



UNIVERSIDADE FEDERAL DO MARANHÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
Doutorado em Ciência da Computação Associação
UFMA/UFPI

Joel de Conceição Nogueira Diniz

Ensemble de Redes Convolucionais para Inspeção
Visual de Estruturas de Concreto Armado

Orientador: Prof. Dr. Anselmo Cardoso de Paiva
Co-orientador: Profa. Dra. Sandra Cristina Alves Pereira
da Silva Cunha

São Luís - MA
Outubro, 2025

Joel de Conceição Nogueira Diniz

**Ensemble de Redes Convolucionais para Inspeção
Visual de Estruturas de Concreto Armado**

TESE DE DOUTORADO

Orientador: Prof. Dr. Anselmo Cardoso de Paiva

Co-orientador: Profa. Dra. Sandra Cristina Alves Pereira da Silva Cunha

São Luís - MA
Outubro, 2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

de Conceição Nogueira Diniz, Joel.

Ensemble de Redes Convolucionais para Inspeção Visual
de Estruturas de Concreto Armado / Joel de Conceição
Nogueira Diniz. - 2025.

105 p.

Coorientador(a) 1: Sandra Cristina Alves Pereira da
Silva Cunha.

Orientador(a): Anselmo Cardoso de Paiva.

Tese (Doutorado) - Programa de Pós-graduação Doutorado
em Ciência da Computação - Associação UFMA/UFPI,
Universidade Federal do Maranhão, São Luís, 2025.

1. Patologias No Concreto. 2. Aprendizado Profundo.
3. Visão Computacional. 4. Detecção. 5. Classificação.
I. Cardoso de Paiva, Anselmo. II. Cristina Alves Pereira
da Silva Cunha, Sandra. III. Título.

Joel de Conceição Nogueira Diniz

Ensemble de Redes Convolucionais para Inspeção Visual de Estruturas de Concreto Armado

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Anselmo Cardoso de Paiva
Orientador
Universidade Federal do Maranhão

Profa. Dra. Sandra Cristina Alves Pereira da Silva Cunha
Co-orientadora
Universidade de Trás-os-Montes e Alto Douro

Prof. Dr. Geraldo Braz Junior
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. João Dallyson Sousa de Almeida
Examinador Interno
Universidade Federal do Maranhão

Profa. Dra. Aura Conci
Examinador Externa
Universidade Federal Fluminense

Prof. Dr. Claudio Baptista
Examinador Externa
Universidade Federal de Campina Grande

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequada para obtenção do título de Doutor em Ciência da Computação.

ANSELMO
CARDOSO DE
PAIVA:37552384387

Assinado de forma digital por
ANSELMO CARDOSO DE
PAIVA:37552384387
Dados: 2025.11.25 14:46:48
-03'00'

Prof. Dr. Anselmo Cardoso de Paiva
Coordenador

São Luís - MA, 31 de Outubro de 2025

*Aos meus amigos e familiares, que sempre estão na torcida pelo meu sucesso, e
sobretudo a Deus!*

Agradecimentos

Agradeço a todos os amigos e familiares que, direta ou indiretamente, contribuíram com todo suporte necessário para eu desenvolver este trabalho.

Agradeço ainda aos amigos e familiares pelo suporte ao que sou, pois tudo que consigo é fruto desta enorme colaboração.

Ao meu Orientador e Professor Doutor Anselmo Cardoso de Paiva pela orientação segura e paciência no desenvolvimento deste trabalho, bem como, pela fundamental contribuição com meu aprendizado, desde a graduação, e atualmente no Doutorado.

À Professora Doutora Sandra Cristina Alves Pereira da Silva Cunha pela co-orientação e todo conhecimento compartilhado, com muita paciência, e fundamental apoio para o doutorado.

A FCT (Fundação Portuguesa de Ciência e Tecnologia) todo o apoio prestado, através dos projetos 2024.AIVLAB.00002.VISION e 2025.AIVLAB.00008.DEUCALION, Laboratório virtual de Inteligência Artificial.

Ao meu Professor Doutor Geraldo Braz Júnior pelas importantes contribuições e palavras de apoio, especialmente pelas aulas de Inteligência Artificial, com especial direcionamento na área de Visão Computacional.

Ao CNPq, à UFMA e à UFPI, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus colegas da UFMA e da UFPI.

Aos professores membros da banca examinadora.

A todos os professores e funcionários do DCCMAPI pelos ensinamentos e pela colaboração.

A todo mundo que, de uma forma ou de outra, me estimulou, ou me ajudou.

"O verdadeiro caminho é sempre o amor, sendo a dor a trilha que reconduz os que se perderam."

(Joel Diniz)

Resumo

Patologias em estruturas de concreto podem ser evidenciadas visualmente em sua superfície, tais como fissuras ou trincas, fragmentação de parte do concreto, eflorescência, manchas de corrosão, e barras de aço expostas, sendo estes dois últimos ocorrência em concreto armado. Portanto, estas patologias podem ser analisadas a partir de imagens de estruturas de concreto armado. Esta tese propõe um ensemble de redes convolucionais para inspeção visual de estruturas de concreto armado. Este método permite agilizar a tarefa de detecção e aumentar sua eficácia, por meio do ganho de tempo nas identificações a serem analisadas, e eliminação, ou redução, de erros, como os advindos de falhas humanas mediante execução massiva de tarefas tediosas nas análises. A tarefa de identificação de patologia pode ocorrer mediante uso de uma rede neural convolucional, caso as imagens estejam recortadas para a patologia específica a ser identificada, ou, com uso de rede de detecção, caso seja uma imagem ampla onde a patologia está inserida em um contexto com diversas classes de patologia, e mesmo área sem patologia. Outra tarefa que possibilita a identificação e a análise destas patologias é a segmentação. O método foi testado nas tarefas de detecção e classificação. As arquiteturas de rede neural utilizadas para detecção foram a YOLO v11 e a TOOD (Task-aligned One-stage Object Detection), na abordagem de redes neurais de um estágio, e a Faster R-CNN, na abordagem de redes neurais de dois estágios. As três redes foram posteriormente combinadas para a aplicação do Ensemble, sendo utilizada Weighted Box Fusion. Já no caso da tarefa de classificação, as arquiteturas de redes utilizadas foram DenseNet121, ResNet50 e MobileNeV3. Realizou-se uma pesquisa para identificar e selecionar conjuntos de dados adequados ao método proposto. O dataset selecionado para classificação é o Ozgnel e o de detecção é o CODEBRIM. A tarefa de detecção consiste em localizar um artefato e classificá-lo. Embora esta abordagem seja eficiente, a pesquisa testou a detecção para a localização de patologias, sem definir inicialmente a qual classe pertencem, e, posteriormente, usar uma rede neural específica de classificação para definir os tipos de patologias e, sobretudo, permitir a eliminação de falsos positivos. A abordagem de combinar rede de detecção e rede dedicada de classificação proporcionou o resultado pretendido, potencializado por Ensemble, que aumentou a sensibilidade na fase de detecção e elevou o número de artefatos localizados. Essa abordagem permite identificar falsos positivos na primeira fase, mas esses são eliminados na última fase.

Palavras-chave: Patologias no Concreto, Aprendizado Profundo, Visão Computacional, Detecção. Classificação.

Abstract

Pathologies in concrete structures can be visually detected on their surface, such as cracks or fissures, fragmentation of concrete, efflorescence, corrosion stains, and exposed steel bars (the latter two occurring in reinforced concrete). Therefore, these pathologies can be analyzed via images of reinforced concrete structures. This thesis proposes an ensemble of convolutional networks for visual inspection of reinforced concrete structures. This method speeds up the detection task and increases its effectiveness by saving time on the identification to be analyzed and by eliminating or reducing errors, such as those arising from human error during the massive execution of tedious tasks in the analysis. The task of identifying pathology can be performed using a convolutional neural network if the images are cropped to the specific pathology to be identified, or using a detection network if the images are broad and the pathology is inserted into a context with several classes of pathology, including areas without pathology. Another task that enables the identification and analysis of these pathologies is segmentation. The method was tested with detection and classification tasks. The neural network architectures used for detection were YOLO v11 and TOOD (Task-aligned One-stage Object Detection) for the single-stage approach, and Faster R-CNN for the two-stage approach. The three networks were then combined for the Ensemble application, using Weighted Box Fusion. For the classification task, the network architectures used were DenseNet121, ResNet50, and MobileNetV3. Research was conducted to locate and select appropriate datasets for the proposed method. The dataset selected for classification is Ozgnet, and the one for detection is CODEBRIM. The detection task allows an artifact to be located and classified. Although this approach is efficient, the research tested using detection to locate pathologies without initially defining which class they belong to, and then using a specific classification neural network to define the types of pathologies and, above all, to eliminate false positives. The approach of combining a detection network and a dedicated classification network achieved the desired result, with Ensemble used to increase sensitivity in the detection phase, thereby increasing the number of artifacts located. This approach allows false positives to pass in the first phase, but they are eliminated in the last phase.

Keywords: Deep Learning, Computer Vision, Detection. Classification, Concrete Pathologies.

Lista de ilustrações

Figura 1 – Fissura/Rachadura. Adaptado de (CODEBRIM 2019).	23
Figura 2 – Barra de aço exposta. Adaptado de (CODEBRIM 2019).	25
Figura 3 – Fragmentação de parte do concreto. Adaptado de (CODEBRIM 2019).	25
Figura 4 – Eflorescência na superfície do concreto. Adaptado de (CODEBRIM 2019).	26
Figura 5 – Mancha de corrosão do aço na superfície do concreto. Adaptado de (CODEBRIM 2019).	26
Figura 6 – Rede Neural Convolutacional. Elaborada pelo autor.	29
Figura 7 – Arquitetura YOLO (Rao 2024).	31
Figura 8 – Modelo YOLO (Redmon et al. 2016).	32
Figura 9 – Arquitetura Faster R-CNN (Ren et al. 2016).	34
Figura 10 – Arquitetura TOOD (Feng et al. 2021).	35
Figura 11 – Arquitetura DenseNet121 (Huang et al. 2017).	37
Figura 12 – Construção de bloco de aprendizagem residual da ResNet50 (He et al. 2016).	38
Figura 13 – Melhoria da MobileNetV3, (He et al. 2016).	40
Figura 14 – Método proposto, com etapa de detecção seguida da etapa de classificação. Elaborada pelo autor.	59
Figura 15 – Conversão da bounding box do dataset CODEBRIM para YOLO. Elaborada pelo autor.	69
Figura 16 – Anotação incorreta devido ao giro na imagem. Elaborada pelo autor.	70
Figura 17 – Anotação imprecisa de mancha de corrosão. Vermelho a anotação original e azul a correção. Elaborada pelo autor.	71
Figura 18 – Anotação de fissura, ora feita por ramo, ora feita por conjunto de fissuras. Vermelho a anotação original e azul a correção. Elaborada pelo autor.	71
Figura 19 – Anotação seccionada de fissura. Vermelho a anotação original e azul a correção. Elaborada pelo autor.	72
Figura 20 – Anotações de fissuras que não foram realizadas. Vermelho a anotação original e azul a correção. Elaborada pelo autor.	72
Figura 21 – Anotação que trata a imagem inteira como um tipo de patologia, confundindo detecção com classificação. Vermelho, a anotação original; azul, a correção. Elaborada pelo autor.	73
Figura 22 – Imagens dos dois tipos de classes do conjunto de dados (Ozgenel 2019), com e sem fissuras. Elaborada pelo autor.	73

Figura 23 – Estrutura e configuração de dados usadas para a CNN. Elaborada pelo autor.	74
Figura 24 – Resultado de classificação para duas imagens das duas classes do conjunto de dados (Ozgenel 2019).	74
Figura 25 – Acurácia obtida pela CNN em uma das etapas de validação cruzada. Elaborada pelo autor.	75
Figura 26 – Valor da função de perda na arquitetura CNN, durante o treinamento. Elaborada pelo autor.	75
Figura 27 – ROC AUC. Elaborada pelo autor.	76
Figura 28 – Classificação negativa para fissura. Elaborada pelo autor.	76
Figura 29 – Classificação positivo para fissura. Elaborada pelo autor.	77
Figura 30 – Imagem de patologias no concreto antes da detecção com YOLO (CODEBRIM 2019). Elaborada pelo autor.	93
Figura 31 – Imagem de patologias no concreto após detecção com YOLO (CODEBRIM 2019). Elaborada pelo autor.	94
Figura 32 – Falso positivo. Elaborada pelo autor.	95
Figura 33 – Detecção positiva. Elaborada pelo autor.	97

Lista de tabelas

Tabela 1 – Artigos relacionados ao tema da tese	55
Tabela 2 – Conjuntos de dados pesquisados.	56
Tabela 3 – Matriz de confusão para 4999 casos positivos e 4999 negativos. . .	75
Tabela 4 – Métricas de classificação e detecção com YOLO no CODEBRIMRev. .	81
Tabela 5 – Matriz de confusão para detecção de manchas de corrosão com YOLO no CODEBRIMRev.	81
Tabela 6 – Matriz de confusão para detecção de fissuras com YOLO no CODEBRIMRev.	81
Tabela 7 – Matriz de confusão para detecção de eflorescências com YOLO no CODEBRIMRev.	81
Tabela 8 – Matriz de confusão para detecção de barras expostas com YOLO no CODEBRIMRev.	82
Tabela 9 – Matriz de confusão para detecção de lascamentos com YOLO no CODEBRIMRev.	82
Tabela 10 – Métricas de classificação e detecção com Rede Faster R-CNN no CODEBRIMRev.	83
Tabela 11 – Matriz de confusão para a detecção de manchas de corrosão com Faster R-CNN no CODEBRIMRev.	83
Tabela 12 – Matriz de confusão para detecção de fissuras com Faster RCNN no CODEBRIMRev.	83
Tabela 13 – Matriz de confusão para detecção de eflorescências com Faster RCNN no CODEBRIMRev.	84
Tabela 14 – Matriz de confusão para detecção de barras expostas com Faster RCNN no CODEBRIMRev.	84
Tabela 15 – Matriz de confusão para detecção de lascamento com Faster RCNN no CODEBRIMRev.	84
Tabela 16 – Métricas de classificação/detecção com TOOD (CODEBRIMRev). .	85
Tabela 17 – Matriz de confusão para detecção de manchas de corrosão com TOOD no CODEBRIMRev.	85
Tabela 18 – Matriz de confusão para detecção de fissuras com TOOD no CODEBRIMRev.	85
Tabela 19 – Matriz de confusão para detecção de eflorescências com TOOD no CODEBRIMRev.	85
Tabela 20 – Matriz de confusão para detecção de barras expostas com TOOD no CODEBRIMRev.	86

Tabela 21 – Matriz de confusão para detecção de lascamento com TOOD no CODEBRIMRev.	86
Tabela 22 – Métricas para classificação com DenseNet121 no CODEBRIMRec.	89
Tabela 23 – Métricas para classificação com ResNet50 no CODEBRIMRec.	90
Tabela 24 – Métricas para classificação com MobileNetV3 no CODEBRIMRec.	90
Tabela 25 – Resumo dos experimentos para a tarefa de detecção.	91
Tabela 26 – Resumo dos experimentos para a tarefa de classificação com detecção.	91
Tabela 27 – Resumo dos experimentos para a tarefa de classificação.	92
Tabela 28 – Aplicação do método duas etapas para identificar fissuras.	92
Tabela 29 – Resultados da tarefa de classificação do tipo de patologia em imagens recortadas.	93

Lista de abreviaturas e siglas

CNN	<i>Convolutional Neural Network</i>
DB	<i>Database</i>
mAP	<i>mean Average Precision</i>
ROI	<i>Region of Interest</i>
YOLO	<i>You Only Look Once</i>

Sumário

1	INTRODUÇÃO	16
2	FUNDAMENTAÇÃO TEÓRICA	22
2.0.1	Patologias de Concreto Armado	22
2.1	Visão Computacional e Aprendizado Profundo	27
2.2	Rede Neural Convolutional	29
2.2.1	YOLO, You Only Look Once	30
2.2.2	Faster R-CNN	32
2.2.3	TOOD, Task-aligned One-stage Object Detection	34
2.2.4	DenseNet121	36
2.2.5	ResNet50	38
2.2.6	MobileNetV3	40
2.2.7	Análise Comparativa e Evolutiva das Arquiteturas de CNN.	41
2.3	Ensemble	42
3	TRABALHOS RELACIONADOS	46
4	METODOLOGIA	58
4.1	Pré-processamento de Imagem	59
4.2	Deteccção e Classificação de Patologias em Estrutura de Concreto Armado	62
4.3	Avaliação	64
5	RESULTADOS	66
5.1	Conjunto de Dados	67
5.1.1	Ajustes dos Conjuntos de Dados	68
5.2	Classificação com conjunto de dados Ozgenel	73
5.3	Deteccção e Classificação de Patologias em uma Etapa	77
5.3.1	Deteccção e Classificação de Patologias com Arquitetura YOLO - Conjunto de dados CODEBRIM	77
5.3.2	Deteccção e Classificação de Patologias com Arquitetura Faster RCNN - Conjunto de dados CODEBRIM	78
5.3.3	Deteccção e Classificação de Patologias com Arquitetura TOOD - Conjunto de dados CODEBRIM	79
5.3.4	Associação dos Resultados com Ensemble	79

5.3.4.1	Detecção com Ensemble Usando Weighted Box Fusion - Conjunto de dados CODEBRIM	79
5.3.5	Detecção e Classificação de Patologias com Arquitetura YOLO - Conjunto de dados CODEBRIMRev	80
5.3.5.1	Experimentos considerando diferentes Thresholds - CODEBRIMRev	81
5.3.6	Detecção e Classificação de Patologias com Arquitetura Faster RCNN - Conjunto de dados CODEBRIMRev	82
5.3.7	Detecção e Classificação de Patologias com Arquitetura TOOD - Conjunto de dados CODEBRIMRev	83
5.4	Identificação de patologias em duas etapas	85
5.4.1	Detecção e Classificação de Patologias com Arquitetura YOLO - Uma Classe - Conjunto de dados CODEBRIMRev	86
5.4.2	Detecção e Classificação de Patologias com Arquitetura Faster RCNN - Uma Classe - Conjunto de dados CODEBRIMRev	87
5.4.3	Detecção e Classificação de Patologias com Arquitetura TOOD - Uma Classe - Conjunto de dados CODEBRIMRev	87
5.5	Detecção com Ensemble - Conjunto CODEBRIMRev	88
5.5.1	Detecção com Ensemble Usando Weighted Box Fusion - Conjunto de dados CODEBRIMRev	88
5.5.2	Detecção com Ensemble Usando Non-maximum Suppression - Conjunto de dados CODEBRIMRev	88
5.6	Classificação de Patologias com Conjunto de Dados Recortados	89
5.6.1	Classificação com DenseNet121 no CODEBRIMRec	89
5.6.2	Classificação com ResNet50 no CODEBRIMRec	89
5.6.3	Classificação com MobileNetV3 no CODEBRIMRec	90
5.7	Resumo dos Resultados Obtidos	90
5.8	Aplicação da Metodologia Duas Etapas para Identificar Fissuras	90
5.9	Discussões	92
6	CONCLUSÃO	98
	REFERÊNCIAS	101

1 Introdução

A construção civil tem grande relevância por sua relação com os centros urbanos e por representar um setor importante para a economia, gerando investimentos e empregos (Wight e MacGregor 2016). Um dos desafios desta área é a manutenção de edifícios. Uma atividade fundamental por sua relação direta com a segurança das pessoas. A manutenção inadequada, ou mesmo inexistente, pode resultar em acidentes com pessoas, até em casos graves e fatais, e na perda de uma edificação.

O concreto armado é o segundo material de construção mais utilizado no mundo, proporcionando uma construção versátil, durável, acessível, funcional e atraente (Taheri 2019). As propriedades de alta resistência à compressão do concreto permitem que esse material resista a essas tensões tão presentes nas estruturas, sendo sua deficiência na resistência à tração compensada por sua associação ao aço (Wight e MacGregor 2016). Ante o concreto, mesmo no local onde será utilizado, confere versatilidade, tornando-o uma das principais alternativas na construção civil. Material de construção durável, mas que pode deteriorar-se devido à exposição a condições severas associadas ao meio ambiente, carregamento, efeitos de ações agressivas, corrosão de metais incorporados, resistência do concreto a mudanças de volume, abrasão/erosão e ações químicas.

Embora o concreto apresente enormes vantagens, ao longo da vida útil da estrutura, está sujeito a patologias que podem comprometer sua resistência. A manutenção é, portanto, essencial para estruturas de concreto armado, dada a sua ampla utilização e os impactos que um eventual colapso destas estruturas pode acarretar. O grande risco decorrente destas patologias é a perda das características estruturais desejadas, como redução da resistência a níveis inferiores aos calculados. Por este motivo, estruturas de concreto armado devem ser monitoradas para prevenir o aparecimento ou agravamento de patologias que podem desenvolver-se ao longo de sua vida útil.

Uma patologia no concreto armado é qualquer manifestação de deterioração, defeito ou anomalia que compromete a integridade estrutural, a durabilidade, a funcionalidade ou os aspectos estéticos da construção, podendo incluir fissuras, corrosão das armaduras, carbonatação, eflorescências, deslocamentos, segregação do concreto ou deformações excessivas, geralmente resultantes de falhas no projeto, na execução, de materiais inadequados ou de falta de manutenção adequada (Souza e Ripper 1998). Estas patologias citadas são detalhadas na seção de revisão da literatura desta tese.

O processo de manutenção de estruturas de concreto armado deve seguir os

requisitos estabelecidos pela NBR 5674:2012 - Manutenção de edificações - Requisitos para o sistema de gestão de manutenção ([ABNT 2012](#)), que define a manutenção como o conjunto de atividades necessárias para conservar ou recuperar a capacidade funcional da edificação e de seus sistemas. Um dos aspectos principais deste processo de manutenção é a inspeção visual da estrutura, que deve ser realizada rotineiramente para avaliar o estado de conservação das estruturas de concreto e fornecer diretrizes para uma manutenção eficaz e de custo-benefício adequado. Na inspeção visual deve-se registrar, quantificar, dimensionar e localizar as possíveis anomalias observadas nos elementos estruturais. Esta inspeção permite identificar danos causados por fatores como corrosão, fadiga, desgaste ou outros eventos adversos, avaliar o desempenho estrutural em relação aos padrões de segurança e às normas aplicáveis, determinar a necessidade de manutenção preventiva ou de reparos e estabelecer um histórico de condições para monitorar mudanças ao longo do tempo. A NBR 5674 exige que as inspeções sejam realizadas por meio de modelos elaborados e ordenados, considerando um roteiro de inspeção dos sistemas e componentes da edificação, as formas de manifestação esperadas da degradação natural associadas à vida útil e as solicitações dos usuários, contribuindo, assim, para a detecção precoce de problemas e para a prevenção de falhas estruturais mais graves.

A inspeção visual de superfícies é, portanto, um importante método para determinar a condição física e funcional de uma estrutura de concreto armado. No entanto, muitos incidentes continuam ocorrendo devido a inspeções e avaliações de condições inadequadas. Se as patologias das estruturas não forem tratadas no início de sua manifestação, alguns dos problemas mais comuns em estruturas podem resultar em graves defeitos, chegando ao colapso total; portanto, é um tema muito relevante. Esta inspeção é baseada na análise visual da superfície da estrutura, na qual se evidencia a maioria das patologias do concreto ([Safiuddin et al. 2018](#)). Diversas patologias, portanto, podem ser analisadas por meio de análise visual da superfície, já que seus efeitos podem aparecer na superfície na forma de fissuras ou rachaduras, fragmentação de parte do concreto, eflorescências, manchas de corrosão e barras de aço expostas.

Existe, portanto, a necessidade de aprimorar os métodos de inspeção visual de estruturas de concreto, visando a detecção e análise de patologias. Entre as técnicas para monitorar estruturas de concreto e identificar a ocorrência de patologias, podemos destacar a inspeção visual por meio da análise de imagens. Neste caso, ocorre a captura de imagens das superfícies da estrutura por um profissional em campo. O diagnóstico da patologia deve ser aprofundado por meio de análise visual de imagens, realizada por um especialista. Muitas vezes, as imagens precisam ser enviadas a profissionais que não estão no local da obra, pois esses profissionais especializados podem não estar disponíveis na região. O recebimento das imagens permite reduzir gastos com deslocamentos ou até suprir a falta de disponibilidade do profissional que

fará o diagnóstico.

Esta análise das imagens, realizada por um humano, consome muito tempo e pode representar um risco de segurança, pois pode resultar em um grande número de imagens, e a análise de todas essas imagens por um profissional pode resultar em erro humano, devido à fadiga em tarefa maçante e repetitiva. Além disso, é um método subjetivo às habilidades, aos conhecimentos e à experiência dos profissionais que a realizam.

A visão computacional pode colaborar em atividades monótonas e repetitivas, resultando em economia de tempo e redução de erros. Em razão disso, métodos de inspeção automatizados podem ser desenvolvidos por meio da aquisição de imagens e de algoritmos de aprendizado de máquina para superar essas limitações. Essa automação pode resultar em um método mais econômico e eficiente e reduzir o risco de acidentes, (Tian et al. 2025).

Além disso, a inspeção automatizada pode ser mais objetiva e confiável, pois, por ser definida por um algoritmo, é reproduzível. Muitos estudos, como os apresentados em (Shamsabadi et al. 2024), investigaram a possibilidade de otimizar a inspeção automatizada de estruturas de concreto por meio de métodos de aprendizagem de máquina, mais especificamente de aprendizagem profunda.

Em geral, esses métodos baseiam-se em aprendizado supervisionado, o que requer a construção de bases de imagens anotadas que permitam a formação de modelos para a detecção de ocorrência de patologias e a classificação do tipo de patologia detectada.

A anotação das patologias em imagens é uma tarefa manual tediosa (CODEBRIM 2019), o que dificulta não apenas a disponibilidade de bases de dados de imagens anotadas, mas também limita o número de imagens anotadas nessas bases. Esta questão é agravada no caso de detecção de patologias, por se tratar de uma tarefa ainda mais difícil e que requer maior cuidado. A anotação correta depende ainda, fundamentalmente, do conhecimento técnico de quem a realiza e da disponibilidade de profissionais especialistas em estruturas de concreto.

