação do conjunto de treino e validação, o treino do modelo YOLO, a avaliação através de métricas de desempenho e, finalmente, a análise crítica dos resultados obtidos. A articulação entre estas fases revela-se essencial, uma vez que cada decisão metodológica influencia diretamente a qualidade dos resultados finais.

A estrutura metodológica que se apresenta neste capítulo estabelece, assim, as bases para a aplicação prática desenvolvida no Capítulo 4, dedicada ao caso de estudo.

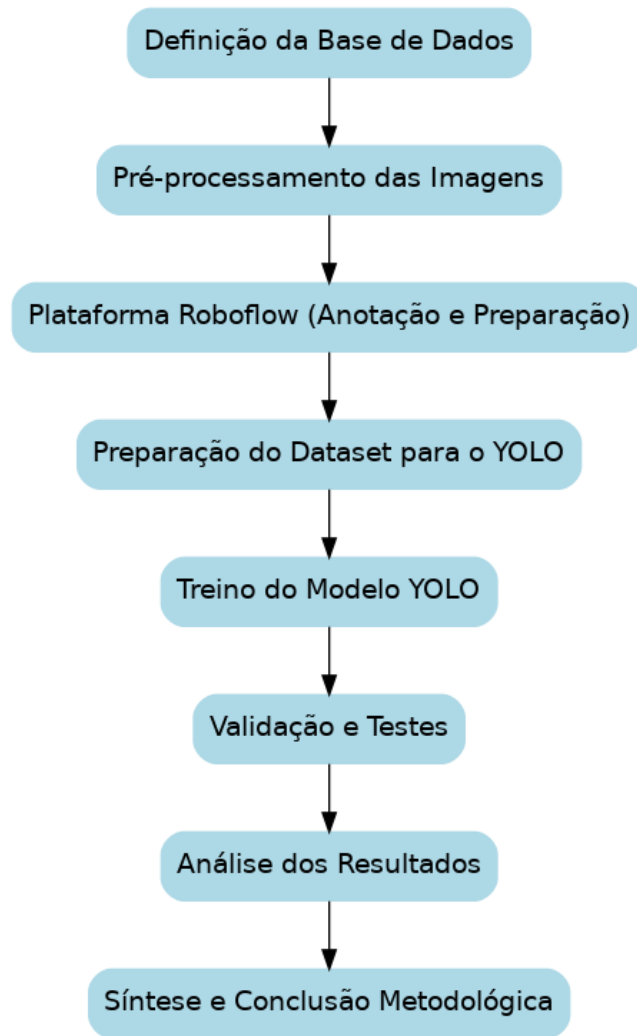


Figura 1 - Fluxograma da metodologia de estudo.

3.1 Definição da base de dados

A definição da base de dados constitui a primeira etapa do processo metodológico, uma vez que a qualidade e a representatividade das imagens utilizadas determinam, em larga medida, a fiabilidade dos resultados obtidos na fase de deteção automática de degradações.

A aquisição das imagens pode realizar-se através de diversos meios tecnológicos. Uma das abordagens mais utilizadas consiste na instalação de câmaras em veículos que percorrem as vias em estudo. Esta metodologia permite recolher imagens contínuas e atualizadas do estado dos pavimentos, garantindo simultaneamente uma cobertura espacial significativa. Outra possibilidade relevante é a utilização de VANTs, que oferecem a vantagem de captar imagens em diferentes ângulos e altitudes, permitindo identificar defeitos que poderiam passar despercebidos a partir do nível do solo. Para além destas alternativas, plataformas digitais como o Google Street

View podem também ser exploradas, funcionando como uma fonte complementar de dados, sobretudo em locais onde a recolha física de imagens seja difícil ou dispendiosa.

Apesar de suas vantagens, cada uma destas formas de captação de imagens apresenta também limitações que importa considerar. A utilização de câmeras em veículos depende de condições de circulação adequadas e pode ser afetada por fatores ambientais, como luminosidade insuficiente ou presença de sombras. A operação de VANTs implica restrições legais relacionadas com o espaço aéreo, e apresenta custos operacionais que podem ser significativos, mas em geral inferiores à captação de imagem com veículos terrestres equipados. Já o recurso ao Google Street View apresenta limitações temporais, dado que as imagens disponibilizadas não são atualizadas em tempo real, assim como espaciais, pois nem todas as regiões dispõem de cobertura completa.

Diversos autores têm utilizado veículos equipados com câmaras especializadas para recolher imagens de pavimentos em diferentes condições de luminosidade e ambientais. O dataset *PaveDistress* [LIU Z et al, 2024], por exemplo, contém imagens de elevada resolução captadas ao longo do troço S315 na China, com condições variadas como sombra, túneis ou tempo nublado, e categoriza defeitos como fendas transversais, longitudinais, covas e reparações. Existe ainda a base de dados de acesso aberto de dados de pavimentos rodoviários do Programa de Desempenho de Pavimentos a Longo Prazo (LTPP) geridos pela U.S. Federal Highway Administration (FHWA) [FHWA-HRT-21-038, 2021].

No contexto da presente investigação, a base de dados é composta por imagens de pavimentos rodoviários flexíveis captadas com recurso a VANT.

Para além das questões técnicas e logísticas, é importante considerar a necessidade de assegurar a diversidade da base de dados, tanto em termos geográficos como de tipos de degradações. Uma base de dados equilibrada, contendo imagens de diferentes tipos e níveis de gravidade de degradações, constitui um requisito essencial para o treino eficaz do modelo de deteção. A ausência de equilíbrio entre classes pode resultar em tendências erradas, comprometendo a capacidade de generalização do algoritmo.

Em resumo, a definição da base de dados não se limita à recolha de imagens, mas envolve também a análise crítica das fontes disponíveis, a ponderação das suas vantagens e limitações, e a garantia de diversidade e representatividade dos dados. Esta etapa é determinante para a robustez do processo metodológico e para a credibilidade dos resultados das fases seguintes.

3.2 Pré processamento das imagens

Após a definição da base de dados, torna-se necessário proceder ao pré-processamento das imagens, etapa que assume particular relevância em tarefas de visão computacional. Este processo tem como objectivo preparar os dados para que possam ser utilizados de forma eficaz na fase de anotação e posteriormente no treino do modelo YOLO. A qualidade do pré-processamento

influencia diretamente o desempenho do algoritmo, uma vez que reduz discrepâncias entre amostras e garante maior consistência na base de dados.

Um dos primeiros passos consiste na normalização e padronização das imagens. Como as bases de dados podem integrar imagens captadas em diferentes condições (resolução variável, distintos dispositivos de recolha, variações de iluminação), é essencial que sejam convertidas para um formato e dimensão uniforme. Normalmente, as imagens são redimensionadas para resoluções compatíveis com o modelo, de modo a assegurar compatibilidade com a arquitetura da rede neural.

Um aspeto de particular importância é a aplicação de técnicas de aumento de dados (*data augmentation*), destinadas a ampliar artificialmente a base de dados inicial. O objetivo passa por introduzir variações controladas nas imagens originais, criando versões alternativas que simulam diferentes condições de captação. Entre os procedimentos aplicados destacam-se:

- Rotações em ângulos variados, que permitem treinar o modelo para reconhecer as mesmas degradações mesmo quando estas surgem em orientações diferentes.
- Espelhamentos horizontais e verticais, úteis para aumentar a diversidade espacial da base de dados.
- Alterações de escala e recortes (*cropping*), que simulam a deteção de defeitos a diferentes distâncias e ângulos de observação.
- Pequenos deslocamentos ou translações, que aumentam a variabilidade posicional dos defeitos dentro da imagem.

Estas transformações não alteram a essência da informação contida na imagem, mas produzem um conjunto adicional de dados que enriquece o treino. Deste modo, mesmo a partir de uma base inicial relativamente reduzida, é possível gerar uma base de dados muito mais ampla, equilibrada e representativa. Este processo contribui de forma decisiva para a robustez do modelo, aumentando a sua capacidade de generalização para contextos reais.

Embora em muitos estudos sejam igualmente utilizados métodos de redução de ruído ou ajuste de luminosidade, essas técnicas não foram aplicadas no presente trabalho. A sua menção serve apenas para contextualizar práticas comuns em outras investigações.

Por fim, procedeu-se à divisão equilibrada dos conjuntos de treino, validação e teste, assegurando a representação das diferentes classes de degradações. Esta repartição é essencial para evitar distorções e permitir uma avaliação confiável do desempenho do modelo, a distribuição das imagens para treino, validação e teste é feita de modo automático e aleatório.

Assim, o pré-processamento das imagens desempenha um papel crucial na metodologia, não apenas pela preparação técnica dos dados, mas sobretudo pelo impacto direto na qualidade e confiança do treino do modelo.

3.3 Anotação e preparação na plataforma Roboflow

Concluído o pré-processamento das imagens, a etapa seguinte consiste na sua importação para a plataforma Roboflow, uma ferramenta amplamente utilizada em projetos de visão computacional pela sua interface intuitiva e pela diversidade de funcionalidades que disponibiliza. O Roboflow permite centralizar todo o processo de anotação e preparação dos dados, assegurando a criação de conjuntos compatíveis com diversos modelos de *deep learning*, incluindo o YOLO.

Um dos aspetos mais relevantes nesta fase é a anotação das imagens, isto é, a marcação manual ou semiautomática das regiões onde se encontram as degradações na superfície dos pavimentos. Este processo é realizado através da criação de caixas limitadoras (bounding boxes) que envolvem cada degradação identificada, sendo depois associadas à caixa o tipo de degradação identificada e respetivo nível de gravidade. A precisão na anotação é determinante para a qualidade do modelo, uma vez que o algoritmo aprende a reconhecer padrões com base na informação fornecida. Anotações imprecisas ou inconsistentes podem induzir distorção nos resultados e comprometer o desempenho do modelo.

Após a anotação das imagens, procede-se à definição dos parâmetros da base de dados, entre os quais se destaca a divisão entre conjuntos de treino, validação e teste. Esta divisão é feita automaticamente pela plataforma, embora permita configuração manual dos percentuais de repartição, por exemplo, 70 % treino, 15 % validação, 15 % teste. A correta segmentação dos dados é fundamental para que o modelo YOLO, evitando sobre ajustamentos (overfitting) ou avaliações irrealistas.

A utilização da plataforma Roboflow permite a exportação da base de dados em múltiplos formatos compatíveis com diferentes arquiteturas de redes neurais. No caso desta investigação, os dados foram tratados nesta plataforma e exportados em formato específico para YOLO, contendo as pastas organizadas para as operações de treino, validação e teste (imagens e etiquetas) e os ficheiros de configuração necessários para iniciar o processo de treino.

Salienta-se ainda que a plataforma Roboflow pode desempenhar um papel essencial na metodologia, ao facilitar a anotação manual das imagens, a definição de parâmetros estruturais da base de dados e a preparação dos ficheiros de dados para integração com o YOLO.

Esta etapa funcionou como uma ponte entre a recolha e pré-processamento das imagens e a fase de treino do modelo, garantindo a organização e consistência em todo o processo.

3.4 Preparação da Base de Dados para o YOLO

Após a anotação e exportação da base de dados através do Roboflow, é necessário proceder à sua adaptação para o treino no YOLO. Esta etapa é fundamental, pois garante que o modelo consegue interpretar corretamente tanto as imagens como os ficheiros de etiquetas associados.

O YOLO requer uma estrutura de diretórios e ficheiros específica, que deve ser respeitada para que o processo de treino decorra sem erros. Para o efeito, os dados devem ser organizados em três pastas principais:

- **Train:** contendo as imagens destinadas ao treino do modelo, acompanhadas dos respetivos ficheiros de etiquetas.
- **Validation:** conjunto de imagens e etiquetas utilizadas para monitorizar o desempenho do modelo durante o processo de treino.
- **Test:** subconjunto independente, reservado para a avaliação final da performance do algoritmo.

Para além dos ficheiros de anotação, é igualmente necessário configurar o ficheiro de classes, geralmente denominado `data.yaml`. Neste ficheiro são listados todos os tipos e níveis de gravidade das degradações a considerar na análise. Este ficheiro assegura que o modelo associa corretamente os valores numéricos presentes nas etiquetas às respetivas designações, por exemplo, classe 0 = fenda longitudinal; classe 1 = cova.

Outro elemento é o ficheiro de configuração do treino, no qual se definem parâmetros como o número de classes, os caminhos para as pastas de treino, validação e teste, bem como as definições de épocas (epochs). A correta parametrização deste ficheiro é decisiva para a eficiência do processo de treino e para a qualidade dos resultados obtidos. Essa configuração pode ser realizada no YOLO de forma direta, programando nas linhas de comando.

A preparação da base de dados para o YOLO exige a criação de uma estrutura organizada e a configuração de ficheiros auxiliares indispensáveis. Esta etapa garante a interoperabilidade entre os dados anotados no Roboflow e o modelo YOLO, constituindo um elo central da metodologia.

3.5 Treino do modelo YOLO

Com a base de dados devidamente estruturada e configurada, a etapa seguinte corresponde ao treino do modelo YOLO. Esta fase assume um papel central na metodologia, pois é neste momento que a rede neural convolucional ajusta os seus pesos internos para aprender a reconhecer corretamente as diferentes classes de degradação dos pavimentos.

O processo de treino deve ser realizado com base nos conjuntos de imagens de treino e validação. O primeiro fornece exemplos que permitem ao modelo aprender os padrões relevantes, enquanto

o segundo possibilita acompanhar o desempenho ao longo das iterações, prevenindo fenômenos de sobre ajustamento (*overfitting*).

Entre os parâmetros mais relevantes a definir destacam-se:

- Número de épocas (epochs): indica quantas vezes o modelo percorre a totalidade do conjunto de treino. Um número reduzido de épocas pode conduzir a sub ajustamento (*underfitting*), enquanto valores excessivos podem originar sobre ajustamento.
- Tamanho do lote (batch size): determina quantas imagens são processadas em simultâneo antes da atualização dos pesos da rede. A escolha depende do equilíbrio entre a disponibilidade de recursos computacionais e a estabilidade do processo de treino.
- Taxa de aprendizagem (learning rate): controla a rapidez com que os parâmetros do modelo são ajustados. Taxas demasiado elevadas podem dificultar a convergência, enquanto taxas muito baixas tornam o processo demasiado lento.

Durante o treino, é possível acompanhar a evolução da função de perda, que quantifica a discrepância entre as previsões do modelo e as anotações reais. Uma diminuição progressiva do das perdas indica que o modelo está a ajustar os seus parâmetros de forma adequada.

Além disso, podem ser avaliadas métricas de desempenho como a precisão, a revocação e a média média da precisão, que permitem medir a capacidade do modelo para identificar corretamente tanto defeitos frequentes como os menos comuns.

Em síntese, o treino do modelo YOLO constitui a fase mais intensiva em termos computacionais, mas também a mais determinante para a qualidade do sistema. A definição criteriosa dos parâmetros e a monitorização contínua dos indicadores de desempenho são elementos essenciais para assegurar a robustez do modelo desenvolvido.

3.6 Avaliação, Testes Finais e Análise dos resultados

Concluído o processo de treino, é fundamental proceder à avaliação do modelo com recurso às métricas específicas de deteção de objetos. Esta fase tem como objetivo verificar a capacidade do sistema em generalizar para novos dados e garantir que o desempenho obtido não resulta apenas de um ajustamento aos exemplos utilizados durante o treino.

Para esse efeito, pode ser utilizado um conjunto de teste independente, composto por imagens que não participaram no treino nem na validação. A avaliação sobre este conjunto permite medir a confiança e a aplicabilidade do modelo em condições reais.

Adicionalmente às métricas já mencionadas, é ainda possível analisar as matrizes de confusão, que permitem identificar classes específicas onde o modelo apresenta maiores dificuldades de classificação, fornecendo pistas para ajustes futuros.

Pode-se concluir que a etapa de avaliação e testes finais é indispensável para validar a metodologia adotada, permitindo não apenas quantificar o desempenho do YOLO, mas também identificar limitações e oportunidades de melhoria. Os resultados obtidos nesta fase fornecem a base para a discussão crítica e para a aplicação prática do modelo em contextos reais, preparando o caminho para a análise apresentada no capítulo 4.

3.7 Síntese e Conclusão Metodológica

O percurso metodológico delineado neste capítulo permitiu estruturar de forma sistemática todas as etapas necessárias para a detecção automática de degradações em pavimentos com recurso ao modelo YOLO. Partindo da definição da base de dados e do pré-processamento das imagens, passando pela anotação e preparação da base de dados no Roboflow, pela sua adaptação à estrutura exigida pelo YOLO e pelo processo de treino, até chegar à avaliação final do modelo, cada fase foi concebida como parte de um processo integrado e sequencial, em conformidade com o fluxograma apresentado.

Importa destacar que, após o treino e avaliação do modelo, torna-se possível proceder à análise dos resultados, etapa em que se examinam em detalhe os indicadores de desempenho, identificam-se pontos fortes e fragilidades e comparam-se diferentes configurações testadas. Esse momento constitui a base para a formulação de uma síntese metodológica, na qual se consolidam as conclusões sobre a eficácia do processo, evidenciando as potencialidades e as limitações da abordagem adotada.

Deste modo, encerra-se a construção metodológica que sustenta o trabalho, estabelecendo um quadro sólido para a aplicação prática. No Capítulo 4 – Caso de Estudo, esta metodologia será implementada em dados concretos, permitindo verificar a sua aplicabilidade em contexto real e analisar os resultados obtidos de forma crítica.

Capítulo 4

4.0 Caso de estudo

O presente capítulo apresenta o caso de estudo desenvolvido no âmbito desta dissertação, centrado num troço de via localizado no Fundão, em Portugal. A escolha desta área deve-se tanto à sua relevância no contexto da infraestrutura rodoviária local quanto à possibilidade de aplicação de tecnologias de captura de imagens para a análise de degradações em pavimentos.

Para a realização do estudo, foram obtidas 459 imagens aéreas através de sobrevoos com VANT, assegurando uma cobertura detalhada do troço em análise. Estas imagens constituem a base de dados utilizada para a aplicação e teste dos modelos de deteção automática de degradações em pavimentos apresentados.

Os dados recolhidos foram posteriormente integrados em plataformas de processamento e treino de modelos de visão computacional, nomeadamente o Roboflow e o YOLO. Ambos os sistemas permitem estruturar a base de dados e realizar o treino com recurso a redes neurais convolucionais (CNN) para deteção de objetos, ainda que com abordagens distintas. A aplicação destes dois ambientes possibilita uma análise comparativa do desempenho de cada um, tendo em vista a robustez, a precisão e a aplicabilidade prática dos resultados obtidos.

A estrutura deste capítulo segue uma lógica progressiva: inicialmente apresenta-se o enquadramento geográfico do troço da via selecionado para o estudo, o equipamento utilizado e as condições do levantamento dos dados e a tipologia da via. Em seguida, apresenta-se o processo de anotação das imagens no Roboflow, fundamentado na norma Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys, 2023 – [ASTM D6433-23], que serviu de referência para a classificação das diferentes degradações observadas e seus respetivos níveis de gravidade. Na sequência, a base de dados é configurada e os modelos de deteção executados no próprio ambiente Roboflow, com a análise dos respetivos resultados. A mesma base de dados é também utilizada na obtenção de modelos preditivos no YOLO, preparando o terreno para a comparação crítica dos resultados.

4.1 Localização e aquisição das imagens para a base de dados

O troço rodoviário em estudo localiza-se em ambiente semi-rural, no quilómetro 49,8 da Estrada Nacional 18 (N18), entre as cidades do Fundão e da Covilhã, em Portugal (Figura 2).

As imagens utilizadas para a constituição da base de dados deste estudo foram obtidas a partir de outro trabalho académico previamente desenvolvido, tendo sido disponibilizadas para a continuidade e conclusão da presente investigação. Essas imagens serviram de suporte fundamental para conclusão deste caso de estudo. Na sequência, serão descritas as principais operações realizadas durante a captação dos dados, bem como os cuidados considerados para garantir a qualidade, representatividade e consistência das imagens utilizadas. [GRAVINHOS, P. et al, 2024], [NERY, D. et al, 2023].



Figura 2 - Municípios do Fundão e Covilhã. Fonte: <https://commons.wikimedia.org/w/index.php?curid=87666110>

O perfil transversal da via é composto por 2 vias de tráfego de 3,2 m de largura divididas por um separador central de 1,0 m de largura, sem bermas, e um sistema de drenagem constituído por coletor enterrado alimentado por sumidouros. O troço foi alvo de beneficiação entre 2005 e 2007, tendo sido aplicada uma camada de base de 20 cm de *tout-venant*, uma camada de ligação (*binder*) de 6 cm de mistura betuminosa densa e uma camada de desgaste de 5 cm em betão betuminoso [GRAVINHOS, P. et al, 2024]. A captação de imagens com VANT foi realizada na via da direita, no sentido SE-NO, abrangendo uma extensão de pavimento de aproximadamente 52 m e uma área de 183 m³, conforme ilustrado na Figura 3. O pavimento da estrada apresenta danos severos: fendas, pele de crocodilo, deformações localizadas, covas e reparações.