Esta tese de doutorado pretende propor, desenvolver e validar uma metodologia computacional para inspeção visual automatizada de estruturas de concreto armado, fundamentada em técnicas de visão computacional e aprendizado profundo por meio de redes neurais convolucionais, que seja capaz de detectar, localizar e classificar automaticamente diferentes tipos de patologias estruturais a partir de imagens fotográficas, contribuindo para o avanço científico na área de monitoramento estrutural e fornecendo uma ferramenta tecnológica para auxiliar engenheiros e profissionais da construção civil nos processos de inspeção, diagnóstico, manutenção preventiva e

garantia de segurança de estruturas de concreto armado, com potencial de aplicação em larga escala para o monitoramento da infraestrutura civil.

Para o alcance do objetivo geral proposto, definem-se os seguintes objetivos específicos que delineiam as etapas metodológicas e investigativas do desenvolvimento desta tese:

- Desenvolver modelos computacionais baseados em redes neurais convolucionais para detecção automática de patologias em superfícies de concreto armado, com foco na identificação precisa das regiões afetadas e na localização espacial das anomalias estruturais, garantindo alta acurácia na delimitação das áreas de interesse para posterior análise detalhada.
- Implementar métodos para a classificação automatizada das regiões de interesse detectadas, categorizando-as conforme os tipos específicos de patologias definidos para a inspeção visual de superfícies de concreto, considerando as particularidades e características visuais de cada tipo de deterioração estrutural.
- Analisar e demonstrar o impacto do uso de redes neurais especializadas em classificação na redução de falsos positivos na etapa de detecção, investigando as vantagens de bases de dados dedicadas exclusivamente à classificação e evidenciando os benefícios desta abordagem em duas etapas para a melhoria da confiabilidade do sistema proposto.
- Implementar e avaliar arquiteturas integradas de redes neurais convolucionais para detecção e classificação simultâneas de múltiplas patologias em imagens de estruturas de concreto armado, otimizando o processo de inspeção visual automatizada por meio da integração de diferentes abordagens de aprendizado profundo em uma única metodologia coesa.
- Investigar e comparar o desempenho de diferentes arquiteturas de redes neurais convolucionais, incluindo modelos de um e de dois estágios, para identificar a abordagem mais eficiente e adequada para a detecção e classificação de patologias específicas do concreto armado, considerando métricas de precisão, recall, velocidade de processamento e recursos computacionais necessários.
- Implementar e avaliar estratégias de transferência de aprendizado e ensemble de classificadores visando a melhoria do desempenho dos modelos propostos, e a mitigação de limitações relacionadas à escassez de conjuntos de dados anotados disponíveis na literatura especializada, maximizando a capacidade de generalização dos modelos desenvolvidos.

- Propor e avaliar melhorias metodológicas para o treinamento das redes neurais mediante revisão crítica de datasets existentes, implementação de tecnologias mais recentes de aprendizado profundo, técnicas de aumento de dados e outras estratégias de otimização que possam contribuir para o avanço do estado da arte na área de inspeção automatizada de estruturas de concreto armado.

O desenvolvimento desta tese resultou em contribuições para a área de visão computacional aplicada à engenharia civil e para a prática profissional de inspeção de estruturas de concreto armado.

A principal contribuição consiste na proposição de ensemble de redes para detecção automatizada de patologias em estruturas de concreto armado, fundamentadas em redes neurais convolucionais de arquitetura two-stages e one-stage, que demonstraram eficácia promissora comparável à dos métodos convencionais de inspeção visual manual. Esta abordagem representa um avanço no estado da arte ao integrar técnicas de aprendizado profundo, especificamente adaptadas às particularidades das patologias estruturais do concreto armado, proporcionando maior precisão e confiabilidade na identificação e na classificação de anomalias.

Do ponto de vista da eficiência operacional, a metodologia desenvolvida proporciona ganhos no tempo de análise de imagens das superfícies das estruturas, contribuindo para a transformação de processos manuais em operações automatizadas que podem ser executadas em menor tempo. Este aspecto contribui não apenas para a otimização dos recursos humanos necessários, mas também viabiliza a implementação de programas de monitoramento estrutural mais frequentes e abrangentes.

A automatização do processo de inspeção contribui significativamente para reduzir erros associados à fadiga e à subjetividade inerentes às análises manuais. A eliminação da necessidade de operações repetitivas e maçantes, características da inspeção visual tradicional quando aplicada a grandes volumes de dados, reduz a ocorrência de falsos negativos e positivos, resultando em diagnósticos mais precisos e confiáveis. Esta melhoria na qualidade dos resultados tem implicações diretas para a segurança estrutural e para a tomada de decisões de manutenção preventiva.

A capacidade de processamento automatizado de grandes quantidades de imagens constitui outra contribuição desta tese. Esta característica torna viável a aplicação da metodologia em cenários de monitoramento de infraestrutura civil em escala, em que o volume de dados tornaria a análise manual impraticável. A escalabilidade da solução proposta abre perspectivas para o desenvolvimento de sistemas de monitoramento estrutural em tempo real e para a implementação de programas de manutenção preditiva baseados em imagens.

Adicionalmente, a pesquisa contribui para o fortalecimento da base científica da

área ao utilizar estrategicamente bases de dados disponíveis na literatura especializada, validando e comparando diferentes abordagens metodológicas. Este trabalho tem potencial para estimular o desenvolvimento de novos recursos computacionais e promover a revisão crítica e aprimoramento dos bancos de dados existentes, fomentando a colaboração científica e o compartilhamento de recursos na comunidade acadêmica.

As contribuições desta tese, além dos aspectos puramente técnicos, oferecem um cenário de aplicação que pode ser incorporado aos processos de trabalho de profissionais da engenharia civil, consultores estruturais e gestores de infraestrutura, promovendo uma cultura de monitoramento baseada em evidências científicas e tecnologia.

A presente tese encontra-se estruturada de modo que o capítulo 2 estabelece os fundamentos teóricos essenciais ao desenvolvimento da pesquisa, abordando de forma abrangente as patologias características de estruturas de concreto armado e os princípios do aprendizado profundo. Particular ênfase é dada às redes neurais convolucionais (CNNs), considerando sua aplicação central nas metodologias propostas. Subsequentemente, o capítulo 3 apresenta uma revisão crítica e sistemática dos trabalhos relacionados ao escopo desta pesquisa, contextualizando o estado da arte e identificando as lacunas científicas que motivaram o desenvolvimento da metodologia aqui proposta. O capítulo 4 detalha as metodologias desenvolvidas nesta tese, com especial destaque para a arquitetura baseada em ensemble de classificadores, descrevendo minuciosamente os procedimentos experimentais, as técnicas implementadas e os critérios de avaliação adotados. A apresentação e a discussão dos resultados obtidos nos experimentos realizados para validar a metodologia proposta constituem o objeto do capítulo 5. Neste capítulo, procede-se à análise comparativa dos resultados alcançados com os reportados na literatura especializada, estabelecendo-se o posicionamento da presente contribuição no contexto científico atual. Finalmente, o capítulo 6 sintetiza as principais conclusões a partir dos resultados obtidos, avalia o cumprimento dos objetivos propostos e delinea perspectivas para trabalhos futuros que dêem continuidade e ampliar as contribuições desta pesquisa.

2 Fundamentação Teórica

Neste capítulo, apresentamos os principais conceitos relacionados à tese, abordando estruturas de concreto armado e suas patologias, bem como as principais arquiteturas de aprendizado profundo utilizadas como base para a tese, com os respectivos modelos de avaliação.

2.0.1 Patologias de Concreto Armado

O concreto é um material muito utilizado na construção civil, especialmente na função estrutural, onde há esforços de compressão, por ser esta a principal função do material. Outra característica que se destaca no concreto é sua capacidade de ser moldado, inclusive *in loco*. Uma deficiência deste material é sua baixa resistência a esforços de tração e, nesse caso, ele é usado em combinação com o aço, que é especialmente resistente à tração. A esta combinação denomina-se concreto armado (Carvalho 2022).

Apesar das vantagens estruturais e construtivas do concreto armado, este material está sujeito a processos de deterioração ao longo de sua vida útil, decorrentes de fatores ambientais, carregamentos, deficiências construtivas ou de projeto. O monitoramento da integridade estrutural torna-se, portanto, fundamental para garantir a segurança e a funcionalidade das edificações e das infraestruturas. Estes processos de deterioração podem evoluir e culminar no surgimento de patologias estruturais, que comprometem não apenas a estética das construções, mas, principalmente, sua capacidade de resistência e durabilidade, exigindo intervenções corretivas que podem ser onerosas quando não identificadas precocemente.

Diversas patologias do concreto são evidenciadas por marcas características expostas à superfície. Isto permite identificar problemas no concreto a partir de sinais característicos em suas superfícies, como manchas e trincas. Cada tipo de manifestação superficial pode levar a conclusões sobre possíveis problemas e até indicar a gravidade da condição já estabelecida, (Souza e Murta 2012).

Dado que essas manifestações apresentam características visuais distintas e específicas, torna-se fundamental não apenas detectar a presença, mas também realizar um diagnóstico que permita identificar o tipo de problema presente na estrutura. Esta etapa de diagnóstico é crucial, pois diferentes patologias demandam abordagens de intervenção na estrutura e adequadas à natureza do defeito identificado. Cada tipo de patologia pode exigir um tratamento diferente, e quanto mais cedo um problema for identificado, maiores serão as chances de resolvê-lo antes que comprometa a estrutura

de forma perigosa.

O registro por meio de fotos e vídeos é uma forma de documentar os problemas identificados em inspeções *in loco*, para análise imediata ou posterior. Esses registros podem ser utilizados para criar um histórico e avaliar a evolução da patologia. O estudo pode servir para a estrutura analisada e fornecer material para embasar ocorrências semelhantes em outras estruturas, além de, conforme citado, possibilitar a análise em outro momento e até mesmo por diferentes profissionais em localidades distintas.

Diversas patologias podem se manifestar no concreto, sendo as principais fissuras ou rachaduras, fragmentação de parte do concreto, eflorescências, manchas de corrosão e barras de aço expostas. É importante ressaltar que estas patologias não raro ocorrem de forma associada, simultaneamente, ou como causadoras de outras, como as fissuras e rachaduras, que favorecem a corrosão das armaduras e que podem evoluir para outra patologia, como o lascamento no concreto.

Dentre as patologias mais comuns, destacam-se as fissuras e rachaduras. Estas consistem em aberturas patológicas com espessuras reduzidas em relação à sua extensão. Rachaduras e fissuras podem ser classificadas de acordo com sua forma e o tamanho da abertura. A norma brasileira NBR 9575 (NBR 9575, 2003) faz distinção entre fissuras e rachaduras, conforme o tamanho da abertura, considerando o limite de 0,5 mm, sendo rachadura acima deste valor. Esta tese utiliza as fissuras e rachaduras (trincas) como uma única patologia.

As fissuras representam uma das principais preocupações no acompanhamento da vida útil do concreto. A Figura 1 apresenta um exemplo de fissura.



Figura 1 – Fissura/Rachadura. Adaptado de (CODEBRIM 2019).

As fissuras podem ser observadas conforme a forma como se manifestam: geométricas ou isoladas; mapeadas ou disseminadas. As fissuras geométricas ou isoladas apresentam-se como aberturas individuais com padrões definidos, geralmente seguindo direções específicas relacionadas aos esforços atuantes na estrutura. Já as fissuras mapeadas ou disseminadas caracterizam-se por formar uma rede interconectada de pequenas aberturas que se espalham pela superfície do concreto, assemelhando-se a um mapa ou a uma teia. Conforme a sua atividade, podem ser ativas (vivas) ou passivas. As fissuras ativas são aquelas que continuam em processo de abertura ou fechamento, indicando que as causas que as originaram ainda estão presentes e atuantes. As fissuras passivas, por sua vez, são aquelas que se estabilizaram e não apresentam mais variações significativas em sua abertura ao longo do tempo. As fissuras também podem ser observadas segundo a variação de sua abertura em "sazonal" ou "progressiva". As fissuras sazonais são aquelas cuja abertura varia de acordo com condições ambientais específicas, como temperatura e umidade, tendendo a se abrir e fechar ciclicamente ao longo das estações do ano. As fissuras progressivas apresentam abertura contínua e crescente ao longo do tempo, independentemente de variações sazonais, o que indica um processo de deterioração em andamento.

A inspeção visual realizada por um profissional é um dos principais métodos para determinar a condição física e funcional de uma estrutura de concreto armado. Esta inspeção deve ser realizada regularmente para garantir que a estrutura continue a funcionar conforme o pretendido. No entanto, muitos incidentes continuam devido a inspeções e avaliações de condições inadequadas. Outras patologias também devem ser identificadas e tratadas para a boa saúde das estruturas de concreto, tais como manchas de corrosão do aço no concreto, fragmentação do concreto, eflorescência do concreto e barras de aço expostas no concreto armado.

Para uma compreensão mais abrangente dos defeitos que podem comprometer a integridade estrutural, apresentam-se, a seguir, as principais tipologias de patologias que afetam as estruturas de concreto armado, detalhando suas características, causas e implicações para a segurança estrutural.

O aço é utilizado no concreto armado para acrescentar características de resistência à tração. O aço do concreto armado precisa estar recoberto pelo concreto, visando a proteção contra a corrosão. A exposição do aço é, portanto, uma patologia e pode ser fruto do processo construtivo ou decorrente de outra patologia, como as rachaduras, que podem resultar no deslocamento da cobertura do aço. A figura 2 mostra um caso de barra de aço exposta.



Figura 2 – Barra de aço exposta. Adaptado de (CODEBRIM 2019).

A fragmentação do concreto pode ter início em outra patologia, como fissura ou rachadura, podendo também ser consequência da expansão do aço submetido a corrosão ou mesmo da perda da estrutura que sofreu algum tipo de impacto. A fragmentação pode originar problemas graves, visto que pode resultar em perda significativa de massa de concreto, reduzindo a massa necessária para resistir à compressão, bem como expondo o aço a intempéries, ambas com risco de colapso da estrutura. A Figura 3 traz uma imagem de concreto com fragmentação.



Figura 3 – Fragmentação de parte do concreto. Adaptado de (CODEBRIM 2019).

A Figura 4 mostra um caso de eflorescência, caracterizada por manchas brancas na superfície do concreto, decorrentes do acúmulo de depósitos cristalininos. Estas manchas são resultantes da evaporação de soluções aquosas salinizadas, produtos de reações químicas, causadas por água que se infiltra e dissolve os sais presentes no cimento e na cal. Além do problema de estética na superfície, sua presença pode, portanto, sinalizar a ocorrência de infiltração de água na estrutura, o que pode resultar em danos à própria estrutura.



Figura 4 – Eflorescência na superfície do concreto. Adaptado de (CODEBRIM 2019).

Na Figura 5 é apresentada a imagem de um exemplo de mancha de corrosão. Esse caso, igualmente, traz um problema de estética, mas evidencia um problema muito grave: a corrosão da armadura da estrutura.

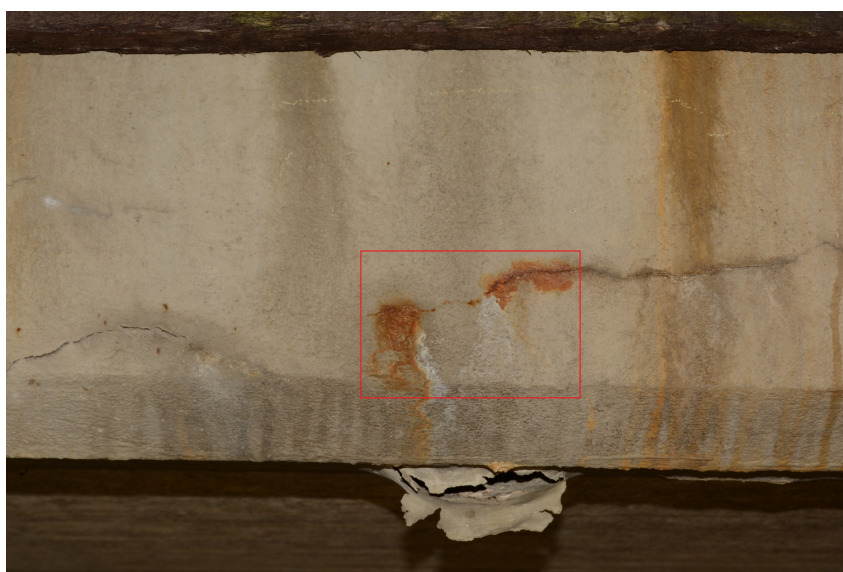


Figura 5 – Mancha de corrosão do aço na superfície do concreto. Adaptado de (CODEBRIM 2019).

As patologias citadas são evidenciadas nas superfícies do concreto, o que possibilita sua identificação por inspeção visual *in loco* ou por análise de imagens. Exatamente esta característica possibilita o uso de técnicas de visão computacional com aprendizado profundo.

2.1 Visão Computacional e Aprendizado Profundo

A área de Visão Computacional tem experimentado avanços significativos nas últimas décadas, impulsionada pelo desenvolvimento de técnicas sofisticadas de processamento de imagens e, mais recentemente, pela consolidação do aprendizado profundo como paradigma dominante para tarefas de reconhecimento e classificação visual. A convergência entre algoritmos de aprendizado de máquina, especialmente as redes neurais convolucionais, e o crescente poder computacional disponível tem possibilitado a aplicação dessas tecnologias em domínios específicos da engenharia civil, particularmente na automatização de processos de inspeção e de diagnóstico estrutural, como os trabalhos apresentados adiante nesta tese e em trabalhos relacionados. Neste contexto, as técnicas de Visão Computacional apresentam-se como ferramentas promissoras para a identificação automática de patologias em imagens de estruturas de concreto armado, oferecendo uma alternativa tecnológica aos métodos tradicionais de inspeção visual manual. A visão computacional foi beneficiada pelo desenvolvimento de arquiteturas neurais profundas, o que abriu possibilidades de aplicações em diversas áreas.

O aprendizado profundo é uma subárea do aprendizado de máquina baseado em redes neurais artificiais com múltiplas camadas ocultas, caracterizada pela capacidade de aprender representações hierárquicas de dados por meio de transformações sucessivas. Esta abordagem permite a modelagem automática de abstrações complexas por meio da aplicação de transformações lineares e não lineares em cascata, nas quais cada camada extrai características de um nível progressivamente mais elevado a partir das representações da camada anterior. A propagação das informações na arquitetura neural ocorre por meio de funções de ativação cuidadosamente selecionadas, que introduzem não linearidades essenciais para a capacidade de aproximação universal da rede. A aplicação de técnicas de aprendizado profundo em visão computacional pode ser classificada em três problemas fundamentais, cada qual com características e desafios específicos: classificação, detecção e segmentação.

A **classificação** consiste na atribuição de uma ou de múltiplas etiquetas categóricas a uma imagem completa, com base em características globais extraídas pela rede neural. Neste problema, considera-se a imagem como uma instância de uma classe predefinida, e o objetivo é determinar a classe ou as classes a que ela pertence, dentre um conjunto finito de classes. No contexto da inspeção de estruturas de concreto

armado, a classificação permitiria determinar se uma imagem representa uma estrutura com ou sem manifestações patológicas, tratando-a como uma unidade indivisível de análise. Também é um exemplo deste problema a classificação de uma imagem como apresentando uma fissura ou não (Krishnan, Karuppan e Khadidos 2025).

A **detecção** é um problema no qual o objetivo é identificar ou determinar a localização espacial dos objetos de interesse na imagem. Esta tarefa requer a delimitação das regiões correspondentes a cada instância detectada, tipicamente representadas por caixas delimitadoras (*bounding boxes*). A detecção de patologias em estruturas de concreto exemplifica esta abordagem, na qual se busca não apenas identificar a presença de defeitos, mas também localizar com precisão suas posições na superfície estrutural, fornecendo informações espaciais fundamentais para a avaliação da extensão e da severidade dos problemas identificados (Kaveh e Alhajj 2024).

A **segmentação** constitui o nível mais refinado de análise visual, com o objetivo de classificar cada *pixel* da imagem. Esta abordagem permite a delimitação precisa dos contornos dos objetos, proporcionando uma representação detalhada da distribuição espacial das diferentes classes presentes na cena. A segmentação semântica associa cada *pixel* a uma classe específica, enquanto a segmentação de instâncias distingue diferentes objetos da mesma classe. No domínio da inspeção estrutural, a segmentação possibilita a quantificação precisa da área afetada por patologias, a caracterização morfológica detalhada dos defeitos e a análise de sua distribuição espacial na superfície do concreto (?).

No âmbito da detecção de objetos em imagens, as arquiteturas de redes neurais profundas podem ser classificadas em duas principais abordagens: arquiteturas de um estágio (*one-stage*) e de dois estágios (*two-stage*) (Bi, Wen e Xu 2023). As arquiteturas de um estágio caracterizam-se por realizar, em uma única passagem pela rede neural, a localização e a classificação de objetos, integrando ambas as tarefas em um processo unificado e otimizado para eficiência computacional.

Estas arquiteturas diferenciam-se fundamentalmente das abordagens de dois estágios por eliminarem a necessidade de uma etapa intermediária de geração de regiões candidatas (*region proposals*). Em vez de identificar primeiro regiões candidatas e, posteriormente, classificá-las, as redes de um estágio processam a imagem integralmente e produzem diretamente as coordenadas das caixas delimitadoras e as probabilidades de classe para cada objeto detectado. Esta abordagem confere vantagens significativas em termos de velocidade de processamento, tornando-a particularmente adequada para aplicações que exigem processamento em tempo real.

Ao longo dos últimos anos, as arquiteturas de um estágio têm experimentado avanços substanciais tanto em eficiência quanto em precisão, superando as limitações

iniciais relacionadas à detecção de objetos pequenos e ao balanceamento entre velocidade e acurácia. Estes desenvolvimentos têm consolidado esta abordagem como uma alternativa competitiva e, em muitos casos, superior às metodologias tradicionais de dois estágios, especialmente em aplicações em que a rapidez de processamento constitui um requisito fundamental.

2.2 Rede Neural Convolutional

As redes neurais convolucionais (*Convolutional Neural Networks* - CNNs) (Li et al. 2022) representam uma arquitetura de aprendizado profundo particularmente eficaz no processamento e na análise de dados com estrutura espacial, especialmente de imagens. Estas redes fundamentam-se na operação matemática de convolução, que permite a detecção de padrões locais por meio de filtros compartilhados espacialmente, conferindo às CNNs propriedades fundamentais, como a invariância translacional e a redução significativa do número de parâmetros em comparação com redes totalmente conectadas. A arquitetura convolucional opera segundo o paradigma *feed-forward*, em que as ativações de cada camada são computadas a partir das saídas da camada precedente, resultando em representações de alto nível adequadas para tarefas de reconhecimento visual.

A arquitetura básica das CNNs é essencialmente a mesma, baseada em um paradigma fixo, composta por uma camada de entrada, seguida de camadas alternadas de convolução e pooling, de uma ou mais camadas completamente conectadas, funções de ativação e uma camada de saída. As primeiras camadas são responsáveis pela extração de características por meio de camadas de convolução e pooling alternadas, nas quais várias operações são realizadas para processar os dados de entrada, gerando representações mais abstratas e de maior nível. As camadas completamente conectadas são usadas em combinação com funções de ativação para executar tarefas tais como classificação baseada nas características extraídas. Várias unidades regulatórias, como batch normalization e dropout são incluídas (Figura 6).

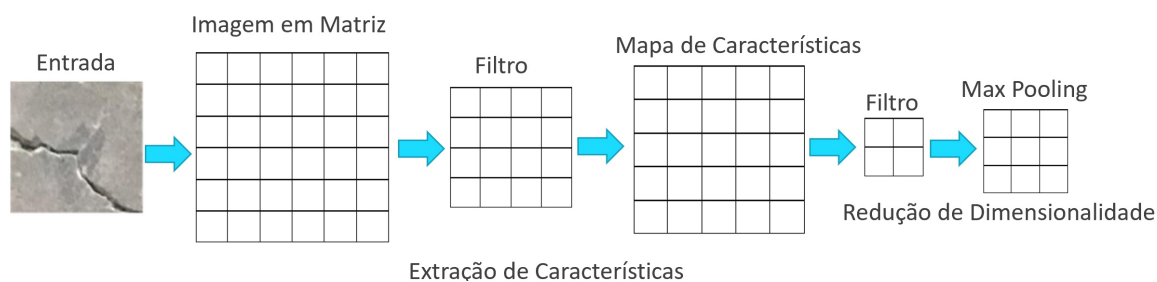


Figura 6 – Rede Neural Convolucional. Elaborada pelo autor.

Os dados resultantes dos processos são distribuídos entre os neurônios artificiais das camadas ocultas da rede, cada um com um peso determinado conforme

sua relevância para o aprendizado. Os neurônios de uma camada são interligados aos da camada seguinte, e os valores são ativados por funções de ativação. Os pesos iniciais são aleatórios, propagados segundo as funções de ativação, seguindo o fluxo da entrada em direção à saída. Na última camada é feita a verificação da resposta predefinida, um erro é calculado e sua correção determina a alteração dos pesos, neste momento seguindo o caminho inverso, retropropagação. Este processo ocorre durante o aprendizado e é definido até que se tenha a correta saída esperada.

Uma rede neural convolucional é uma arquitetura de aprendizado profundo especializada no processamento de dados com estrutura de grade, como imagens. Seu funcionamento baseia-se em camadas convolucionais que aplicam filtros deslizantes para extrair características hierárquicas, das mais simples (bordas, texturas) às mais complexas (formas, padrões específicos), seguidas por camadas de pooling que reduzem a dimensionalidade, preservando características essenciais. Essas operações de convolução e pooling permitem que a CNN aprenda representações espaciais invariantes a translações e variações, culminando em camadas totalmente conectadas para classificação final. A arquitetura é particularmente eficaz para tarefas de visão computacional, pois captura automaticamente padrões locais e hierárquicos nos dados, simulando conceitos neurocientíficos do sistema visual humano.

Nas seções seguintes, exploraremos as arquiteturas de redes neurais convolucionais que fundamentam nosso trabalho, abordando-as tanto em sua cronologia quanto em sua evolução conceitual. Começaremos pela ResNet50 (2016), que se estabeleceu como a base fundamental das arquiteturas modernas ao introduzir as conexões residuais, uma inovação aprimorada posteriormente pela DenseNet121 (2017), que otimizou ainda mais a reutilização de características. No campo da detecção de objetos, analisaremos o pioneirismo da rede YOLO (2015), que revolucionou a área com sua abordagem de estágio único para detecção em tempo real, contrastando-a com a Faster R-CNN (2015), que priorizou a alta precisão por meio de um processo de dois estágios. Para endereçar a eficiência computacional, investigaremos a MobileNetV3 (2019), uma arquitetura projetada especificamente para otimizar o desempenho em dispositivos com recursos limitados. Por fim, avançaremos para a rede TOOD (2021), que representa uma evolução significativa dos detectores de estágio único.

2.2.1 YOLO, You Only Look Once

A arquitetura YOLO é amplamente utilizada na detecção em visão computacional. A denominação YOLO, acrônimo de "You Only Look Once" (Você Só Olha Uma Vez, em tradução livre), refere-se a uma arquitetura de rede neural convolucional projetada especificamente para detecção de objetos em tempo real. Esta arquitetura distingue-se por sua velocidade de processamento e precisão na detecção, características que a

tornaram amplamente adotada na comunidade científica. Esta arquitetura tem evoluído continuamente por meio de múltiplas versões até a presente data. Uma das versões mais recente é a YOLO V11 ([Wang et al. 2024](#)).

A proposta fundamental da arquitetura YOLO consiste em uma rede neural completamente conectada, projetada para realizar predições de caixas delimitadoras (bounding boxes) e suas respectivas probabilidades de classe em uma única passagem pela rede. Enquanto métodos como o Faster R-CNN utilizam redes de propostas de regiões (Region Proposal Networks) seguidas de classificação separada dessas regiões em múltiplas etapas, o YOLO executa tanto a proposição quanto a classificação de regiões de interesse em uma única interação com a imagem de entrada.

Do ponto de vista arquitetural, o algoritmo YOLO processa uma imagem de entrada por meio de uma rede neural convolucional relativamente simples para a detecção de objetos. A Figura 7 apresenta detalhadamente a arquitetura da rede CNN empregada pelo modelo YOLO original. A implementação da rede incorpora estratégias de transferência de aprendizado em sua concepção. Especificamente, as vinte camadas convolucionais iniciais são pré-treinadas no conjunto de dados ImageNet, utilizando pooling de média, seguido de uma camada totalmente conectada, durante a fase de pré-treinamento. O desempenho da rede é aprimorado por meio da adição de quatro camadas convolucionais adicionais e de duas camadas totalmente conectadas, especificamente adaptadas para a tarefa de detecção. A arquitetura culmina com uma camada totalmente conectada responsável pela predição das probabilidades de classe e pela localização das caixas delimitadoras.

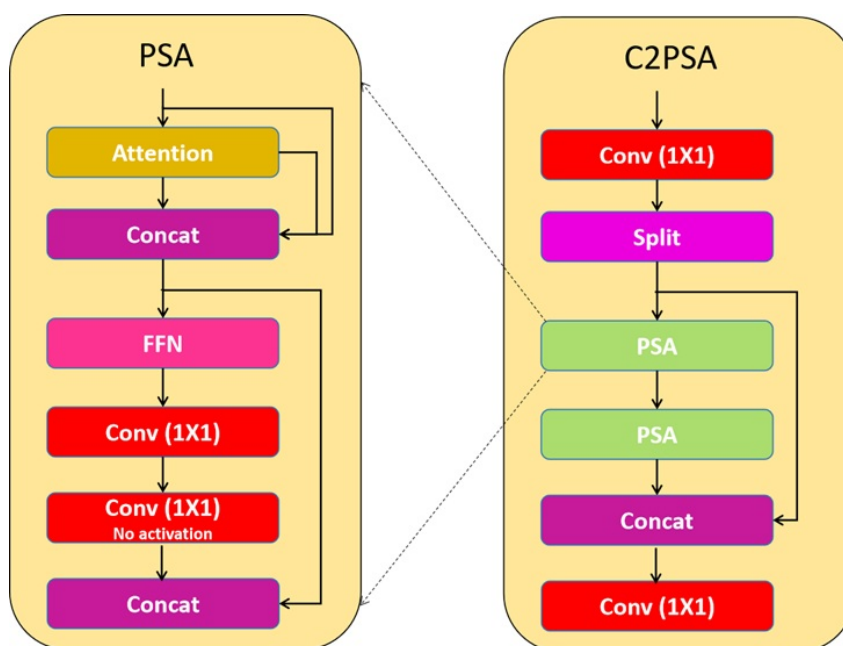


Figura 7 – Arquitetura YOLO ([Rao 2024](#)).

O processo de detecção inicia-se com a divisão da imagem de entrada em uma grade de dimensões $S \times S$, conforme demonstrado na Figura 8. Cada célula da grade assume a responsabilidade de detectar objetos cujo centro geométrico esteja localizado dentro de seus limites espaciais, estabelecendo, assim, uma correspondência direta entre as regiões da imagem e as células da grade.

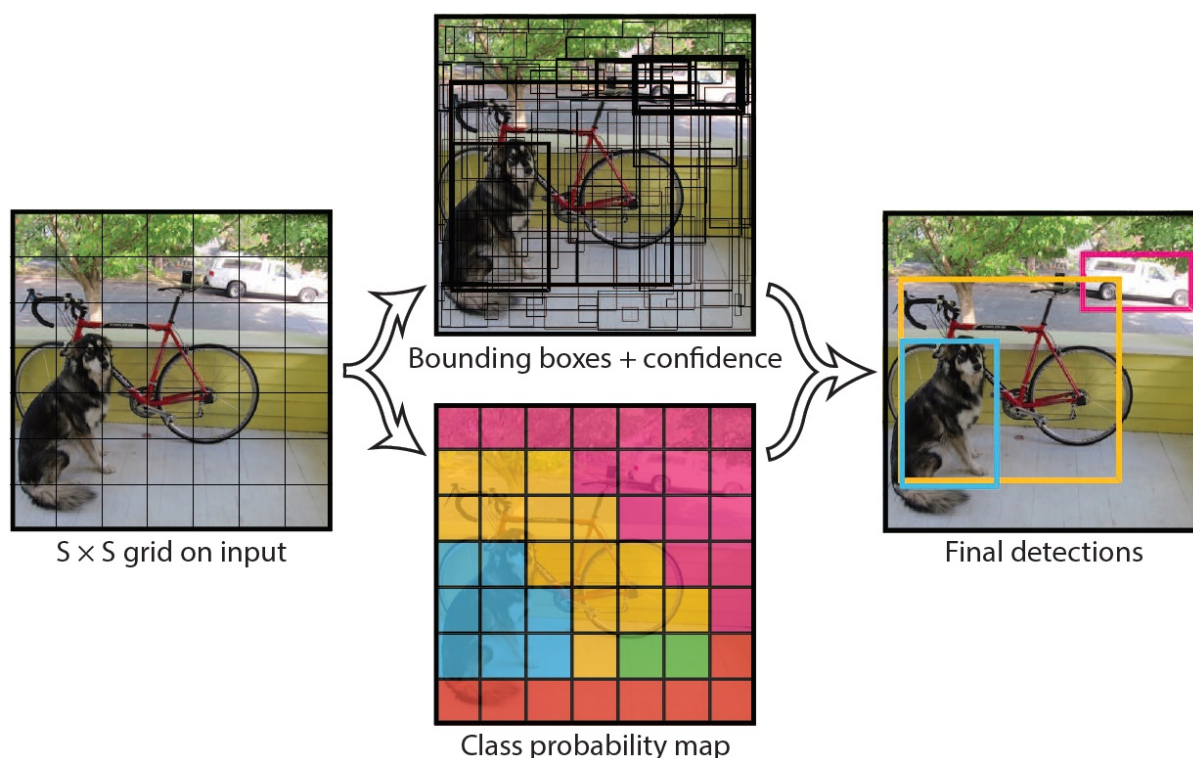


Figura 8 – Modelo YOLO (Redmon et al. 2016).

Como etapa de pós-processamento, a arquitetura YOLO emprega a técnica de supressão não máxima (Non-Maximum Suppression - NMS). Esta operação é fundamental para o refinamento da precisão e da eficiência na detecção, uma vez que múltiplas caixas delimitadoras podem ser preditas para um mesmo objeto, resultando em sobreposições com variações mínimas na localização e nas dimensões. A aplicação da NMS permite a eliminação de detecções redundantes ou incorretas, assegurando que cada objeto seja representado por uma única caixa delimitadora otimizada.

2.2.2 Faster R-CNN

A arquitetura Faster R-CNN, apresentada em (Ren et al. 2016), introduziu inovações que superaram algumas limitações das abordagens anteriores. Esta proposta constitui um avanço em relação aos modelos R-CNN e Fast R-CNN, consolidando a geração de propostas de regiões e a classificação de objetos em uma única arquitetura neural unificada, proporcionando ganhos em eficiência computacional e precisão de detecção. A arquitetura do Faster R-CNN, ilustrada na Figura 9, fundamenta-se em

uma estrutura modular composta por dois componentes principais integrados. O primeiro módulo, denominado Region Proposal Network (RPN), consiste em uma rede neural convolucional especializada na geração autônoma de regiões candidatas diretamente a partir dos mapas de características extraídos da imagem de entrada. O segundo componente, Fast R-CNN Detector, utiliza as propostas de regiões de interesse geradas pela RPN para realizar a classificação final dos objetos e o refinamento das coordenadas das caixas delimitadoras correspondentes.

A arquitetura Faster R-CNN compartilha uma rede de características convolucionais comum a ambos os módulos, tipicamente implementada com base em arquiteturas estabelecidas como VGG16 ou ResNet. Esta estratégia de compartilhamento não apenas otimiza a eficiência computacional, reduzindo o custo de processamento, mas também promove a consistência na extração de características entre os diferentes estágios de processamento.

O mecanismo de âncoras (anchors) constitui outro elemento inovador da RPN, permitindo prever propostas com múltiplas escalas e proporções aspectuais de forma sistemática. A abordagem baseada em âncoras supera as limitações inerentes aos métodos de janela deslizante tradicionais, possibilitando a captura eficiente de objetos com diferentes dimensões e formas geométricas por meio de um conjunto predefinido de escalas e proporções aspectuais. Este mecanismo sistemático elimina a necessidade de varreduras exaustivas da imagem em múltiplas escalas.

Durante o processo de treinamento, a arquitetura Faster R-CNN emprega uma estratégia de otimização conjunta, na qual tanto a RPN quanto o detector Fast R-CNN são otimizados simultaneamente por meio de um esquema de treinamento end-to-end. Esta abordagem integrada permite que ambos os componentes se adaptem mutuamente, resultando em maior convergência e desempenho global do sistema.

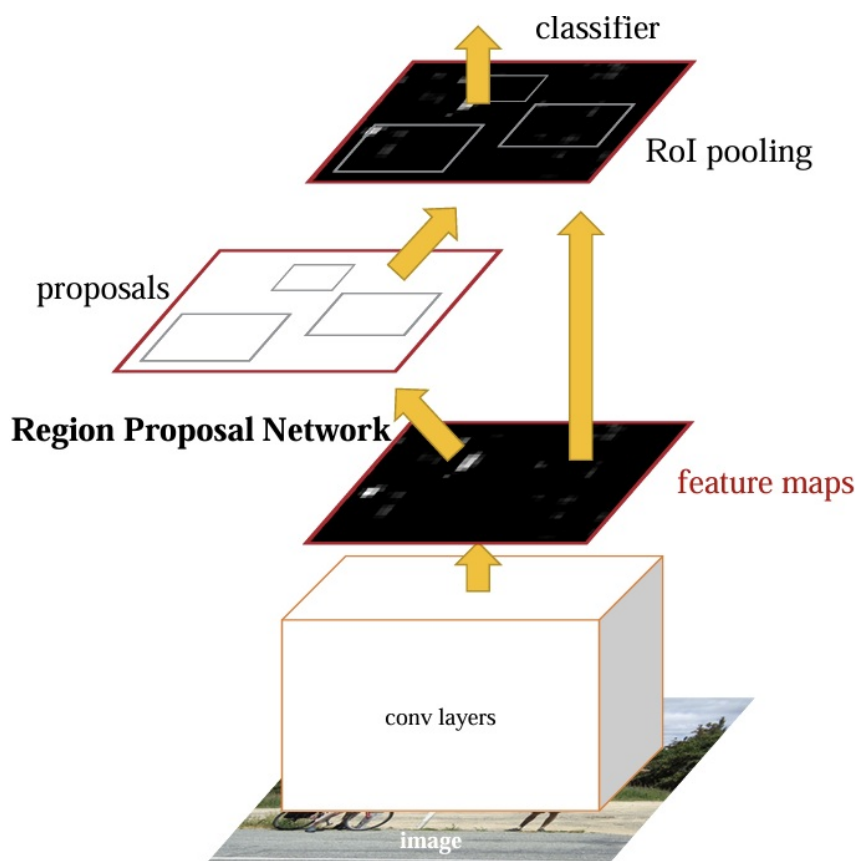


Figura 9 – Arquitetura Faster R-CNN (Ren et al. 2016).

Em termos de desempenho, o Faster R-CNN estabeleceu novos padrões de referência para tarefas de detecção de objetos, demonstrando superioridade significativa em relação aos métodos anteriores. Avaliações experimentais conduzidas no conjunto de dados PASCAL VOC evidenciaram que o modelo alcança mais de 70% de Mean Average Precision (mAP), representando um avanço substancial em precisão de detecção. Adicionalmente, a arquitetura consegue operar em velocidades próximas ao tempo real, conciliando eficiência computacional e alta precisão (Ren et al. 2015).

A arquitetura Faster R-CNN teve importância em desenvolvimentos subsequentes na área de visão computacional. Extensões diretas desta abordagem, como o Mask R-CNN (He et al. 2017) para segmentação de instâncias, demonstram a versatilidade e extensibilidade dos princípios arquiteturais estabelecidos por esta proposta, consolidando sua posição como uma contribuição seminal no campo da detecção de objetos baseada em aprendizado profundo.

2.2.3 TOOD, Task-aligned One-stage Object Detection

A arquitetura TOOD (Task-Aligned One-Stage Object Detection), proposta em (Feng et al. 2021), aborda o problema do desalinhamento entre as tarefas de classificação e localização que caracteriza os métodos de estágio único.

O problema central que a arquitetura TOOD se propõe a resolver reside na otimização independente das tarefas de predição de classe e de regressão de caixas delimitadoras em detectores convencionais. Esta abordagem desacoplada frequentemente resulta em inconsistências, em que objetos com alta confiança de classificação podem apresentar localizações imprecisas ou, inversamente, em que localizações precisas são acompanhadas de baixas pontuações de confiança. Tal desalinhamento compromete a eficácia geral do sistema de detecção, particularmente em cenários que exigem alta precisão simultânea em ambas as tarefas.

A arquitetura TOOD, ilustrada na Figura 10, fundamenta-se em uma estrutura de estágio único que incorpora dois componentes inovadores principais: o Task-Aligned Head (T-Head) e o Task-Aligned Learning (TAL). Esta configuração substitui os cabeçalhos convencionais de classificação e regressão por uma abordagem integrada que promove o alinhamento explícito entre as duas tarefas ao longo de todo o processo de treinamento e inferência.

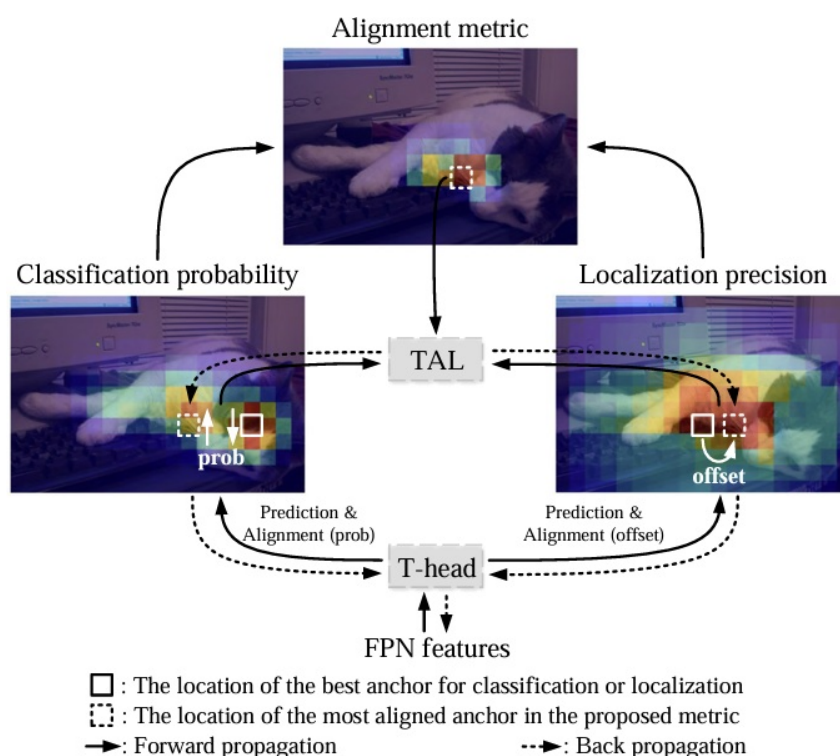


Figura 10 – Arquitetura TOOD (Feng et al. 2021).

O Task-Aligned Head implementa um mecanismo de interação dinâmica entre as características utilizadas para classificação e regressão; isto é feito por meio de camadas de atenção que ajustam dinamicamente as representações de características com base no grau de alinhamento entre as tarefas. Este mecanismo permite que a rede adapte automaticamente a importância relativa das características para cada tarefa, baseando-se na qualidade da sobreposição (IoU) e na confiança na classificação. Esta abordagem incorpora simultaneamente métricas de classificação e de precisão de

localização na definição de uma métrica de alinhamento de tarefas. Consequentemente, durante o processo de otimização, as regiões que apresentam maior consistência entre a classificação e a localização recebem maior ponderação, promovendo a convergência em soluções mais equilibradas.

A implementação arquitetural emprega uma rede backbone convencional, tipicamente baseada em arquiteturas estabelecidas, como ResNet, para a extração de características hierárquicas. Essas características são subsequentemente processadas através de uma Feature Pyramid Network (FPN) para capturar informações multiescala.

Em termos de desempenho experimental, a arquitetura TOOD demonstrou eficácia superior quando avaliada no conjunto de dados MS COCO, alcançando 51,1% de Average Precision (AP) com backbone ResNet-101 ([Feng et al. 2021](#)). Estes resultados evidenciam melhorias substanciais em relação aos detectores de estágio único convencionais, validando a efetividade da abordagem de alinhamento de tarefas proposta.

A versatilidade da arquitetura TOOD manifesta-se na capacidade de integração com diferentes backbones, incluindo ResNet e CSPDarknet. Esta flexibilidade arquitetural permite adaptar a metodologia a diferentes requisitos computacionais e de desempenho, mantendo os benefícios do alinhamento de tarefas em diversas configurações.

2.2.4 DenseNet121

A arquitetura DenseNet, proposta por ([Huang et al. 2017](#)), otimiza o fluxo de informações e de gradientes ao longo da rede. A inovação reside na implementação de conexões densas, que estabelecem conectividade direta entre cada camada e todas as camadas subsequentes em blocos densamente conectados, em contraste com abordagens como as ResNets, que empregam conexões residuais apenas entre camadas adjacentes. No paradigma DenseNet, a entrada de cada camada é constituída pela concatenação dos mapas de características de todas as camadas precedentes, promovendo a reutilização extensa de características e facilitando a propagação de gradientes por caminhos mais curtos. A arquitetura DenseNet121, ilustrada na Figura 11, exemplifica uma implementação específica desta família arquitetural, composta por 121 camadas organizadas em uma estrutura hierárquica bem definida. A rede inicia-se com uma camada convolucional de entrada, seguida por pooling para redução dimensional inicial, estabelecendo uma representação de características de base para o processamento subsequente.

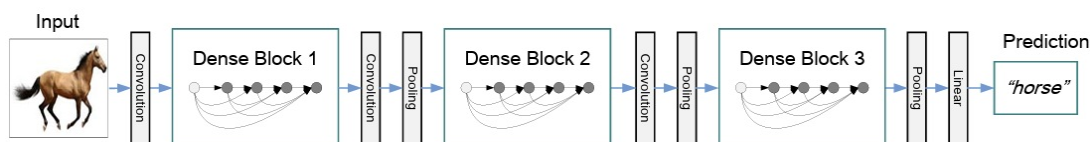


Figura 11 – Arquitetura DenseNet121 (Huang et al. 2017).

A estrutura principal da arquitetura organiza-se em quatro Dense Blocks intercalados por Transition Layers. Cada Dense Block implementa o princípio de conectividade densa, em que cada camada convolucional recebe como entrada a concatenação de todos os mapas de características gerados pelas camadas anteriores no mesmo bloco. Esta configuração facilita a propagação de informações e de gradientes, mitigando o desvanecimento de gradientes que caracteriza redes muito profundas.

Dentro de cada Dense Block, as camadas convolucionais seguem uma sequência operacional padronizada, que consiste em normalização por lotes (batch normalization), função de ativação ReLU e convolução 3×3 . Esta sequência, denominada pre-activation design, demonstrou-se eficaz na promoção da convergência durante o treinamento e na manutenção da estabilidade numérica.

As Transition Layers desempenham papel crucial na regulação da complexidade computacional e dimensional da rede. Posicionadas entre Dense Blocks consecutivos, estas camadas implementam convoluções 1×1 seguidas de pooling de média 2×2 , reduzindo tanto o número de mapas de características quanto suas dimensões espaciais. Esta estratégia previne o crescimento exponencial da complexidade computacional decorrente da concatenação contínua de características.

A conectividade densa implementada no DenseNet estabelece um total de $\frac{L(L+1)}{2}$ conexões diretas em uma rede com L camadas, em contraste com as L conexões das arquiteturas convencionais. Esta proliferação de conexões diretas promove diversos benefícios: facilita a propagação de gradientes, favorece a reutilização de características, reduz o número de parâmetros necessários e melhora a regularização implícita da rede.

O mecanismo de growth rate constitui um hiperparâmetro fundamental que controla a taxa de crescimento dos mapas de características. Cada camada de um Dense Block produz k mapas de características adicionais, onde k representa o growth rate. Esta parametrização permite um controle preciso da complexidade da rede e do número total de parâmetros.

A arquitetura DenseNet121 organiza-se em Dense Blocks com 6, 12, 24 e 16 camadas, respectivamente, totalizando 121 camadas paramétricas. Esta configuração demonstrou um equilíbrio eficaz entre capacidade representacional e eficiência

computacional, estabelecendo-se como uma das variantes mais utilizadas da família DenseNet.

As arquiteturas DenseNet demonstraram eficiência superior quando comparadas às alternativas contemporâneas. No conjunto de dados ImageNet, o DenseNet121 atinge desempenho competitivo em relação à ResNet-50, utilizando aproximadamente metade do número de parâmetros. Esta eficiência paramétrica traduz-se em menores requisitos de memória e de processamento, características particularmente valiosas para aplicações com restrições computacionais.

2.2.5 ResNet50

A arquitetura ResNet, proposta em (He et al. 2016), propõe uma solução para a degradação de desempenho observada em redes neurais extremamente profundas, em que o aumento do número de camadas resulta paradoxalmente em maior erro de treinamento e de teste. Este fenômeno deriva da dificuldade de otimização associada à anulação dos gradientes e à propagação de informações em arquiteturas muito profundas. Esta arquitetura introduz blocos residuais que implementam skip connections para facilitar o fluxo de informações e de gradientes por toda a rede. Estes blocos residuais reformulam o problema de aprendizagem, permitindo que as camadas aprendam mapeamentos residuais em relação à entrada, em vez de aprender mapeamentos completos. Esta reformulação matemática facilita significativamente a otimização de redes muito profundas.

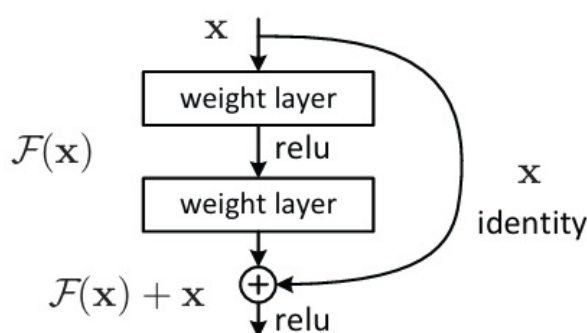


Figura 12 – Construção de bloco de aprendizagem residual da ResNet50 (He et al. 2016).

A arquitetura ResNet50, ilustrada na Figura 12, exemplifica uma implementação específica desta família arquitetural, composta por 50 camadas organizadas em uma estrutura hierárquica bem definida. A rede inicia-se com uma camada convolucional 7×7 , seguida por max pooling 3×3 , estabelecendo uma extração inicial de características de baixo nível, com redução dimensional adequada. A estrutura principal organiza-se em quatro estágios residuais, cada um contendo múltiplos blocos com configurações específicas. Os estágios contêm, respectivamente, 3, 4, 6 e 3 blocos residuais,

totalizando 16 blocos que, combinados com as camadas iniciais e finais, resultam em uma arquitetura de 50 camadas parametrizadas. Cada bloco residual implementa uma arquitetura de gargalo (bottleneck) composta por três operações convolucionais sequenciais: convolução 1×1 para redução dimensional, convolução 3×3 para extração de características espaciais e convolução 1×1 para expansão dimensional. Esta configuração otimiza a eficiência computacional ao reduzir o número de parâmetros e de operações necessárias, mantendo a capacidade representacional da rede.

Cada bloco residual computa $\mathcal{F}(x) + x$, onde $\mathcal{F}(x)$ representa o mapeamento residual aprendido pelas camadas convolucionais e x representa a entrada do bloco. Esta formulação permite que o bloco aprenda o mapeamento residual $\mathcal{F}(x) = \mathcal{H}(x) - x$, em que $\mathcal{H}(x)$ é o mapeamento desejado.

As skip connections fornecem caminhos diretos para o fluxo de gradientes desde as camadas finais até as iniciais. Esta característica mitiga o problema da anulação do gradiente, permitindo o treinamento eficaz de redes com profundidade substancialmente maior do que as arquiteturas predecessoras.

A implementação de blocos de gargalo otimiza a eficiência computacional por meio da redução sistemática da dimensionalidade dos mapas de características. As convoluções 1×1 iniciais reduzem o número de canais antes da operação computacionalmente custosa de convolução 3×3 , enquanto as convoluções 1×1 finais restauram a dimensionalidade original, mantendo a compatibilidade com as conexões residuais.

Quando as dimensões de entrada e de saída de um bloco residual diferem, empregam-se conexões de projeção por meio de convoluções 1×1 , com o stride apropriado. Estas conexões asseguram compatibilidade dimensional para a operação de adição elemento a elemento, mantendo a integridade do mapeamento residual em diferentes escalas de resolução.