Figura 3 - Localização do trecho rodoviário em estudo [GRAVINHOS, P. et al, 2024]

As condições específicas do local foram previamente analisadas e o plano de voo foi definido em conformidade com as restrições legais em vigor. Procedeu-se também à verificação da existência de eventuais obstáculos, como árvores, postes de iluminação ou cabos aéreos, tendo-se concluído que estes não interferiam com a trajetória estabelecida.

O levantamento da informação foi realizado com um drone DJI Mavic 2 Pro (peso: 907 g), enquadrado na “categoria aberta”, isto é, uma aeronave não tripulada de baixo peso (inferior a 25 kg), operada a baixas altitudes (até 120 m), sempre dentro do campo visual do operador, fora de zonas restritas e a uma distância segura de pessoas e infraestruturas.

Não tendo sido identificadas limitações adicionais, o plano de voo foi elaborado com especial atenção à mitigação de riscos para pessoas e bens. Assim, a inspeção foi programada para o período diurno, num dia com menor intensidade de tráfego (sábado, 15 de maio de 2022, durante a manhã). As condições meteorológicas locais foram verificadas antecipadamente, confirmando-se céu limpo e vento fraco, de modo a garantir a estabilidade do VANT ao longo do voo. Em cumprimento da regulamentação, procedeu-se ainda à restrição temporária do tráfego na via em análise.

O voo foi realizado a uma altitude de 3 metros, sendo operado manualmente. A área de solo capturada em cada imagem, que depende da altitude do voo e da distância focal da lente da câmara usada, afetando o nível de detalhe da imagem, foi de 3,97 m por 2,65 m de e resolução em pixels de 0,73 mm (lado do pixel). A figura 4 mostra uma das 459 imagens obtidas, apresentando boa qualidade para a deteção de degradações e seus respetivos níveis de gravidade.

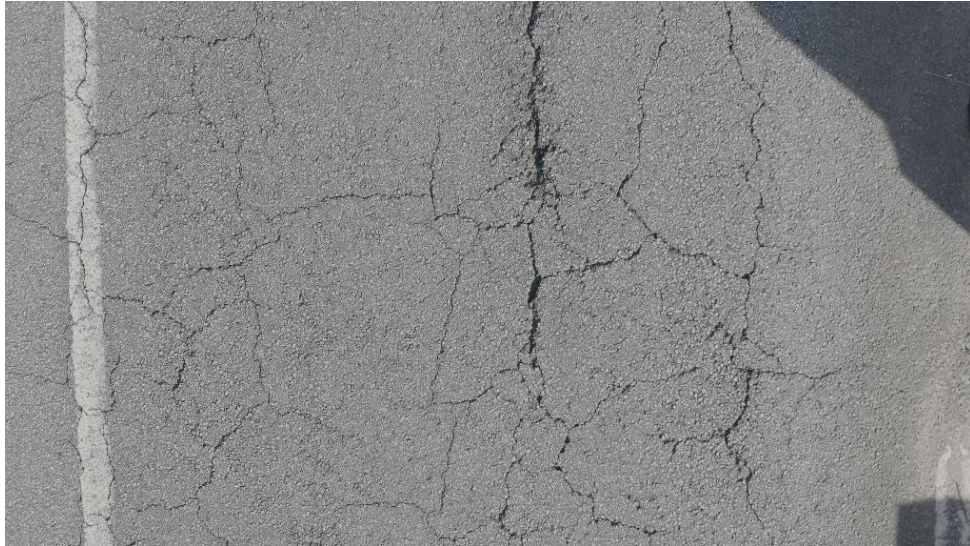


Figura 4 - Exemplo de imagem usada no estudo

4.2 Anotações das imagens

Após a recolha das imagens, tornou-se necessário proceder à sua anotação manual, de modo a preparar o conjunto de dados para o treino do modelo de deteção. Este processo consiste em identificar e delimitar, em cada imagem, as regiões correspondentes às diferentes degradações do pavimento rodoviário, atribuindo-lhes a respetiva classe. Para tal, recorreu-se à plataforma Roboflow, que disponibiliza um ambiente intuitivo e eficiente para a anotação e organização de dados visuais.

No total, foram anotadas 459 imagens, nas quais se assinalaram as áreas de interesse de acordo com as degradações observadas. Ao todo foram realizadas 933 anotações, uma média 2,7 anotações por imagem. Para garantir a consistência e a clareza do processo, as classes de degradação e os seus respetivos níveis foram recodificados em formato numérico, simplificando a estrutura do conjunto de dados e assegurando uma melhor integração com o algoritmo de treino. Esta codificação facilita ainda a padronização da informação e contribui para reduzir ambiguidades durante a fase de processamento.

A tabela seguinte apresenta a relação entre as classes consideradas no estudo e a sua codificação numérica, constituindo a base para a aprendizagem supervisionada do modelo.

Tabela 5 - Classificação das degradações.

Nova codificação das classes	Classes de degradações	Número de anotações
100	Pele de crocodilo em nível severo	289
101	Pele de crocodilo em nível baixo	10
102	Pele de crocodilo em nível médio	494
103	Fendas longitudinais em nível baixo	125
104	Fendas longitudinais em nível médio	46
106	Covas em nível baixo	60
107	Covas em nível médio	95
108	Fendas transversais em nível baixo	102
109	Fendas transversais em nível médio	31

A Tabela 5 apresenta a classificação das classes de degradações identificadas nas imagens utilizadas. A classe 100 refere-se à pele de crocodilo em nível severo, contabilizando 289 anotações, representando uma das categorias mais expressivas do conjunto de dados. A classe 101, correspondente à pele de crocodilo em nível baixo, registou 10 anotações, enquanto a classe 102, que descreve a pele de crocodilo em nível médio, apresentou 494 anotações, sendo a mais representativa entre todas as classes. A classe 103 diz respeito às fendas longitudinais em nível baixo, com 125 anotações, e a classe 104, que representa as fendas longitudinais em nível médio, contabilizou 46 anotações. As classes 106 e 107 correspondem às covas em diferentes níveis de severidade: nível baixo com 60 anotações e nível médio com 95 anotações, respectivamente. As classes 108 e 109 representam as fendas transversais, sendo 102 anotações para o nível baixo e 31 anotações para o nível médio. Esses valores mostram uma predominância das degradações do tipo pele de crocodilo, especialmente em níveis médios, refletindo a maior incidência desse tipo de dano no pavimento analisado.

No início da fase de anotação, considerando que a base de dados disponível era relativamente reduzida, composta por apenas 459 imagens, procurou-se explorar estratégias que permitissem maximizar a sua utilização. Verificou-se que a maioria das degradações presentes nas imagens correspondia ao padrão conhecido como pele de crocodilo. De forma a diversificar as classes e aumentar a representatividade do conjunto, foi inicialmente tentada a subdivisão dessas áreas de pele de crocodilo em categorias distintas, tais como fissuras longitudinais e fissuras transversais.

Apesar de esta abordagem ter permitido a criação de um conjunto de dados aparentemente mais variado, os resultados obtidos nos primeiros testes revelaram-se insatisfatórios. O modelo apresentou dificuldades em distinguir de forma clara entre as classes, gerando uma performance global abaixo do esperado. Esta situação conduziu à conclusão de que a estratégia inicial de subdivisão artificial das degradações não era adequada, uma vez que criava ambiguidades na anotação e confundia o processo de treino do algoritmo.

Face a esta limitação, optou-se por reanotar integralmente todas as imagens, desta vez seguindo rigorosamente as categorias de degradações estabelecidas pela norma ASTM D6433-23 [ASTM, 2023], que fornece diretrizes específicas para a classificação de defeitos em pavimentos rodoviários. Esta decisão garantiu maior consistência na rotulagem, melhor alinhamento com padrões internacionais e, conseqüentemente, uma melhoria significativa na qualidade da base de dados.

4.3 Testes preliminares no Roboflow

Concluída a fase de anotação, procedeu-se à realização de testes preliminares diretamente na plataforma Roboflow, com o objetivo de avaliar a consistência do conjunto de dados e obter uma primeira percepção acerca do desempenho do modelo. Esta etapa assume particular relevância, uma vez que permite identificar eventuais problemas na base de dados como erros de anotação, classes pouco representadas ou desequilíbrios na distribuição das imagens antes da implementação em ambientes externos de treino.

Ao todo, foram efetuados 20 testes distintos, nos quais se exploraram diferentes combinações de parâmetros de treino, número de épocas e variações de pré-processamento das imagens. A realização destes ensaios experimentais possibilitou não apenas compreender melhor o comportamento do modelo em condições controladas, mas também ajustar gradualmente a estratégia metodológica, de forma a otimizar os recursos disponíveis e aumentar a fiabilidade dos resultados nas fases subsequentes.

Durante a fase de testes no Roboflow, uma das estratégias exploradas para mitigar a limitação do número reduzido de imagens disponíveis consistiu na aplicação de técnicas de ampliação de dados (*data augmentation*). Estas técnicas visam aumentar artificialmente a dimensão e a diversidade do conjunto de dados, através da introdução de variações nas imagens originais. No entanto, verificou-se que a aplicação destas transformações deve ser cuidadosamente controlada, uma vez que algumas delas podem introduzir distorções que comprometem a fiabilidade do processo de anotação.

Constatou-se, por exemplo, que a rotação das imagens em 90 graus não se revelou adequada, dado que altera a natureza das fendas: fendas longitudinais passam a ser interpretadas como transversais e vice-versa, gerando ambigüidade que prejudica a classificação correta das degradações. De forma semelhante, a utilização de zoom excessivo mostrou-se problemática, pois

pode modificar a percepção da gravidade ou do nível da degradação, levando o modelo a interpretações incorretas. Também o espelhar verticalmente (*vertical flip*) não apresentou resultados satisfatórios, uma vez que altera a disposição natural das imagens de pavimentos rodoviários, afastando-as das condições reais de recolha.

Assim, optou-se por restringir as transformações a operações de espelhamento horizontal (*horizontal flip*), que mantêm a consistência espacial das degradações e ainda aumenta a variabilidade do conjunto de treino sem comprometer a fiabilidade das classes.

No que diz respeito à divisão do conjunto de dados (*dataset split*), os melhores resultados foram obtidos com a configuração aproximada de 70% para treino (*train set*), e 15% tanto para validação (*validation set*) como para teste (*test set*). Esta distribuição permitiu assegurar um volume suficiente de dados para o treino do modelo, ao mesmo tempo que garantiu a disponibilidade de subconjuntos adequados para a validação durante o processo e para a avaliação final do desempenho.

Ao realizar o treino de modelos na plataforma Roboflow, é possível selecionar diferentes arquiteturas de referência, designadas por *type*. Entre as opções mais utilizadas destacam-se o Roboflow 3.0 e a família YOLO, que neste estudo foi representada pelo YOLOv11.

O Roboflow 3.0 corresponde a um modelo proprietário desenvolvido pela própria plataforma, otimizado para uma utilização mais direta e com menor necessidade de ajustes técnicos por parte do utilizador. Trata-se de uma solução acessível para prototipagem rápida, embora ofereça menos controlo sobre os parâmetros internos da arquitetura. Já o YOLOv11 é uma versão mais recente da conhecida família de modelos *You Only Look Once*, amplamente aplicada em tarefas de deteção em tempo real. A sua inclusão no Roboflow permite treinar diretamente numa arquitetura robusta e consolidada, oferecendo maior flexibilidade e, em muitos casos, melhor desempenho quando comparado com soluções genéricas.

Considerando estas opções, optou-se por apresentar nesse caso de estudo a análise em quatro testes principais no Roboflow:

- Versão 1 - Treino com o modelo Roboflow 3.0 utilizando a base de dados original (sem ampliação de dados).
- Versão 2 - Treino com o modelo YOLOv11, com a base de dados original (sem ampliação de dados).
- Versão 3 - Treino com o modelo Roboflow 3.0 aplicando técnicas de ampliação de dados.
- Versão 4 - Treino com o modelo YOLOv11, aplicando técnicas de ampliação de dados.

Esta abordagem permitiu não só avaliar o impacto da ampliação artificial do conjunto de dados no desempenho final, mas também comparar diretamente duas arquiteturas distintas de deteção

de objetos, verificando a sua adequação à tarefa de identificação de degradações em imagens de pavimentos rodoviários.

4.3.1 Roboflow – Versão 1

A primeira análise do caso de estudo teve como objetivo avaliar o desempenho do modelo utilizando o conjunto original de 459 imagens, sem aplicação de técnicas de ampliação de dados. A opção por não recorrer a essa etapa permitiu observar de forma direta como o modelo se comportaria apenas com os dados disponíveis, sem influência de imagens modificadas.

Para a divisão do conjunto de dados, foi estabelecida uma repartição tradicional de 70% para treino e 15% simultaneamente para validação e teste. Essa configuração buscou assegurar que a maior parte das imagens fosse dedicada ao aprendizado do modelo, mantendo, contudo, uma parcela equilibrada destinada à verificação da sua capacidade de generalização e à avaliação final.

Na etapa de configuração do modelo, foi selecionada a opção Roboflow 3.0, com o tipo Extra Large (XL), o qual levou cerca de 4 horas para ser processado, num computador portátil com GPU (unidade de processamento gráfico) Intel(R) UHD Graphics 620. Esta versão é reconhecida por oferecer o melhor nível de precisão dentre as alternativas disponíveis. Há 5 tipos diferentes de versões, Fast, Accurate, Médio, Large e Extra Large, variando entre o tempo decorrido e a precisão dos resultados. A versão utilizada, Extra Large, embora exija maior poder computacional e, conseqüentemente, maior tempo de processamento. A escolha por esse tipo de análise teve como objetivo maximizar a qualidade das detecções, mesmo que em detrimento da rapidez de execução.

O processo de treinamento foi consideravelmente exigente em termos de recursos, com um tempo total de aproximadamente quatro horas até a disponibilização dos resultados. Esse intervalo reflete o custo computacional inerente a modelos de maior complexidade, mas também evidencia o potencial de alcançar métricas mais consistentes em termos de precisão e robustez. Os resultados das métricas dos modelos são apresentados na figura 5.



Figura 5 - Métricas para versão 1 Roboflow

Em termos de métricas, este modelo apresentou um mAP50 de 61,1%, uma revocação de 65,6% e uma precisão de 69,7%. Esses valores revelam um desempenho considerado satisfatório, sobretudo tendo em conta a limitação de não ter sido utilizada nenhuma técnica de ampliação de dados. Além disso, é importante destacar que a base de dados disponível é composta por um

número reduzido de imagens e que a área de estudo se restringe a uma via de circulação de aproximadamente 150 metros. Esse contexto reforça a relevância dos resultados obtidos, pois, mesmo com uma amostra espacial e quantitativamente limitada, o modelo conseguiu alcançar métricas consistentes. A precisão relativamente elevada de 69,7% indica que, na maioria dos casos, as predições realizadas pelo modelo corresponderam corretamente às classes identificadas, com uma baixa taxa de falsos positivos. Por outro lado, o recall de 65,6% sugere que o modelo ainda apresentou dificuldades em identificar todos os objetos existentes nas imagens, resultando em alguns falsos negativos. O valor de mAP50 de 61,1%, embora moderado, reforça que há margem de melhoria, principalmente através da introdução de técnicas de ampliação de dados ou do ajuste de parâmetros de treino.

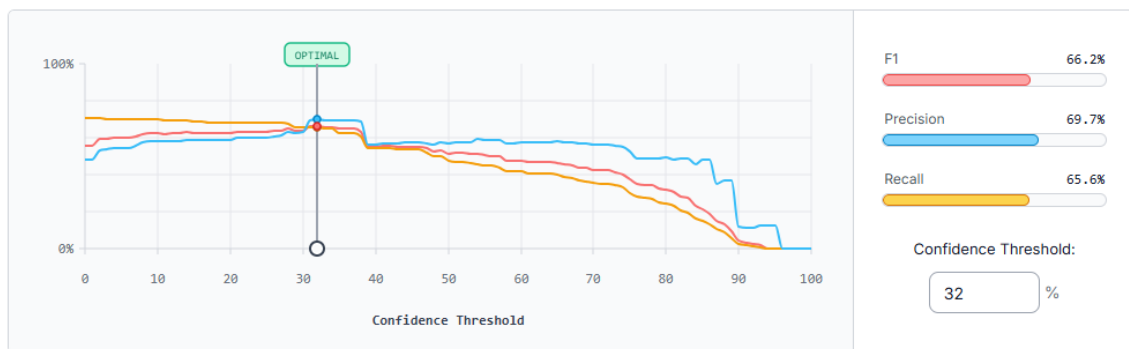


Figura 6 – Limiar de confiança para versão 1 Roboflow

A Figura 6 mostra o comportamento das métricas do modelo em função do limiar de confiança. Esse parâmetro define a probabilidade mínima que a rede precisa atribuir a uma detecção para considerá-la válida. Em outras palavras, quanto maior o limiar de confiança, mais “seguro” o modelo precisa estar de que um objeto realmente pertence à classe antes de registrá-lo como detecção. Por exemplo, se o limiar de confiança estiver em 80%, apenas as detecções com alta confiança serão aceitas, reduzindo falsos positivos, mas aumentando o risco de deixar de detectar objetos (falsos negativos).

No gráfico observa-se que o ponto ótimo de confiança foi estabelecido em 32%, ou seja, nesse valor o modelo alcança o melhor equilíbrio entre precisão e recall. Nesse ponto, o desempenho registrado foi de F1-score de 66,2%, precisão de 69,7% e recall de 65,6%. Assim, percebe-se que o modelo apresenta boa capacidade de identificar corretamente os objetos (alta precisão), mantendo também uma taxa satisfatória de detecção global (recall). Nota-se ainda que, conforme o limiar de confiança aumenta, tanto a precisão quanto o recall tendem a cair, mostrando que exigir níveis altos de confiança pode comprometer o desempenho geral do modelo.

Tabela 6 - Matriz de confusão da versão 1 Roboflow

	100	101	102	103	104	106	107	108	109	Falsos negativos
100	38	0	3	0	0	0	0	0	0	2
101	0	0	0	0	0	0	0	0	0	0
102	5	0	64	1	0	0	0	0	0	5
103	0	0	1	9	0	0	0	0	0	7
104	0	0	0	0	0	0	0	0	0	2
106	0	0	0	0	0	6	0	0	0	0
107	0	0	0	0	0	0	11	0	0	4
108	0	0	0	0	0	0	0	12	0	4
109	0	0	0	0	0	0	0	1	1	0
Falsos Positivos	3	1	9	8	1	2	1	3	0	0

A matriz de confusão apresentada na Tabela 6 permite uma análise detalhada do desempenho do modelo em cada classe. Os valores dispostos na diagonal principal representam os acertos, ou seja, os casos em que a classe prevista coincidiu com a classe real. Nota-se, por exemplo, que a classe 102 obteve 64 acertos, seguida da classe 100 com 38 acertos, evidenciando um bom desempenho nessas categorias. Por outro lado, algumas classes apresentaram resultados mais modestos, como a 103, com apenas 9 acertos, além de 7 falsos negativos e 8 falsos positivos, revelando maior dificuldade do modelo em reconhecer corretamente esses padrões.

Os valores fora da diagonal representam as confusões entre classes, quando o modelo atribuiu incorretamente um rótulo a uma instância. Esse comportamento pode ser observado, por exemplo, quando a classe 102 foi classificada 5 vezes como 100 e 1 vez como 103, ou quando a classe 109 foi classificada 1 vez como 108. Já a última coluna mostra os falsos negativos, isto é, os objetos que pertenciam à classe, mas não foram detectados pelo modelo. De forma complementar, a última linha indica os falsos positivos, quando o modelo atribuiu determinada classe de forma equivocada.

De modo geral, a matriz evidencia que o modelo apresenta bom desempenho em classes mais do conjunto de dados, mas encontra maior dificuldade em classes menos frequentes, onde o número de exemplos corretos é reduzido. Esse comportamento pode estar associado ao desequilíbrio na distribuição das imagens de treino e da base de dados, o que sugere a necessidade de alimentar a base de dados com mais imagens e anotações das classes menos representadas ou a aplicação de estratégias de balanceamento, para melhorar a capacidade de generalização do modelo.



Figura 7 - Curvas de perda da versão 1 Roboflow

A Figura 7 apresenta as curvas de perda referentes às três componentes do treinamento: *Box Loss*, *Class Loss* e *Object Loss*.