A arquitetura conclui com uma camada de pooling de média global que reduz cada mapa de características a um único valor, seguida por uma camada totalmente conectada para classificação. Esta configuração elimina a necessidade de camadas totalmente conectadas de grande dimensionalidade, reduzindo o número de parâmetros e mitigando o risco de overfitting.

Em termos de desempenho experimental, a ResNet50 demonstrou melhorias substanciais em benchmarks estabelecidos. No conjunto de dados ImageNet, a arquitetura alcançou 23,9% de erro top-1, superando significativamente arquiteturas predecessoras com menor complexidade computacional, estimada em 3,8 bilhões de operações de ponto flutuante por inferência.

Variantes como ResNet101 e ResNet152 mantêm os princípios fundamentais, escalando a profundidade por meio da adição de blocos residuais. Esta escalabilidade

demonstra a robustez do paradigma residual em arquiteturas de complexidade variável.

2.2.6 MobileNetV3

A arquitetura MobileNetV3, proposta em (Howard et al. 2019), integra técnicas de busca de arquitetura neural, Neural Architecture Search (NAS), com otimizações manuais. A fase inicial emprega o algoritmo MnasNet para a exploração do espaço arquitetural, otimizando simultaneamente precisão e latência por meio de uma função objetivo multicritério. Subsequentemente, otimizações manuais específicas são aplicadas para refinar componentes críticos da arquitetura, resultando em uma solução híbrida que capitaliza as vantagens de ambas as abordagens.

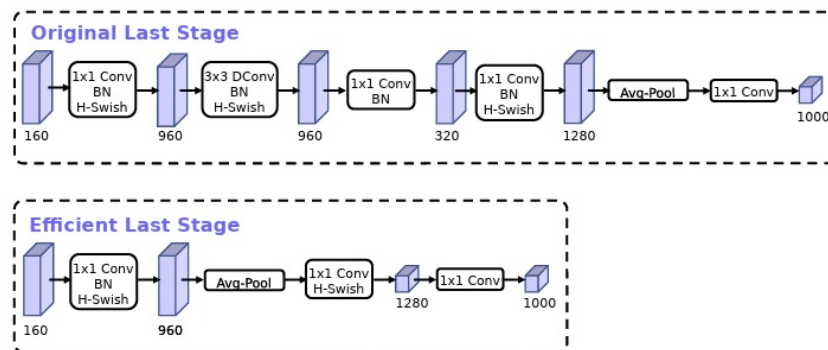


Figura 13 – Melhoria da MobileNetV3, (He et al. 2016).

A arquitetura MobileNetV3, ilustrada na Figura 13, fundamenta-se em blocos Mobile Inverted Bottleneck Convolution (MBConv), incorporando refinamentos substanciais que otimizam tanto a eficiência computacional quanto a capacidade representacional. Estes blocos implementam convoluções separáveis em profundidade precedidas por expansão dimensional e seguidas por projeção linear.

A arquitetura introduz a função de ativação h-swish, uma aproximação computacionalmente eficiente da função swish, que demonstra desempenho superior ao do ReLU6 tradicionalmente empregado em arquiteturas móveis. A h-swish é definida como $h\text{-swish}(x) = x \cdot \frac{\text{ReLU6}(x+3)}{6}$, oferecendo propriedades de suavidade superiores com overhead computacional mínimo em dispositivos móveis otimizados.

Além disso, propõe a integração seletiva de módulos Squeeze-and-Excitation (SE) nos blocos MBConv. Estes módulos implementam mecanismos de atenção por canal que adaptam dinamicamente à importância relativa dos canais de características, baseando-se no conteúdo global da entrada. A integração destes módulos, determinada pelo processo de NAS, otimiza o trade-off entre ganhos de precisão e overhead computacional.

As camadas finais são otimizadas para reduzir a latência por meio da eliminação

de operações computacionalmente custosas nas últimas camadas, onde o impacto na precisão é mínimo. Especificamente, a última convolução 1×1 e a camada de pooling médio global são reorganizadas para reduzir o número de operações de ponto flutuante.

A família MobileNetV3 compreende duas variantes principais otimizadas para diferentes pontos de operação no espaço precisão-eficiência. A MobileNetV3-Large visa aplicações que priorizam a precisão máxima sob restrições de eficiência moderadas, enquanto a MobileNetV3-Small destina-se a cenários de eficiência extrema, em que recursos computacionais são severamente limitados.

A variante Large emprega uma arquitetura mais profunda, com um maior número de canais nos blocos MBConv, o que resulta em capacidade representacional superior. Esta configuração incorpora módulos SE em localizações estratégicas definidas pelo processo de NAS, maximizando os ganhos de precisão com um overhead computacional controlado.

A variante Small implementa uma arquitetura substancialmente mais compacta, otimizada para a redução de latência e de consumo energético. Esta versão emprega um número reduzido de blocos MBConv, com menor profundidade de canais, mantendo módulos SE apenas nas posições de maior impacto, determinadas experimentalmente.

O processo de otimização por meio de NAS emprega uma metodologia multi-objetiva que considera simultaneamente a precisão no conjunto de dados ImageNet e a latência em dispositivos móveis reais. Esta abordagem assegura que as otimizações arquiteturais sejam relevantes para cenários de implementação prática, evitando soluções que otimizam métricas teóricas sem benefícios reais.

Em termos de desempenho experimental, a MobileNetV3-Large demonstra eficácia superior, alcançando 75,2% de precisão top-1 no ImageNet com apenas 219 milhões de operações de multiplicação-acumulação (MACs). A variante Small atinge 67,0 % de precisão com overhead computacional significativamente reduzido, superando a MobileNetV2 em 3,2% de precisão com latência equivalente.

A versatilidade das arquiteturas MobileNetV3 manifesta-se em sua aplicabilidade como backbone para diversas tarefas de visão computacional. Implementações bem-sucedidas incluem detecção de objetos através do EfficientDet, segmentação semântica em tempo real, e reconhecimento facial em dispositivos móveis, demonstrando a transferibilidade das otimizações realizadas.

2.2.7 Análise Comparativa e Evolutiva das Arquiteturas de CNN.

Esta seção apresenta uma tabela com uma análise comparativa das arquiteturas utilizadas nesta tese, organizando suas características principais, métricas de desempenho e aplicações preferenciais.

A ResNet50 foi introduzida em 2016 para a tarefa principal de classificação, sem utilizar um detector específico. Suas características principais incluem: skip connections, blocos bottleneck, 50 camadas, e soluciona o problema de degradação com uma taxa de erro de 23,9%. O desempenho reportado foi de 3,8 bilhões de FLOPs. Sua aplicação foco é a classificação geral, servindo como backbone para outras tarefas.

DenseNet121 foi proposta em 2017, também focada em classificação, sem um detector específico (N/A). Suas características principais são: conexões densas, 121 camadas, controle de growth rate e reutilização de características. Apresentou desempenho competitivo com ResNet-50, utilizando aproximadamente 50% menos parâmetros. Sua aplicação foca na classificação eficiente, funcionando como backbone para outras tarefas.

YOLO ("You Only Look Once") surgiu em 2015 com a tarefa principal de detecção de objetos em um único passo, utilizando o detector YOLO. Suas características principais incluem: detecção em tempo real, grade $S \times S$ e uma única passagem pela rede. É conhecida pela velocidade em tempo real. Sua aplicação foca na detecção rápida em tempo real.

Faster R-CNN foi apresentada em 2015 para detecção de objetos em dois estágios, empregando a Region Proposal Network (RPN) como tipo de detector. Suas características principais são: mecanismo de âncoras, treinamento end-to-end, compartilhamento de características, com mais de 70% mAP no conjunto PASCAL VOC. Apresentou velocidade próxima ao tempo real. Sua aplicação foco está na detecção de alta precisão, em dispositivos móveis e em aplicações com restrições computacionais.

MobileNetV3 foi lançada em 2019 com foco em classificação, utilizando Neural Architecture Search (NAS) como abordagem de detecção. Suas características principais incluem: blocos MBConv, H-Swish e Squeeze-and-Excitation. Reportou desempenho de 75,2% top-1 na versão Large (ImageNet) e 67,4% top-1 na versão Small. Sua aplicação está voltada para dispositivos com 219M MACs (Large).

TOOD foi proposta em 2021 para detecção de objetos em estágio único, empregando Task-Aligned Head com Task-Aligned Learning. Suas características principais são: alinhamento, classificação-localização e mecanismo de atenção dinâmico. Alcançou 51,1% AP no MS COCO com ResNet-101 como backbone. Sua aplicação foco é a detecção precisa de estágio único.

2.3 Ensemble

Ensemble Learning, ou Aprendizado por Agrupamento, é uma metodologia que busca aumentar a capacidade de previsão para dados não vistos por meio da combinação de múltiplos modelos de predição simples treinados para a mesma tarefa.

Esta abordagem visa produzir, por meio do agrupamento destes modelos, um modelo mais complexo, com maior capacidade de predição, em que cada modelo individual possui eficiências e deficiências específicas, e a associação combina as melhores capacidades dos modelos utilizados, resultando em um preditor mais capaz e menos suscetível a fraquezas individuais, como vieses ([Shaikh et al. 2024](#)).

A eficácia de um ensemble pode superar a de um único aprendiz por três razões distintas: razões estatísticas, onde pode não haver informação suficiente nos dados de treinamento para identificar o aprendiz ótimo; razões computacionais, onde o método de busca do aprendiz pode ser falho, resultando em hipóteses subótimas; e razões de representação, onde o espaço de hipóteses pode não incluir a função objetivo verdadeira, sendo que o ensemble pode fornecer uma aproximação mais confiável, estendendo o espaço de função representável.

A decomposição viés-variança constitui uma ferramenta essencial para compreender o comportamento dos modelos, em que o viés refere-se ao erro causado por modelos excessivamente simples e a variância indica o grau de desvio dos valores reais em relação aos valores projetados. Modelos individuais frequentemente têm dificuldade em equilibrar viés e variância, e o ensemble pode reduzir ambos, especialmente ao empregar modelos diferentes.

A diversidade entre os classificadores de base é um fator crítico para a eficácia das abordagens de ensemble, definida como uma medida de desacordo que avalia as diferenças entre os modelos gerados. Uma maior diversidade geralmente beneficia o desempenho do ensemble, ajudando a mitigar o overfitting enquanto captura os padrões subjacentes dos dados. Esta diversificação pode ser alcançada por meio de diferentes algoritmos, arquiteturas, hiperparâmetros ou subconjuntos de dados de treinamento para cada aprendiz base. Em relação à similaridade dos modelos, podemos ter dois casos de combinação: modelos homogêneos, em que o mesmo tipo de modelo é treinado em diferentes conjuntos da base de dados, e modelos heterogêneos, que combinam modelos distintos para formar um sistema mais complexo.

O conceito de ensemble não representa uma estrutura fixa, mas sim uma combinação maleável que resulta em diferentes possibilidades de implementação. Os modelos mais utilizados podem ser agrupados em três principais combinações de classificação: Bagging, Boosting e Stacking. No Bagging (Bootstrap Aggregating), os preditores são geralmente homogêneos e independentes, utilizados de forma paralela, e o resultado é obtido pela média dos resultados dos modelos base. Esta abordagem visa produzir um conjunto de novas observações estatisticamente indistinguíveis dos dados originais, dividindo um conjunto de dados maior em subconjuntos menores e repetindo o treinamento por meio de bootstrapping. Uma aplicação específica é o Weighted Box Fusion, que combina previsões de modelos de detecção de objetos

utilizando a pontuação de confiança de todas as caixas delimitadoras propostas para construir caixas médias.

Na abordagem Boosting, realizada geralmente com preditores homogêneos, os modelos são utilizados de forma sequencial, em que um modelo alimenta o posterior até se obter o resultado final. Esta estratégia consegue uma previsão melhor ao combinar linearmente aprendizes fracos, em que cada modelo sucessivo é ajustado com um peso maior atribuído às observações que os modelos anteriores trataram inadequadamente. O Stacking, também conhecido como Stacked Generalization, utiliza geralmente preditores heterogêneos em paralelo, onde a saída destes modelos é combinada através de um metamodelo que aprende a integrar as previsões dos modelos de base de forma otimizada.

As estratégias de fusão de decisão constituem um componente fundamental após o treinamento dos aprendizes de base, e suas saídas são combinadas de acordo com regras predeterminadas. O ensemble learning aborda desafios como a redução do overfitting, a melhoria da generalização, o aumento da robustez, o incremento da precisão preditiva, a otimização da eficiência computacional e a melhoria da interpretabilidade, sendo transformador em termos de desempenho em modelos de deep learning e em outras abordagens de machine learning. Entre as principais abordagens estão a média ingênua, a média ponderada dos modelos, métodos de votação (incluindo a máxima, a média e a média ponderada), o classificador ótimo de Bayes e técnicas mais avançadas, como o Super Learner e o Query-by-Committee. A maioria dos modelos utiliza média ingênua, que não é adaptativa aos dados e apresenta desempenho subótimo devido à sensibilidade aos aprendizes enviesados (Ju, Bibaut e Laan 2018). A média não ponderada de modelos é o método mais comum, obtendo-se predições pela média dos aprendizes primários. A média simples melhora a generalização ao reduzir a variância, ao utilizar a função softmax para as probabilidades de classe. Em (Szegedy C. 2015), é indicado que a média não ponderada é adequada quando o desempenho dos aprendizes de base é semelhante. O método de votação baseia-se em predições de rótulos finais, como o rótulo majoritário. Em (?), mostra-se que a dependência pareada entre aprendizes influencia significativamente a votação majoritária. O método é benéfico para classificação multirótulo, com implementação típica no algoritmo RAKEL (?). Os ensembles de votação compreendem: (A) Votação Máxima (hard/soft voting), frequentemente usada em bagging; (B) Votação por Média, mais forte em termos de poder preditivo e mais precisa do que a votação majoritária; (C) Votação Média Ponderada, atribuindo pesos diferentes aos aprendizes, indicando a importância na predição. Como outra abordagem, podemos citar o classificador ótimo de Bayes, que combina as hipóteses de cada aprendiz de base com a distribuição condicional do rótulo-alvo, usando a regra de Bayes, embora a escolha das probabilidades a priori seja difícil. Em (Wolpert 1992), é proposta a Generalização

Empilhada, que infere vieses dos generalizadores sobre o conjunto de treinamento, determinando os pesos ideais por meio de validação cruzada e de mínimos quadrados. O Super Learner, proposto por (Laan, Polley e Hubbard 2007), é um composto ponderado das predições dos aprendizes de base, escolhendo pesos apropriados por meio de validação cruzada como alternativa à seleção do melhor classificador. Query-by-Committee (QBC) é uma técnica de aprendizado ativo em que um comitê de modelos seleciona os pontos mais informativos para a rotulagem. As possibilidades de combinação não se limitam às abordagens mencionadas, visto que dependem do conhecimento e da criatividade dos desenvolvedores, devendo ser testadas e validadas em conjuntos de dados com características específicas, com a escolha baseada em testes anteriores realizados em bases de dados semelhantes ao que se pretende utilizar. Bagging, boosting, stacking e votação são considerados métodos mais influentes devido à versatilidade, aos fundamentos teóricos sólidos e ao sucesso prático. Bagging reduz o overfitting por meio de paralelização eficiente, o boosting lida com viés e variância para predições precisas, o stacking combina modelos heterogêneos por meio de meta-aprendiz e a votação oferece simplicidade computacional.

3 Trabalhos Relacionados

A compreensão do estado da arte na detecção automatizada de patologias em estruturas de concreto armado constitui um importante elemento para a proposição de soluções inovadoras e eficazes na área de monitoramento da integridade estrutural. Este capítulo apresenta uma análise de trabalhos recentes e relevantes que empregam técnicas de aprendizado profundo, particularmente redes neurais convolucionais, para a identificação e segmentação de defeitos, como fissuras, fragmentação, lascamento e outras manifestações patológicas, em elementos estruturais de concreto.

A estruturação metodológica desta revisão bibliográfica fundamenta-se em uma taxonomia que contempla a evolução conceitual e tecnológica das abordagens de aprendizado de máquina aplicadas à detecção de patologias em estruturas de concreto. A organização proposta segue uma progressão hierárquica que se inicia com as metodologias de aprendizado supervisionado tradicional, progride por meio de abordagens que integram informações multidimensionais e multimodais, aborda técnicas emergentes de aprendizado supervisionado reduzido e culmina em estratégias desenvolvidas para lidar com as limitações impostas pela escassez de dados anotados. Esta estruturação possibilita não apenas a identificação de tendências evolutivas no campo de pesquisa, mas também facilita a análise comparativa entre trabalhos que compartilham fundamentos metodológicos similares, contribuindo para a identificação de lacunas científicas e de oportunidades de avanço na tarefa de detecção automatizada de patologias em estruturas de concreto armado.

As pesquisas que empregam metodologias tradicionais de aprendizado supervisionado com dados bidimensionais representam uma parcela significativa dos trabalhos recentes na área, utilizando arquiteturas consolidadas de redes neurais convolucionais para tarefas de classificação e detecção de patologias. Os estudos focam principalmente na aplicação de modelos estabelecidos como InceptionV3, MobileNet e YOLO, buscando otimizações específicas para o domínio da inspeção de estruturas de concreto.

Em ([Arafin, Billah e Issa 2024](#)), é apresentada a detecção visual de danos em infraestruturas por meio de aprendizado profundo, o que atesta a redução da subjetividade e o aumento da precisão dos diagnósticos no monitoramento de estruturas. Este trabalho criou um conjunto de dados diversificado, com imagens de fissuras de concreto (4087) e de fragmentação (1100) para avaliação com CNN. O dataset permite tarefas de classificação e segmentação. Visual Geometry Group (VGG)19, ResNet50 e InceptionV3, baseados em CNN, são usados para classificação. A segmentação

semântica é realizada por meio de dois modelos codificador-decodificador. U-Net e a arquitetura de rede de análise de cena em pirâmide são desenvolvidas com base em quatro modelos de backbone, incluindo VGG19, ResNet50, InceptionV3 e EfficientNetB3. Os modelos do classificador CNN são analisados em dois otimizadores (estocástico, descida gradiente (SGD)), propagação quadrática média (RMSprop) e taxas de aprendizagem de 0,1, 0,001 e 0,0001. Uma vez que a segmentação é analisada para SGD e para estimativa de momento adaptativo, foram realizados treinamentos com três taxas de aprendizado distintas - 0,1, 0,01 e 0,0001. O InceptionV3 alcança o melhor desempenho na classificação de defeitos, com precisão de 91,98%, usando o otimizador RMSprop. A segmentação foi realizada para fissuras usando U-Net baseado em EfficientNetB3, e para fragmentação usando InceptionV3. O modelo U-Net superou todos os demais algoritmos, com pontuações F1 de 95,66% e 89,43%, respectivamente. Este trabalho explora a tarefa de segmentação, que não foi contemplada em nossa pesquisa.

O estudo apresentado em ([Flotzinger et al. 2024](#)) enfatiza que o reconhecimento de defeitos de concreto armado (RCDs) é um elemento crucial para determinar a integridade estrutural, a segurança do tráfego e a durabilidade das pontes. No entanto, a maioria dos conjuntos de dados existentes no domínio RCD é derivada de um pequeno número de pontes adquiridas em posições de câmera específicas, em condições de iluminação e com hardware fixo. Essas limitações questionam a usabilidade de modelos treinados com tais dados de código aberto em cenários do mundo real. Este trabalho aborda esse problema testando tais modelos em nosso conjunto de dados "dacl1k", um conjunto de dados RCD altamente diverso para classificação de vários rótulos com base em inspeções de construção, incluindo 1.474 imagens. Assim, os modelos foram treinados em diferentes combinações de dados de código aberto (metaconjuntos de dados) que foram posteriormente avaliados tanto extrinsecamente quanto intrinsecamente. Durante a avaliação extrínseca, são relatadas métricas no dacl1k e nos metakonjuntos de dados. A análise de desempenho em dacl1k mostra a usabilidade prática dos metadados, em que o melhor modelo mostra uma taxa de correspondência exata de 32%. Além disso, foi conduzida uma avaliação intrínseca, agrupando os recursos de gargalo do melhor modelo derivado da avaliação extrínseca, para verificar se o modelo aprendeu a distinguir conjuntos de dados ou classes (RCDs), que foi o objetivo almejado. O conjunto de dados dacl1k e nossos modelos treinados serão disponibilizados publicamente, permitindo que pesquisadores e profissionais coloquem seus modelos em teste no mundo real. Este trabalho apresenta os resultados parciais das métricas, não apenas a classificação final, que nos permite comparar igualmente nossas métricas de pré-classificação.

Os autores em ([Tian et al. 2025](#)) enfatizam que a identificação completa e precisa de vários defeitos em superfícies de concreto é crucial para garantir a

segurança estrutural e prolongar a vida útil. No entanto, em inspeções de engenharia reais, as diversas formas e complexidades dos defeitos estruturais de concreto desafiam a robustez e a generalização insuficientes dos modelos tradicionais, levando frequentemente a detecções errôneas e subdetecções, o que, em última análise, compromete a segurança estrutural. Para superar as desvantagens mencionadas, os autores propõem um modelo eficiente de detecção de defeitos em concreto, denominado YOLOv11-EMC (detecção eficiente de defeitos em concreto multicategoria). Primeiramente, a convolução ordinária é substituída por uma convolução deformável modificada para extrair com eficiência características de defeitos irregulares, o que aprimora significativamente a robustez e a generalização do modelo. Em seguida, o módulo C3k2 é integrado a um módulo de convolução dinâmica revisado, que reduz cálculos desnecessários e, ao mesmo tempo, aprimora a flexibilidade e a representação de características. Os experimentos mostram que, em comparação com o YOLOv11, o YOLOv11-EMC melhorou a precisão, a recuperação, o mAP50 e o F1 para 8,3%, 2,1%, 4,3% e 3%, respectivamente. Os resultados dos testes de campo com drones mostram que o YOLOv11-EMC reduz com sucesso as falsas e subdetecções, aumentando simultaneamente a precisão da detecção, fornecendo uma metodologia superior para tarefas que exigem a identificação de falhas tangíveis em aplicações práticas de engenharia.

Os trabalhos focados em segmentação semântica avançada representam uma evolução natural das técnicas de classificação e detecção, buscando não apenas identificar a presença de patologias, mas também delinear com precisão suas formas e extensões. Estes estudos concentram-se no desenvolvimento de arquiteturas especializadas e de técnicas de refinamento que permitem a segmentação pixel-level de alta qualidade, essencial para análises quantitativas detalhadas das manifestações patológicas.

A publicação ([Deng et al. 2024](#)) igualmente destaca que a extração precisa de rachaduras é importante, mas desafiadora na inspeção de pontes, particularmente de pequenas rachaduras capturadas em imagens de alta resolução (HR). Este artigo apresenta uma estrutura de refinamento de limites de rachaduras (CBRF) para segmentação meticulosa de imagens de rachaduras HR. Primeiro, um módulo de extração de características de escala tripla é projetado para aprimorar a representação de pixels de rachaduras minúsculas. Em seguida, uma operação em cascata, envolvendo etapas globais e locais, é adotada para conduzir o refinamento. Além disso, uma perda de limite ativa é introduzida no processo de treinamento para resolver a inconsistência semântica das áreas de limite de rachaduras. O primeiro conjunto de dados de imagens de rachaduras HR é estabelecido para avaliar de forma completa o CBRF. Finalmente, um estudo de caso baseado em veículo aéreo não tripulado (UAV) é conduzido na Ponte Yinpenling, o que confirma ainda mais a praticidade do CBRF na

melhoria da segurança e eficiência da detecção de pontes baseada em UAV, garantindo a precisão. Este trabalho reforça a dificuldade na detecção de rachaduras e adota uma abordagem semelhante para preclassificar imagens que possam conter essa patologia.

O estudo (Shi et al. 2024) enfatiza que a segmentação automatizada de fissuras a partir de imagens digitais é crucial para o monitoramento da integridade da infraestrutura por meio de técnicas de avaliação não destrutivas. No entanto, capturar com precisão padrões complexos de fissuras com vários ramos continua sendo um desafio. Este artigo apresenta um estudo comparativo de quatro modelos de aprendizado profundo de última geração, UNet, FPN, PSPNet e DeepLabV3, para segmentação de fissuras a partir de dados de vídeo. Vários backbones de codificadores, incluindo ResNet, ResNeXt, SEnet e DPN, são examinados sob condições de carregamento de compressão manual e automática.

As fissuras foram induzidas em materiais à base de cimento por meio de um sistema de carregamento de compressão uniaxial equipado com imagens sincronizadas, que apresentam diversos fenômenos de fissuração, que variam de fissuras únicas e discretas a redes complexas de fissuras com ramificações e interseções. Os conjuntos de dados de segmentação foram rotulados manualmente para treinamento e teste do modelo. Experimentos foram conduzidos para avaliar o desempenho da segmentação de fissuras com quatro métricas amplamente utilizadas: Interseção sobre União (IoU), pontuação F1, Recall e Precisão.

Os resultados mostraram que a U-Net, pareada com a S-Net como backbone do codificador, obteve o melhor desempenho geral em IoU, pontuação F1 e precisão para carregamento manual, enquanto a FPN, alavancando um codificador DPN68, obteve o rastreamento mais equilibrado e robusto de propagações de trincas intrincadas. Este trabalho compara soluções de última geração e analisa casos de falha, orientando o design do modelo e as prioridades de pesquisa contínua para enfrentar desafios persistentes nas tarefas de monitoramento de trincas. Este trabalho utiliza imagens de fissuras induzidas por carregamento, o que pode diferir das imagens, mais naturais, que utilizamos em nossa pesquisa.

O artigo (Xu et al. 2024) destaca que o conjunto de dados é crucial para os resultados da segmentação de fissuras em aprendizado profundo. No entanto, a quantidade e a qualidade das anotações em conjuntos de dados usados para segmentação de fissuras são desiguais, e há uma falta de benchmarks de conjuntos de dados. Para abordar essa questão, este artigo propõe um benchmark de conjunto de dados para segmentação de fissuras de concreto ao nível de píxel, com base em um modelo de aprendizado profundo.