- *Box Loss*: relacionada à precisão das caixas delimitadoras (bounding boxes). Observa-se uma queda acentuada nas primeiras épocas, seguida de estabilização em torno de 1,1 após aproximadamente 100 épocas, indicando que o modelo conseguiu ajustar de forma consistente a localização dos objetos.
- *Class Loss*: avalia os erros na classificação das instâncias detetadas. Assim como no *Box Loss*, verificou-se uma redução significativa nas épocas iniciais, estabilizando-se entre 1,0 e 1,2. Essa oscilação mostra que, embora o modelo tenha aprendido a classificar, ainda apresenta variação entre classes, o que pode estar relacionado com o desequilíbrio no número de exemplos por categoria.
- *Object Loss*: mede a capacidade do modelo em identificar a presença ou ausência de objetos em uma dada região da imagem. Nota-se uma queda acentuada no início do treinamento e posterior estabilização em torno de 1,2. No entanto, nas últimas épocas há uma leve tendência de crescimento, sugerindo pequenas dificuldades na generalização da presença dos objetos, embora sem indícios críticos de sobreajuste.

De forma geral, as três curvas apresentam o comportamento esperado: queda acentuada inicial e posterior estabilização, indicando que o modelo convergiu de forma adequada. As flutuações observadas podem estar associadas à quantidade reduzida de dados de treinamento e ao desequilíbrio entre classes, fatores que limitam a estabilidade plena do processo de aprendizado.

A partir dos resultados apresentados, é possível concluir que o modelo da Versão 1 atingiu um desempenho global satisfatório, considerando as limitações da base de dados. As métricas de avaliação indicaram valores consistentes, com mAP@50 de 61,1%, precisão de 69,7% e revocação de 65,6%, refletindo um equilíbrio adequado entre a capacidade de identificar corretamente os objetos e a de evitar classificações equivocadas. O gráfico do limiar de confiança evidenciou que o ponto ótimo de operação foi alcançado em 32%, garantindo o melhor compromisso entre minimizar falsos positivos e não perder demasiadas detecções. Além disso, a análise do desempenho por épocas mostrou uma estabilização das métricas de mAP ao longo do

treinamento, bem como uma convergência adequada das curvas de perda (*Box*, *Class* e *Object Loss*), sem sinais relevantes de sobreajuste.

A matriz de confusão, por sua vez, permitiu uma análise detalhada das fragilidades do modelo em classes específicas. A classe 103 - fendas longitudinais em nível baixo) destacou-se negativamente, apresentando o maior número de falsos negativos, o que indica que é frequentemente perdida. Para melhorar esse resultado, seria necessário aumentar o número de anotações dessa classe, garantindo maior representatividade no processo de treinamento. Outro ponto crítico foi a confusão recorrente entre as classes 102 e 100 (pele de crocodilo em nível médio e pele de crocodilo em nível severo), sendo que a classe 102 foi incorretamente classificada como 100 em cinco ocasiões. Esse comportamento evidencia a semelhança entre os padrões dessas classes e reforça a necessidade de ampliar a base de dados e realizar mais anotações dessas degradações. Já a classe 104 - fendas longitudinais em nível médio apresentou baixo desempenho devido à reduzida quantidade de exemplos disponíveis, o que dificultou o aprendizado do modelo para essa classe.

É importante ressaltar que o estudo foi conduzido em um cenário restrito, com apenas 459 imagens disponíveis e cobrindo um trecho de via de aproximadamente 150 metros. Essas limitações impactaram diretamente a capacidade de generalização do modelo, especialmente em classes menos representadas. Ainda assim, os resultados obtidos podem ser considerados positivos, demonstrando que, mesmo em um contexto de dados limitados, o modelo foi capaz de identificar as degradações de forma consistente, servindo como uma base sólida para as próximas etapas de experimentação. A figura 8 apresenta um exemplo de uma das imagens obtidas na detecção das degradações com recurso a versão 1 da plataforma Roboflow.



Figura 8 - Exemplo de detecção das degradações presentes numa imagem com recurso à versão 1 – Roboflow – Roboflow 3.0

4.3.2 Roboflow –Versão 2

Na segunda análise efetuada, o objetivo manteve-se o mesmo que na versão 1: avaliar a capacidade do modelo, com a mesma base de dados. A principal alteração introduzida nesta versão foi a

escolha do tipo de modelo no Roboflow: enquanto a versão 1 utilizou o modelo padrão da plataforma Roboflow, o Roboflow 3.0, na versão 2 optou-se pelo modelo YOLOv11, reconhecido por suas melhorias em detecção rápida e maior precisão em classes com menor representatividade.

Essa mudança permitiu investigar se a utilização de um modelo mais recente e especializado poderia reduzir problemas observados na versão anterior, como falsos negativos elevados na classe 103 (fendas longitudinais em nível baixo) e confusões frequentes envolvendo a classe 102 (pele de crocodilo em nível médio). As restantes condições foram mantidas, isto é, sem ampliação de dados, com a mesma divisão do conjunto de dados (70% para treino, 15% para validação e 15% para teste). Foi ainda selecionada a opção Extra Large (XL), o qual levou cerca de 4 horas para ser processado.

Esta análise visa comparar o desempenho do YOLOv11 em relação à versão 1 (Roboflow 3.0), fornecendo informações sobre ganhos potenciais em precisão, revocação e capacidade de identificação de classes críticas, contribuindo assim para a evolução do modelo final.



Figura 9 - Métricas para versão 2 Roboflow

Na versão 2, utilizando o modelo YOLOv11 na configuração XL e com um processo de balanceamento dos dados igual ao utilizado na versão 1, com 70% de train set, 15% valid set e 15% de test set, foram obtidas métricas que mostram um desempenho consistente, ainda que com algumas diferenças em relação à versão 1. O valor de mAP@50 atingiu 59,5%, a precisão ficou em 59,8% e a revocação alcançou 68,6%. Nota-se que, embora o mAP e a precisão tenham apresentado uma leve redução quando comparados à versão 1, o aumento da revocação indica que o modelo se tornou mais eficiente em detectar as instâncias corretas, reduzindo a ocorrência de falsos negativos.

Esse comportamento sugere que o balanceamento dos dados e a maior capacidade do modelo YOLOv11 XL contribuíram para aumentar a revocação da rede, tornando-a mais abrangente na identificação das classes. Entretanto, o ganho em revocação veio acompanhado de uma diminuição da precisão, o que revela a necessidade de ajustes adicionais para reduzir confusões entre classes e minimizar falsos positivos. Assim, os resultados da versão 2 reforçam que o processo de treinamento avançou em termos de abrangência da detecção, mas ainda há espaço para melhorias na qualidade das predições.

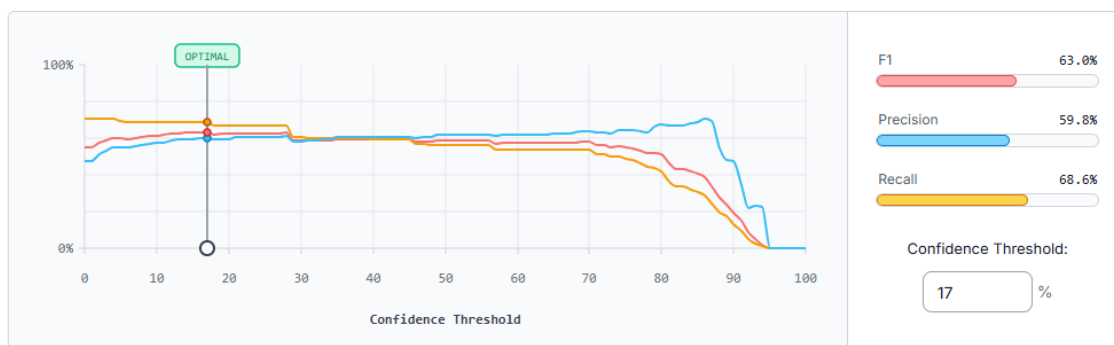


Figura 10 – Limiar de confiança para versão 2 Roboflow

A análise da curva de desempenho em função do limiar de confiança (ver Figura 10) indica que o ponto ótimo de operação do modelo foi encontrado em 17% de confiança, no qual se obteve um equilíbrio entre a precisão e a revocação. Nesse limiar, os valores estabilizaram em F1-score de 63,0%, precisão de 59,8% e revocação de 68,6%, confirmando que o modelo tende a favorecer a revocação na detecção das classes, mesmo à custa de um número maior de falsos positivos.

Conforme os valores do limiar de confiança aumentam, melhor é considerado o resultado. O valor de 17% apresentado nesta versão é considerado baixo, o que influencia negativamente em seus resultados e na precisão do modelo.

O comportamento das curvas mostra que, conforme o limiar de confiança aumenta, há uma queda mais acentuada da revocação em relação à precisão, o que reforça que este modelo beneficia de limiares mais baixos para garantir que as instâncias sejam detectadas. Isso está diretamente alinhado com os resultados observados no conjunto de métricas globais, em que a revocação se destacou em relação à precisão. Assim, a versão 2 revela que, embora haja espaço para otimização em termos de redução de falsos positivos, o modelo se mostrou eficaz em manter um nível elevado de abrangência nas detecções, mas à custa de um limiar de confiança reduzido.

Tabela 7 - Matriz de confusão da versão 2 Roboflow

	100	101	102	103	104	106	107	108	109	Falsos negativos
100	38	0	4	0	0	0	0	0	0	1
101	0	0	0	0	0	0	0	0	0	0
102	4	0	68	1	0	0	0	0	0	3
103	0	0	0	9	0	0	0	0	0	8
104	0	0	0	0	0	0	0	0	0	2
106	0	0	0	0	0	6	0	0	0	0
107	0	0	0	0	0	0	11	0	0	4
108	0	0	0	0	0	0	0	15	1	0
109	0	0	0	0	0	0	0	1	1	0
Falsos Positivos	8	3	9	7	1	3	1	3	2	0

A matriz de confusão da versão 2, apresentada na Tabela 7, permite observar em detalhe o comportamento do modelo frente às diferentes classes. Nota-se que as classes 100 - pele de crocodilo em nível severo (38 acertos), 102 - pele de crocodilo em nível médio (68 acertos) e 108 - fendas transversais em nível baixo (15 acertos) apresentaram os melhores desempenhos, refletindo maior consistência na detecção. Entretanto, algumas classes ainda apresentaram elevado número de erros, como a 103 - fendas longitudinais em nível baixo, que apesar de ter registrado 9 acertos, acumulou 8 falsos negativos, evidenciando que continua sendo uma das classes mais problemáticas.

Além disso, observa-se que ocorreram confusões relevantes entre classes, como no caso da classe 106 - covas em nível baixo, que teve 6 detecções corretas, mas apresentou erros que a confundiram com outras classes, e a classe 107 - covas em nível médio, que contou com 11 acertos, mas também 4 falsos negativos. Os falsos positivos se distribuíram de forma relativamente uniforme entre as várias classes, indicando que a queda de precisão já observada nas métricas globais também se reflete diretamente na matriz.

De forma geral, a matriz confirma o diagnóstico obtido a partir das métricas: o modelo se mostra abrangente na detecção, ou seja, alta revocação, mas ainda sofre com confusões entre classes específicas e perda de precisão. Esses resultados reforçam a necessidade de ampliar o número de imagens e anotações, especialmente para classes com maior índice de falsos negativos, como a 103, além de adotar estratégias de refinamento que possam reduzir os falsos positivos.

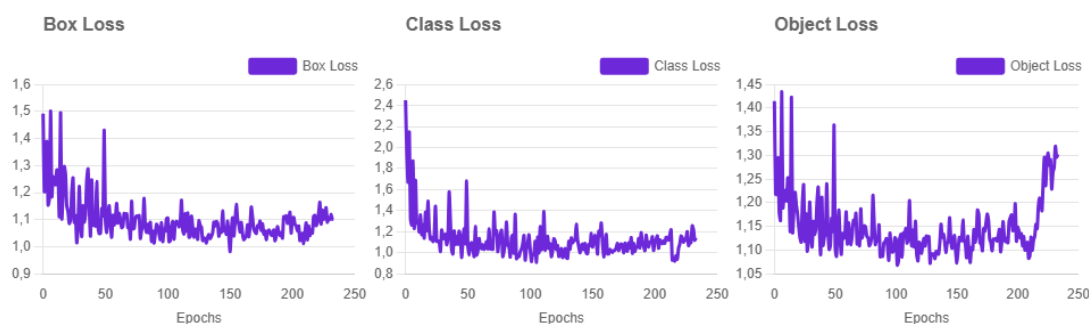


Figura 11 - Curvas de perdas da versão 2 Roboflow

A avaliação das curvas de perda durante o treinamento da versão 2 mostra um comportamento estável após as primeiras iterações, ainda que com oscilações relevantes em determinados pontos (ver Figura 11). O *Box Loss* apresentou queda rápida nas primeiras épocas, estabilizando em torno de 1,1 após a metade do treinamento, o que indica que o modelo conseguiu aprender a ajustar razoavelmente bem as delimitações das caixas de detecção.

Já o *Class Loss* demonstrou uma redução consistente, alcançando valores próximos de 1,0, o que reflete uma boa capacidade do modelo em distinguir as classes, ainda que persistam confusões

em categorias específicas, como evidenciado na matriz de confusão. O *Object Loss*, por sua vez, apresentou maior instabilidade, principalmente nas últimas épocas, chegando a crescer novamente após a época 200. Esse comportamento pode estar associado à dificuldade do modelo em diferenciar entre regiões contendo objetos e o fundo da imagem, o que ajuda a explicar a queda da precisão observada nas métricas globais.

De forma geral, os gráficos de loss confirmam que o modelo foi capaz de aprender os padrões do conjunto de treino, mas também revelam limitações na convergência final, especialmente no reconhecimento dos objetos em si.

A versão 2, treinada com o modelo YOLOv11 XL e com balanceamento dos dados, apresentou resultados consistentes, mas com características distintas da versão 1. As métricas globais registraram um mAP@50 de 59,5%, precisão de 59,8%, revogação de 68,6% e F1-score de 63%, evidenciando uma maior sensibilidade do modelo, ainda que à custa de um aumento dos falsos positivos. A matriz de confusão confirmou esse comportamento: classes como 100, 102 e 108 tiveram bom desempenho, mas outras, como a 103, ainda concentraram elevado número de falsos negativos, mostrando que a distribuição desigual das anotações segue impactando o desempenho do modelo. Já a análise das curvas de perda revelou que o modelo conseguiu estabilizar o Box Loss e o Class Loss, mas manteve instabilidades no Object Loss, sobretudo nas últimas épocas, o que ajuda a explicar as dificuldades em reduzir falsos positivos. Assim, a versão 2 se destacou pela maior abrangência na detecção e pelo avanço na revocação, mas também expôs a necessidade de estratégias adicionais para refinar a precisão e melhorar a diferenciação entre classes próximas. A Figura 12 apresenta um exemplo de degradações detetadas em uma imagem, com recurso à versão 2 da plataforma Roboflow.



Figura 12 - Exemplo de detecção das degradações presentes numa imagem com recurso à versão 2 – Roboflow – YOLOv11

4.3.3 Roboflow – Versão 3

Na versão 3 do estudo, manteve-se o mesmo conjunto de dados utilizado nas versões anteriores, porém com a aplicação de técnicas de ampliação de dados, tendo como objetivo analisar o impacto dessa estratégia sobre o desempenho do modelo Roboflow 3.0. Para esta etapa, foi empregue a configuração Roboflow 3.0 XL, a mesma adotada na versão 1, mas com uma base enriquecida por transformações artificiais. As operações de ampliação de dados realizadas incluíram espelhamento horizontal, recorte com variação de zoom entre 0% e 10% e rotações entre -15° e $+15^\circ$, de modo a gerar maior diversidade de amostras durante o processo de treino. Após a aplicação destas técnicas de ampliação de dados, o total de imagens passou a ser de 781, distribuídas segundo um rebalanceamento automático de 62% para treino e 19% para validação e teste. O processo de treinamento desta versão teve duração aproximada de 6 horas.

Essa versão não teve como propósito substituir os dados originais, mas sim investigar se a introdução de variações artificiais poderia influenciar a capacidade de generalização do modelo. Assim, a versão 3 constitui um passo importante para verificar os efeitos da ampliação de dados na tarefa de detecção, reconhecendo desde o início que os resultados podem não necessariamente superar aqueles obtidos sem o uso dessa técnica, mas ainda assim fornecem subsídios relevantes para compreender o comportamento do modelo em diferentes condições de treinamento.



Figura 13 - Métricas para versão 3 Roboflow

Como pode ser observado na Figura 13, o desempenho obtido na versão 3 apresentou um mAP@50 de 59,2%, indicando que, em média, o modelo detectou corretamente cerca de 60% dos objetos, considerando a sobreposição adequada entre a detecção e o objeto real. A precisão foi de 54,2%, mostrando que pouco mais da metade das detecções realizadas eram corretas, enquanto o recall de 67,4% revela que a maioria dos objetos reais foi identificada, embora ainda houvesse perdas. Esses resultados sugerem que o modelo tende a gerar muitos falsos positivos, refletindo um comportamento de alta revocação, mas menor precisão. Apesar do aumento da base de dados, os indicadores não superaram a versão anterior, o que evidencia que a ampliação das imagens nem sempre garante uma melhoria significativa na precisão, especialmente diante do número total limitado de imagens e da presença de classes sub-representadas.

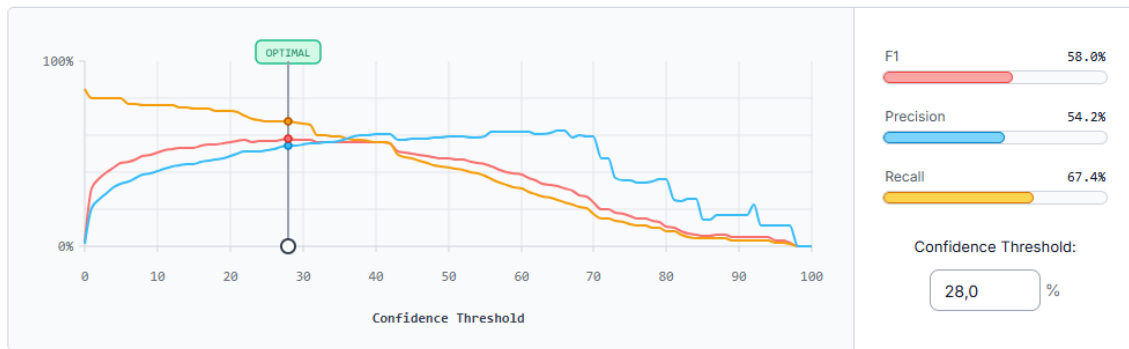


Figura 14 – Limiar de confiança da versão 3 Roboflow

A Figura 14 apresenta o gráfico do limiar de confiança referente à versão 3 do modelo treinado no Roboflow. Observa-se que o limiar ótimo de confiança foi definido em 28%, valor relativamente baixo, o que indica que o modelo necessita de um nível reduzido de confiança para alcançar o melhor equilíbrio entre precisão e revocação. Isso sugere que as previsões realizadas não apresentam alta consistência, uma vez que seria esperado um limiar mais elevado em modelos com maior robustez.

Quanto aos resultados quantitativos, o modelo apresentou um F1-score de 58,0%, refletindo um desempenho moderado no equilíbrio entre precisão e revocação. A precisão atingiu 54,2%, evidenciando a ocorrência de falsos positivos em número significativo, ou seja, o modelo por vezes classifica incorretamente objetos como pertencentes a uma determinada classe. Por outro lado, a revocação foi de 67,4%, indicando que o modelo conseguiu identificar a maioria das instâncias presentes, ainda que com erros de classificação associados. Essa diferença entre precisão e revocação mostra que o modelo tem maior capacidade de detectar objetos, mas ainda carece de maior seletividade para reduzir classificações equivocadas.

Tabela 8 – Matriz de confusão da versão 3 Roboflow

	100	101	102	103	104	106	107	108	109	Falsos negativos
100	49	0	22	0	0	0	0	0	0	15
101	0	2	0	0	0	0	0	0	0	0
102	3	0	144	0	0	0	0	0	0	23
103	0	0	0	22	3	0	0	0	0	13
104	0	0	0	0	9	0	0	0	0	2
106	0	0	0	0	0	9	1	0	0	7
107	0	0	1	0	0	0	24	0	0	6
108	0	0	0	0	0	0	0	27	0	2
109	0	0	0	0	0	0	0	6	0	0
Falsos Positivos	24	3	78	19	11	5	1	6	6	0

A Tabela 8 apresenta a matriz de confusão da versão 3 do modelo, permitindo uma análise mais detalhada do desempenho por classe. Observa-se que algumas classes tiveram desempenho mais consistente, como a classe 102 - pele de crocodilo em nível médio, que registrou 144 acertos, mas também apresentou 23 falsos negativos e 78 falsos positivos, indicando que, apesar de ser a mais bem representada, ainda sofre com erros relevantes de classificação. A classe 100 - pele de crocodilo em nível severo também obteve um número expressivo de acertos (49), porém com 15 falsos negativos e 24 falsos positivos, revelando dificuldades na diferenciação frente às demais categorias.

Em contrapartida, algumas classes apresentaram baixo número de amostras corretamente identificadas, o que evidencia limitações importantes do modelo. A classe 103 - fendas longitudinais em nível baixo, por exemplo, foi corretamente detectada em 22 casos, mas acumulou 13 falsos negativos e 19 falsos positivos, sendo uma das mais problemáticas do conjunto. Já classes como 106 - covas em nível baixo e 107 - covas em nível baixo tiveram desempenho modesto, com apenas 9 e 6 acertos, respectivamente, acompanhados de falsos negativos e positivos que impactam negativamente sua confiabilidade.

De modo geral, a matriz de confusão confirma a tendência já observada nos indicadores globais: o modelo apresenta uma revocação relativamente elevada, conseguindo identificar várias instâncias presentes, mas ao custo de gerar classificações equivocadas, sobretudo nas classes 100, 102 e 103, que concentram os maiores índices de erro. Isso reforça a interpretação de que o aumento da base de dados nesta versão não resultou em ganhos substanciais de robustez, uma vez que os erros de classificação permanecem significativos.



Figura 15 - Curvas de perda da versão 3 Roboflow

A Figura 15 apresenta as curvas de perda obtidas durante o treinamento da versão 3 com o modelo Roboflow 3.0, correspondentes a *Box Loss*, *Class Loss* e *Object Loss* ao longo de 100 épocas. Nota-se que as três curvas exibem grande oscilação ao longo do processo, sem um padrão de convergência suave. Esse comportamento indica que o modelo apresentou dificuldades em estabilizar o aprendizado, mesmo após um número elevado de épocas.

No caso da *Box Loss*, observa-se variação frequente entre valores próximos de 1,0 e picos acima de 1,8, sugerindo inconsistência na capacidade de ajustar corretamente as coordenadas das caixas delimitadoras (*bounding boxes*). A *Class Loss*, por sua vez, também se manteve instável, oscilando em torno de 1,2 a 1,7, o que evidencia dificuldade do modelo em diferenciar adequadamente as classes do conjunto de dados. A *Object Loss* apresentou comportamento semelhante, variando em torno de 1,3 a 1,6, indicando que o modelo teve problemas em distinguir com clareza a presença ou ausência de objetos nas imagens.

Esse padrão de instabilidade nas curvas sugere que, apesar do aumento da base de dados por meio de técnicas de ampliação de dados, o modelo não alcançou uma melhora significativa em termos de robustez e generalização. Pelo contrário, a falta de suavidade nas curvas reforça que o treinamento apresentou dificuldades de convergência, o que pode estar associado tanto à complexidade das classes quanto ao desequilíbrio e limitações da base de dados utilizada.

Ao concluir a análise da versão 3, observa-se que o modelo ainda apresenta limitações importantes, especialmente no reconhecimento de algumas classes específicas. A classe 102 - pele de crocodilo em nível médio foi frequentemente perdida, mesmo sendo uma das mais representadas no conjunto de dados, o que indica a necessidade de ampliar significativamente o número de amostras anotadas dessa categoria. Além disso, a classe 100 - Pele de crocodilo em nível severo apresentou confusões recorrentes com a classe 102, demonstrando que o modelo encontra dificuldades em distinguir ambas, possivelmente pela semelhança visual entre elas ou pela carência de exemplos suficientemente diversificados que permitam melhorar a diferenciação. Outro aspecto crítico é a classe 101 - pele de crocodilo em nível baixo, que possui apenas duas instâncias em todo o conjunto (representando cerca de 0,5% do total). Essa sub-representação severa compromete de forma direta o aprendizado do modelo, que não dispõe de exemplos suficientes para generalizar padrões e realizar previsões consistentes. Essa discrepância entre classes, com algumas super-representadas e outras praticamente ausentes, gera um desequilíbrio estrutural nos dados, afetando negativamente tanto a precisão quanto a revocação.

De modo geral, a versão 3 evidencia que o simples aumento da base de dados por meio da ampliação de dados não foi suficiente para corrigir os problemas fundamentais do conjunto de dados. Para melhorar o desempenho em versões futuras, será imprescindível coletar e anotar um número maior de instâncias, em especial das classes mais críticas, a fim de reduzir o desequilíbrio e possibilitar um treinamento mais robusto e consistente.

A Figura 16 apresenta degradações no pavimento detetadas com recurso à versão 3 do programa Roboflow.



Figura 16 - Exemplo de detecção das degradações presentes numa imagem com recurso à versão 3 – Roboflow – Roboflow 3.0 com base de dados ampliada

4.3.4 Roboflow – Versão 4

A quarta versão do estudo foi concebida a partir das mesmas bases estabelecidas na versão 3, mantendo o objetivo de ampliar a diversidade do conjunto de imagens e, consequentemente, aumentar a capacidade de generalização do modelo. Para isso, foram novamente aplicadas técnicas de ampliação de dados, detalhadas da seguinte forma: espelhamento horizontal, que permite simular a ocorrência de objetos em orientações invertidas; recorte, com variação de 0% de zoom mínimo e 10% de zoom máximo, que introduz alterações subtis no enquadramento e escala dos objetos; e rotações entre -15° e $+15^\circ$, cuja finalidade é expor o modelo a variações angulares que podem ocorrer em situações reais. A utilização combinada dessas técnicas possibilita que o modelo seja treinado com imagens mais variadas, reduzindo a probabilidade de sobreajuste (*overfitting*) e preparando-o para lidar com cenários menos controlados.

Assim como na versão 3, manteve-se o balanceamento da base de dados com 62% das imagens destinadas ao treinamento, 19% para validação e 19% para teste. Essa proporção foi escolhida de modo a garantir consistência metodológica entre as diferentes versões do experimento, permitindo que as comparações de desempenho sejam feitas de forma controlada.

A principal modificação introduzida nesta versão está relacionada ao tipo de modelo adotado. Enquanto a versão 3 foi executada com o modelo Roboflow 3.0, a versão 4 foi treinada no YOLOv11, na configuração XL, que representa uma arquitetura mais robusta e com maior capacidade de aprendizado. Essa escolha teve como objetivo avaliar se a utilização de uma rede mais recente e poderosa poderia proporcionar ganhos significativos de desempenho em relação às versões anteriores, mesmo mantendo as estratégias de ampliação de dados. O processo de treinamento demandou aproximadamente seis horas.

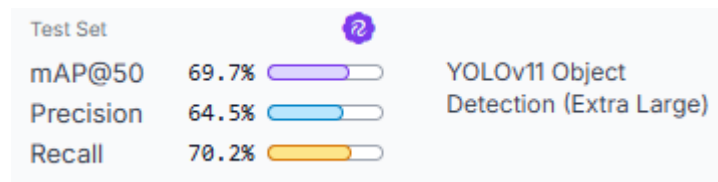


Figura 17 - Métricas para versão 4 Roboflow

Na versão 4, utilizando o modelo YOLOv11 XL, os resultados demonstraram um desempenho global superior em relação à versão 3. O valor de mAP@50 atingiu 69,7%, representando um aumento significativo na métrica de avaliação principal. Este resultado indica que o modelo conseguiu melhorar a capacidade de identificar corretamente os objetos no conjunto de teste, aproximando-se de valores considerados satisfatórios em estudos de detecção automática.

A precisão alcançou 64,5%, o que revela uma redução no número de falsos positivos relativamente às versões anteriores (2 e 3), sugerindo maior seletividade do modelo na classificação. Paralelamente, a revocação atingiu 70,2%, demonstrando uma boa capacidade de identificar a maioria das instâncias presentes nas imagens, ainda que com algumas perdas associadas. A combinação destes dois indicadores reflete um modelo mais equilibrado, capaz de aumentar a taxa de detecções corretas sem comprometer excessivamente a seletividade.

Estes resultados evidenciam que a utilização do YOLOv11 na versão XL trouxe ganhos de desempenho em relação à versão 3, sobretudo pela maior robustez da arquitetura, que se mostrou mais eficiente na exploração da base de dados aumentada. Contudo, ainda se observam margens para melhoria, uma vez que tanto a precisão como a revocação permanecem abaixo de valores ideais para aplicações em cenários reais mais exigentes.

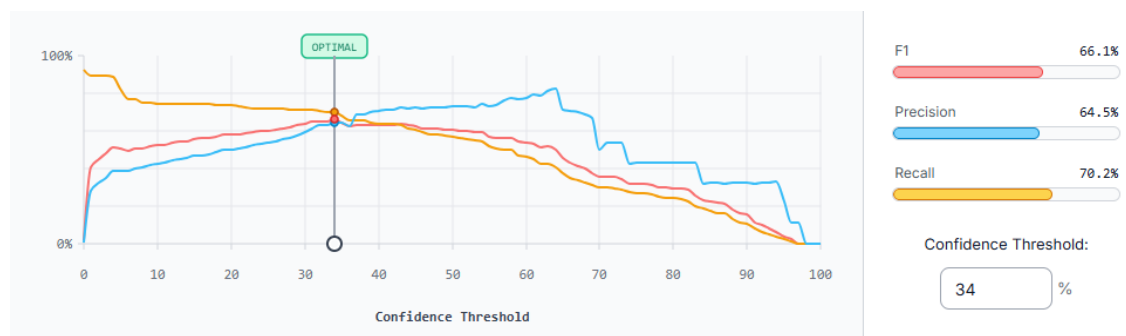


Figura 18 – Limiar de confiança da versão 4 Roboflow

A Figura 18 apresenta o gráfico do confidence threshold relativo à versão 4. O ponto ótimo foi identificado em 34%, valor mais elevado do que o registrado na versão 3 (28%), sugerindo que nesta versão o modelo apresenta maior confiança nas suas predições. Este aumento do limiar ótimo é um indicador positivo, uma vez que evidencia maior consistência nas classificações, reduzindo a necessidade de aceitar predições com baixos níveis de confiança.

Em termos de métricas associadas, o modelo alcançou um F1-score de 66,1%, superior ao obtido na versão 3 (58%), confirmando um melhor equilíbrio entre precisão e revocação. A precisão manteve-se em 64,5%, revelando uma redução moderada de falsos positivos, enquanto a revocação chegou a 70,2%, o que indica que o modelo conseguiu identificar uma proporção mais elevada das instâncias presentes no conjunto de teste.

A análise conjunta destes resultados mostra que o YOLOv11 XL não só conseguiu manter uma boa capacidade de detecção (revocação), como também aumentou a seletividade das suas predições, refletindo-se num valor mais robusto de F1-score. Estes ganhos sugerem que a adoção desta arquitetura mais recente contribuiu para uma melhoria efetiva do desempenho face às versões anteriores.

Tabela 9 - Matriz de confusão da versão 4 Roboflow

	100	101	102	103	104	106	107	108	109	Falsos negativos
100	81	0	3	0	0	0	0	0	0	2
101	0	0	0	0	0	0	0	0	0	2
102	5	0	156	1	1	0	0	0	0	7
103	0	0	1	23	3	0	0	0	0	11
104	0	0	0	0	9	0	0	0	0	2
106	0	0	0	0	0	15	1	0	0	1
107	0	0	1	0	0	0	25	0	0	5
108	0	0	0	0	0	0	0	33	0	6
109	0	0	0	0	0	0	0	3	3	0
Falsos Positivos	17	3	15	17	10	4	2	8	1	0

A Tabela 9 apresenta a matriz de confusão correspondente à versão 4, permitindo avaliar em detalhe o desempenho por classe. De forma geral, observa-se uma melhoria em relação à versão 3, com um aumento considerável no número de predições corretas em várias categorias. A classe 102 - pele de crocodilo em nível médio, que anteriormente registava perdas significativas, alcançou 156 acertos, ainda que tenha apresentado 7 falsos negativos e 15 falsos positivos, revelando que, embora tenha havido progresso, continua a ser uma das categorias mais desafiantes para o modelo.

A classe 100 - pele de crocodilo em nível severo também demonstrou evolução, com 81 classificações corretas, contra 49 na versão 3, e apenas 2 falsos negativos, embora ainda apresente 17 falsos positivos, confirmando a persistência de alguma confusão com outras classes. A classe 103 - fendas longitudinais em nível baixo, por sua vez, manteve um desempenho modesto, com 23 acertos, mas acumulou 11 falsos negativos e 17 falsos positivos, permanecendo como uma das classes mais problemáticas, tal como observado em versões anteriores.

Outras classes, como a 104 - fendas longitudinais em nível médio, 106 - covas em nível baixo e 107 - covas em nível médio, tiveram desempenho mais equilibrado, ainda que limitado pelo reduzido número de amostras disponíveis. Destaca-se também a classe 108 - fendas transversais em nível baixo, que atingiu 33 predições corretas, mas ainda apresentou 8 falsos positivos e 6 falsos negativos, evidenciando alguma inconsistência no seu reconhecimento. A classe 101 - pele de crocodilo em nível baixo, devido à sua sub-representação (apenas duas instâncias no total do conjunto), não obteve qualquer detecção correta, reforçando a ideia de que o desequilíbrio de dados compromete de forma significativa a capacidade de aprendizagem do modelo.

De forma global, a matriz de confusão da versão 4 confirma os ganhos já observados nas métricas globais, nomeadamente na redução de erros em classes com maior representatividade, como a 100 e a 102. No entanto, o desempenho continua a ser limitado pelo desequilíbrio do conjunto de dados, especialmente nas classes minoritárias, como a 101, que carecem de exemplos suficientes para treinar adequadamente o modelo. Assim, embora o YOLOv11 XL tenha contribuído para uma melhoria significativa, o problema estrutural da base de dados permanece como principal fator limitante.

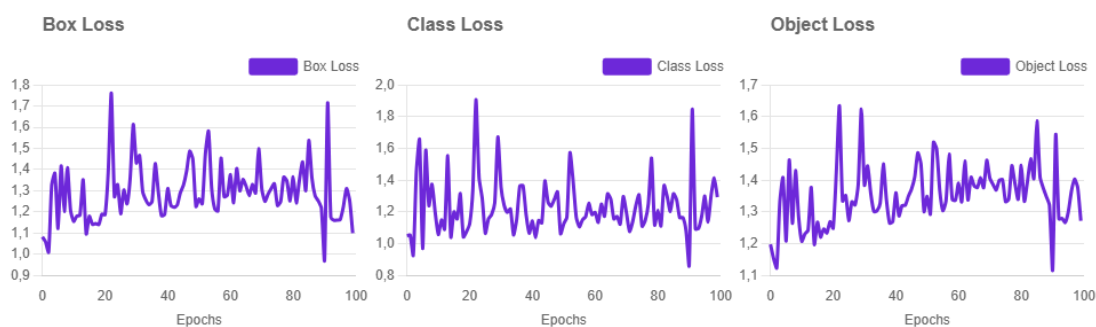


Figura 19 - Curvas de perdas da versão 4 Roboflow

Na Figura 19 apresentam-se as curvas de perda da versão 4, correspondentes às métricas *Box Loss*, *Class Loss* e *Object Loss*, ao longo das 100 épocas de treino. De forma geral, observa-se que as três curvas apresentam oscilações regulares, sem tendência clara de convergência para valores próximos de zero.

A *Box Loss*, associada à precisão na localização das caixas delimitadoras, variou entre aproximadamente 1,0 e 1,7, estabilizando maioritariamente em torno de 1,3. Apesar de uma descida inicial, verificou-se a manutenção de flutuações significativas ao longo de todo o processo, o que indica que o modelo teve dificuldade em refinar consistentemente os limites espaciais das detecções.

No caso da *Class Loss*, que representa os erros na classificação das instâncias detectadas, registaram-se valores entre 0,9 e 1,8, com um valor médio próximo de 1,4. Esta oscilação evidencia que, mesmo após várias épocas de treino, o modelo não conseguiu reduzir de forma estável os erros de atribuição de classe, o que se reflete diretamente nos problemas identificados em classes

minoritárias, como a 101 - pele de crocodilo em nível baixo e a 103 - fendas longitudinais em nível baixo.

Por fim, a *Object Loss*, relacionada com a capacidade de distinguir objetos em relação ao fundo da imagem, apresentou valores próximos da *Box Loss*, oscilando entre 1,1 e 1,6, estabilizando em torno de 1,3. Este comportamento sugere que o modelo manteve um nível de incerteza relativamente elevado na diferenciação entre regiões com e sem objetos, o que contribui para o número de falsos positivos registados em várias classes.

O desempenho obtido na versão 4, embora superior às versões iniciais, continua a evidenciar limitações relevantes. A análise da matriz de confusão confirma que a classe 103 permanece frequentemente perdida, o que pode estar associado ao facto de corresponder a fendas de menor dimensão e, conseqüentemente, mais difíceis de detetar pelo modelo nas imagens. Esta dificuldade reflete-se tanto nos falsos negativos elevados como nas incertezas registadas ao longo do treino, visíveis nas métricas de perda.

A classe 102 - pele de crocodilo em nível médio manteve-se como uma das mais problemáticas, apresentando confusões recorrentes com a classe 100 - pele de crocodilo em nível severo. Esta situação indica que as duas categorias possuem semelhanças visuais consideráveis, o que reforça a necessidade de ampliar o número de exemplos de cada uma, permitindo ao modelo distinguir melhor as diferenças entre ambas. Por sua vez, a classe 101 - pele de crocodilo em nível baixo continua a necessitar de um número mais robusto de exemplos no conjunto de treino, uma vez que a sua representação atual é insuficiente e compromete a capacidade do modelo em aprender padrões consistentes para esta categoria.

De forma global, os resultados obtidos com o YOLOv11 XL, após seis horas de treino, mostram ganhos em métricas como a revocação, mas deixam claro que o desequilíbrio das classes e a variabilidade limitada da base de dados ainda condicionam a performance final. Assim, o progresso desta versão reforça a importância de estratégias direcionadas para o aumento e diversificação da base de dados, sobretudo para as classes minoritárias ou mais complexas, de forma a reduzir as confusões observadas e melhorar a generalização do modelo em cenários reais.

A Figura 20 apresenta um exemplo de uma das deteções realizadas com recurso à versão 4.

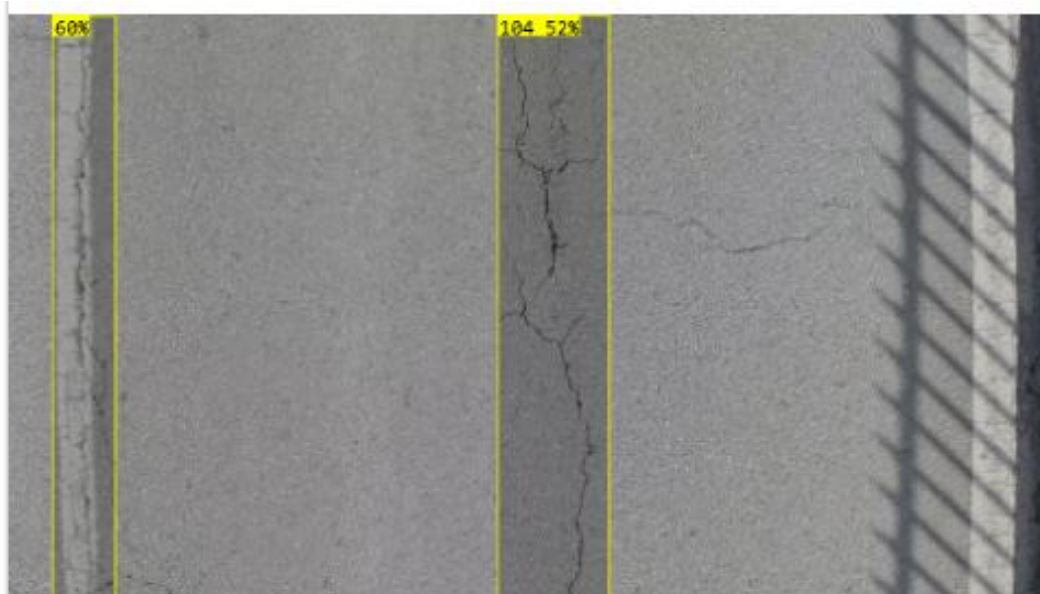


Figura 20 - Exemplo de detecção das degradações presentes numa imagem com recurso à versão 4 – Roboflow – YOLOv11 com base de dados ampliada

4.4 Testes no modelo YOLO

Após a conclusão da etapa experimental realizada na plataforma Roboflow, tornou-se necessário avançar para os testes diretamente no modelo YOLO, de forma a obter uma avaliação mais rigorosa e completa do desempenho do algoritmo em ambiente local. Para este fim, foi utilizada uma licença académica disponibilizada pela *Ultralytics*, empresa responsável pelo desenvolvimento do YOLO, especificamente direcionada a estudantes. Esta licença possibilitou o acesso integral às funcionalidades do algoritmo, permitindo conduzir os ensaios de forma autónoma e sem as limitações de uma plataforma online.