Primeiramente, a capacidade do Resnet-101, DRN, Xception e Mobilenetv2 de extrair características de fissuras é avaliada para selecionar preferencialmente uma

rede de backbone apropriada para o modelo DeepLabV3+. Em segundo lugar, com base no modelo de aprendizado profundo e no conjunto de dados de fissuras, propõe-se um modelo de cálculo universal para estimar o benchmark da quantidade de amostras anotadas de fissuras de alta qualidade, bem como a proporção de mistura entre amostras anotadas de alta e baixa qualidade. Além disso, um método de caracterização de visualização em nível de pixel é construído para inferir detalhes das bordas de fissuras. Os resultados mostram que o modelo com Resnet-101 como rede de backbone apresenta desempenho significativo, com precisão de pixel (PA), intersecção média sobre união (mIoU) e intersecção ponderada por frequência sobre união (FWIoU) de 99,76%, 95,05% e 99,52%, respectivamente.

A taxa de combinação de imagens de fissuras anotadas de baixa e de alta qualidade tem efeitos distintos no desempenho do modelo em conjuntos de dados com diferentes números de amostras. O benchmark proposto para o conjunto de dados de fissuras de concreto tem valor e significância de engenharia e pode ser aplicado para melhorar o desempenho do modelo selecionando com precisão métodos apropriados, como aumentar o número de amostras no conjunto de dados ou otimizar a arquitetura do modelo, ao mesmo tempo em que minimiza o tempo de anotação do conjunto de dados. A pesquisa reafirma o problema de referência de anotações para esta pesquisa.

Um grupo emergente de pesquisas explora a integração de informações provenientes de diferentes modalidades de dados e dimensões espaciais, buscando superar as limitações inerentes às abordagens baseadas exclusivamente em imagens bidimensionais. Estes trabalhos incorporam informações tridimensionais, dados de sensores LiDAR e metodologias de modelagem BIM para proporcionar análises mais completas e precisas das condições estruturais.

Os autores de (Liu et al. 2024) afirmam que os métodos de segmentação de trincas de última geração ainda enfrentam desafios no reconhecimento de trincas rasas de baixo contraste e na compatibilidade multicena. Portanto, este artigo propõe uma solução baseada em uma rede de fusão de informações de estrutura multidimensional (MDSIFNet). Esta rede consiste em dois ramos, um para extrair informações de características espaciais bidimensionais (2D) e outro para adquirir informações geométricas tridimensionais (3D). No primeiro, o módulo de restrição de estrutura de curva 2D, projetado com base em conhecimento prévio combinado a um módulo de extração de características 2D pré-treinado, pode reduzir o número de amostras de aprendizado e realizar a segmentação de trincas em cenas não vistas.

No último, uma transformação one-hot e um módulo de extração de informações de estrutura geométrica 3D são projetados para que a rede preste atenção às características geométricas de trincas rasas e melhore sua precisão de segmentação. Finalmente, um módulo de fusão acopla as informações de características de estrutura

multidimensional desses dois ramos e gera resultados de segmentação pixel a pixel, que podem ser usados para a medição de trincas e para a análise quantitativa. Os resultados experimentais mostram que a rede modelo projetada MDSIFNet apresenta melhor desempenho do que os métodos de estado da arte, tanto em termos de desempenho quanto de resultados visuais. Essa extração de informações geométricas 3D é algo diferenciado nesta pesquisa.

O estudo apresentado em (Bahreini e Hammad 2024) afirma que obter informações precisas sobre áreas defeituosas de infraestruturas ajuda a executar ações de reparo de forma mais eficiente. Recentemente, scanners LiDAR têm sido usados para a inspeção de defeitos de superfície. Além disso, métodos de aprendizado de máquina têm atraído a atenção de pesquisadores para a segmentação semântica e a classificação por meio de nuvens de pontos. Embora tenha sido feito muito trabalho para processar informações visuais por meio de imagens, a pesquisa sobre métodos de aprendizado de máquina para segmentação semântica de dados brutos de nuvem de pontos ainda está em seus estágios iniciais. Além disso, a tecnologia LiDAR é comumente usada para criar modelos BIM no estado em que se encontram. Portanto, o modelo BIM precisa ser integrado aos resultados da segmentação semântica de defeitos após a inspeção baseada em LiDAR.

Abordando as questões acima, este artigo tem os seguintes objetivos: (1) Desenvolver um método de segmentação semântica de defeitos de superfície de concreto com base em nuvens de pontos; e (2) Desenvolver um processo semiautomatizado para modelagem conforme inspecionada. Os desafios relacionados ao tamanho do conjunto de dados e às classes desbalanceadas são abordados. A análise de sensibilidade é aplicada para identificar a combinação de hiperparâmetros mais adequada e investigar seus efeitos no desempenho da rede. O método proposto resultou em recalls de 98,56% e 96,50% para a segmentação semântica de fissuras e lascas, respectivamente. Além disso, o pós-processamento dos resultados da segmentação semântica de defeitos de superfície de concreto é realizado para semiautomatizar a modelagem conforme inspecionada.

O BIM, conforme inspecionado, inclui as informações atualizadas das instalações no momento da coleta de dados. Este processo semiautomatizado permitiu gerenciar e visualizar os defeitos detectados, extraíndo suas dimensões e identificando as condições no modelo 3D. Este trabalho utiliza a metodologia BIM, Building Information Modelling, favorecendo a compreensão da estrutura inspecionada em âmbito 3D.

Metodologias de Aprendizado com Supervisão Reduzida

Uma tendência nas pesquisas recentes é o desenvolvimento de metodologias que reduzem a dependência de grandes volumes de dados anotados, reconhecendo as dificuldades práticas e os elevados custos associados à rotulagem manual de imagens

para tarefas de segmentação. O aprendizado semisupervisionado emerge como uma solução promissora, permitindo o aproveitamento de grandes quantidades de dados não rotulados em conjunto com um conjunto menor de dados rotulados.

Os autores em (Shamsabadi et al. 2024) afirmam que, apesar dos resultados promissores na detecção de trincas superficiais baseadas em visão, abordagens orientadas por dados ainda enfrentam a escassez de conjuntos de dados ricos e rotulados. Tal limitação tem dificultado uma aplicação prática mais ampla de modelos de detecção. Para abordar essa questão, uma estrutura semissupervisionada é proposta, capaz de aprender com uma quantidade substancial de dados não rotulados e atingir alta precisão, mesmo quando os conjuntos de dados rotulados disponíveis são de tamanho limitado. A estrutura é projetada por meio do treinamento supervisionado, da regularização de consistência semissupervisionada e do autotreinamento com pseudo-rotulagem baseada em certeza, resultando em uma abordagem simples, porém eficaz.

Apesar de usar apenas 2% do total de conjuntos de dados rotulados de Concreto e Asfalto, o mIoU resultante foi de 2,6% e 4,7%, respectivamente, menor do que o melhor desempenho do modelo treinado em 100% dos dados rotulados. Notavelmente, a estrutura projetada auxiliou o modelo a se aproximar e até mesmo a exceder os níveis de saturação com apenas 20% e 25% dos conjuntos de dados de concreto e de asfalto. Esta pesquisa se diferencia pelo uso de aprendizado semisupervisionado, exatamente para resolver a questão que igualmente destacamos: a escassez de datasets para detecção.

O aprendizado fracamente supervisionado representa uma outra abordagem para a redução da supervisão necessária, utilizando anotações menos precisas ou completas, como rótulos de imagem inteira ou caixas delimitadoras, em vez de segmentações pixel-level detalhadas. Esta metodologia é particularmente atrativa para aplicações práticas em que a anotação precisa é extremamente custosa ou impraticável.

O estudo apresentado em (Mishra et al. 2024) reforça que a detecção ou segmentação de rachaduras em estruturas de concreto é um processo vital no monitoramento da saúde estrutural (structural health monitoring, SHM). Embora as técnicas de aprendizado de máquina supervisionado tenham obtido grande sucesso neste domínio, a coleta e a anotação de dados continuam sendo desafiadoras.

A coleta de dados de imagem é desafiadora, tediosa e trabalhosa, incluindo o acesso a conjuntos de dados representativos e a rotulagem manual de dados de treinamento no domínio do SHM. Segundo a literatura, há problemas significativos na anotação manual de dados de imagem. Para abordar essa lacuna, este artigo propõe uma estrutura de aprendizado supervisionado fracamente em dois estágios, utilizando uma nova rede de atenção a rachaduras (CrANET) com mecanismo de atenção para

detectar e segmentar rachaduras em imagens sem anotações humanas, com rótulos de nível de pixel.

Esta estrutura classifica imagens de superfície de concreto em rachaduras ou sem rachaduras e, em seguida, utiliza a visualização de mapeamento de ativação de classe de gradiente para gerar a segmentação das rachaduras. Profissionais e especialistas no domínio avaliam posteriormente esses resultados de segmentação por meio de um estudo de validação conduzido por especialistas. Como a literatura sugere que o aprendizado fracamente supervisionado é uma prática limitada em SHM, este título de pesquisa motivará pesquisadores em SHM a investigar e a desenvolver uma abordagem de aprendizado fracamente supervisionado como estado da arte. Esta abordagem reforça nossa tese sobre as dificuldades de anotação e busca separar as etapas das tarefas para melhorar a eficiência na identificação de patologias.

A publicação (Shi et al. 2024) enfatiza que a segmentação automática de agregados de concreto baseada em visão facilita a análise quantitativa rápida da forma, da gradação e da uniformidade dos agregados de concreto. No entanto, os métodos existentes baseiam-se principalmente no aprendizado profundo supervisionado. Como as imagens de superfície de concreto geralmente contêm vários agregados com tamanhos e limites irregulares, isso leva a um processo de anotação de dados demorado e trabalhoso. Para reduzir os custos de anotação, este estudo propõe um método de segmentação de agregados de concreto com aprendizado profundo fracamente supervisionado (WSDL), que pode treinar modelos precisos de segmentação de agregados de concreto em nível de pixel usando apenas anotação de nível de caixa delimitadora de baixo custo. Um conjunto de dados de segmentação de agregados de concreto, com 1400 imagens, é criado e experimentos são conduzidos.

Os resultados mostram que o método proposto pode atingir a segmentação de agregados de concreto em nível de pixel com um AP50 de 95,27%. Em comparação com o método baseado em aprendizado profundo supervisionado, o custo de anotação é reduzido em 86,24%. Este estudo também avalia a viabilidade do método proposto para a quantificação do grau de segregação de agregados de concreto. Sistemas especializados para a análise automática da granulometria e da uniformidade dos agregados de concreto podem ser desenvolvidos com base neste método no futuro. Este trabalho foca na patologia de um dos componentes do concreto, o agregado.

Um desafio recorrente identificado na literatura é a escassez de conjuntos de dados adequados para treinamento de modelos robustos. Esta seção apresenta abordagens inovadoras que buscam contornar esta limitação por meio de técnicas de geração de dados sintéticos e de estratégias de transfer learning, representando soluções práticas para um problema fundamental na área.

Outros trabalhos relacionados foram realizados antes de 2024, dentre os quais

(Jamshidi, El-Badry e Nourian 2023) destaca que um elemento-chave em um sistema automatizado de inspeção visual para estruturas de concreto é a identificação das propriedades geométricas dos defeitos superficiais, como fissuras. Demonstrou-se que redes neurais totalmente convolucionais (FCNs) são ferramentas poderosas para segmentação de fissuras em imagens de inspeção. No entanto, o desempenho dos FCNs depende do tamanho do conjunto de dados com que são treinados.

Na ausência de grandes conjuntos de dados de imagens rotuladas para a segmentação de fissuras no concreto, essas redes podem perder sua excelente precisão de previsão ao serem testadas em um novo conjunto de dados-alvo com diferentes condições de imagem. Neste estudo, em primeiro lugar, uma abordagem de Aprendizagem por Transferência é desenvolvida para permitir que as redes distingam melhor as rachaduras dos pixels de fundo. Um conjunto de dados sintético é gerado e utilizado para ajustar uma U-Net pré-treinada com um conjunto de dados público. Na abordagem de síntese de dados proposta, baseada no aumento de dados CutMix, as imagens de crack do conjunto de dados público são combinadas com as imagens de fundo de um conjunto de dados de destino potencial.

Em segundo lugar, uma vez que as fissuras se propagam ao longo do tempo, propõe-se, para imagens sequenciais de superfícies de concreto, uma nova técnica de fusão de dados temporais. Nessa técnica, as previsões da rede de vários intervalos de tempo são agregadas para melhorar a recuperação das previsões. É mostrado que a aplicação das melhorias propostas aumentou a pontuação F1 e mIoU em 28,4% e 22,2%, respectivamente, o que representa uma melhoria significativa no desempenho da rede de segmentação. Este trabalho permite considerar a forma das fissuras e gera um conjunto de dados sintéticos para minimizar o problema da escassez de dados, com um grande número de imagens. Esta pesquisa difere por desenvolver um conjunto sintético para resolver o problema da escassez de dados na segmentação.

Os trabalhos relacionados, entre os principais, os mais recentes sintetizados na tabela 1, mostram que diversas arquiteturas de redes neurais têm sido utilizadas para promover a identificação de patologias em concreto armado. Não apenas melhorias de métricas, mas também os estudos buscam soluções que possam propiciar ganhos nas análises, como no caso de avaliar o progresso de patologias ao longo do tempo. Um ponto verificado diz respeito às bases de imagens utilizadas. Estas bases podem conter imagens já trabalhadas e melhoradas, que podem apresentar métricas de valor mais elevadas, mas é importante ter imagens que retratem o que é comum em imagens reais de estruturas de concreto armado. As questões relacionadas às bases disponíveis levam alguns trabalhos a produzir suas próprias bases de dados para seus estudos. Importante notar que os artigos existentes mostram que o estudo de redes neurais para identificação de patologias mostra que este é um tema ainda em foco. Os trabalhos

inicialmente usavam apenas tarefas de classificação e, depois, evoluíram para detecção e, agora, são utilizados com segmentação.

Os trabalhos relacionados, entre os principais, os mais recentes sintetizados na tabela 1, mostram que diversas arquiteturas de redes neurais têm sido utilizadas para promover a identificação de patologias em concreto armado. Não apenas melhorias de métricas, mas também os estudos buscam soluções que possam propiciar ganhos nas análises, como no caso de avaliar o progresso de patologias ao longo do tempo.

Artigo	Arquitetura	Dataset	Métrica
Jamshidi 2023	U-Net	Próprio	F1-Score 28.4
Arafin 2024	InceptionV3 EfficientNetB3 (U-Net) InceptionV3 (U-Net)	Próprio	Acurácia 91.98 F1-Score 95.66
Shamsabadi 2024	ImageNet U-Net (ResNet-18)	Ozgenel Crack500	mAP 60.1
Liu 2024	VGG16 ResNet50 Swim-t MDSIFNet	CFD CRACK500	mAP 79.3
Mishra 2024	GradCAM CrANET CBAM	Ozgenel	Acurácia 75 Precisão 90 Sensibilidade 55
Bahreini 2024	DGCNN	Próprio	Sensibilidade 98.56
Deng 2024	ResNet-50 Deeplab V3+	Deepcrack	mAP 87.33
Shia 2024	Unet FPN PSPNet DeepLabV3	Próprio	Precisão 84.5 Sensibilidade 77.4 F1-Score 74.9
Xu 2024	Resnet-101 DRN Xception Mobilenetv2 DeepLabv3 +	Próprio	mAP 95.05
Shi 2024	ResNet101 BiFPN	Próprio	mAP 95.27
Flotzingera 2024	MobileNetV3-Large	CODEBRIM	Sensibilidade 15.00(Crack) 34.09(Effl) 24.44(Spall) 9.43(BExp) 38.39(Rust)
Zhuang 2025	YOLO v11	CODEBRIM Próprio	mAP 66.6

Tabela 1 – Artigos relacionados ao tema da tese .

Além disso, como resultado da pesquisa de trabalhos relacionados, é possível obter referências para conjuntos de dados que propiciam experimentos de uso de redes

neurais convolucionais para a detecção de patologias em concreto armado. A tabela 2 abaixo resume os principais conjuntos de dados pesquisados para a identificação de patologias no concreto armado estudadas.

Dataset	Tarefa	Anotação
Phi-Net - Pacific Earthquake Engineering Research Center Hub ImageNet	Classificação	Fissuras e danos estruturais
Ozgenel - Concrete Crack Images for Classification	Classificação	Fissuras em concreto
CODEBRIM - COncrete DEfect BRidge IMage	Detecção	Patologias em concreto
COCO-Bridge - Common Objects in Context for bridge inspection	Detecção	Identificação estrutural
RDD2020 - Dataset for Smartphone-based Road Damage	Classificação Detecção	Patologias em estradas
Automated Road Crack Detection	Detecção	Rachadura de asfalto
Automatic Detection of Concrete Spalling	Classificação	Fragmentação
Automated Bridge Crack Detection	Classificação	Fissuras em concreto
Multi-Classifer for Reinforced Concrete Bridge Defects	Classificação	Patologias em concreto
CSSC - CCNY Concrete Structure Spalling and Crack	Classificação	Rachadura e lascamento do concreto
Cambridge Bridge Inspection	Classificação	Defeitos em pontes
SDNET2018 - Concrete crack image	Classificação	Rachadura em concreto
Crack Detection in Images of Masonry	Classificação	Fissuras em alvenarias

Tabela 2 – Conjuntos de dados pesquisados.

A análise dos trabalhos relacionados evidencia lacunas que fundamentam e justificam a abordagem metodológica proposta nesta tese. Primeiramente, observa-se que a maioria dos estudos concentra-se em soluções de etapa única, seja para detecção ou classificação, não explorando adequadamente a separação dessas tarefas como estratégia para reduzir falsos positivos, conforme identificado em (Tian et al. 2025) como um problema recorrente em aplicações práticas. Adicionalmente, embora alguns trabalhos abordem individualmente questões como a escassez de dados anotados (Shamsabadi et al. 2024, Mishra et al. 2024) e a necessidade de conjuntos de dados mais realistas (Flotzinger et al. 2024), não se identifica na literatura uma abordagem integrada que combine estratégias de ensemble com metodologias de duas etapas para maximizar simultaneamente a detecção e a precisão da classificação. A criação de um novo conjunto de dados por meio de recortes de imagens realistas, contendo interferências comuns em cenários práticos, representa uma contribuição

metodológica que endereça as limitações identificadas por ([Flotzinger et al. 2024](#)) quanto à aplicabilidade de modelos treinados em dados com condições controladas.

Finalmente, a aplicação de técnicas de ensemble especificamente para a detecção de patologias em concreto armado, combinando arquiteturas de um e dois estágios por meio de média ponderada, constitui uma lacuna metodológica não explorada nos trabalhos analisados, representando um avanço rumo a soluções mais robustas e aplicáveis em contextos reais de inspeção estrutural. Desta forma, a metodologia proposta posiciona-se como uma resposta integrada às principais limitações identificadas no estado da arte, contribuindo para o avanço do campo por meio de uma abordagem que combina rigor científico com aplicabilidade prática.

4 Metodologia

A concepção e o desenvolvimento de um método para identificação automatizada de patologias em estruturas de concreto armado constituem o ponto central desta tese, representando a implementação de uma solução prática e aplicável. Este capítulo apresenta, de forma detalhada, a arquitetura do método proposto, que se fundamenta na integração de técnicas de visão computacional e de ensemble de redes neurais convolucionais para abordar os desafios inerentes à inspeção estrutural automatizada.

O método desenvolvido propõe duas etapas complementares que exploram diferentes estratégias de detecção e classificação, buscando otimizar simultaneamente a capacidade de identificação de manifestações patológicas e a precisão na categorização dos defeitos detectados. Por meio da implementação de técnicas de ensemble e da criação de conjuntos de dados específicos, o método visa superar as limitações identificadas no estado da arte, particularmente quanto à redução de falsos positivos e à aplicabilidade em cenários reais de inspeção. A estruturação metodológica abrange desde os aspectos fundamentais de aquisição e pré-processamento de imagens até as etapas específicas de detecção, classificação e avaliação, estabelecendo um framework sistemático que contribui tanto para o avanço científico na área quanto para a viabilização de aplicações práticas no monitoramento de integridade estrutural.

Este método contribui para a análise de imagens de patologias em concreto ao poupar tempo na análise de imagens com recurso da visão computacional, reduzindo erros, que se devem especificamente a erros humanos originados pela necessidade de repetir a tediosa operação manual de um grande volume de dados, fornecendo operação com uma enorme quantidade de dados de imagem, e fazer uso de conjuntos de dados presentes na literatura, o que pode estimular o surgimento de novas bases. O método é organizado nas seguintes etapas: Pré-processamento das imagens; Detecção de Patologias; Classificação das Patologias; e Avaliação.

Inicialmente, a imagem pode passar por uma etapa de pré-processamento para melhorar sua qualidade. Em seguida, a imagem pré-processada é submetida a uma etapa de detecção que visa identificar as regiões da imagem (ROIs - Region of Interest), que contenham características visuais associadas a uma patologia de estrutura de concreto. As ROIs são então submetidas à etapa seguinte, que classifica a imagem da ROI em classes de patologias identificáveis pelo método. Para avaliar o método proposto nos experimentos, foi acrescida uma etapa de avaliação da assertividade desse método. O método é proposto com as duas etapas, detecção e classificação, realizadas de maneira independente por dois modelos específicos propostos. O primeiro

modelo (detecção) utiliza um ensemble de redes de detecção. Na etapa seguinte, redes de classificação são utilizadas com base nos resultados das caixas delimitadoras obtidas no ensemble da primeira fase. A Figura 14 apresenta a organização do método.

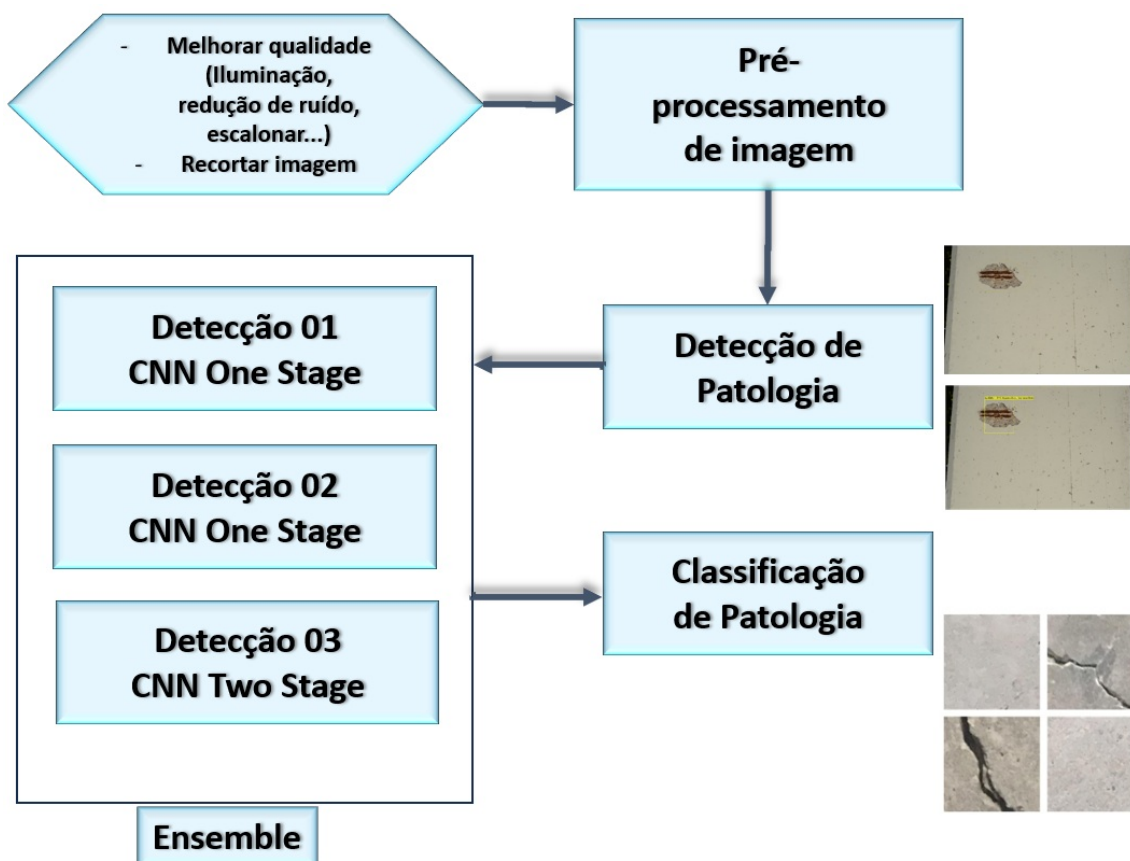


Figura 14 – Método proposto, com etapa de detecção seguida da etapa de classificação. Elaborada pelo autor.

As seções a seguir descrevem em detalhes as etapas do método.

4.1 Pré-processamento de Imagem

O método proposto fundamenta-se na utilização de dois conjuntos de dados distintos para indução dos modelos de detecção e classificação. Cada um desses conjuntos de dados deve ser adequado às tarefas computacionais correspondentes. Para selecionar estes conjuntos, devem ser utilizados critérios estabelecidos com base na análise de trabalhos anteriores na literatura, considerando parâmetros fundamentais, como o volume de imagens disponível, a qualidade técnica das aquisições e a abrangência das anotações fornecidas. Os critérios de seleção definidos contemplaram não apenas aspectos quantitativos, relativos ao número de amostras e anotações, mas também aspectos qualitativos, concernentes à diversidade tipológica das manifestações patológicas representadas nos conjuntos de dados, assegurando, assim, a adequação

dos datasets às demandas específicas das tarefas de detecção e classificação propostas na presente tese.

O primeiro conjunto de dados destina-se especificamente à detecção de patologias em estruturas de concreto. Conforme previamente estabelecido, embora o volume de imagens e a densidade informacional constituam fatores determinantes, priorizou-se a identificação de conjuntos de dados que contemplassem anotações para o espectro mais amplo possível de patologias do concreto armado. A delimitação do escopo concentrou-se em estruturas de concreto, excluindo-se deliberadamente datasets relacionados à pavimentação que, embora amplamente disponíveis devido à sua utilização em diversos países, apresentam menor representatividade no contexto brasileiro. Adicionalmente, estabeleceu-se como critério fundamental a busca por conjuntos de dados que preservassem o maior grau de realismo possível em suas aquisições. Esta abordagem justifica-se pela identificação de limitações inerentes a conjuntos de dados caracterizados por imagens excessivamente processadas, desprovidas de interferências ambientais típicas do contexto real de aplicação, tais como pichações e outras ocorrências urbanas, que, embora propiciem métricas de desempenho superiores durante o treinamento, demonstram transferibilidade limitada quando aplicadas a cenários reais de inspeção estrutural.