O algoritmo selecionado para esta fase do estudo foi o YOLOv8, uma das versões mais recentes e robustas, amplamente reconhecida pela sua eficiência na deteção de objetos em diferentes contextos. Para a realização dos testes foram definidos dois cenários: o primeiro recorrendo à base de dados original, sem qualquer aplicação de técnicas adicionais de pré-processamento das imagens, e o segundo integrando técnicas de ampliação de dados, a fim de avaliar o impacto direto dessa estratégia na melhoria dos indicadores de desempenho.

Para garantir a consistência e compatibilidade entre as diferentes etapas do estudo, foi realizado o *download* da base de dados anotada preparada no Roboflow, no formato “type YOLOv11”, que assegura a plena integração com a versão do algoritmo utilizada. Assim, manteve-se a coerência entre os conjuntos de dados explorados anteriormente na plataforma Roboflow e aqueles aplicados nos testes em ambiente local no YOLOv8, o que facilita a comparação entre os resultados obtidos nos dois contextos.

É igualmente importante salientar que a utilização do YOLO nesta etapa não se processa através de uma interface gráfica intuitiva, como sucede no Roboflow. O YOLO não é uma aplicação de uso imediato, mas sim uma ferramenta que exige a execução em ambiente de programação não-visual.

No presente estudo, a execução foi feita por meio do Anaconda PowerShell, com a introdução de comandos e linhas de código específicas para treinar, validar e testar os modelos. Este aspeto confere um carácter mais técnico e detalhado ao processo, exigindo maior domínio das ferramentas de programação e do funcionamento interno do algoritmo.

4.4.1 YOLO – Versão 1

A versão 1 de testes no YOLOv8 foi realizada com a base de dados original, sem aplicação de técnicas de ampliação de dados. Para este fim, o conjunto de imagens anotadas foi descarregado no formato type YOLOv11, de forma a garantir a compatibilidade com o algoritmo e assegurar a correta leitura das anotações durante o processo de treino.

O treino desta versão exigiu um tempo total de execução de aproximadamente 20 horas e 20 minutos, evidenciando o elevado custo computacional associado ao processo, sobretudo quando se trata de modelos complexos e de grande escala como o YOLOv8. Esta duração prolongada ilustra a exigência técnica da tarefa e a necessidade de recursos computacionais adequados para suportar experiências de deteção de objetos em contextos de investigação.

Esta versão assume um papel essencial no estudo, uma vez que estabelece a linha de referência inicial do desempenho do YOLOv8 quando aplicado diretamente à base de dados tal como foi construída. Os resultados obtidos permitirão compreender de que forma o modelo responde ao conjunto de imagens sem qualquer enriquecimento adicional, servindo de base para posteriores comparações com versões em que o aumento de dados será introduzido.

```
100 epochs completed in 20.206 hours.
Optimizer stripped from C:\Users\User\runs\detect\train15\weights\last.pt, 6.3MB
Optimizer stripped from C:\Users\User\runs\detect\train15\weights\best.pt, 6.3MB

Validating C:\Users\User\runs\detect\train15\weights\best.pt...
Ultralytics 8.3.203 Python-3.10.18 torch-2.8.0+cpu CPU (Intel Core i5-8265U 1.60GHz)
Model summary (fused): 72 layers, 3,007,403 parameters, 0 gradients, 8.1 GFLOPs

```

Class	Images	Instances	Box(P	R	mAP50	mAP50-95)	100%	3/3	0.2it/s	13.0s
all	69	171	0.44	0.555	0.546	0.325				
100	38	43	0.768	0.977	0.965	0.718				
101	1	1	0	0	0.0199	0.00721				
102	55	70	0.518	0.9	0.814	0.601				
103	14	17	0.466	0.412	0.49	0.206				
104	7	7	0.101	0.714	0.272	0.106				
106	4	4	0.518	0.5	0.736	0.385				
107	7	7	0.536	0.857	0.858	0.43				
108	11	11	0.52	0.545	0.545	0.321				
109	10	11	0.53	0.0909	0.217	0.149				

```
Speed: 3.0ms preprocess, 157.2ms inference, 0.0ms loss, 2.7ms postprocess per image
```

Figura 21 - Resultados da versão 1 YOLO

Os resultados mostram um mAP50 global de 54,6% e um mAP50-95 de 32,5%, uma precisão de 44% e revocação de 55,5%, evidenciando um desempenho razoável do modelo, considerando a complexidade do problema e a limitação do conjunto de dados.

Ao efetuar uma análise por classe, observou-se que a classe 100 - pele de crocodilo em nível severo obteve o melhor desempenho, alcançando um mAP50 de 96,5%, demonstrando que o modelo consegue detetá-la com elevada precisão. Por outro lado, a classe 103 - fendas longitudinais em nível baixo apresentou os valores mais baixos (mAP50 de 19,9%), o que confirma a sua maior dificuldade de detecção. Já as classes 101 - pele de crocodilo em nível baixo e 102 - pele de crocodilo em nível médio situaram-se num nível intermédio, com desempenhos aceitáveis, mas ainda aquém da classe 100.

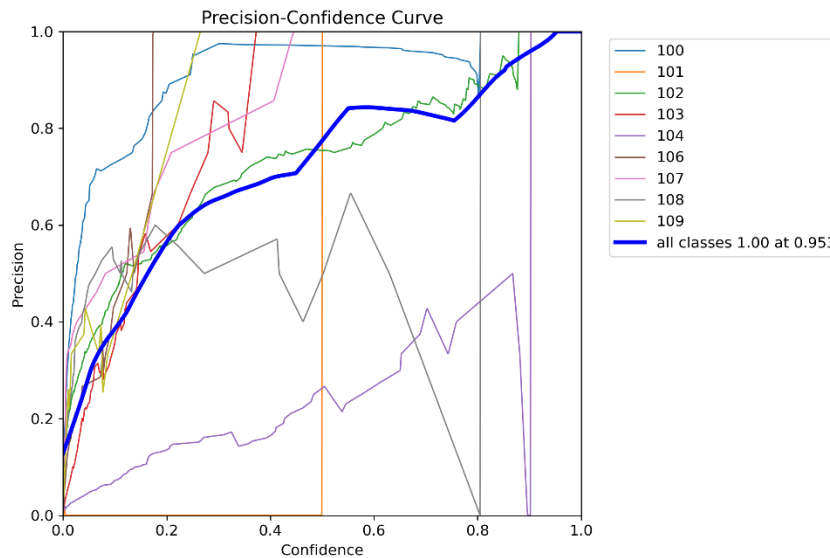


Figura 22 - Curva de precisão-confiança para o teste versão 1 YOLO

A curva Precisão-Confiança revela que, de forma geral, a precisão do modelo aumenta à medida que o nível de confiança cresce. Considerando todas as classes, observa-se que a precisão global alcança 1.00 quando o limiar de confiança está próximo de 0,953, indicando que, em previsões altamente confiantes, os erros são praticamente eliminados.

Ao analisar individualmente, a classe 100 - pele de crocodilo em nível severo mantém um desempenho muito consistente em todos os níveis de confiança, confirmando a sua robustez. Já classes como a 102 - pele de crocodilo em nível médio e a 107 - covas em nível médio também evidenciam bom comportamento, mostrando estabilidade e tendência a altos valores de precisão com limites de confiança mais elevados.

Em contrapartida, classes como a 104 - fendas longitudinais em nível médio e a 108 - fendas transversais em nível baixo, revelam maior oscilação, com quedas significativas na precisão mesmo em níveis de confiança elevados, sugerindo dificuldades do modelo em distinguir corretamente estas categorias.

No geral, o gráfico da Figura 22 confirma que o ajuste de um limiar de confiança mais elevado pode ser útil para reduzir falsos positivos, embora isso possa implicar a perda de algumas deteções de classes mais frágeis.

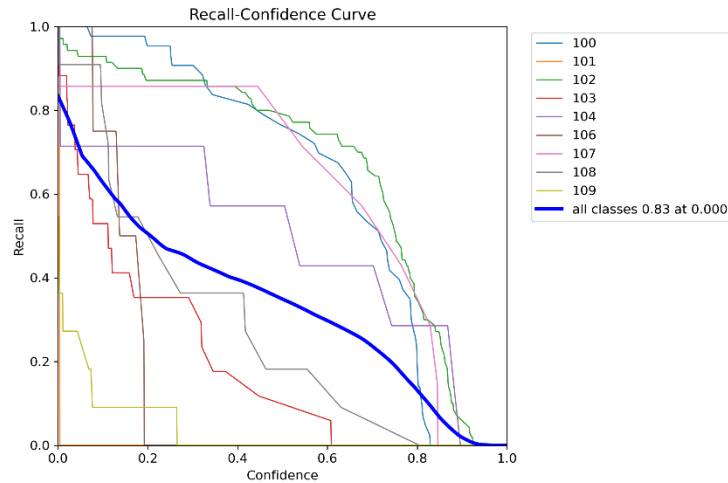


Figura 23 - Curva de Revocação-Confiança para o teste versão 1 YOLO

A curva Revocação-Confiança evidencia que o modelo alcança um valor médio de revocação global de 0,83 quando considerado todos os limites de confiança. Este resultado indica que o modelo consegue detetar a maior parte dos objetos presentes nas imagens, embora a sua eficácia varie bastante entre classes.

De forma particular, observa-se que classes como a 100 - pele de crocodilo em nível severo, 102 - pele de crocodilo em nível médio e 107 - covas em nível médio mantêm valores de revocação elevados ao longo de diferentes níveis de confiança, o que mostra que estas categorias são detetadas de forma relativamente consistente. Por outro lado, classes como a 103 - fendas longitudinais em nível baixo, 108 - fendas transversais em nível baixo e sobretudo a 109 - fendas transversais em nível médio apresentam desempenhos muito inferiores, com quebras abruptas de revocação à medida que o nível de confiança aumenta, refletindo uma maior dificuldade do modelo em reconhecer estas classes.

Em termos práticos, o gráfico demonstra que reduzir o limite de confiança pode aumentar a revocação global permitindo capturar mais objetos, mas à custa de uma maior probabilidade de falsos positivos. Já limites mais elevados reduzem drasticamente a revocação em classes frágeis, limitando o número de deteções.

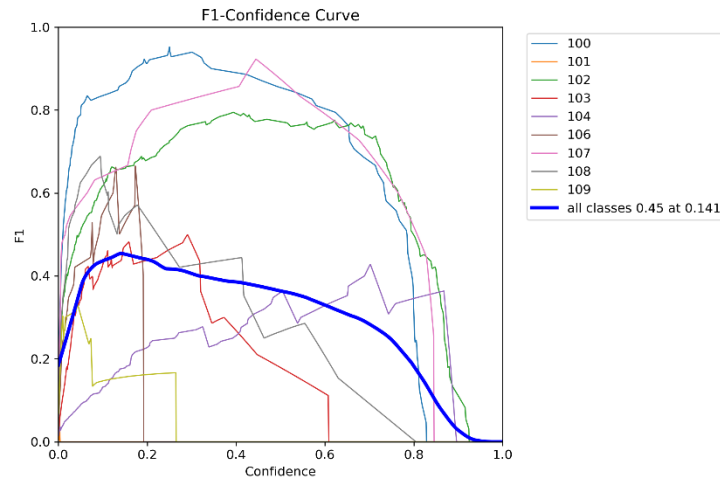


Figura 24 - Curva de F1-Confiança para o teste versão 1 YOLO

O gráfico da curva F1-Confiança apresenta a relação entre o valor do limite de confiança e o desempenho do modelo em termos da métrica F1. Essa métrica, que resulta da média harmónica entre precisão e revocação, é particularmente relevante, uma vez que equilibra a capacidade de identificar corretamente as classes (reduzindo os falsos negativos) com a capacidade de evitar classificações incorretas (reduzindo os falsos positivos).

Observa-se que o melhor valor médio de F1 foi de 0,45, alcançado com um limite de confiança em torno de 0,14. Isto indica que, neste ponto, o modelo apresenta o equilíbrio mais favorável entre deteções corretas e erros de classificação. Para valores muito baixos de confiança, há um aumento de previsões incorretas, enquanto valores muito elevados tornam o modelo demasiado restritivo, reduzindo as deteções.

No que diz respeito às classes específicas, destacam-se as classes 100 - pele de crocodilo em nível severo, 102 - pele de crocodilo em nível médio e 107 - covas em nível médio, que alcançaram valores de F1 mais elevados, evidenciando maior consistência e fiabilidade na deteção. Por outro lado, a classe 103 - obteve resultados mais baixos, refletindo maior dificuldade do modelo em identificá-la de forma precisa. Já as classes 104 - fendas longitudinais em nível médio, 106 - covas em nível baixo, 108 - fendas transversais em nível baixo e 109 - fendas transversais em nível médio apresentaram curvas bastante instáveis e valores reduzidos de F1, sugerindo limitações acrescidas na sua deteção.

De forma global, este gráfico permite concluir que, embora o modelo apresente bom desempenho em determinadas classes, existem categorias que continuam a constituir um desafio para a deteção automática, especialmente quando o limiar de confiança é ajustado para valores mais exigentes.

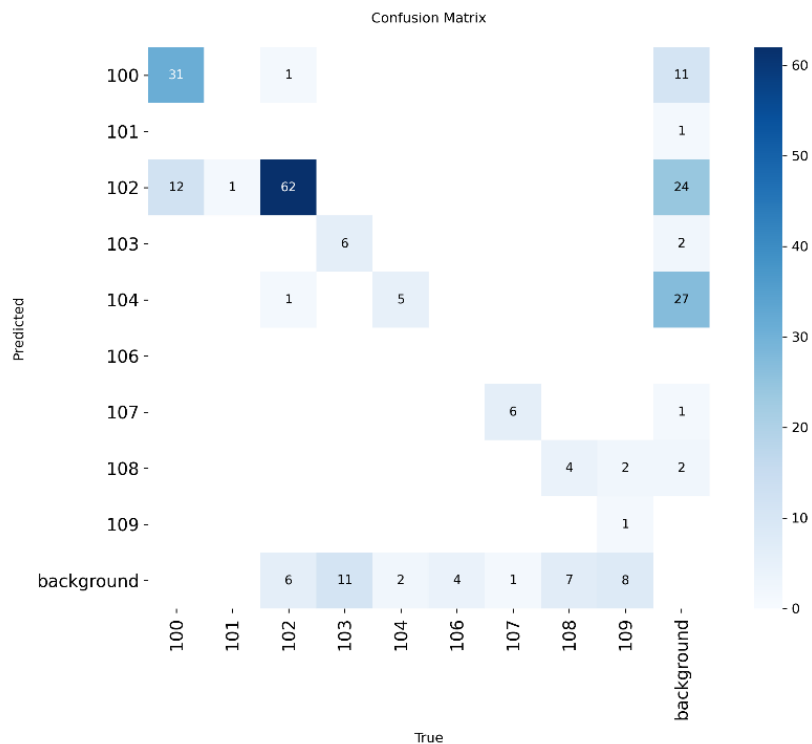


Figura 25 - Matriz de confusão da versão 1 YOLO

Na matriz de confusão apresentada na Figura 25, nota-se que a classe 100 – pele de crocodilo em nível severo, apresenta um número considerável de acertos (31), embora existam também erros de detecção, incluindo 11 instâncias que não foram detetadas pelo modelo. A classe 102 – pele de crocodilo em nível médio, destaca-se pelo elevado número de verdadeiros positivos (62), o que confirma uma boa capacidade de detecção, mas, ao mesmo tempo, registra 24 casos não detetados, o que demonstra fragilidade na sua consistência. A classe 103 – fendas longitudinais em nível baixo, surge apenas com seis instâncias corretamente identificadas e duas não detetadas, revelando maiores dificuldades do modelo nesta categoria. O mesmo ocorre com a classe 104 – fendas longitudinais em nível médio, que conta com cinco acertos e 27 instâncias não detetadas, sugerindo baixa capacidade de generalização. Já a classe 107 – covas em nível médio, apresenta desempenho mais equilibrado, com seis acertos e apenas uma instância não detetada, enquanto as classes 108 – fendas transversais em nível baixo e 109 fendas transversais em nível médio, revelam desempenho fraco, com pouquíssimas instâncias corretamente classificadas e uma proporção elevada de não detecções.

É importante salientar que, na matriz, o rótulo “background” não corresponde a uma classe anotada na base de dados. Ele representa apenas os casos em que o modelo não conseguiu detetar a instância verdadeira, ou seja, falhas completas de detecção (falsos negativos). Assim, a presença desta coluna/linha é essencial para identificar o número de ocorrências ignoradas pelo modelo.

No geral, a matriz de confusão evidencia que, embora algumas classes (como a 100 e a 102) revelem bom desempenho, ainda existem muitas instâncias não detetadas ou confundidas, em especial nas classes minoritárias, comprometendo a consistência global do modelo.

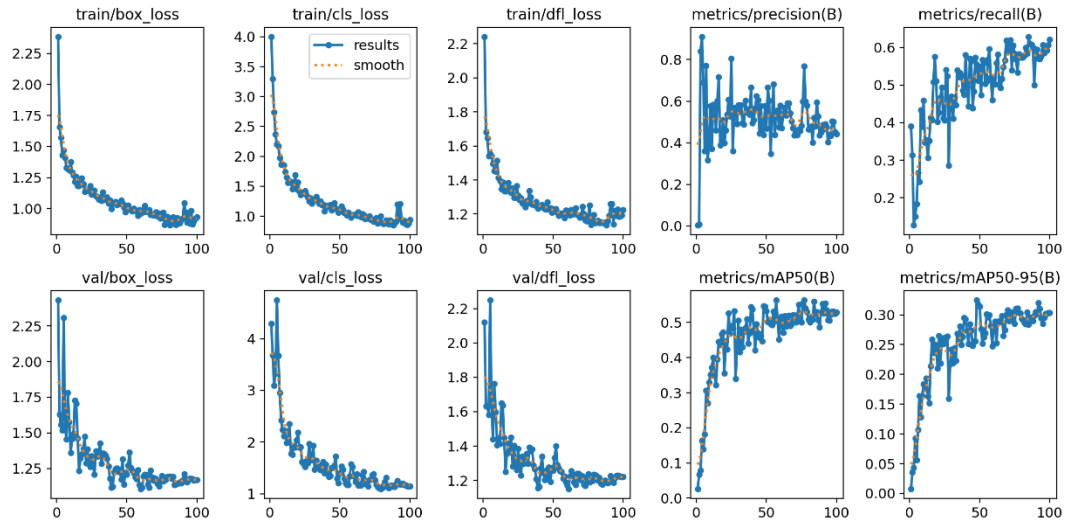


Figura 26 - Métricas da versão 1 YOLO

No primeiro caso de estudo utilizando o YOLOv8 (versão 1), observou-se uma evolução consistente do processo de treinamento ao longo das 100 épocas. Os gráficos de perda evidenciam que tanto no conjunto de treino quanto no de validação houve uma redução progressiva dos erros, indicando aprendizado estável do modelo. Nas fases de treino e validação, a *box loss*, relacionada à localização das caixas delimitadoras, apresentou queda significativa nas primeiras épocas e estabilizou em torno de valores próximos a 1.0, sugerindo que o modelo conseguiu ajustar adequadamente a detecção espacial dos objetos. A *cls loss*, que mede os erros de classificação, também diminuiu de forma contínua, estabilizando próximo a 1.0, o que demonstra boa capacidade de distinção entre as classes anotadas.

Já a *df_l loss* (*Distribution Focal Loss*), associada ao refinamento da localização das caixas, seguiu o mesmo comportamento, diminuindo de maneira consistente.

Em relação às métricas de desempenho, verificou-se um crescimento expressivo da revocação, que atingiu valores próximos a 0.6, indicando que o modelo foi capaz de identificar a maioria dos objetos presentes nas imagens. A precisão apresentou maior oscilação ao longo do treinamento, mas estabilizou em torno de 0.5–0.6, o que revela certa dificuldade em evitar falsos positivos. Os indicadores de qualidade global, expressos pelo mAP50 e mAP50-95, também mostraram evolução progressiva. Como pode ser observado nos últimos dois gráficos da Figura 26, o mAP50 alcançou valores próximos a 0.45–0.50, enquanto o mAP50-95, métrica mais rigorosa, estabilizou em torno de 0.27–0.30. Esses resultados são coerentes com a complexidade da tarefa e demonstram que o modelo conseguiu aprender padrões relevantes para a detecção, embora

ainda exista margem para aprimoramento, especialmente no aumento da precisão e na melhoria do desempenho sob critérios mais restritivos de sobreposição (IoU).

De forma geral, os resultados da versão 1 indicam que o modelo YOLO conseguiu aprender de maneira estável e generalizar bem entre treino e validação, sem sinais marcantes de sobreajuste. Contudo, a performance final, principalmente em termos de mAP mais elevado, ainda está limitada pelo conjunto de dados disponível e pela complexidade intrínseca do problema.

A Figura 27 apresenta um exemplo de detecção obtido com o modelo versão 1 do YOLO.

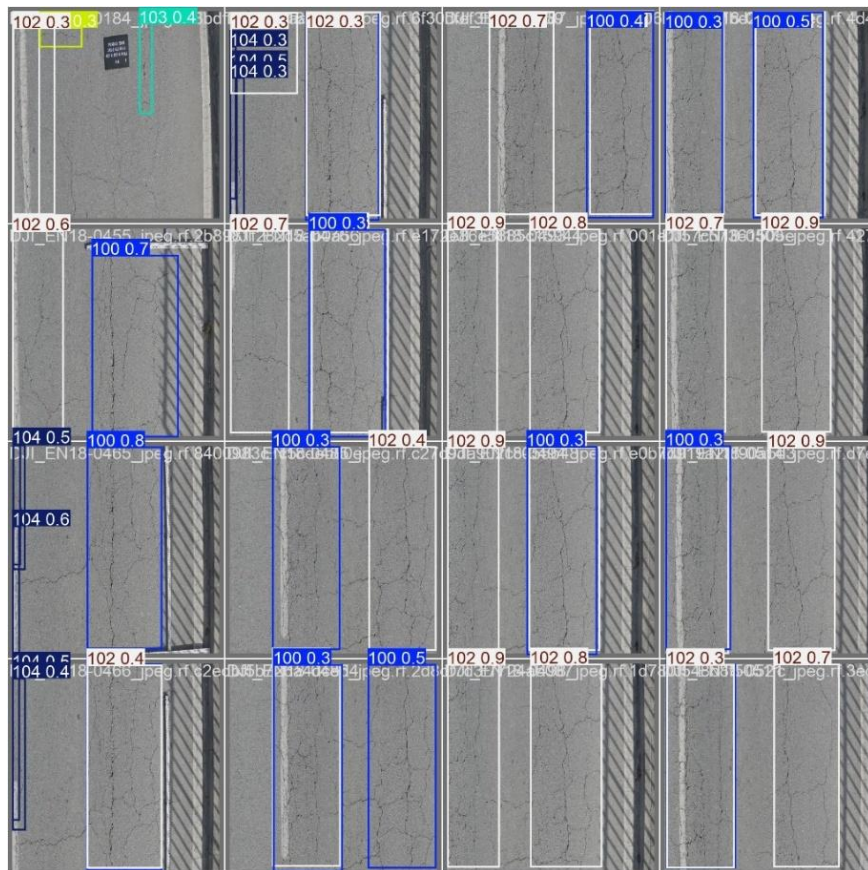


Figura 27 - Exemplo de detecção com o modelo versão 1 do YOLO.

4.4.2 Versão 2 YOLO

Na versão 2 desenvolvida com o modelo YOLO, manteve-se a mesma base de dados utilizada na Versão1, porém com a aplicação de técnicas de ampliação de dados, com o fim de aumentar a variabilidade das amostras e, conseqüentemente, aumentar a robustez do treinamento. As estratégias de aumento de dados incluíram o uso de espelhamento horizontal, recorte (com 0% de minimum zoom e 10% de maximum zoom) e rotações entre -15° e $+15^\circ$, que permitiram gerar variações artificiais das imagens originais. Com isso, o número total de imagens passou de 459 para 781, representando um incremento considerável no volume de dados disponíveis.

A divisão da base de dados seguiu uma proporção equilibrada, sendo 62% destinada ao treino, 19% à validação e 19% ao teste, de modo a garantir uma avaliação consistente do desempenho do modelo. O treinamento deste caso demandou maior tempo computacional, totalizando aproximadamente 27 horas de execução, reflexo do aumento no número de imagens e da complexidade adicional trazida pelas transformações aplicadas.

Este caso tem como objetivo avaliar em que medida o uso de técnicas de ampliação de dados pode contribuir para melhorar a capacidade de generalização do modelo, possibilitando uma comparação direta com os resultados obtidos na Versão 1, em que a base de dados foi utilizada em sua forma original.

```

100 epochs completed in 27.143 hours.
Optimizer stripped from C:\Users\User\runs\detect\train16\weights\last.pt, 6.3MB
Optimizer stripped from C:\Users\User\runs\detect\train16\weights\best.pt, 6.3MB

Validating C:\Users\User\runs\detect\train16\weights\best.pt...
Ultralytics 8.3.203 Python-3.10.18 torch-2.8.0+cpu CPU (Intel Core i5-8265U 1.60GHz)
Model summary (fused): 72 layers, 3,007,403 parameters, 0 gradients, 8.1 GFLOPs
Class      Images  Instances  Box(P  R  mAP50  mAP50-95): 100% ██████████ 5/5 0.2it/s 21.9s
  all         149      403    0.587  0.529  0.522  0.282
  100          92      103    0.566  0.981  0.924  0.714
  101           6         6         1     0     0     0
  102         118      149    0.545  0.859  0.824  0.562
  103          33      42    0.591  0.241  0.34  0.114
  104          14      14     0.5  0.429  0.38  0.101
  106          12      14     0.31  0.5  0.464  0.136
  107          32      32    0.804  0.906  0.939  0.488
  108          26      28    0.644  0.643  0.608  0.284
  109          14      15    0.323  0.2  0.217  0.141
Speed: 3.0ms preprocess, 126.8ms inference, 0.0ms loss, 1.6ms postprocess per image

```

Figura 28 - Resultados da versão 2 YOLO

Em termos de resultados globais, o modelo alcançou valores satisfatórios. A revocação situou-se em 52.9%, revelando que o sistema conseguiu identificar cerca de 53% dos objetos presentes nas imagens. A precisão global foi de 58.7% o que indica que, embora tenha conseguido localizar corretamente uma parte relevante dos objetos, ainda apresentou falsos positivos em quantidade considerável. O mAP@50 foi de 52.2%, ligeiramente superior ao obtido na Versão 1, o que sugere que o uso de técnicas de ampliação de dados contribuiu para um melhor desempenho geral. No entanto, o mAP@50-95 manteve-se em 28.2% resultado semelhante ao da Versão 1, o que demonstra que o modelo continua a apresentar limitações quando avaliado em critérios mais rigorosos de sobreposição (IoU elevados).

A análise por classe permite observar diferenças relevantes. A classe 100 – pele de crocodilo severa, destacou-se com um desempenho muito elevado, atingindo valores de revocação de 98.1% e mAP@50 de 92.4%, demonstrando que o modelo aprendeu de forma consistente os padrões desta categoria. Situação semelhante ocorreu na classe 102 – pele de crocodilo em nível médio, com recall de 85.9% e mAP@50 de 82.4%, também evidenciando um bom ajuste. Outras classes, como a 103 – fendas longitudinais em nível baixo e a 108 – fendas transversais em nível baixo, apresentaram desempenhos medianos, com valores de precisão e revocação aceitáveis, mas inferiores às melhores classes. Por outro lado, algumas categorias com menor número de instâncias, como as classes 101 – pele de crocodilo em nível baixo, 104 – fendas longitudinais em nível médio, 106 – covas em nível baixo, e 109 fendas transversais em nível médio, registaram desempenhos fracos, com valores reduzidos de revocação e mAP, demonstrando a dificuldade do modelo em aprender adequadamente classes com poucos exemplos.

De forma geral, os resultados da Versão 2 indicam que a ampliação dos dados contribuiu positivamente para aumentar o mAP@50 global e melhorar o desempenho em classes mais representativas. Contudo, classes pouco representadas na base de dados continuaram a apresentar resultados insatisfatórios, evidenciando que o aumento artificial das imagens não foi suficiente para compensar a escassez de exemplos reais. Estes resultados reforçam a importância do equilíbrio entre as classes no processo de anotação e sugerem que, para atingir melhorias mais significativas, seria necessária uma base de dados maior e mais homogênea.

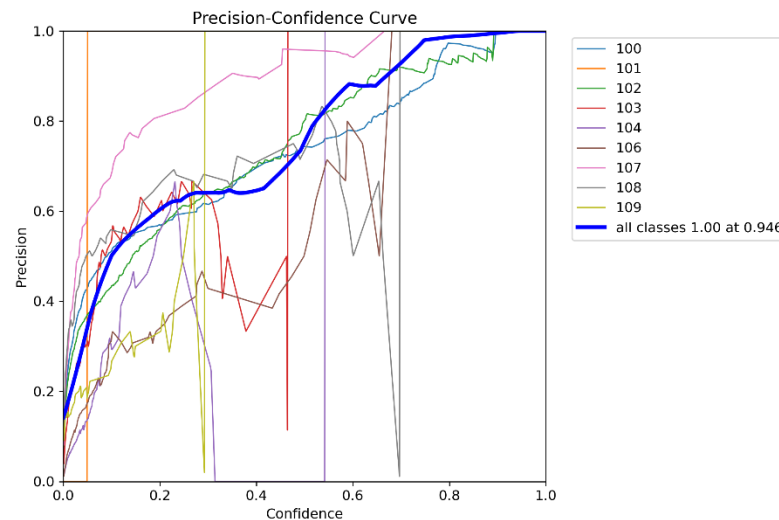


Figura 29 - Curva de Precisão-Confiança para a versão 2 YOLO

A Figura 29 apresenta a curva de precisão-confiança em função do limite de confiança. Em termos gerais, a curva agregada de todas as classes (linha azul espessa) mostra uma tendência crescente, atingindo valores de precisão próximos de 1.0 quando o nível de confiança ultrapassa 0.9, o que demonstra que, ao considerar apenas previsões de alta confiança, o modelo reduz significativamente a ocorrência de falsos positivos.

Ao observar o desempenho por classe, nota-se que, à semelhança dos resultados da Versão 1, algumas categorias apresentam curvas bem definidas e estáveis. É o caso da classe 100 – pele de crocodilo em nível severo e da classe 102 – pele de crocodilo em nível médio, cujas curvas atingem valores de precisão elevados, superiores a 0.9, quando o limite de confiança se aproxima de 1.0, confirmando os bons resultados já identificados nos indicadores de mAP. Também a classe 107 – covas em nível médio, evidenciou uma curva bastante consistente, alcançando rapidamente níveis altos de precisão mesmo com limites de confiança médios.

Por outro lado, algumas classes revelam maior instabilidade, como a classe 103 – fendas longitudinais em nível baixo, que apresenta oscilações acentuadas ao longo da curva, sugerindo inconsistência na detecção. As classes 104 – fendas longitudinais em nível médio, 106 – covas em nível baixo e 109 – fendas transversais em nível médio, menos representadas na base de dados, também evidenciam comportamento irregular, com valores de precisão bastante reduzidos para

limites de confiança baixos, o que indica dificuldade do modelo em aprender os padrões destas categorias.

De forma global, o gráfico confirma a tendência observada nos resultados numéricos: as classes com maior número de instâncias anotadas apresentam curvas de precisão mais estáveis e de melhor desempenho, enquanto as classes minoritárias sofrem com maior variabilidade e baixa confiabilidade nas previsões. Este comportamento reforça a conclusão de que a escassez de exemplos em determinadas classes limita a capacidade do modelo, mesmo com o uso de técnicas de ampliação de dados.

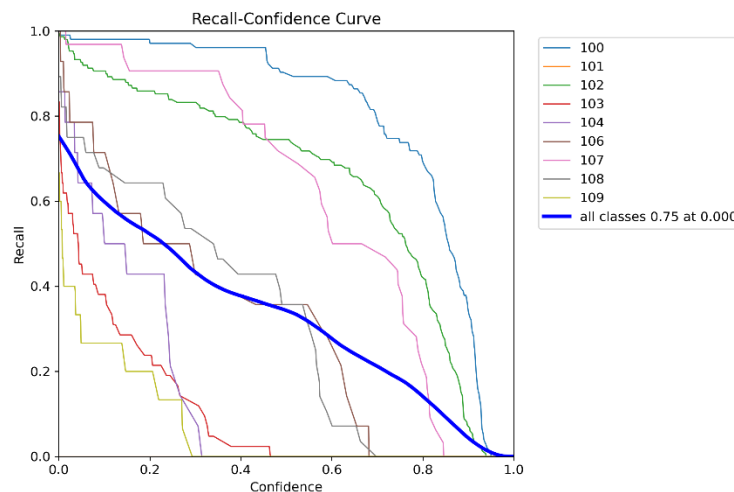


Figura 30 - Curva de Revocação-Confiança para a versão 2 YOLO

A Figura 30 mostra a curva de revocação-confiança em função do nível de confiança, para as diferentes classes do modelo. Observa-se que a revocação tende a diminuir quando a confiança aumenta, evidenciando o equilíbrio natural entre reter mais detecções, alta revocação, e garantir maior precisão, para limites altos de confiança.

A linha azul representa o valor médio global, indicando que o modelo apresenta uma revocação média de 0.75 quando o limite de confiança é zero, ou seja, ao considerar todas as predições possíveis. À medida que o limite de confiança aumenta, a revocação global desce gradualmente, atingindo valores próximos de zero para limites de confiança próximos de 1.0.

Ao analisar classe por classe, nota-se que a classe 100 – pele de crocodilo em nível severo, apresenta a curva mais robusta, conseguindo manter níveis elevados de revocação, superiores a 80%, mesmo até valores intermédios de limite de confiança. Também a classe 102 – pele de crocodilo em nível médio, e a classe 107 – covas em nível médio, demonstraram um desempenho satisfatório, com curvas consistentes e revocação relativamente elevada em diferentes níveis de confiança.

Em contrapartida, algumas classes apresentam desempenho bastante fraco, como a classe 103 – fendas longitudinais em nível baixo, e a classe 109 – fendas transversais em nível médio, cujas curvas descem abruptamente para valores baixos de limite de confiança, revelando grande

dificuldade do modelo em identificar instâncias dessas classes. O mesmo comportamento instável pode ser observado nas classes 104 – fendas longitudinais em nível médio, 106 – covas em nível baixo, e 108 – fendas transversais em nível baixo, confirmando a tendência já observada nos gráficos anteriores: classes com menos exemplos anotados resultam em curvas de revocação menos estáveis e valores globais reduzidos.

Assim, este gráfico reforça a conclusão de que o modelo apresenta boa capacidade de detecção nas classes mais representadas da base de dados, mas baixa generalização nas categorias minoritárias, o que limita o equilíbrio global entre revocação e precisão.

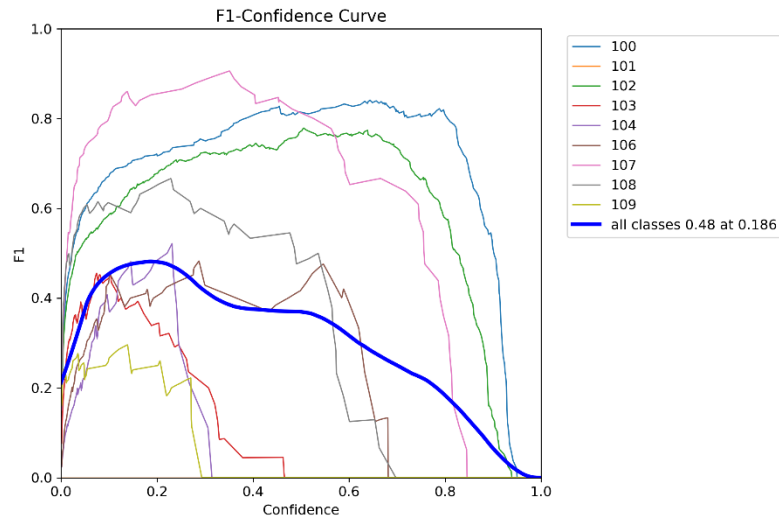


Figura 31 - Curva de F1-Confiança para a versão 2 YOLO

A Figura 31 apresenta a curva F1–Confiança, que mostra como a métrica F1 varia em função do nível de confiança atribuído pelo modelo às suas predições. Recorde-se que o F1 resulta do equilíbrio entre precisão e revocação, sendo, portanto, uma métrica bastante útil para avaliar o desempenho global de um modelo de detecção.

A linha azul, que representa a média de todas as classes, evidencia que o valor máximo de F1 atingido pelo modelo foi de 48%, obtido quando o limite de confiança se encontrava em 18.6%. Isto significa que, para este ponto de decisão, o modelo conseguiu encontrar o melhor compromisso entre identificar corretamente as instâncias e evitar falsos positivos.

Quando analisamos as curvas individuais, observa-se que algumas classes apresentam resultados bastante satisfatórios, nomeadamente a classe 100 – pele de crocodilo em nível severo, classe 102 – pele de crocodilo em nível médio, e classe 107 – covas em nível médio, que atingiram valores de F1 acima de 75% em determinados limites de confiança. Estas classes destacam-se como aquelas em que o modelo demonstra maior robustez e equilíbrio entre as métricas.

Por outro lado, várias classes revelam baixo desempenho, como a classe 103 – fendas longitudinais em nível baixo, classe 104 – fendas longitudinais em nível médio, e classe 109 – fendas transversais em nível médio, cujas curvas permanecem consistentemente em limites de confiança reduzidos, raramente ultrapassando valores de F1 acima de 40%. Este comportamento

reflete tanto as dificuldades de generalização do modelo nessas categorias como a influência da distribuição desigual de exemplos na base de dados, com algumas classes a apresentarem poucas instâncias anotadas.

De modo geral, a curva confirma que o modelo funciona de forma aceitável para as classes mais representadas, mas mantém limitações importantes nas classes minoritárias, evidenciando a necessidade de melhorar o balanceamento e a dimensão dos dados ou aplicar técnicas adicionais de ampliação da base de dados de treino para corrigir tais discrepâncias.

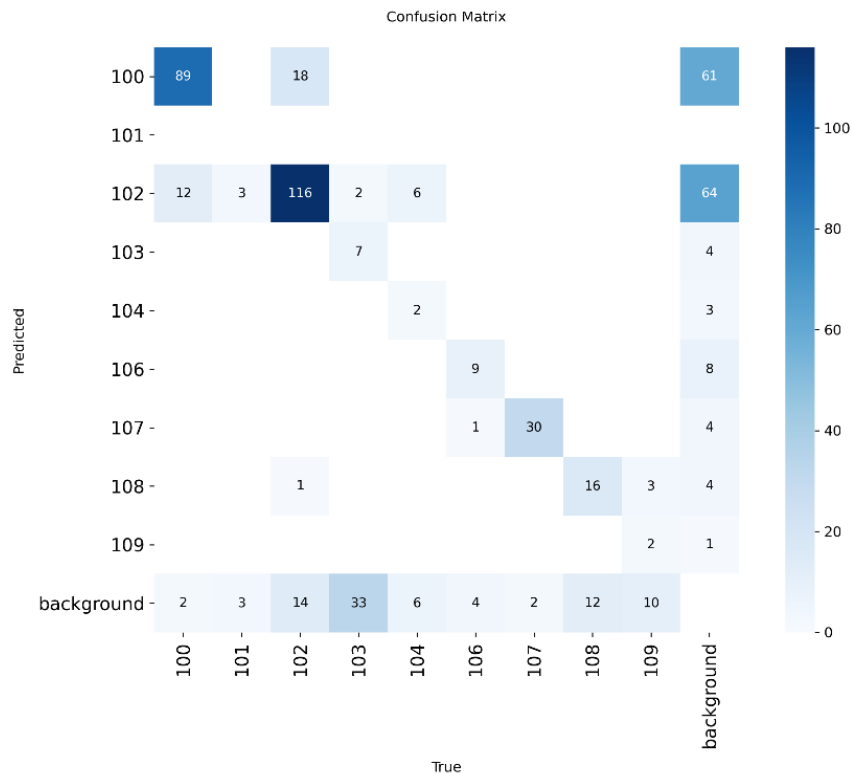


Figura 32 - Matriz de confusão da versão 2 YOLO

Na matriz de confusão da Figura 32, observa-se que algumas classes tiveram bom desempenho, como a classe 102 – pele de crocodilo em nível médio, que alcançou 116 acertos diretos, embora ainda apresente confusões com as classes 100 – pele de crocodilo em nível severo, 101 – pele de crocodilo em nível baixo, e 109 – fendas transversais em nível médio. De forma semelhante, a classe 100 obteve 89 predições corretas, mas também sofreu erros consideráveis, sobretudo confundindo-se com a o background, tendo 61 instâncias.