Outro conjunto de dados necessário é o de imagens para a tarefa de classificação. Este conjunto é usado para o caso de imagens já recortadas para a patologia a serem classificadas para a patologia de concreto. Neste conjunto, também é importante a questão de grande número de imagens, e que possam estar mais balanceadas possíveis, em relação às classes, para evitar tendência do aprendizado para determinada patologia. Estes conjuntos estão mais disponíveis devido à sua maior facilidade de anotação.

O segundo conjunto de dados requerido pela metodologia destina-se à classificação de patologias em estruturas de concreto. Este conjunto de dados é empregado no processamento de imagens previamente segmentadas, nas quais as regiões de interesse contendo patologias específicas do concreto foram delimitadas. Para este conjunto, estabelece-se como parâmetro crítico não apenas a amplitude quantitativa das amostras disponíveis, mas também a distribuição equilibrada entre as classes representadas, visando mitigar potenciais vieses de aprendizado que possam favorecer determinadas categorias de patologias em detrimento de outras. Observa-se que conjuntos de dados desta natureza apresentam maior disponibilidade na literatura científica, atribuível à relativa simplicidade do processo de anotação quando comparados aos datasets destinados à detecção.

A investigação sistemática e contínua de conjuntos de dados constitui um elemento fundamental para assegurar as condições adequadas de aprendizado das

arquiteturas neurais empregadas na presente tese. A caracterização detalhada dos dois conjuntos de dados utilizados, bem como a descrição pormenorizada do novo dataset desenvolvido por meio de técnicas de segmentação de imagens, são apresentadas na seção que descreve os experimentos realizados nesta tese, na qual se demonstram os procedimentos metodológicos específicos adotados para a criação e validação destes recursos fundamentais à investigação proposta.

A qualidade das imagens adquiridas é fator fundamental para a correta detecção e classificação de possíveis patologias presentes na estrutura de concreto a ser avaliada. Imagens de baixa qualidade podem resultar na não detecção ou até mesmo na classificação incorreta da patologia.

A qualidade das imagens constitui um parâmetro determinante para a eficácia dos processos de detecção e classificação de patologias em estruturas de concreto armado. Imagens de baixa qualidade podem comprometer a acurácia dos algoritmos, resultando em falhas de detecção ou em classificações incorretas das patologias identificadas.

Embora a aquisição controlada de imagens seja desejável, as condições de campo frequentemente impõem limitações que demandam intervenções corretivas (Ai e Kwon 2020), tornando imperativo o emprego de técnicas de pré-processamento para melhorar sua qualidade. As técnicas comumente empregadas para melhoramento incluem correção de iluminação, realce de nitidez, redução de ruído e correção de imperfeições inerentes ao processo de aquisição.

Quando as imagens capturadas apresentam inadequações para as tarefas computacionais propostas, a implementação de etapas de pré-processamento constitui uma estratégia fundamental para viabilizar aplicações em cenários reais. Tais procedimentos propiciam melhorias na precisão das etapas subsequentes de detecção, e as regiões de interesse identificadas também se beneficiam desses aprimoramentos ao longo do processo de classificação patológica.

Várias técnicas podem ser empregadas para o aprimoramento de imagens; a escolha depende do problema que a imagem contém, relacionado às condições em que a imagem será tirada, bem como às condições do ambiente onde se encontra a estrutura a ser analisada.

O arsenal de técnicas disponíveis para o aprimoramento de imagens é extenso, e a seleção metodológica é determinada pelas características específicas das deficiências identificadas, intrinsecamente relacionadas às condições de aquisição e ao ambiente da estrutura analisada. Entre as abordagens disponíveis, destacam-se os filtros de melhoramento, os algoritmos de realce que enfatizam bordas e detalhes estruturais, os ajustes de intensidade para otimizar brilho e contraste, as técnicas

de redução de ruído adaptadas às características específicas das interferências identificadas e os procedimentos de correção cromática para normalizar a distribuição espectral.

Cada imagem constitui um artefato com características determinadas por múltiplas variáveis, incluindo a metodologia de aquisição, as especificações do equipamento utilizado e as condições ambientais vigentes durante a captura, exigindo, assim, abordagens personalizadas para cada contexto específico de aplicação.

O método proposto na presente investigação não incorpora procedimentos de pré-processamento de imagens, fundamentando-se na premissa de que se utilizam bases de dados de qualidade adequada. Entretanto, reconhece-se que aplicações em ambientes com condições de aquisição menos controladas demandarão, necessariamente, a implementação das técnicas de aprimoramento supracitadas para assegurar a eficácia do método proposto.

4.2 Detecção e Classificação de Patologias em Estrutura de Concreto Armado

Subsequentemente à etapa de aquisição, procede-se à identificação de regiões de interesse que apresentem características visuais indicativas da presença de patologias na estrutura analisada. Esta fase resulta na delimitação espacial das áreas correspondentes a cada patologia detectada, por meio da geração de caixas delimitadoras.

Uma abordagem comumente utilizada consiste em realizar simultaneamente a detecção e a classificação de patologias em uma única etapa, na qual uma rede neural de detecção de um estágio identifica espacialmente as patologias por meio de caixas delimitadoras e as classifica em tipos específicos em uma única passagem computacional. As regiões de interesse são determinadas por pontos de ancoragem predefinidos, distribuídos sistematicamente sobre a imagem de entrada, e a imagem processada é rotulada de acordo com as classes de patologia detectadas.

Para o treinamento supervisionado, o conjunto de dados requer anotações que delimitem espacialmente cada patologia por meio de caixas delimitadoras, acompanhadas dos respectivos rótulos de classe (fissura, eflorescência, carbonatação, entre outras), envolvendo a demarcação manual das coordenadas espaciais que circunscrevem cada região de interesse e constituindo o ground truth utilizado durante o treinamento.

Embora esta configuração proporcione significativa eficiência temporal para aplicações em tempo real, apresenta limitações importantes: demonstra capacidade

de reconhecimento em classes específicas, mas possui restrições na identificação exhaustiva de todas as manifestações presentes, visto que o aumento da sensibilidade para melhorar a identificação estabelece um trade-off com a precisão, resultando na proliferação de falsos positivos.

Para superar estas limitações, propõe-se uma metodologia sequencial que dissocia as tarefas de localização e de classificação. Na primeira fase, todas as patologias são anotadas em uma única classe genérica denominada "patologia", eliminando as distinções tipológicas. O modelo concentra-se exclusivamente na localização espacial das regiões de interesse.

Nesta etapa inicial, empregam-se múltiplas arquiteturas neurais integradas por meio de uma estratégia de ensemble baseada em média ponderada. Os pesos atribuídos a cada modelo são determinados pelas métricas de precisão individuais obtidas na validação. Esta diversificação arquitetural aproveita capacidades complementares de diferentes abordagens, produzindo uma identificação mais robusta que supera o desempenho individual de cada modelo constituinte.

Na segunda fase, as regiões identificadas são extraídas das imagens originais por meio do recorte das áreas delimitadas, resultando em um conjunto de imagens recortadas. Estas imagens são processadas por um classificador especializado, treinado para identificar os tipos de patologia e eliminar falsos positivos oriundos da etapa de detecção.

O conjunto de dados de classificação é desenvolvido por meio da extração sistemática de regiões de interesse do dataset de detecção original. Como este caracteriza-se por imagens realistas, contendo interferências ambientais típicas (pichações e outras ocorrências urbanas), esta estratégia assegura a preservação das características visuais e da diversidade tipológica necessárias ao treinamento eficaz do módulo classificador.

Quando a etapa de detecção apresenta elevada incidência de falsos positivos, fenômeno atribuível à limitação quantitativa dos dados de treinamento, à adoção de limiares de sensibilidade excessivamente permissivos ou à combinação de ambos, a implementação de uma segunda rede neural dedicada à classificação constitui uma estratégia eficaz para o refinamento dos resultados e a eliminação de detecções espúrias. Esta abordagem fundamenta a segunda configuração metodológica proposta, visando maximizar a identificação correta de manifestações patológicas.

O método incorpora múltiplas arquiteturas neurais para a tarefa de detecção, integradas por meio de estratégias de ensemble, o que propicia um incremento significativo na capacidade de identificação precisa de patologias.

4.3 Avaliação

A avaliação da metodologia proposta fundamenta-se em um conjunto abrangente de métricas quantitativas aplicadas tanto às tarefas de classificação quanto às de detecção de patologias. Para a avaliação das redes neurais convolucionais empregadas na classificação de imagens recortadas, utiliza-se um conjunto de teste independente, constituído por imagens previamente separadas do conjunto de dados original para esta finalidade específica. O processo de avaliação resulta na geração de uma matriz de confusão, a partir da qual derivam-se as métricas fundamentais para análise de desempenho.

A Acurácia é definida como a proporção de predições corretas em relação ao total de amostras avaliadas, expressa pela equação:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}$$

Onde TP representa os verdadeiros positivos, TN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos.

A Precisão quantifica a proporção de predições positivas que são efetivamente corretas:

$$\text{Precisão} = \frac{TP}{TP + FP}$$

A Sensibilidade (ou Recall) mede a capacidade do modelo em identificar corretamente todas as instâncias positivas, com patologia:

$$\text{Sensibilidade} = \frac{TP}{TP + FN}$$

O F1-Score constitui a média harmônica entre Precisão e Sensibilidade, fornecendo uma métrica balanceada:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$$

Para a avaliação das tarefas de detecção, emprega-se a métrica Mean Average Precision (mAP), que representa a média das precisões médias calculadas para cada classe de patologia. A mAP é computada considerando-se diferentes limiares de Intersection over Union (IoU), definido como:

$$\text{IoU} = \frac{\text{Área de Interseção}}{\text{Área de União}}$$

Onde a área de interseção corresponde à sobreposição entre a caixa delimitadora predita e a verdadeira, e a área de união representa a área total coberta por ambas as caixas.

A Average Precision (AP) para cada classe é calculada como a área sob a curva Precisão-Recall:

$$AP = \int_0^1 \text{Precisão}(r) dr$$

sendo a mAP definida como:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i$$

onde N representa o número total de classes de patologias.

5 Resultados

Este capítulo apresenta os resultados obtidos nos experimentos realizados para validar o método proposto para detecção e classificação automatizadas de patologias em estruturas de concreto armado. O conjunto de experimentos conduzidos visa demonstrar o desempenho da abordagem proposta e proporcionar fundamentação empírica para justificar a aplicabilidade do método e compreender seus desempenhos em cenários de aplicação.

A condução dos experimentos fundamenta-se em um método que contempla a seleção de conjuntos de dados adequados às tarefas propostas, seguida da implementação de experimentos específicos de detecção e classificação, com variações controladas nas arquiteturas neurais empregadas. Os experimentos incluem análises de combinações entre diferentes arquiteturas em uma mesma tarefa computacional, bem como a integração sequencial de tarefas, visando otimizar o fluxo da metodologia e maximizar os resultados obtidos, aplicando-se tais procedimentos a ambas as abordagens propostas.

Os resultados experimentais são apresentados de forma comparativa, incluindo análises de desempenho com base nas métricas estabelecidas e discussões críticas sobre as vantagens e limitações identificadas. Adicionalmente, este capítulo apresenta a comparação dos resultados alcançados com trabalhos identificados na revisão bibliográfica, posicionando as contribuições da presente investigação no contexto do estado da arte atual e demonstrando os avanços científicos e práticos proporcionados pelo método desenvolvido.

Os experimentos computacionais foram realizados em uma infraestrutura híbrida composta por dois ambientes distintos. O ambiente local consistiu em uma estação de trabalho equipada com processador Intel Core i7-14700K, 32 GB de memória RAM e duas unidades de processamento gráfico NVIDIA GeForce RTX 4060 Ti, cada uma com 16 GB de memória dedicada. Para experimentos que demandaram maior capacidade computacional, utilizou-se um ambiente de computação do Laboratório AlVLAB@Deucalion – Laboratório virtual de Inteligência Artificial da Fundação para a Ciência e Tecnologia FCT de Portugal disponibilizado através dos projetos 2024.AIVLAB.00002.VISION e 2025. AIVLAB.00008. DEUCALION, no qual foram disponibilizadas quatro unidades NVIDIA A100 com 40 GB de memória VRAM cada. Esta configuração permitiu a execução eficiente dos algoritmos propostos, atendendo tanto às demandas de processamento local quanto aos requisitos de alta performance para experimentos de maior escala.

Para a implementação e o treinamento das arquiteturas de redes neurais profundas empregadas neste trabalho, adotou-se o framework PyTorch, versão 2.7.1 (Paszke et al. 2019), como plataforma principal de desenvolvimento. O PyTorch, construído sobre a biblioteca Torch (Collobert et al., 2011), oferece flexibilidade e eficiência para prototipagem e experimentação com modelos de aprendizado profundo. Complementarmente, foram utilizados frameworks especializados, incluindo o Ultralytics para a implementação do YOLO v11 e o MMDetection (Chen et al., 2019), uma toolbox abrangente que disponibiliza implementações de referência para diversos algoritmos de estado da arte em visão computacional. A exceção foi feita para os modelos da família YOLO, que foram implementados utilizando sua biblioteca nativa específica, garantindo compatibilidade total com os pesos pré-treinados oficiais e com as otimizações específicas da arquitetura.

O método de divisão dos dados adotado em todos os experimentos seguiu a estratégia de particionamento estratificado, com 70% das amostras destinadas ao conjunto de treinamento, 20% à validação e 10% ao teste. Esta proporção foi mantida de forma consistente em cada classe do problema, garantindo uma representatividade estatística adequada em todos os subconjuntos. O particionamento estratificado assegurou que a distribuição original das classes fosse preservada em cada divisão, mitigando potenciais vieses de amostragem.

5.1 Conjunto de Dados

Entre os conjuntos de dados disponíveis publicamente, selecionaram-se dois conjuntos principais para realizar os experimentos de validação do método, cada qual especificamente adequado às tarefas computacionais propostas. Para a tarefa de classificação, o conjunto de dados selecionado caracteriza-se pelo alinhamento direto com os objetivos da pesquisa e pela significativa amplitude quantitativa das amostras disponíveis. Para a tarefa de detecção, a escolha fundamentou-se na identificação de um conjunto de dados com a abrangência adequada das patologias anotadas, maximizando, assim, as possibilidades de detecção automatizada. Adicionalmente, desenvolveu-se um terceiro conjunto de dados destinado especificamente à tarefa de classificação, por meio da definição de ROIs sobre o conjunto de dados de detecção, visando preservar as características das imagens e equiparar o desafio computacional às condições encontradas em aplicações práticas.

O conjunto de dados OZGENEL, proposto em (Ozgenel 2019), é especializado na classificação binária de presença ou ausência de fissuras em estruturas de concreto. Esta base de dados contém 40 mil imagens balanceadas para as duas categorias, 20 mil imagens contendo estruturas de concreto com fissuras e 20 mil imagens que não apresentam fissuras visíveis na estrutura.

O conjunto de dados CODEBRIM (CODEBRIM 2019) é utilizado especificamente para a detecção de patologias em estruturas de concreto, contendo, adicionalmente, a classe da patologia anotada. Este conjunto de dados apresenta cinco categorias de patologias distintas: fissuração, fragmentação, eflorescência, exposição de armaduras e manchas de corrosão. As imagens apresentam dimensões variáveis, com codificação RGB, refletindo condições realistas de aquisição em campo.

A distribuição quantitativa das patologias no dataset CODEBRIM (Mundt et al. 2019) apresenta-se da seguinte forma: 2.507 ocorrências de trincas, 1.898 casos de fragmentação (*spalling*), 833 casos de eflorescência, 1.507 situações de barras expostas e 1.559 casos de manchas de corrosão. Esta distribuição heterogênea reflete a variabilidade natural na ocorrência de diferentes manifestações patológicas em estruturas reais.

5.1.1 Ajustes dos Conjuntos de Dados

A preparação adequada dos conjuntos de dados constitui uma etapa fundamental para assegurar a qualidade e confiabilidade dos modelos desenvolvidos, demandando frequentemente ajustes e correções nas anotações disponíveis para maximizar a eficácia do processo de treinamento. A identificação e correção de inconsistências nas bases de dados selecionadas revelaram-se necessárias para garantir a integridade dos dados utilizados e otimizar o aproveitamento das informações disponíveis nos conjuntos de dados empregados.

A realização dos experimentos demandou a adaptação dos arquivos de anotação de caixas delimitadoras (bounding boxes) para compatibilização com a arquitetura YOLO, conforme ilustrado na Figura 15. O conjunto de dados CODEBRIM emprega um sistema de anotação baseado em coordenadas cartesianas dos vértices extremos dos retângulos delimitadores, especificamente os pontos correspondentes aos cantos opostos da diagonal principal que definem geometricamente a região de interesse.

Em contrapartida, a arquitetura YOLO utiliza um sistema de representação alternativo para as caixas delimitadoras, fundamentado nas coordenadas do centroide geométrico do retângulo, acrescidas das dimensões espaciais correspondentes à altura e à largura da região delimitada. Esta divergência de representação geométrica entre os sistemas de anotação tornou necessária a implementação de um processo de conversão sistemática dos arquivos de anotação.

A transformação entre os sistemas de coordenadas foi realizada através da aplicação das seguintes relações matemáticas:

$$x_{center} = \frac{x_{min} + x_{max}}{2}$$

$$y_{center} = \frac{y_{min} + y_{max}}{2}$$

$$width = x_{max} - x_{min}$$

$$height = y_{max} - y_{min}$$

onde (x_{min}, y_{min}) e (x_{max}, y_{max}) representam as coordenadas dos vértices extremos no formato CODEBRIM, e $(x_{center}, y_{center}, width, height)$ constituem os parâmetros requeridos pelo formato YOLO.

Para automatizar este processo de conversão, desenvolveu-se um algoritmo específico que processa sistematicamente todos os arquivos de anotação, realizando a transformação geométrica necessária e gerando os arquivos de saída no formato compatível com a arquitetura YOLO. Esta abordagem automatizada assegurou não apenas a eficiência do processo de conversão, mas também a consistência e precisão das transformações aplicadas ao conjunto completo de anotações disponíveis no conjunto de dados.

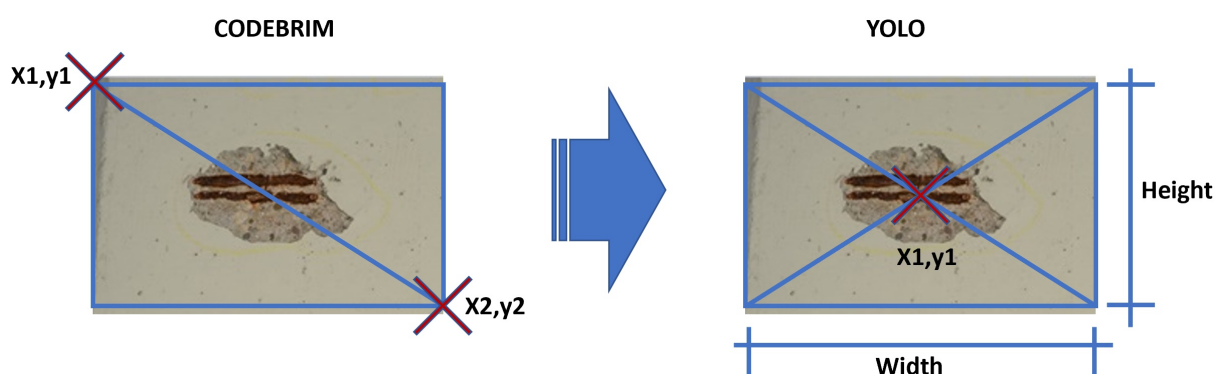


Figura 15 – Conversão da bounding box do dataset CODEBRIM para YOLO. Elaborada pelo autor.

Procedeu-se também a uma revisão do conjunto de dados ([CODEBRIM 2019](#)) visando identificar e corrigir as inconsistências identificadas analisando as anotações relacionadas à patologia de fissuração do concreto. Durante este processo de análise, foram identificados casos em que as coordenadas das caixas delimitadoras apresentavam posicionamento completamente inadequado em relação às patologias efetivamente presentes nas imagens. A análise detalhada destes casos revelou que tais inconsistências provavelmente decorriam de operações de rotação aplicadas às imagens após o processo de anotação, resultando na dessincronização entre as coordenadas das caixas delimitadoras e a localização real das patologias.

Foi então realizada uma revisão das anotações, observando cada imagem. Todas as imagens foram revisitadas para que, visualmente, fosse possível constatar a correteza das anotações. Os casos identificados visualmente como com caixas

erroneamente colocadas fora da patologia ou com definição de patologia equivocada foram corrigidos.

A Figura 16 exemplifica um caso característico de anotação incorreta atribuível à rotação da imagem, demonstrando a discrepância entre a localização real da patologia e o posicionamento da caixa delimitadora correspondente. Tais correções mostraram-se fundamentais para assegurar a qualidade do dataset e, conseqüentemente, a confiabilidade dos resultados experimentais.



Figura 16 – Anotação incorreta devido ao giro na imagem. Elaborada pelo autor.

Adicionalmente, procedeu-se a uma estratégia de otimização do aproveitamento das informações disponíveis, por meio da identificação e anotação de patologias de fissuração presentes nas imagens que não haviam sido contempladas no processo de anotação original. Esta abordagem permitiu incrementar significativamente o volume de dados anotados disponíveis para treinamento, maximizando, assim, o potencial de aprendizado das arquiteturas neurais empregadas.

Uma revisão técnica das anotações foi realizada pelo autor, tendo em vista sua formação em engenharia civil e seu mestrado em estruturas de concreto armado. Nesta etapa de ajuste das anotações pelo autor, as fissuras ramificadas foram meticulosamente verificadas. Alguns grupos de anotações consideravam toda a área da fissura como uma única patologia, enquanto outros anotavam por subgrupos de ramos. Essas anotações foram padronizadas de modo a que uma única caixa de delimitação fosse utilizada para cada fissura contendo os subramos.

Por fim, foi realizado um esforço para anotar patologias que não constavam no conjunto de dados utilizado. Este processo de revisão das anotações gerou uma nova versão do conjunto de dados que denominamos CODEBRiMRev.

Alguns exemplos de inconsistências corrigidas são apresentados nas figuras a seguir. A Figura 17 exemplifica um caso de anotação que não se encaixa na patologia de corrosão. A Figura 19 apresenta um caso em que a fissura foi anotada de forma descontínua. A Figura 18 exibe casos em que as fissuras se ramificam, formando árvores, subconjuntos, e a anotação é feita de forma distinta. A Figura 20 mostra casos de anotações que não foram realizadas. A Figura 20 apresenta um exemplo em que a imagem é classificada como um tipo de patologia, confundindo detecção com classificação. Estes problemas, se não corrigidos, irão ter direta influência nas métricas obtidas para detecção e classificação de patologias nas imagens.

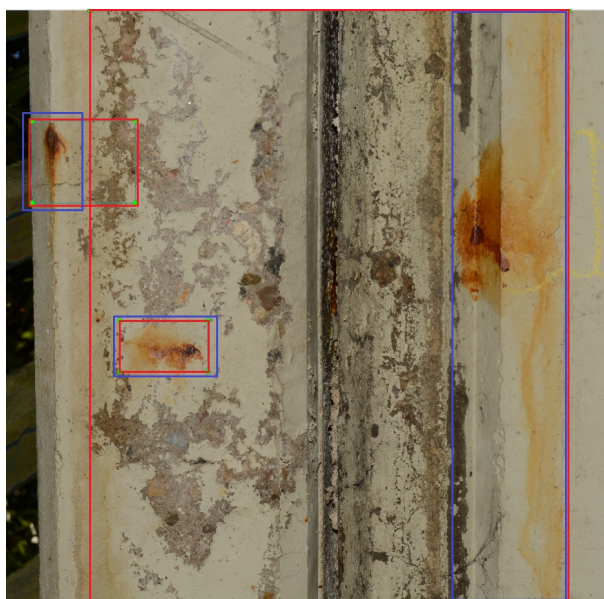


Figura 17 – Anotação imprecisa de mancha de corrosão. Vermelho a anotação original e azul a correção. Elaborada pelo autor.

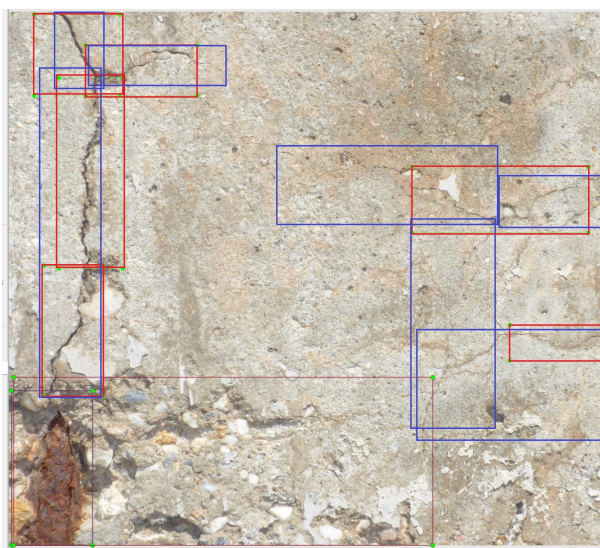


Figura 18 – Anotação de fissura, ora feita por ramo, ora feita por conjunto de fissuras. Vermelho a anotação original e azul a correção. Elaborada pelo autor.

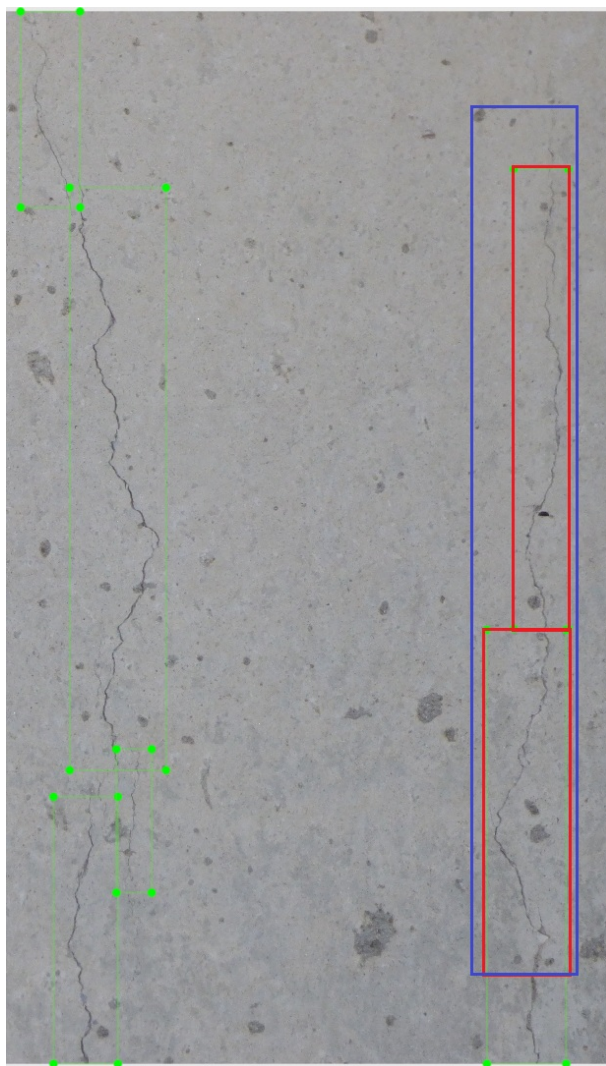


Figura 19 – Anotação seccionada de fissura. Vermelho a anotação original e azul a correção. Elaborada pelo autor.

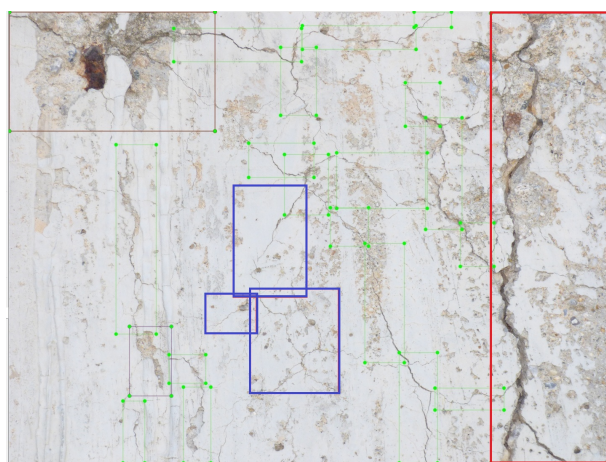


Figura 20 – Anotações de fissuras que não foram realizadas. Vermelho a anotação original e azul a correção. Elaborada pelo autor.



Figura 21 – Anotação que trata a imagem inteira como um tipo de patologia, confundindo detecção com classificação. Vermelho, a anotação original; azul, a correção. Elaborada pelo autor.

5.2 Classificação com conjunto de dados Ozgenel

Para avaliar a possibilidade de uso de redes neurais para classificar patologias em estruturas de concreto, realizamos experimentos iniciais com o conjunto de dados (Ozgenel 2019), que apresenta somente as ROIs de imagens de estruturas anotadas conforme a classe da patologia. A Figura 22 apresenta alguns exemplos de imagens do conjunto de dados, contendo fissuras e imagens sem fissuras. Conforme ilustrado na figura, estas imagens são recortadas para a patologia de fissuras no concreto.



Figura 22 – Imagens dos dois tipos de classes do conjunto de dados (Ozgenel 2019), com e sem fissuras. Elaborada pelo autor.

O modelo foi construído com o objetivo de reconhecer as duas classes presentes no conjunto de dados (Ozgenel 2019), estruturas com e sem fissuras. A arquitetura da CNN é apresentada na Figura 23. Na Figura 24 é possível ver o resultado de uma classificação com o conjunto de dados (Ozgenel 2019).

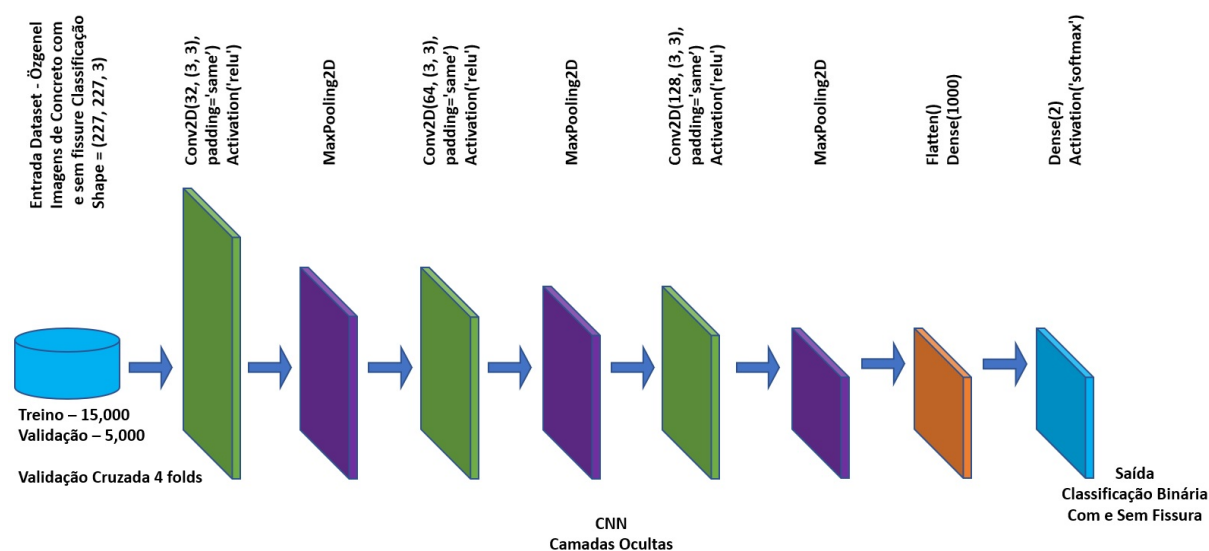


Figura 23 – Estrutura e configuração de dados usadas para a CNN. Elaborada pelo autor.



Positive test: This image is 100.00 percent Positive and 0.00 percent Negative.



Negative test: This image is 0.00 percent Positive and 100.00 percent Negative.

Figura 24 – Resultado de classificação para duas imagens das duas classes do conjunto de dados (Özgenel 2019).

Elaborada pelo autor.

Com o intuito de testar uma rede mais robusta, o segundo experimento foi realizado utilizando a rede YOLO (?). A acurácia obtida é de 0,98. Nesta fase inicial, as demais análises de métricas não foram aprofundadas, mas apenas foi avaliada a possibilidade de aplicar redes neurais para classificar patologias no concreto. A Tabela 3 apresenta a matriz de confusão obtida a partir do experimento realizado.

	01	02
01 - Positivo	4969	30
02 - Negativo	22	4977

Tabela 3 – Matriz de confusão para 4999 casos positivos e 4999 negativos.

O resultado obtido teve precisão de 99,43%, utilizando validação cruzada. A Figura 25 apresenta o resultado de uma seção. Também nesta figura é possível observar que o treinamento e a validação convergiram após cerca de 20 épocas. O experimento foi realizado com validação cruzada utilizando quatro partes de treinamento, levando a resultados de 99,45%, 99,43%, 99,41% e 99,44%. Isso resultou em uma precisão média de 99,43% e um desvio padrão de 0,017078.

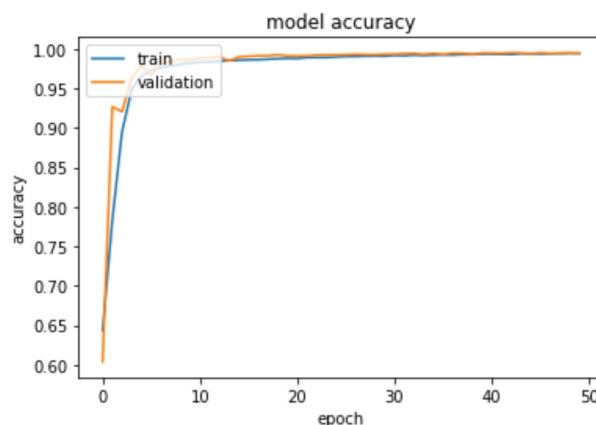


Figura 25 – Acurácia obtida pela CNN em uma das etapas de validação cruzada. Elaborada pelo autor.

A Figura 26 mostra a evolução da função de perda ao longo de uma das etapas da validação cruzada.

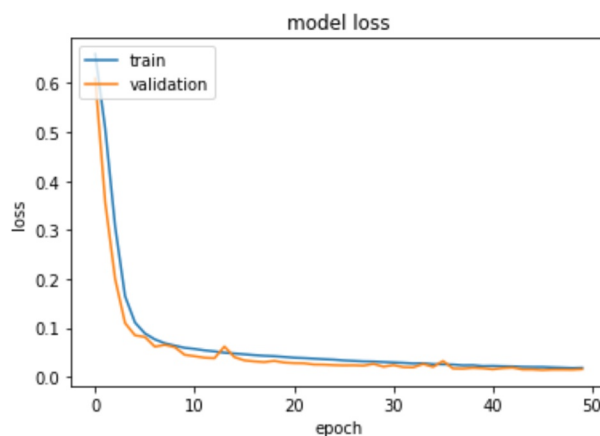


Figura 26 – Valor da função de perda na arquitetura CNN, durante o treinamento. Elaborada pelo autor.

Os testes realizados permitem avaliar o desempenho da rede. Inicialmente, utilizou-se um conjunto de mil imagens para avaliar o desempenho. A tabela 3 apresenta a matriz de confusão, a partir da qual se inferiu um desempenho de 99,48% para a rede neural. O parâmetro ROC AUC é 0,999876. AUC corresponde à área sob a curva ROC, Figura 27.

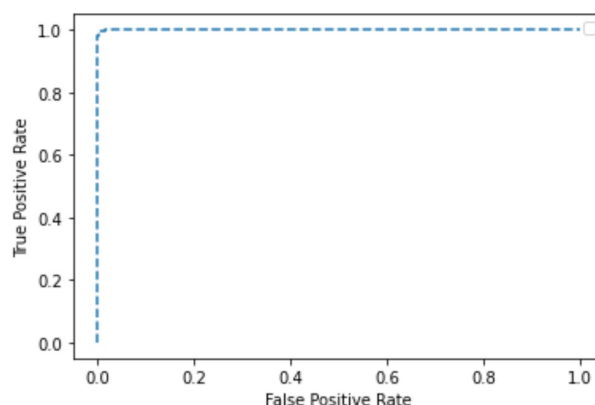


Figura 27 – ROC AUC. Elaborada pelo autor.

A Figura 24 com duas imagens, uma com e outra sem fissura, ilustra a capacidade da rede de classificar uma imagem individualmente. Este teste visa simular a situação em que uma imagem individual precisa ser classificada.



Figura 28 – Classificação negativa para fissura. Elaborada pelo autor.



Figura 29 – Classificação positivo para fissura. Elaborada pelo autor.

5.3 Detecção e Classificação de Patologias em uma Etapa

A detecção e classificação simultâneas de patologias em estruturas de concreto armado representam uma possibilidade importante, consolidando em um único processo computacional as tarefas de localização espacial e identificação tipológica dos defeitos. Neste capítulo, avaliamos o desempenho de diferentes arquiteturas de redes neurais convolucionais aplicadas à detecção e classificação unificadas de patologias, comparando abordagens de estágio único e múltiplos estágios.

Estes experimentos foram realizados utilizando três arquiteturas diferentes, duas com estágio único (YOLO e TOOD - Task-aligned One-stage Object Detection) e uma com dois estágios (Faster RCNN). Para efeito de comparação com outros trabalhos, realizamos o experimento com o conjunto de dados original e com o conjunto de dados após as correções descritas nesta tese.

O aprendizado para detecção das patologias é utilizado para classificar qual classe de patologia a imagem dada possui. As redes detectoras são treinadas para a tarefa de detecção, e utilizadas para indicar qual patologia dada imagem possui.

5.3.1 Detecção e Classificação de Patologias com Arquitetura YOLO - Conjunto de dados CODEBRIM

Este experimento foi realizado utilizando o conjunto de dados CODEBRIM ([CODEBRIM 2019](#)), para permitir comparação com outros trabalhos que utilizam a mesma versão do conjunto de dados e para avaliar a influência das anotações incorretas nos modelos desenvolvidos.

Na etapa de redimensionamento das imagens, todas foram ajustadas ao tamanho padrão de 640×640 pixels. Essa resolução foi escolhida por ser o padrão da arquitetura YOLO utilizada.

Foi utilizada a arquitetura YOLO v11, com backbone DarkNet, treinada para detectar patologias e classificar as regiões detectadas em cinco classes distintas.

O treinamento foi realizado com os seguintes parâmetros: Sigmoid como função de ativação; processo de otimização com algoritmo Adam e learning rate inicial de 0.008; processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

O algoritmo Adam é usado para ajustar a taxa de aprendizado para cada parâmetro da rede, usando estimativas de primeiro e segundo momentos (média e variância dos gradientes). Isso permite treinamento rápido e estável mesmo com dados esparsos ou ruidosos. É o otimizador padrão em muitos frameworks como PyTorch e Keras, devido à sua eficácia empírica. Foi realizada transferência de aprendizagem usando o modelo YOLO11x (extra large) disponibilizado pela ultralytics, este modelo é treinado na base COCO, Common Objects in Context. COCO é um conjunto de dados de objetos anotados para detecção e segmentação em larga escala, 330K imagens (>200K anotações).

Durante a fase de inferência, foi definido o limiar de confiança de 0,5 para a medida de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

5.3.2 Detecção e Classificação de Patologias com Arquitetura Faster RCNN - Conjunto de dados CODEBRIM

O segundo experimento desta fase foi realizado com rede neural Faster RCNN, empregando como backbone a arquitetura ResNet50 para extração de características. O conjunto de dados utilizado foi igualmente o CODEBRIM ([CODEBRIM 2019](#)). A ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de dois estágios.

As imagens do conjunto de dados foram redimensionadas para o menor lado de 800 pixels, e a outra dimensão foi calculada para manter a razão de aspecto da imagem.

O modelo foi configurado para reconhecer cinco classes distintas, utilizando Softmax como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam, com learning rate inicial de 0,005, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de 0,5 para o limiar de confiança de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

5.3.3 Detecção e Classificação de Patologias com Arquitetura TOOD - Conjunto de dados CODEBRIM

O terceiro experimento desta fase foi realizado com TOOD. O conjunto de dados utilizado foi igualmente (CODEBRIM 2019). A ideia deste experimento foi testar o desempenho da tarefa de detecção com uma rede neural de um estágio, de modo a ter outra arquitetura de um estágio que corroborasse os resultados do experimento realizado com a rede YOLO.

A rede neural foi ajustada para processar imagens com resolução de entrada de 640 pixels, empregando a arquitetura ResNet50 como backbone para extração de características.

O treinamento realizado com a seguinte configuração: O modelo foi configurado para reconhecer igualmente cinco classes distintas, utilizando Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo SGD, com learning rate inicial de 0,01, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

O SGD (Stochastic Gradient Descent, ou Gradiente Descendente Estocástico) é um método de otimização para atualizar os parâmetros da rede neural usando o gradiente da função de perda calculado em mini-lotes (subconjuntos aleatórios de dados). Ele move os pesos na direção oposta ao gradiente em cada interação (descendendo o gradiente). Ele pode ser mais ruidoso e oscilatório do que o Gradiente Descendente, o que levou a variações, como SGD com Momentum e Adam, para melhorar a convergência.

Durante a fase de inferência, foram definidos thresholds criteriosos para garantir a qualidade das detecções, estabelecendo-se o valor de 0,5 para o threshold de confiança e 0,5 para o threshold de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

5.3.4 Associação dos Resultados com Ensemble

Visando obter um melhor resultado, que pudesse inclusive utilizar as redes experimentadas, as três redes utilizadas foram associadas em um Ensemble.

5.3.4.1 Detecção com Ensemble Usando Weighted Box Fusion - Conjunto de dados CODEBRIM

A estratégia adotada, com o uso de Ensemble, foi utilizar os resultados das redes neurais treinadas e processadas para uma solução otimizada, neste caso, Weighted

Box Fusion. Este método combina previsões de modelos de detecção de objetos. A pontuação da confiança de todas as caixas delimitadoras propostas é utilizada para construir caixas médias. Essa abordagem já foi usada para vencer desafios do Open Images e COCO Object Detection. As coordenadas da caixa de detecção são definidas pela média ponderada das caixas encontradas nos diferentes modelos utilizados, para obter um posicionamento mais bem encaixado no artefato detectado.

Os resultados dos experimentos, YOLO, TOOD e Faster RCNN, foram então combinados para obter um melhor desempenho, o esperado com a aplicação de Ensemble. Os parâmetros utilizados nas redes neurais são os descritos nas seções anteriores.

A revisão do conjunto de dados ([CODEBRIM 2019](#)) resulta em um novo conjunto com erros corrigidos e novas anotações ou alterações de anotações para melhoria do aprendizado, conforme citado na secção de conjunto de dados. Após a revisão, uma nova rodada de experimentos foi realizada na busca de melhores resultados.

Foram refeitos os experimentos utilizando os mesmos três modelos empregados na primeira fase de experimentos desta tese, de detecção. A ideia dos três modelos foi confrontar não apenas o desempenho de modelos diferentes, mas igualmente o desempenho de redes de um estágio, como YOLO, e de dois estágios, como Faster RCNN, e, desta vez, usar o conjunto de dados revisado.

5.3.5 Detecção e Classificação de Patologias com Arquitetura YOLO - Conjunto de dados CODEBRIMRev

O primeiro experimento desta nova fase foi realizado com YOLO v11. O conjunto de dados utilizado foi ([CODEBRIM 2019](#)), desta vez revisado, CODEBRIMRev. A ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de um estágio.

Na etapa de redimensionamento das imagens, todas foram ajustadas para o tamanho padrão de 640 pixels. Essa resolução foi escolhida por ser o padrão da Arquitetura YOLO utilizada.

O treinamento foi realizado com a seguinte configuração: Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam com learning rate inicial estabelecido em 0.008, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de confiança de 0,5 a medida de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

As métricas obtidas são apresentadas na tabela 4.

Patologia	Acurácia	Sensibilidade	Precisão	F1 score
Mancha Corrosão	66.99	77.63	77.63	77.63
Fissura	52.43	87.72	54.35	67.11
Eflorescência	66.02	77.14	50.00	60.67
Barra Exposta	83.50	94.55	78.79	85.95
Lascamento	74.76	90.16	73.33	80.88

Tabela 4 – Métricas de classificação e detecção com YOLO no CODEBRIMRev.

As tabelas 5, 6, 7, 8 e 9 mostram as matrizes de confusão obtidas dos experimentos realizados, no caso das cinco patologias experimentadas, mancha de corrosão, fissuras, lascamento, eflorescência e barra exposta.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	09.71%	16.50%
	Positivo	16.50%	57.28%

Tabela 5 – Matriz de confusão para detecção de manchas de corrosão com YOLO no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	03.88%	40.78%
	Positivo	06.80%	48.54%

Tabela 6 – Matriz de confusão para detecção de fissuras com YOLO no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	39.81%	26.21%
	Positivo	07.77%	26.21%

Tabela 7 – Matriz de confusão para detecção de eflorescências com YOLO no CODEBRIMRev.

5.3.5.1 Experimentos considerando diferentes Thresholds - CODEBRIMRev

Uma questão bastante em foco neste trabalho é o aumento da sensibilidade da rede testada para a detecção, em busca da identificação do maior número possível de patologias. Uma possibilidade para aumentar a sensibilidade é a adoção de um limiar mais baixo. Em contrapartida, a melhor sensibilidade, obtém-se uma precisão menor, então estes testes podem mostrar o comportamento destas duas importantes métricas.

Na etapa de redimensionamento das imagens, todas foram ajustadas ao tamanho padrão de 640 pixels.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	33.01%	13.59%
	Positivo	02.91%	50.49%

Tabela 8 – Matriz de confusão para detecção de barras expostas com YOLO no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	21.36%	19.42%
	Positivo	05.83%	53.40%

Tabela 9 – Matriz de confusão para detecção de lascamentos com YOLO no CODEBRIMRev.

O experimento foi implementado utilizando a arquitetura YOLO v11, com backbone DarkNet, treinada para detectar patologias e classificar as regiões detectadas em cinco classes distintas.

O treinamento foi realizado com a seguinte configuração: Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam com learning rate inicial estabelecido em 0.008, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de confiança de 0,5 a medida de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

O pensamento é que esta abordagem aumenta o número de detecções, com o prejuízo de muitos falsos positivos, mas isso pode ser corrigido com uma rede de detecção especializada para classificação. Estes experimentos foram conduzidos com YOLO v11, com o conjunto de dados ([CODEBRIM 2019](#)), revisado, para os limites de 0,1, a 0,9, variando a cada 0,1.

5.3.6 Detecção e Classificação de Patologias com Arquitetura Faster RCNN - Conjunto de dados CODEBRIMRev

O segundo experimento desta nova fase foi realizado com Faster RCNN. O conjunto de dados utilizado foi igualmente ([CODEBRIM 2019](#)), desta vez revisado como CODEBRIMRev. A ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de dois estágios.

As imagens do conjunto de dados foram redimensionadas para imagens retangulares, com o menor lado de 800 pixels, e a outra dimensionada para manter a razão de aspecto da imagem.

O treinamento realizado com a seguinte configuração: O modelo foi configurado para igualmente reconhecer cinco classes distintas, utilizando Softmax como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam com learning rate inicial estabelecido em 0.005, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de 0,5 para o limiar de confiança de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

As métricas obtidas são apresentadas na tabela 10.

Patologia	Acurácia	Sensibilidade	Precisão	F1 score
Mancha Corrosão	66.02	72.37	79.71	75.86
Fissura	57.28	84.21	57.83	68.57
Eflorescência	82.52	85.71	69.77	76.92
Barra Exposta	88.35	94.55	85.25	89.66
Lascamento	75.73	83.61	77.27	80.31

Tabela 10 – Métricas de classificação e detecção com Rede Faster R-CNN no CODEBRIMRev.

As tabelas 11, 12, 13, 14 e 15 mostram as matrizes de confusão obtidas a partir dos experimentos realizados, para as cinco patologias experimentadas: mancha de corrosão, fissuras, lascamento, eflorescência e barra exposta.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	12.62%	13.59%
	Positivo	20.39%	53.40%

Tabela 11 – Matriz de confusão para a detecção de manchas de corrosão com Faster R-CNN no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	10.68%	33.98%
	Positivo	08.74%	46.60%

Tabela 12 – Matriz de confusão para detecção de fissuras com Faster RCNN no CODEBRIMRev.

5.3.7 Detecção e Classificação de Patologias com Arquitetura TOOD - Conjunto de dados CODEBRIMRev

O terceiro experimento desta nova fase foi realizado com TOOD. O conjunto de dados utilizado foi igualmente (CODEBRIM 2019), desta vez revisado, CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	53.40%	12.62%
	Positivo	04.85%	29.13%

Tabela 13 – Matriz de confusão para detecção de eflorescências com Faster RCNN no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	37.86%	08.74%
	Positivo	02.91%	50.49%

Tabela 14 – Matriz de confusão para detecção de barras expostas com Faster RCNN no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	26.21%	14.56%
	Positivo	09.71%	49.51%

Tabela 15 – Matriz de confusão para detecção de lascamento com Faster RCNN no CODEBRIMRev.

Destacando que a ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de um estágio, visto que a primeira, YOLO, apresentou melhor desempenho, e esta análise permitiu verificar se era a implementação em particular ou a arquitetura de um estágio que realmente tinha bom desempenho para esta pesquisa.

A rede neural foi ajustada para processar imagens com resolução de entrada de 640 pixels, empregando como backbone a arquitetura ResNet50 para extração de características.

O treinamento realizado com a seguinte configuração: o modelo foi configurado para reconhecer igualmente cinco classes distintas, utilizando Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo SGD, com learning rate inicial de 0,01, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foram definidos thresholds criteriosos para garantir a qualidade das detecções, estabelecendo-se o valor de 0.5 para o threshold de confiança de 0.5 para o threshold de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

As métricas obtidas são apresentadas na tabela 16.

Patologia	Acurácia	Sensibilidade	Precisão	F1 score
Mancha Corrosão	72.82	96.05	74.49	83.91
Fissura	55.34	94.74	55.67	70.13
Eflorescência	34.95	94.29	33.67	49.62
Barra Exposta	60.19	96.36	57.61	72.11
Lascamento	58.25	95.08	59.18	72.96

Tabela 16 – Métricas de classificação/detecção com TOOD (CODEBRIMRev).

As tabelas 17, 18, 19, 20 e 21 mostram as matrizes de confusão obtidas a partir dos experimentos realizados, para as cinco patologias experimentadas: mancha de corrosão, fissuras, lascamento, eflorescência e barra exposta.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	01.94%	24.27%
	Positivo	02.91%	70.87%

Tabela 17 – Matriz de confusão para detecção de manchas de corrosão com TOOD no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	02.91%	41.75%
	Positivo	02.91%	52.43%

Tabela 18 – Matriz de confusão para detecção de fissuras com TOOD no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	02.91%	63.11%
	Positivo	01.94%	32.04%

Tabela 19 – Matriz de confusão para detecção de eflorescências com TOOD no CODEBRIMRev.

5.4 Identificação de patologias em duas etapas

O segundo método propõe o uso de duas etapas para a identificação de patologias no concreto armado. A primeira etapa é dedicada à localização de patologias, sem necessidade de determinar a qual classe a patologia pertence. A segunda etapa é dedicada à classificação da patologia previamente detectada, incluindo a eliminação de falsos positivos.

A primeira fase é realizada com três tipos de arquitetura, duas de um estágio e uma de um estágio. Os resultados são associados por meio de Ensemble.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	08.74%	37.86%
	Positivo	01.94%	51.49%

Tabela 20 – Matriz de confusão para detecção de barras expostas com TOOD no CODEBRIMRev.

		Previsto:	
		Negativo	Positivo
Real:	Negativo	01.94%	38.83%
	Positivo	02.91%	56.31%

Tabela 21 – Matriz de confusão para detecção de lascamento com TOOD no CODEBRIMRev.

A última fase é igualmente realizada com três arquiteturas de rede, neste caso, específicas para classificação. Esta última fase entrega o resultado desejado da identificação das patologias.

5.4.1 Detecção e Classificação de Patologias com Arquitetura YOLO - Uma Classe - Conjunto de dados CODEBRIMRev

O primeiro experimento desta nova fase foi realizado com YOLO v11. O conjunto de dados utilizado foi ([CODEBRIM 2019](#)), revisado. A ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de um estágio e considerando apenas o caso genérico de patologia, em vez de categorizar as patologias em cinco classes, como nos experimentos anteriores, estas cinco categorias foram consideradas como apenas uma classe, patologia.