Por outro lado, várias classes apresentam poucas instâncias corretamente classificadas, evidenciando a dificuldade do modelo em generalizar para categorias menos representadas. É o caso da classe 103 – fendas longitudinais em nível baixo, com apenas 7 acertos, e da classe 104 – fendas longitudinais em nível médio, com apenas 2 acertos, ambas frequentemente confundidas com outras classes ou mesmo identificadas como “background”.

Um aspecto importante a destacar é a frequência com que instâncias de diferentes classes foram confundidas com o “*background*”. Este padrão é visível em várias categorias, por exemplo, 33 instâncias da classe 103 e 14 da classe 102, indicando que o modelo, em alguns casos, não reconhece a presença do objeto e opta por classificá-lo como ausência de classe. De forma geral, a matriz de confusão mostra que o modelo apresenta bom desempenho em algumas classes dominantes, mas ainda sofre com a elevada taxa de confusões em classes minoritárias e com a tendência de atribuir falsos negativos ao “*background*”. Estes resultados reforçam a conclusão de que seria necessário aumentar o número de exemplos anotados para as classes mais fracas e melhorar a diversidade da base de treino, de forma a reduzir os enviesamentos do modelo.

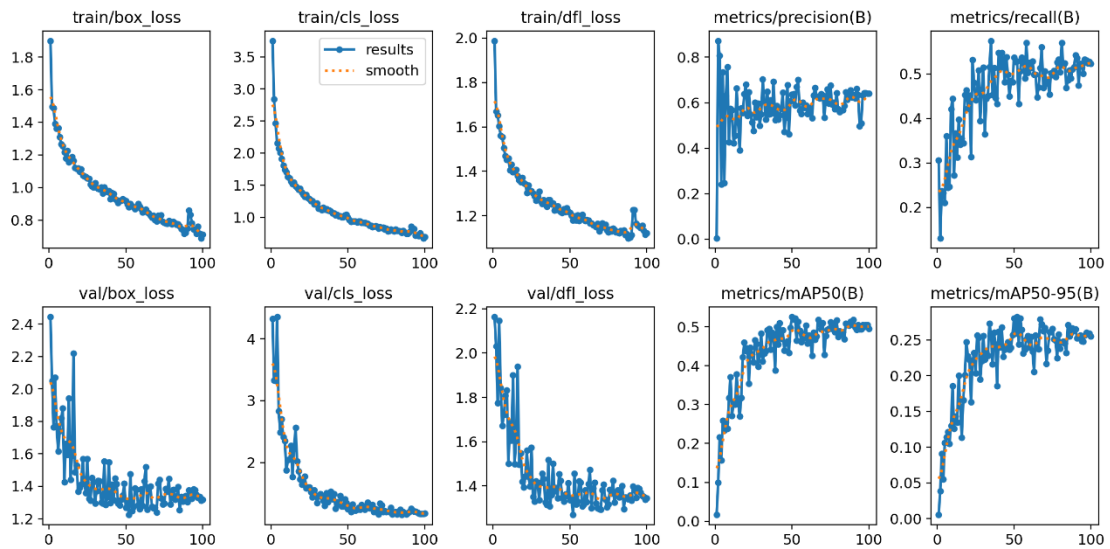


Figura 33 - Métricas da versão 2 YOLO

Nos três primeiros gráficos da parte superior da Figura 33 (*train/box_loss*, *train/cls_loss* e *train/dfl_loss*), observa-se uma tendência decrescente consistente, indicando que o modelo foi aprendendo a ajustar as caixas delimitadoras, a classificar as instâncias e a melhorar a regressão da distribuição das localizações. O mesmo padrão pode ser verificado nos gráficos de validação (*val/box_loss*, *val/cls_loss* e *val/dfl_loss*), embora com uma maior oscilação, o que é esperado, dado que a validação trabalha com dados não vistos durante o treino. A descida gradual destes valores sugere que não ocorreu sobreajuste significativo, pois não há divergência acentuada entre treino e validação.

Nas métricas de desempenho (precisão, revocação, mAP50 e mAP50-95), nota-se uma evolução positiva ao longo do treino. A precisão (precision) e o recall aumentaram gradualmente até estabilizar em valores intermédios, o que indica que o modelo conseguiu melhorar tanto a sua capacidade de identificar corretamente os objetos (precisão) como de recuperar a maioria das instâncias presentes (recall).

Em relação ao mAP50, a métrica atingiu valores próximos de 60% o que confirma que o modelo conseguiu obter um desempenho razoável quando considerada a sobreposição de 50%. Já no mAP50-95, os valores finais ficaram em torno de 25%, o que é significativamente mais baixo,

mostrando que o modelo ainda tem dificuldades em realizar previsões consistentes em limiares de IoU mais exigentes.

No geral, este gráfico evidencia que o treino foi estável e progressivo, sem sinais de degradação acentuada. Ainda assim, as oscilações presentes nas métricas de validação e o valor relativamente baixo do mAP50-95 reforçam a ideia de que o modelo tem limitações nas classes minoritárias e nos casos mais complexos de detecção, o que já foi também observado na matriz de confusão.

Os resultados demonstraram que o modelo alcançou um desempenho estável, com descida consistente das perdas e evolução gradual das métricas. Contudo, os valores finais de precisão, revocação e mAP mantiveram-se apenas em níveis moderados, sem ganhos expressivos em relação à Versão 1. Embora classes como a 100 – pele de crocodilo em nível severo e a 102 – pele de crocodilo em nível médio, tenham apresentado melhor desempenho, outras permaneceram problemáticas, com elevada confusão ou detecções não detectadas.

Assim, conclui-se que a aplicação da técnica de ampliação de dados contribuiu para uma maior robustez do treino, mas não foi suficiente para superar as limitações impostas pelo reduzido número de imagens originais e pelo desequilíbrio entre classes. O modelo mostrou-se funcional e consistente, mas ainda distante de uma performance ideal em cenários complexos.

A Figura 34 mostra resultados obtidos da versão 2 do YOLO.

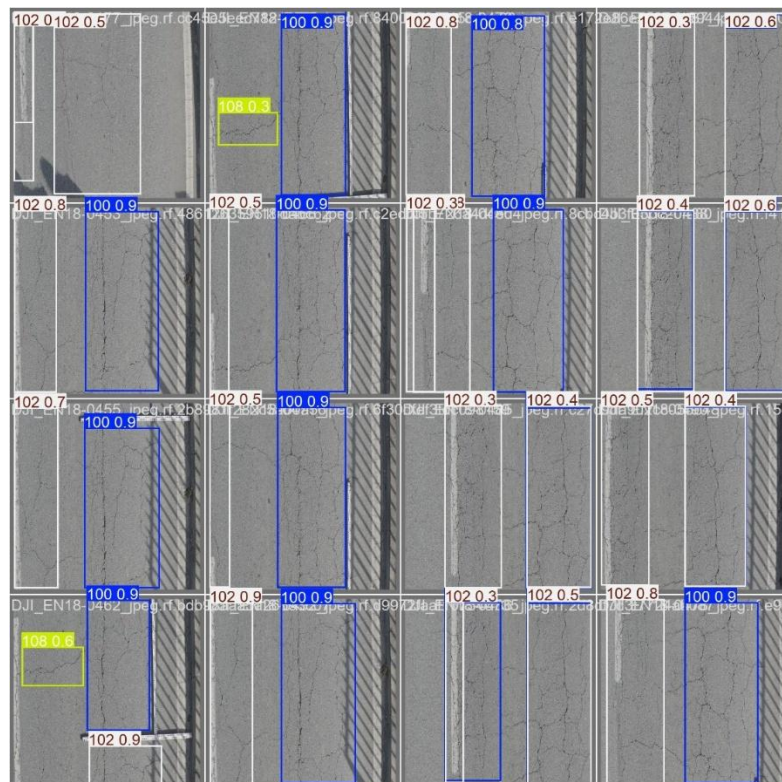


Figura 34 - Exemplo de detecção com o modelo versão 2 do YOLO.

4.5 Comparação dos 6 modelos

Nesta seção são apresentados e discutidos, comparativamente, os resultados obtidos nas diferentes versões de treinamento do modelo YOLO e na plataforma Roboflow. O objetivo dessa análise é avaliar o impacto das configurações de treinamento, do balanceamento dos dados e do tipo de arquitetura utilizada no desempenho final do modelo. Por meio dessa comparação, busque identificar quais ajustes contribuíram de forma mais significativa para a melhoria das métricas de detecção e para a eficiência computacional do processo.

A Tabela 11 apresenta uma síntese dos resultados obtidos para os 6 modelos.

Tabela 10 – Síntese dos resultados obtidos para os 6 modelos treinados.

Roboflow	BD	Nº de imagens	mAP@50	Precision (%)	Recall (%)	Modelo (%)	Tempo de processamento (h)
Versão 1	Original	459	61.1	69.7	65.6	Roboflow 3.0 Object detection (XL)	4
Versão 2	Original	459	59.5	59.8	68.6	YOLOv11 Roboflow (XL)	4
Versão 3	Ampliada	781	59.2	54.2	67.4	Roboflow 3.0 Object detection (XL)	6
Versão 4	Ampliada	781	69.7	64.5	70.2	YOLOv11 Roboflow (XL)	6
YOLOv8	BD	Nº de imagens	mAP@50	Precision (%)	Recall (%)	Modelo (%)	Tempo de processamento (h)
Versão 1	Original	459	54.6	44.0	55.5	YOLOv8	20
Versão 2	Ampliada	781	52.2	58.7	52.2	YOLOv8	27

4.5.1 Testes realizados na plataforma Roboflow

A plataforma Roboflow apresentou-se como uma ferramenta acessível e bastante prática para o desenvolvimento dos modelos de detecção. Disponível em versão gratuita, permite que qualquer utilizador experimente funcionalidades básicas, como a anotação de imagens e a realização de pequenos treinos. No entanto, para a execução dos testes mais complexos desenvolvidos neste

trabalho, que exigiram um número elevado de experiências, foi necessário adquirir de uma licença paga.

Um dos pontos fortes do Roboflow é a sua interface intuitiva e visual, que facilita tanto a organização da base de dados como o acompanhamento das etapas de treino. Mesmo utilizadores com menos experiência em visão computacional conseguem rapidamente compreender o funcionamento da plataforma, o que facilita o processo de utilização da plataforma.

Outro aspeto a destacar é a rapidez de processamento. Os testes realizados, incluindo os que recorreram a arquiteturas mais pesadas com a versão XL, foram concluídos em tempos significativamente mais curtos quando comparados com o tempo necessário para o treino no YOLO. Essa agilidade revelou-se uma vantagem importante, sobretudo em fases de experimentação, permitindo avaliar diferentes hipóteses de forma eficiente. No entanto, é provável que a arquitetura dos modelos YOLO tenha sido simplificada resultando em tempos mais reduzidos de processamento.

Por fim, a plataforma também se mostrou bastante prática na apresentação de métricas e resultados finais, disponibilizando gráficos e indicadores de forma clara e organizada. Não apenas facilita a interpretação dos dados, como também permite comparar rapidamente o desempenho entre diferentes testes.

A tabela 12 apresenta a variação dos resultados obtidos nos modelos desenvolvidos no Roboflow, evidenciando de forma clara o impacto da utilização ou não de técnicas de ampliação de dados e da escolha entre os modelos Roboflow 3.0 e YOLOv11.

Tabela 11 - Comparação dos resultados obtidos na plataforma Roboflow

Versão	Modelo	Ampliação de dados	Precisão (%)	Revocação (%)	mAP@50 (%)	Tempo (h)
V1	Roboflow 3.0	Não	69.7	65.6	61.1	4
V3	Roboflow 3.0	Sim	54.2	67.4	59.2	6
$\Delta(V3-V1)$	-	-	-15.5	+1.8	-1.9	+2
V2	YOLOv11	Não	59.8	68.6	59.5	4
V4	YOLOv11	Sim	64.5	70.2	69.7	6
$\Delta(V2-V4)$	-	-	+4.7	+1.6	+10.2	+2
$\Delta(V1-V2)$	-	-	+9.9	-3.0	+1.6	0
$\Delta(V3-V4)$	-	-	-10.3	-2.8	-10.5	0

Ao comparar a Versão 1 com a Versão 3 (ambas desenvolvidas com o modelo Roboflow 3.0), observa-se que a introdução de técnicas de ampliação de dados não trouxe melhorias, pelo contrário, comprometeu o desempenho em vários aspetos. A precisão sofreu uma descida de 15.5

pontos percentuais (pp). Assim, uma queda de 15.5 pp significa que o modelo passou a cometer significativamente mais erros ao classificar corretamente as instâncias. A revocação, por sua vez, subiu ligeiramente, representando uma melhoria de apenas 1.8 pp, e o mAP₅₀ apresentou uma redução de -1.9 pp. Estes números mostram que, apesar de o modelo ter identificado um pouco mais de instâncias, a sua capacidade de acertar nas previsões diminuiu consideravelmente, levando a mais falsos positivos. Para além disso, o tempo de treino aumentou de 4 para 6 horas, o que, face à perda de qualidade nos resultados, torna esta versão menos eficiente.

Já na comparação entre a Versão 2 e a Versão 4 desenvolvidas com o modelo YOLOv11, o efeito da ampliação de dados revelou-se positivo. Neste caso, todas as métricas registaram melhorias: a precisão aumentou +4.7 pp, revocação aumentou +1.6 pp e o mAP₅₀ + 10.2 pp. Este último valor foi um dos melhores de todo o conjunto de testes no Roboflow, indicando que, além de detetar mais objetos, o modelo conseguiu fazê-lo com maior consistência. O tempo de treino também foi de cerca de 6 horas, mas neste caso o aumento foi compensado pelo ganho claro em desempenho.

De forma geral, pode-se concluir que o impacto da ampliação dos dados depende fortemente do tipo de modelo utilizado. No Roboflow 3.0, o aumento da base de dados resultou numa queda expressiva da precisão e não trouxe ganhos significativos no mAP₅₀, demonstrando que nem sempre aumentar a diversidade da base é benéfico. Pelo contrário, no YOLOv11, os aumentos tiveram um efeito positivo, refletindo-se em melhorias simultâneas de precisão, revocação e mAP₅₀. Este contraste evidencia a importância de avaliar cada arquitetura de forma independente e de adaptar as estratégias de treino às suas especificidades, em vez de assumir que os mesmos métodos vão produzir melhorias universais.

Em relação à comparação entre os modelos Roboflow 3.0 e YOLOv11 para a mesma base de dados, comparando as versões 1 e 2, ambas sem aumento da base de dados, pode-se notar que a precisão da versão 1 apresenta mais 9.9 pontos percentuais em relação à versão 2. Em relação à revocação, verifica-se para a versão 1, em relação à versão 2, uma queda de 3 pontos percentuais e um mAP@50 com mais 1.6 pontos percentuais, indicando que em termos gerais a versão 1 teve melhor desempenho (modelo Roboflow 3.0). Já as versões 3 e 4, onde houve aumento da base de dados, todas as métricas da versão 3 apresentaram piores resultados em relação à versão 4, a precisão reduziu 10.3 pontos percentuais, a revocação diminuiu em 2.8 pontos percentuais e um mAP@50 de menos 10.5 pontos percentuais. Para este caso o modelo YOLOv11 apresentou melhor desempenho.

4.5.2 Testes realizados com o modelo YOLO

O YOLO representou uma alternativa importante no âmbito deste estudo, sobretudo pelo facto de ser possível aceder gratuitamente a uma licença de utilização. No entanto, a execução dos testes apresenta características diferentes daquelas observadas na plataforma Roboflow.

Em primeiro lugar, para treinar os modelos no YOLO e na plataforma Roboflow, é necessário que as imagens já estejam previamente anotadas, o que implica realizar esse processo manualmente ou recorrer a outra ferramenta antes da sua utilização. Essa etapa, embora fundamental, acrescenta uma carga de trabalho adicional, uma vez que o YOLO não possui um ambiente próprio e visual para este fim, ao contrário do Roboflow, onde essa etapa pode ser feita dentro do próprio ambiente da plataforma. Assim, o utilizador precisa preparar cuidadosamente os dados de entrada, o que pode tornar o processo mais demorado e técnico.

Outro ponto relevante é a interface do modelo. Diferentemente do Roboflow, que oferece uma abordagem gráfica e intuitiva, o YOLO é baseado em código, exigindo dos utilizadores conhecimentos básicos de programação e alguma familiaridade com linhas de comando. Esta característica pode ser um desafio para iniciantes, mas ao mesmo tempo oferece maior flexibilidade e controlo sobre os parâmetros de treino, permitindo personalizações que vão além do que uma interface simplificada normalmente oferece.

No que diz respeito ao tempo de processamento para obtenção dos modelos, os testes realizados demonstraram que o YOLO requer um esforço computacional significativamente maior em relação ao Roboflow, o que pode estar relacionado com uma arquitectura mais complexa. Os treinos prolongaram-se por mais de 20 horas, mesmo quando a base de dados não era muito extensa. Ainda assim, o YOLO tem a vantagem de ser altamente reconhecido na comunidade científica e profissional, com documentação vasta, suporte contínuo da Ultralytics e resultados consistentes em tarefas de visão computacional.

Assim, pode-se afirmar que o YOLO, embora mais exigente em termos técnicos e de tempo de processamento, constitui uma ferramenta poderosa e flexível, que privilegia a profundidade e a personalização sobre a simplicidade e a rapidez. A tabela 13 apresenta os resultados obtidos no YOLO, comparando as Versões 1 e 2 desenvolvidas.

Tabela 12 - Comparação dos resultados obtidos com o modelo YOLO

Versão	Augmentations	Nº de imagens	Tempo treino (h)	Precisão (%)	Revocação (%)	mAP@50 (%)
V1	Não	459	20	44.0	55.5	54.6
V2	Sim	781	27	58.7	52.9	52.2
$\Delta(V1-V2)$	-	+322	+7	+14.7	-2.6	-2.4

Na versão 1 do modelo, obteve-se 44,0% de precisão, 55,5% de revocação e um mAP@50 de 54,6%. Observa-se um equilíbrio moderado entre precisão e revocação, indicando que o modelo conseguiu identificar corretamente um número considerável de instâncias, mesmo com um conjunto de dados limitado.

Já na versão 2, em que foi utilizada uma base de dados aumentada, os resultados apresentaram uma melhora na precisão, que passou para 58,7%, mostrando que o modelo se tornou mais rigoroso e cometeu menos falsos positivos. No entanto, a revocação diminuiu para 52,9%, o que demonstra que o modelo deixou de identificar parte das instâncias verdadeiras. Apesar dessa variação, o mAP@50 aumentou ligeiramente para 52,2%, evidenciando uma melhoria discreta na capacidade global de detecção.

Esses resultados mostram que o aumento da base de dados contribuiu para um modelo mais seletivo e preciso, embora com uma leve perda na revocação. Em síntese, o modelo treinado com dados aumentados apresentou melhor desempenho geral, ainda que o ganho não tenha sido linear em todas as métricas.

Esse comportamento pode estar relacionado tanto com a forma como os dados adicionais foram gerados, como com possíveis desequilíbrios introduzidos no processo.

4.5.3 Comparação entre testes YOLO (Roboflow YOLOv11 versus YOLOv8)

Ao comparar as duas plataformas utilizadas neste estudo, Roboflow e YOLO, é possível identificar diferenças significativas tanto em termos de acessibilidade como de funcionamento prático.

Para que? todo o processo seja conduzido de forma didática e acessível, mesmo para utilizadores com pouca experiência em programação. Para anotações e testes simples, existe uma versão gratuita disponível, contudo, para os testes realizados neste trabalho, devido ao número de execuções necessárias e ao tamanho da base de dados, foi necessário recorrer a uma licença paga. Uma das vantagens mais evidentes do Roboflow é a rapidez de processamento: mesmo em modelos de maior complexidade, como o XL, os tempos de treino mantiveram-se relativamente curtos, entre 4 e 6 horas, o que possibilitou obter resultados de forma célere e eficiente.

Por outro lado, o YOLO, apesar de também estar acessível gratuitamente graças à licença disponibilizada pela Ultralytics, exige que as imagens estejam previamente anotadas em ferramentas externas, o que representa uma etapa adicional do processo. Além disso, toda a interação com a plataforma é feita através de linha de comando, sem recurso a uma interface visual, o que pode tornar o uso mais difícil para quem não tem familiaridade com programação. Outro fator a destacar é o tempo de execução, bastante superior ao registado no Roboflow: cada treino completo ultrapassou as 20 horas, chegando a quase 27 horas em alguns casos.

Enquanto o Roboflow privilegia a usabilidade e rapidez, o YOLO apresenta-se como uma solução mais flexível e personalizável, mas que exige maior conhecimento técnico e recursos computacionais mais robustos.

Para avaliar o impacto do ambiente de treino sobre o desempenho, procedeu-se a uma comparação direta entre os modelos obtidos na plataforma Roboflow (configurado com tipo de modelo YOLOv11) e o modelo YOLO (YOLOv8) para a mesma base de imagens em dois cenários: sem ampliação de dados (base de dados original) e com ampliação de dados. Esta comparação de métricas e desempenho é apresentada na Tabela 14.

Tabela 13 - Comparação dos resultados entre os modelos obtidos no Roboflow-YOLOv11 e no modelo YOLO-YOLOv8

Plataforma e versão de treino	Configuração	Nº de imagens	Tempo de treino (h)	Precision (%)	Recall (%)	mAP@50 (%)
Roboflow V1	YOLOv11 - sem ampliação de dados	459	4	59.8	68.6	59.5
YOLO V1	YOLOv8 - sem ampliação de dados	459	20	44.0	55.5	54.6
$\Delta(\text{Roboflow V1}-\text{YOLO V1})$	-	-	-16	-15.8	-13.1	-4.9
Roboflow V2	YOLOv11 - com ampliação de dados	781	6	64.5	70.2	69.7
YOLO V2	YOLOv8 - com ampliação de dados	781	27	58.7	52.9	52.2
$\Delta(\text{Roboflow V2}-\text{YOLO V2})$	-	-	-21	-5.8	-17.3	-17.5

A comparação entre os testes realizados na plataforma Roboflow e no modelo YOLO permitiu obter conclusões relevantes para este estudo, fornecendo uma visão clara sobre o desempenho e as limitações de cada ambiente no contexto da deteção automática. De forma geral, verificou-se que o Roboflow manteve um desempenho superior em termos de eficiência e rapidez de processamento, alcançando métricas mais elevadas tanto nos testes sem ampliação de dados como naqueles com aumento da base de imagens. Por outro lado, o YOLO destacou-se pela flexibilidade e autonomia que oferece ao utilizador, permitindo maior controlo sobre os parâmetros de treino e uma exploração mais aprofundada do comportamento do modelo, ainda que com maior exigência técnica e maior tempo de execução. Assim, as diferenças observadas, especialmente quanto ao tempo de processamento e à estabilidade dos resultados, sugerem que trabalhos futuros avaliem o grau de complexidade e otimização dos algoritmos utilizados em cada plataforma.

Nos testes sem ampliação de dados, utilizando 459 imagens, o Roboflow-YOLOv11 atingiu uma precisão de 59,8%, revocação de 68,6% e um mAP@50 de 59,5%, com tempo de treino de apenas 4 horas. Já o YOLOv8, sob as mesmas condições, apresentou 44,0% de precisão, 55,5% de revocação e 54,6% de mAP@50, com tempo de treino de 20 horas. Nota-se, portanto, uma diferença expressiva de desempenho e eficiência a favor do Roboflow, tanto na qualidade dos resultados quanto no tempo computacional.

Quando aplicada à base de dados aumentada (781 imagens), o Roboflow-YOLOv11 obteve novamente resultados superiores, com 64,5% de precisão, 70,2% de revocação e mAP@50 de 69,7%, em 6 horas de treino. Em comparação, o YOLOv8 apresentou 58,7% de precisão, 52,9% de revocação e 52,2% de mAP@50, necessitando 27 horas de processamento. A diferença entre as duas plataformas mostra-se ainda mais acentuada nesta fase, indicando que o Roboflow consegue tirar melhor proveito da ampliação da base de dados, enquanto o YOLOv8 demonstrou maior sensibilidade à complexidade acrescida, resultando em tempos de treino muito mais longos e desempenho inferior nas métricas globais.

Esta comparação final encerra o ciclo do estudo, mostrando que, apesar de limitações como o número restrito de imagens e a curta extensão do trecho analisado, foi possível obter resultados satisfatórios e extrair conclusões importantes sobre o comportamento e a eficiência das duas plataformas utilizadas na análise.

Capítulo 5

5.0 Conclusões

O presente trabalho constituiu a etapa final de um processo de investigação dedicado à análise e avaliação do desempenho de modelos de detecção automática de objetos, para identificar degradações da superfície de pavimentos rodoviários, recorrendo a duas plataformas distintas: o Roboflow e o modelo YOLO. Trata-se de uma abordagem ainda inovadora e em fase inicial de desenvolvimento, conforme evidenciado pela revisão da literatura, que revelou um número ainda limitado de estudos aplicando técnicas de *deep learning* neste domínio específico. Essa constatação demonstra que, embora promissora, esta metodologia ainda enfrenta desafios significativos para a sua consolidação e aplicação generalizada no contexto da gestão e conservação de pavimentos rodoviários.

O estudo foi estruturado em seis testes experimentais, dos quais quatro decorreram na plataforma Roboflow e dois no ambiente YOLO. As condições definidas em cada versão permitiram compreender de forma sistemática como variáveis como a ampliação de dados, o balanceamento entre conjuntos de treino, validação e teste, e a seleção de diferentes arquiteturas influenciam diretamente o desempenho e a eficiência computacional dos modelos.

Nos primeiros ensaios, testou-se o comportamento dos modelos com a base de dados original obtida, sem ampliação de dados, o que possibilitou estabelecer um ponto de referência inicial. Numa fase posterior, aplicaram-se técnicas de ampliação de dados, tais como rotações, zooms e inversões horizontais das imagens, as quais aumentaram em 70.15% o número de imagens disponíveis e permitiram avaliar de forma comparativa o impacto da maior variabilidade dos dados no desempenho final dos modelos. Paralelamente, foram exploradas arquiteturas de maior dimensão (XL), o que implica um acréscimo considerável no tempo de treino, mas também revelou de que forma o aumento da complexidade estrutural se traduz em ganhos de precisão e generalização.

A análise dos resultados evidenciou de forma clara a relevância da qualidade e representatividade do conjunto de dados, demonstrando que o desempenho dos modelos é fortemente condicionado pela distribuição equilibrada e pela nitidez das instâncias identificadas nas imagens. Algumas classes apresentaram maior suscetibilidade a erros, quer por escassez de exemplos, quer por semelhança visual com outras classes, o que limitou parcialmente a obtenção de métricas mais elevadas. Esta constatação reforça a ideia de que, em aplicações de visão computacional, a qualidade da base de dados é tão determinante quanto a arquitetura do modelo.

De um modo geral, os seis casos de estudo proporcionaram uma compreensão abrangente tanto do funcionamento dos modelos e das métricas de avaliação, como das potencialidades e restrições de cada ambiente de treino. Enquanto o Roboflow se destacou pela praticidade e automatização dos processos, o YOLO demonstrou maior flexibilidade e controle técnico, permitindo uma exploração mais profunda e personalizável. Assim, este conjunto de experimentações não só consolidou os conhecimentos adquiridos, como também forneceu uma base sólida e comparativa para a formulação das conclusões finais apresentadas nos tópicos seguintes, encerrando de forma coerente o percurso analítico desenvolvido ao longo deste trabalho.

5.1 Síntese dos principais resultados e tendências

Os resultados obtidos ao longo das diferentes etapas experimentais permitiram identificar tendências consistentes quanto ao comportamento dos modelos e às condições de treino. De modo geral, verificou-se que a plataforma Roboflow, com a versão YOLOv11, apresentou desempenho globalmente superior para a base de dados aumentada e a versão Roboflow 3.0 para a base de dados original, em termos de suas métricas. Essa vantagem manteve-se estável entre as configurações testadas, o que demonstra a eficiência da integração automática e otimizada oferecida pela plataforma.

Por outro lado, o YOLOv8, executado em ambiente local, mostrou-se uma ferramenta de grande valor sobretudo pela flexibilidade que oferece. A possibilidade de ajustar parâmetros, personalizar o processo de treino e compreender em detalhe o funcionamento interno do modelo torna-o particularmente indicado para utilizadores com conhecimentos técnicos mais avançados ou que pretendam explorar configurações específicas. Contudo, essa liberdade vem acompanhada de maior exigência computacional e de um tempo de processamento consideravelmente superior, o que limita a sua aplicabilidade em contextos com restrições de recursos e meios computacionais.

No que diz respeito à identificação das degradações da superfície dos pavimentos, a ampliação da base de dados, de 459 para 781 imagens, visou melhorar a detecção e a generalização dos modelos, mas foi realizada de forma aleatória, sem aplicação de técnicas de reequilíbrio entre classes. Os resultados mostraram que essa expansão produziu efeitos distintos nas plataformas analisadas. No Roboflow, observou-se uma melhoria consistente em todas as métricas, com o $mAP@50$ a aumentar de 61,1% para 69,7%, a precisão a manter-se estável com 69,7% e a revocação a subir de 65,6% para 70,2%, indicando maior capacidade de detecção global. Já no YOLOv8, os resultados com a base ampliada não demonstraram o mesmo comportamento: o $mAP@50$ manteve-se praticamente inalterado de 54,6% para 54,2%, a precisão aumentou de 44,0% para 58,7%, mas a revocação decresceu de 55,5% para 52,2%, sugerindo menor equilíbrio entre detecções corretas e falsas negativas. Esses resultados evidenciam que o simples aumento da base de dados não garante ganhos proporcionais de desempenho, sendo a qualidade das anotações e o rigor do julgamento humano fatores determinantes para a eficácia do processo de aprendizagem. Entre os seis modelos testados, o Roboflow – Versão 4 (modelo YOLOv11 XL com base ampliada)

apresentou o melhor desempenho geral, combinando as melhores métricas e um tempo de processamento significativamente inferior ao do YOLOv8, confirmando-se como a opção mais eficiente e equilibrada.

No que se refere ao desempenho por classe, verificou-se que as categorias 102 – pele de crocodilo em nível médio, e 100 - pele de crocodilo em nível severo, foram as que apresentaram os melhores resultados de detecção. Esse comportamento deve-se, em grande parte, ao facto de essas degradações possuírem maior dimensão e contraste visual, o que facilita a sua identificação pelo modelo, além de representarem a classe com maior número de ocorrências na base de dados, contribuindo para uma aprendizagem mais robusta. Ademais, as suas características geométricas bem definidas e facilmente distinguíveis reduzem a probabilidade de confusão com outras classes de defeitos, o que explica o melhor desempenho obtido nessas categorias.

De forma global, conclui-se que o Roboflow é uma solução mais acessível, rápida e eficiente para estudos aplicados e com recursos limitados, enquanto o YOLOv8 constitui uma opção mais flexível e investigativa, adequada a cenários que exijam maior controlo e experimentação. Assim, o conjunto dos resultados aponta para uma relação de complementaridade entre as plataformas, em que a escolha ideal depende diretamente dos objetivos, da disponibilidade computacional e do nível de profundidade técnica pretendido no estudo.

5.2 Limitações do caso de estudo

O caso de estudo apresentou limitações que impactaram diretamente o desempenho dos modelos e a sua capacidade de generalização. A principal restrição esteve associada ao tamanho reduzido e desequilíbrio entre classes de degradação, da base de dados, composta inicialmente por apenas 459 imagens. Estes aspetos limitaram a diversidade de exemplos disponíveis para o treinamento e dificultou o aprendizado robusto de padrões visuais, especialmente em classes com menor representatividade.

A extensão do trecho analisado, com de cerca de 150 metros, também contribuiu para a baixa variabilidade espacial, restringindo a variedade de cenários, degradações e condições ambientais captadas. Essa homogeneidade pode ter levado o modelo a se ajustar excessivamente às características específicas do local, comprometendo seu desempenho em contextos distintos.

Adicionalmente, a presença de ruídos visuais como sombras, variações de iluminação, reflexos e diferenças de contraste interferiu na nitidez das imagens e, conseqüentemente, na qualidade das anotações e das predições. Esses fatores dificultaram a distinção entre as áreas de deterioração e o fundo da via, resultando em detecções imprecisas ou até na omissão de defeitos, ocasionando falsos negativos, sobretudo em regiões com baixa iluminação ou com muitas sombras.

Essas limitações evidenciam a necessidade de ampliar significativamente a base de dados, não apenas em quantidade total de imagens, mas também na distribuição equilibrada entre as

diferentes classes de degradação. É fundamental aumentar o número de amostras nas categorias menos representadas, de forma a permitir que o modelo aprenda de maneira mais consistente os padrões visuais de cada tipo de defeito. Além disso, a inclusão de imagens captadas sob diferentes condições de luminosidade e variações naturais do pavimento contribuirá para uma representação mais realista e abrangente do fenómeno, fortalecendo a robustez e a capacidade de generalização dos modelos treinados.

5.3 Indicações para trabalhos futuros

Com base nos resultados alcançados, é possível delinear várias direções promissoras para o aprofundamento deste tipo de investigação. Uma das principais recomendações consiste em ampliar e diversificar o uso das técnicas de aumento da base de dados, não apenas através do aumento do número de imagens, mas também pela incorporação de transformações mais variadas como ajustes de brilho, contraste, saturação, rotação, recorte e distorções geométricas.

Estas estratégias podem contribuir para a criação de bases de dados mais robustas e realistas, capazes de melhorar a capacidade de generalização dos modelos face a diferentes condições de captação.

Outra vertente relevante de investigação envolve a análise do desempenho dos modelos sob distintas condições ambientais e temporais. A inclusão de imagens captadas em diferentes períodos do dia, estações do ano ou sob condições atmosféricas adversas, como chuva, neblina ou variações acentuadas de luminosidade, permitiria avaliar a resiliência dos algoritmos e a sua adaptabilidade a contextos mais desafiantes e representativos da realidade.

Recomenda-se igualmente o desenvolvimento de bases de dados mais extensas e heterogêneas, englobando um número maior de trechos de vias e uma distribuição mais equilibrada entre as classes de degradação. Tal expansão permitirá não só melhorar o desempenho das métricas dos modelos, mas também validar a sua aplicabilidade prática em diferentes contextos geográficos e operacionais.

Por fim, uma direção de particular interesse seria a integração das potencialidades das duas plataformas analisadas. O Roboflow poderá continuar a ser explorado como ferramenta de prototipagem rápida, facilitando a anotação, o treino inicial e a visualização dos resultados, enquanto o YOLO, pela sua flexibilidade e controlo técnico, poderá ser utilizado em fases mais avançadas, destinadas ao refinamento e personalização dos parâmetros de treino. Essa complementaridade entre ferramentas pode representar um caminho eficiente para otimizar o equilíbrio entre usabilidade e desempenho.

Dessa forma, os trabalhos futuros poderão consolidar e expandir o conhecimento desenvolvido neste estudo, contribuindo para o avanço de sistemas de deteção mais precisos, adaptáveis e

aplicáveis a cenários reais cada vez mais complexos, aproximando a investigação científica das exigências práticas da gestão e monitorização de pavimentos rodoviários.

REFERENCIAS

- BUCKO, B.; LIESKOVSKA, E.; ZÁBOVSKA, K.; ZÁBOVSKY, M. Computers Vision Based pothole detection under challenging conditions. Slovakia, 2020. Sensors.
- CONTI, P.; LIRD, A.; CID, H. Automated distress detection, classification and measurement for asphalt urban pavements using YOLO. Chile, 2023. Engineering proceeding.
- DU, Y.; PAN, N.; XU, Z.; DENG, F.; SHEN, Y.; KANG, H. Pavement distresses detection and classification based on YOLO network. China, 2020. International journal of pavement engineering.
- ESPÍNOLA, A.; FREITAS, G.; JÚNIOR, E. Pothole and patch detection on asphalt pavement using deep convolutional neural network. CILAMCE-PANACM-2021. Brazil, 2022.
- FEITOSA, I.; SANTOS, B.; ALMEIDA, P. Pavement Inspection in Transport Infrastructures Using Unmanned Aerial Vehicles (UAVs). Sustainability MDPI. Portugal, 2024.
- GANESHAN, D.; SHARIF, M.; APEAGYEI, A. Road deterioration detection: A machine learning based system for automated paved pavement crack identifications and analysis. England, 2023. International conference on innovation and intelligence for information, computing and technologies (3ICT).
- GONG, H.; TESIC, J.; TAO, J.; LUO, X.; WANG, F. Automated pavement crack detection with deep learning methods: What are the main factors and how to improve the performance? United States, 2023. TRR Journal of the transportation research board.
- HALIM, M.; IBRAHIM, A.; OSMAN, M.; KADER, M.; TERMIZI, M.; ABU, A. Detection of pothole for repair works of asphalt flexible pavement optimization using YOLO. Malaysia, 2022. AIP conference proceedings.
- HUSSAIN, M.; YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines. Inglaterra, 2023.
- LIU, Z.; WU, W.; CUI, B. PaveDistress: A comprehensive dataset of pavement distresses detection. Elsevier. China, 2024.
- MAJIDIFARD, H.; ADU-GYAMF, Y.; BUTTLAR, W. Deep machine learning approach to develop a new asphalt pavement condition index. United States, 2020. Construction and Building materials. ScienceDirect.
- MAJIDIFARD, H.; ADU-GYAMF, Y.; BUTTLAR, W. Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses. United States, 2020. TRR Journal of the transportation research board.
- NAFFA, S.; ASHOUR, K.; MOHAMED, R.; ESSAM, H., EMAD, D.; ELHENAWY, M., ASHQAR, H., HASSAN, A.; ALHADIDI, T. Automated pavement cracks detection and classification using deep learning. In: International Conference on Computing and Machine Intelligence (ICMI). Egypt, 2024.
- NERY, D.; SANTOS, B.; ALMEIDA, P. Otimização de métodos de auscultação de pavimentos rodoviários com recurso à geomática e veículos aéreos não tripulados (VANT). Portugal, 2023.
- OLIVEIRA, M.; SANTOS, B. REVIEW ON ROAD PAVEMENT DISTRESS DETECTION WITH YOLO. ICEUBI International Congress on Engineering 2024 Sustainability and Resilience innovation and solutions. Portugal, 2024.
- PAN, Y.; ZHANG, X.; CERVONE, G.; YANG, L. Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume: 11, Issue: 10. China, 2018.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, real-time object detection. In: Conference on Computer Vision and Pattern Recognition. United States, 2016.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection. arXiv preprint. Estados Unidos, 2015.

REN, M.; ZHANG, X.; CHEN, X.; ZHOU, B.; FENG, Z. A deep learning network model for road pavement damage detection from urban street view imagery. China, 2023. *Internacional journal of Applied Earth observation and geoinformation*.

ROBOFLOW. About Roboflow. Disponível em: <https://roboflow.com/about>. Acesso em: 5 set. 2025.

SANTOS, B.; GRAVINHOS, P.; ALMEIDA, P.; NERY, D. In book: Proceedings of the 5th International Conference on Transportation Geotechnics (ICTG) 2024, Volume 1. Portugal, 2024.

SILVA, L.; BLAS, H.; GARCÍA, D.; MENDES, A.; GONZÁLEZ, G. Na architectural multi-agent system for a pavement monitoring system with potholes recognition in UAV images. *Sensors MDPI*. Espanha, Brasil, 2020.

SPRITAN, S. Understanding IoU, Precision, Recall, and mAP for Object Detection Models. *Medium*. Estados Unidos, 2023.

SUN, Y.; SONG, W.; ZHANG, J.; TENG, Q. Crack YOLO: Rural pavement distress detection model with complex scenarios. *Electronics*. China, 2024

TERVEN; J.; ESPARZA, D.; GONZÁLEZ, J.; A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine learning & knowledge extraction*. México, 2023.

U.S. Federal Highway Administration (FHWA). Long-Term Pavement Performance Information Management System User Guide. Publication No. FHWA-HRT-21-038. Washington, D.C.: U.S. Department of Transportation, May 2021.

UKHWAH, E.; YUNIARNO, E.; SUPRAPTO, Y. Asphalt pavement pothole detection using deep learning method based on YOLO neural network. *International seminar on intelligent technology and its application (ISITIA)*. Indonesia, 2019.

VALAPPIL, N.; JHON, M.; NAJIHA, D. Pavemen crack detection and localization using convolutional neural networks (CNNs). *ResearchGate*. Emirados Arabes Unidos, 2019.

XU, L.; FU, K.; MA, T.; TANG, F.; FAN; J. Automatic detection of urban pavement distress and dropped objects with a compressive dataset collected via smartphone.. *Buildings*. China, 2024.

YA'ACOB, N.; ZURAIMI, M.; RAHMAN, A.; YUSOF, A.; ALI, D. Real time pavement crack detection based on artificial intelligence. *Malaysia*, 2024.

Journal of advanced research in applied sciences and engineering technology.

YAO, H.; FAN, Y.; WEI, X.; LIU, Y.; CAO, D.; YOU, Z. Research and optimization of YOLO based method for automatic pavement detect detection. *Electronic research archive AIMS*. China, 2024.

ZHOU, H.; KONG, H.; WEI, L.; CREIGHTON, D.; NAHAVANDI, S. Efficient road detection and tracking for unmanned aerial vehicle. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, VOL. 16, NO. 1. Singapura, 2015.