Na etapa de redimensionamento das imagens, todas foram ajustadas ao tamanho padrão de 640 pixels.

O treinamento foi realizado com a seguinte configuração: Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam com learning rate inicial estabelecido em 0.008, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de confiança de 0,5 a medida de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

5.4.2 Detecção e Classificação de Patologias com Arquitetura Faster RCNN - Uma Classe - Conjunto de dados CODEBRIMRev

O segundo experimento desta nova fase foi realizado com Faster RCNN. O conjunto de dados utilizado foi, igualmente ([CODEBRIM 2019](#)), revisado. A ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de dois estágios.

As imagens do conjunto de dados foram redimensionadas para imagens retangulares, com o menor lado de 800 pixels, e a outra dimensionada para manter a razão de aspecto da imagem.

O treinamento realizado com a seguinte configuração: O modelo foi configurado para igualmente reconhecer patologia como classe única, utilizando Softmax como função de ativação. Para o processo de otimização, adotou-se o algoritmo Adam com learning rate inicial estabelecido em 0.005, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foi definido o limiar de 0,5 para o limiar de confiança de IoU (Intersection over Union) no processo de Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados obtidos.

5.4.3 Detecção e Classificação de Patologias com Arquitetura TOOD - Uma Classe - Conjunto de dados CODEBRIMRev

O terceiro experimento desta nova fase foi realizado com TOOD. O conjunto de dados utilizado foi igualmente revisado ([CODEBRIM 2019](#)). Destacando que a ideia deste experimento foi entender o desempenho da tarefa de detecção com uma rede neural de um estágio, como referência ao uso de YOLO, que apresentou melhor desempenho.

A rede neural foi ajustada para processar imagens com resolução de entrada de 640 pixels, empregando como backbone a arquitetura ResNet50 para extração de

Capítulo 5. Resultados 80 características.

O treinamento realizado com a seguinte configuração: o modelo foi configurado para reconhecer igualmente cinco classes distintas, utilizando Sigmoid como função de ativação. Para o processo de otimização, adotou-se o algoritmo SGD, com learning rate inicial de 0,01, processando os dados em lotes de tamanho 16 (batch size) ao longo de 1500 épocas de treinamento.

Durante a fase de inferência, foram definidos thresholds criteriosos para garantir a qualidade das detecções, estabelecendo-se o valor de 0.5 para o threshold de confiança de 0.5 para o threshold de IoU (Intersection over Union) no processo de

Non-Maximum Suppression (NMS), visando eliminar detecções redundantes e otimizar a precisão dos resultados.

5.5 Detecção com Ensemble - Conjunto CODEBRIMRev

Foi realizado um experimento com a abordagem de ensemble dessas arquiteturas, utilizando o conjunto de dados revisado. Este experimento foi realizado utilizando as abordagens Weighted Box Fusion e Non-maximum Suppression. Este experimento visou avaliar o comportamento de ambas as abordagens na tarefa.

5.5.1 Detecção com Ensemble Usando Weighted Box Fusion - Conjunto de dados CODEBRIMRev

Neste experimento, foi utilizada a abordagem Weighted Box Fusion para combinar os resultados obtidos pelas arquiteturas YOLO, TOOD e Faster RCNN, para a obtenção de melhor desempenho, como esperado com a aplicação de Ensemble. Os resultados combinados, desta vez, foram os obtidos com as arquiteturas citadas aplicadas ao conjunto de dados revisado CODEBRIMRev ([CODEBRIM 2019](#)).

Os resultados dos experimentos, YOLO, TOOD e Faster RCNN, foram então combinados para obter um melhor desempenho, o esperado com a aplicação de Ensemble. Os parâmetros utilizados nas redes neurais são os descritos nas seções anteriores.

5.5.2 Detecção com Ensemble Usando Non-maximum Suppression - Conjunto de dados CODEBRIMRev

Desta vez, para ter um parâmetro de comparação com Ensemble utilizado com Weighted Box Fusion, foi utilizada uma segunda abordagem de Ensemble, combinando igualmente os resultados dos experimentos YOLO, TOOD e Faster RCNN, para obter melhor desempenho, como esperado com a aplicação de Ensemble. Os resultados combinados, igualmente aos obtidos com WBF, foram os obtidos com as arquiteturas citadas aplicadas ao dataset ([CODEBRIM 2019](#)), revisado, CODEBRIMRev.

Os resultados dos experimentos, YOLO, TOOD e Faster RCNN, foram então combinados para obter um melhor desempenho, o esperado com a aplicação de Ensemble. Os parâmetros utilizados nas redes neurais são os descritos nas seções anteriores.

5.6 Classificação de Patologias com Conjunto de Dados Recortados

Por fim, as imagens geradas com detecção são classificadas por classificadores específicos para definir a qual classe de patologia pertencem. A identificação de patologia em estruturas de concreto, agora é classificada em cinco classes, fissuras, mancha de concreto, eflorescência, lascamento e armadura exposta.

Três arquiteturas de classificação foram testadas: DenseNet121, ResNet50 e MobileNetV3. O treino foi realizado no dataset criado a partir da base de dados (CODEBRIM 2019) revisada, CODEBRIMRec, inicialmente anotada para detecção, mas foi recortada e anotada para produzir um conjunto de dados de classificação.

5.6.1 Classificação com DenseNet121 no CODEBRIMRec

O primeiro experimento desta fase de classificação foi realizado com a arquitetura DenseNet121, sobre o novo dataset de classificação criado a partir do dataset CODE-BRIM (CODEBRIM 2019), CODEBRIMRec, inicialmente anotado para detecção, mas recortado e anotado para produzir um conjunto de dados de classificação.

Foi utilizada na rede Dense Blocks. Função de ativação ReLU. Otimizador SGD. Learning Rate 0.1. Batch Size 32. Resolução de entrada 224 x 224.

As métricas obtidas são apresentadas na tabela 22.

Precisão	Sensibilidade	Acurácia	AUC
0.88	0.90	0.89	0.93

Tabela 22 – Métricas para classificação com DenseNet121 no CODEBRIMRec.

5.6.2 Classificação com ResNet50 no CODEBRIMRec

O segundo experimento desta fase de classificação foi realizado com a arquitetura ResNet50, sobre o novo dataset de classificação CODEBRIMRec, criado a partir do conjunto de dados (CODEBRIM 2019).

Foi utilizada para a rede de Blocos Residuais. Função de ativação ReLU. Otimizador SGD. Learning Rate 0.1. Batch Size 32. Resolução de entrada 224 x 224.

As métricas obtidas são apresentadas na tabela 23.

Precisão	Sensibilidade	Acurácia	AUC
0.91	0.85	0.87	0.74

Tabela 23 – Métricas para classificação com ResNet50 no CODEBRIMRec.

5.6.3 Classificação com MobileNetV3 no CODEBRIMRec

O último experimento de classificação foi realizado com a arquitetura MobileNetV3, sobre o novo dataset de classificação CODEBRIMRec, criado a partir do conjunto de dados ([CODEBRIM 2019](#)).

Foi utilizada para a rede de Blocos Residuais. Função de ativação ReLU. Otimizador SGD. Learning Rate 0.001. Batch Size 32. Resolução de entrada 224 x 224.

As métricas obtidas são apresentadas na tabela [24](#).

Precisão	Sensibilidade	Acurácia	AUC
0.88	0.93	0.90	0.95

Tabela 24 – Métricas para classificação com MobileNetV3 no CODEBRIMRec.

5.7 Resumo dos Resultados Obtidos

Nesta seção, apresentamos uma tabela-resumo com todos os resultados obtidos com a metodologia de uma etapa.

A Tabela [25](#) apresenta um resumo de todos os resultados obtidos na tarefa de detecção inicial, que deram origem ao ensemble realizado.

A Tabela [26](#) apresenta um resumo de todos os resultados obtidos na tarefa de detecção. A Tabela [27](#) apresenta um resumo de todos os resultados obtidos na tarefa de classificação.

5.8 Aplicação da Metodologia Duas Etapas para Identificar Fissuras

Nesta seção, apresentamos a aplicação da metodologia de duas etapas para identificar fissuras em estruturas de concreto. A Tabela [28](#) apresenta os resultados obtidos.

Este experimento foi realizado sobre o conjunto de teste do CODEBRIMRev, uma vez que a divisão adotada nos experimentos foi 70-20-10, este experimento é realizado em 10

Arquitetura	Base Dados	Classes Patologia	Precisão	Sensibilidade	F1 score	Acurácia	mAP (0.5)
YOLO	CODEBRIM	5	0.49	0.33	0.39	0.24	0.51
Faster RCNN	CODEBRIM	5	0.35	0.20	0.23	0.16	0.18
TOOD	CODEBRIM	5	0.37	0.21	0.27	0.15	0.28
Ensemble WBF	CODEBRIM	5	0.55	0.21	0.30	0.18	0.30
YOLO	CODEBRIMRev	5	0.14	0.34	0.19	0.09	0.34
Faster RCNN	CODEBRIMRev	5	0.21	0.33	0.25	0.14	0.35
TOOD	CODEBRIMRev	5	0.05	0.49	0.10	0.05	0.42
YOLO	CODEBRIMRev	1	0.14	0.31	0.19	0.15	0.27
Faster RCNN	CODEBRIMRev	1	0.26	0.37	0.30	0.21	0.33
TOOD	CODEBRIMRev	1	0.11	0.40	0.17	0.16	0.36
Ensemble WBF	CODEBRIMRev	1	0.08	0.43	0.14	0.17	0.44
Ensemble NMS	CODEBRIMRev	1	0.08	0.43	0.14	0.14	0.39

Tabela 25 – Resumo dos experimentos para a tarefa de detecção.

Arquitetura	Patologia	Acurácia	Sensibilidade	Precisão	F1 score
YOLO	Mancha Corrosão	66.99	77.63	77.63	77.63
	Fissura	52.43	87.72	54.35	67.11
	Eflorescência	66.02	77.14	50.00	60.67
	Barra Exposta	83.50	94.55	78.79	85.95
	Lascamento	74.76	90.16	73.33	80.88
Faster RCNN	Mancha Corrosão	66.02	72.37	79.71	75.86
	Fissura	57.28	84.21	57.83	68.57
	Eflorescência	82.52	85.71	69.77	76.92
	Barra Exposta	88.35	94.55	85.25	89.66
	Lascamento	75.73	83.61	77.27	80.31
TOOD	Mancha Corrosão	72.82	96.05	74.49	83.91
	Fissura	55.34	94.74	55.67	70.13
	Eflorescência	34.95	94.29	33.67	49.62
	Barra Exposta	60.19	96.36	57.61	72.11
	Lascamento	58.25	95.08	59.18	72.96

Tabela 26 – Resumo dos experimentos para a tarefa de classificação com detecção.

As detecções são realizadas com o ensemble WBF, que resulta das detecções obtidas com as redes YOLO, Faster R-CNN e TOOD. Essas detecções são recortadas para classificação.

O classificador utilizado é o MobileNetV3. A Tabela 27 resumiu os experimentos com três redes de classificação diferentes e, em comparação com DenseNet121 e ResNet50, a rede que obteve melhor desempenho foi a que utilizou a técnica de aprendizado de máquina.

Arquitetura	Base Dados	Classes Patologia	Precisão	Sensibilidade	AUC	Acurácia
DenseNet121	CODEBRIMRec	5	0.88	0.90	0.93	0.89
ResNet50	CODEBRIMRec	5	0.91	0.85	0.74	0.87
MobileNetV3	CODEBRIMRec	5	0.88	0.93	0.95	0.90

Tabela 27 – Resumo dos experimentos para a tarefa de classificação.

Confiança	Fissuras	Não Fissuras	Total
0.3	204	10	214
0.4	133	2	135
0.5	70	0	70

Tabela 28 – Aplicação do método duas etapas para identificar fissuras.

5.9 Discussões

A seguir, discutimos os resultados obtidos nesta tese: identificação de patologias em estruturas de concreto armado em uma etapa, conforme apresentado na literatura, e em duas etapas, propostas nesta tese. As tarefas envolvidas são a detecção e a classificação.

A Tabela 29 apresenta os resultados da classificação dos tipos de patologias. Os resultados apresentados nesta tese utilizaram o conjunto de dados CODEBRIM. Os resultados apresentados pelos demais autores utilizam o conjunto de dados Ozgenel. Os trabalhos que utilizaram conjuntos de dados Ozgenel lidam com bases de dados balanceadas e com imagens recortadas de fissura, o que caracteriza uma tarefa em condições mais controladas. Essas condições resultam em conjuntos que permitem métricas muito elevadas. A base usada nesta tese possui imagens que retratam a realidade de imagens de patologias em estruturas de concreto armado, com ruídos, tais como pichações, que melhor representam casos realistas, que foi o objetivo da pesquisa.

Inicialmente, foi aplicado o método que utiliza a identificação das patologias em uma única etapa. Para esta tarefa, foram utilizadas três arquiteturas de redes diferentes. Esta abordagem demonstrou que é possível identificar as patologias de forma mais direta, porém sem uma análise específica de cada uma.

Este problema analisado apresenta ainda a particularidade de que mais de uma

Autor	Métrica Acurácia %
Resultado DenseNet121 (Tese autor)	89.00
Resultado ResNet50 (Tese autor)	87.00
Resultado MobileNetV3 (Tese autor)	90.00
Paper publicado (Diniz et al. 2023)	99.43
Alipour (Alipour e Harris 2020)	98.60
Arafin2024 (Arafin, Billah e Issa 2024)	91.98
Bai (Bai et al. 2020)	95.92
Jitendra (Jitendra et al. 2020)	98.60
Kim (Kim, Kim e Lee 2020)	99.98
Pal (Pal et al. 2021)	98.00
Ugne (Orinaité et al. 2022)	99.00

Tabela 29 – Resultados da tarefa de classificação do tipo de patologia em imagens recortadas.

patologia pode estar presente na mesma região. A Figura 30 exibe casos de patologia de concreto. É possível visualizar barras de aço expostas, sendo que, neste mesmo ponto, há fragmentação do concreto e as barras de aço apresentam corrosão. Esta figura foi utilizada na rede de detecção treinada para testar quais patologias poderiam ser identificadas.



Figura 30 – Imagem de patologias no concreto antes da detecção com YOLO ([CODEBRIM 2019](#)). Elaborada pelo autor.

A Figura 31 mostra o resultado da identificação com uma das redes utilizadas em uma única etapa. Um retângulo amarelo, a caixa delimitadora, indica o local das patologias identificadas. A rede conseguiu detectar a exposição de armadura, neste caso é exibida a confiança da detecção de 70%, a fragmentação, e, por fim, a corrosão da armadura.

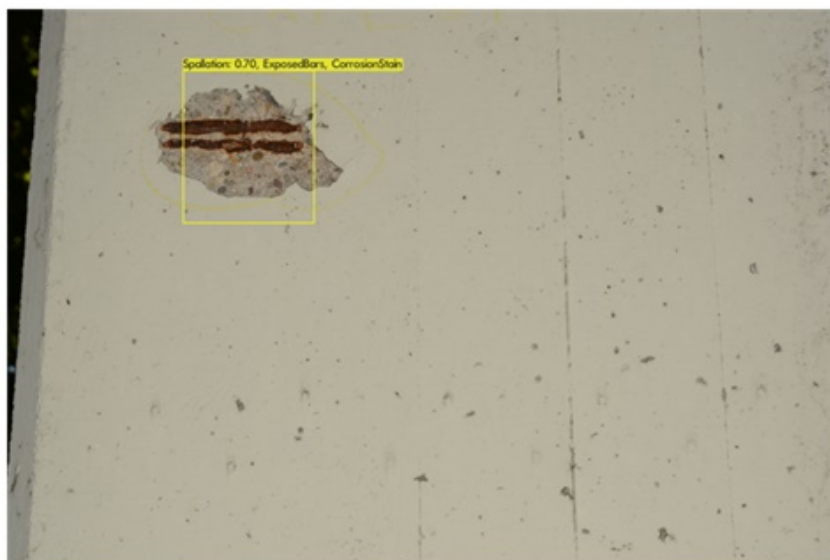


Figura 31 – Imagem de patologias no concreto após detecção com YOLO (CODEBRIM 2019). Elaborada pelo autor.

Uma ação importante consiste em revisar sempre o conjunto de dados utilizado, etapa que faz parte das ações de treinamento de redes neurais. Particularmente, o conjunto de dados, CODEBRIM, requer algumas análises quanto à sua revisão. Identificar possíveis inconsistências e corrigi-las para permitir um melhor aprendizado da rede neural, mas também entender a padronização dos critérios de anotação.

O aprendizado de uma rede neural está diretamente relacionado à qualidade das anotações do conjunto de dados. Uma anotação com erros resulta em um aprendizado deficiente. A capacidade da rede neural de detectar um determinado objeto dependerá dos dados utilizados no aprendizado. Por exemplo, se apresentarmos a patologia de corrosão em vez de fissuração, será detectada a corrosão em vez da fissuração.

Uma revisão completa do conjunto de dados foi realizada para detectar possíveis inconsistências nas anotações sobre a patologia de fissuras no concreto. Alguns números estavam completamente errados quanto ao posicionamento da caixa delimitadora; esses casos foram possivelmente devidos à rotação da imagem após a anotação. Outra melhoria foi o melhor aproveitamento das fissuras presentes nas imagens, que não haviam sido anotadas.

É possível aumentar a sensibilidade de detecção para identificar mais casos de patologias, embora com prejuízo à precisão. Em contrapartida, essa abordagem aumenta o número de falsos positivos na fase de detecção. Este não é um problema, visto que o objetivo é identificar o maior número possível de patologias, desde que seja possível eliminar os falsos positivos. A eliminação de falsos positivos é possível com a abordagem adotada no método apresentado nesta tese, que envolve uma rede especialmente treinada para a tarefa de classificação.

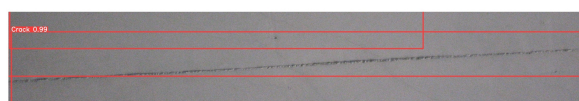
Existem casos em que as imagens contêm elementos que podem levar a falsas identificações. É importante analisar estes casos e é por isso que este critério foi utilizado quando da análise do conjunto de dados a ser utilizado. A Figura 32 - A mostra um caso de manchamento do concreto que foi identificado como fissuração do concreto. Há casos em que o formato dessas manchas pode dificultar a identificação. A Figura 32 - B mostra casos em que o aço exposto foi confundido com fissuras no concreto. Embora o aço exposto possa vir acompanhado de fissuras, neste caso, a patologia mais importante a identificar seria o próprio aço exposto. A Figura 32 - C mostra uma junta confundida com uma trinca. Este caso é um tanto complexo, pois há casos de fissuras indesejadas em juntas, mas também pode ser caso de junta específica para dilatação, por exemplo. Um especialista precisa avaliar este caso.



(A) Mancha de Corrosão



(B) Barra Exposta



(C) Junta

Figura 32 – Falso positivo. Elaborada pelo autor.

A tarefa de detecção, de localizar um artefato em uma imagem mais geral, é mais complexa do que a simples classificação. A classificação costuma ser menos complexa, pois consiste simplesmente em classificar toda a imagem a ser processada; por outro lado, a detecção tem essa mesma atividade, mas em um recorte da imagem que precisa ser localizada, exigindo, portanto, mais processamento. A anotação de classificação e detecção também é diferente.

A classificação basta informar a classe de classificação a que pertence uma imagem. Em caso de detecção, é necessário localizar o artefato a ser classificado na imagem para gerar a marcação, e a quantidade de artefatos em uma mesma imagem pode ser numerosa. A marcação de artefatos na anotação para detecção pode ser subjetiva, sendo importante que um profissional qualificado a realize e, em muitos casos, pode ser difícil, dependendo de como o artefato a ser classificado aparece na imagem. Estas últimas dificuldades podem resultar em erros ou imprecisões nas anotações.

O maior grau de dificuldade nas anotações de detecção torna esta tarefa mais demorada e dispendiosa, o que pode levar à disponibilidade de menos conjuntos de dados. Também pode resultar em conjuntos de dados de qualidade inferior, o que afeta a qualidade do treinamento de uma rede neural e a capacidade de detecção da rede treinada.

A detecção de artefatos em imagens gera partes identificadas pela caixa delimitadora, que podem ser recortadas para gerar novas imagens que, uma vez detectadas, podem ser classificadas. Esta classificação será tão precisa quanto a da rede neural treinada e utilizada.

Como a detecção não é totalmente precisa, algumas imagens resultantes são classificadas incorretamente. Esses erros podem ser identificados por meio de uma nova classificação com uma rede neural melhor treinada para esta tarefa. Por exemplo, isto é possível se a rede de classificação for treinada com um conjunto de dados maior. Como a anotação para classificação é mais fácil do que a detecção, não é incomum que o conjunto de dados de classificação contenha uma quantidade muito maior de dados.

Ainda que diversos fatores sejam desafiadores para o uso de redes neurais, é perfeitamente possível seu uso para a identificação de patologias. A Figura 33 mostra um caso de sucesso de detecção positiva.

O experimento descrito na seção 5.8 apresenta um exemplo de aplicação da metodologia de duas etapas para a identificação de fissuras em estruturas de concreto. O senso comum adota confiança de 0,5 e evita valores menores. Este pensamento é justificado para evitar um número maior de falsos positivos. Os resultados nos mostram que a adoção de confiança de 0,3, um exemplo de valor menor que 0,5, permite a identificação de três vezes mais objetos, neste caso, patologia, mais especificamente fissuras. Este exemplo, em contrapartida, resulta em maior número de falsos positivos.



Figura 33 – Detecção positiva. Elaborada pelo autor.

6 Conclusão

Esta tese apresentou um método promissor com potencial para contribuir para a área de inspeção visual de estruturas de concreto armado. A metodologia proposta, fundamentada em um ensemble de redes neurais convolucionais para detecção e classificação de patologias, demonstrou capacidade de agilizar a obtenção de resultados em comparação com abordagens tradicionais. O trabalho evidencia como técnicas de aprendizado profundo podem ser aplicadas de forma efetiva neste domínio, apresentando resultados que sugerem sua viabilidade para implementação em cenários práticos. O método desenvolvido possui características que o tornam adequado para aplicação em inspeções reais de estruturas, podendo auxiliar profissionais da área na identificação precoce de patologias, na priorização de intervenções de manutenção e na elaboração de laudos técnicos mais precisos e fundamentados. Adicionalmente, a abordagem proposta tem potencial para ser integrada a sistemas de monitoramento contínuo de infraestruturas, contribuindo para uma gestão mais eficiente do patrimônio construído e para a extensão da vida útil das estruturas de concreto armado.

Os resultados apresentados nesta tese indicaram a possibilidade de ganhos significativos na área e sugerem que é possível desenvolver recursos efetivos com materiais e técnicas existentes, podendo estimular o surgimento de novos conjuntos de dados e, conseqüentemente, de soluções inovadoras para o campo da inspeção de estruturas de concreto armado.

É possível observar nos experimentos realizados, mais próximos da realidade das imagens reais, que o método apresenta boa capacidade de identificação de patologias no concreto, igualmente indicada com testes de imagem individualizados em boas condições, mas permite o questionamento sobre a melhoria da precisão e a eliminação de falsos positivos. O modelo proposto permite detectar robustamente as patologias presentes nas imagens, mesmo ao custo de muitos falsos positivos. E, no segundo passo, temos uma forma eficiente de eliminar quase a totalidade dos falsos positivos com o uso de um classificador específico treinado para as classes de patologias pretendidas.

Há casos em que a detecção da patologia e a conseqüente classificação podem ser difíceis ou mesmo impossíveis por não estarem claramente evidenciadas na imagem. É o caso das manchas que podem se sobrepor a casos de manchas de corrosão e/ou de eflorescência. Alguns casos de fissuras, que apresentam formato geométrico reto, também podem ser confundidos com juntas estruturais ou fissuras.

Na prática, o método pode automatizar a identificação de patologias em

estruturas de concreto. A estrutura a ser analisada precisa ser devidamente fotografada, produzindo imagens da mais alta qualidade possível. As imagens geradas devem ser analisadas para verificar sua qualidade e, conseqüentemente, se é necessário realizar algum aprimoramento de imagem. Neste ponto de aprimoramento da imagem, este método pode ser aprimorado por meio de mecanismos automáticos que identifiquem quais aprimoramentos precisam ser realizados e apliquem, conseqüentemente, a melhoria necessária. As imagens aprimoradas prosseguem então para a detecção, com as regiões de possível patologia indicadas em toda a estrutura. A classificação é então realizada em regiões separadas por detecção. O conjunto de imagens identificadas e classificadas como patologias e a respectiva classificação poderão então seguir para análise pelo perito, que definirá as ações a tomar.

O método proposto exemplifica como é possível aprimorar tarefas de identificação de patologias em imagens de estruturas de concreto, mas muitas melhorias podem ser exploradas para alcançar resultados ainda mais robustos. As técnicas de visão computacional têm feito cada vez mais parte do cotidiano humano, nas mais diversas áreas, onde podem ser aplicadas. Os recursos dessa área estão presentes no dia a dia das pessoas, embora muitas vezes passem despercebidos. A expansão dessas técnicas para diferentes áreas resulta em um ganho real para o desenvolvimento da humanidade.

Uma importante etapa na sequência da inspeção visual de patologias em concreto amorado é a quantificação dos danos na superfície das estruturas. Neste caso, deixamos como sugestão o avanço dos estudos para analisar e experimentar a tarefa de segmentação.

Devido a complexidade, as anotações para classificação, detecção e segmentação, possuem maior desafio, tempo e dificuldade de anotação, sendo a classificação mais fácil, e a segmentação a mais desafiadora. Como consequência, o número de conjuntos de dados para segmentação é menor que o de classificação, sendo detecção intermediária. Mesmo assim, já é possível encontrar conjuntos anotados de patologias em estruturas de concreto, que podem ser analisados para trabalhos futuros.

Uma abordagem relevante a ser também estudada em trabalhos futuros é a importância das características. Analisar como os modelos utilizados focam, por exemplo, em padrões de fissuras, áreas de corrosão e manchas de umidade. Investigar qual modelo contribui mais para cada tipo de patologia. Detectar possíveis vieses, como se o modelo estivesse usando características irrelevantes.

Adicionalmente, sugerimos o uso de ativação de gradiente, combinando o que a rede enxerga (ativações) com o que a rede aprende (gradientes), para identificar regiões importantes nas imagens. Isto pode ajudar na interpretabilidade, mostrando exatamente onde a rede está olhando. Validação, confirmando que as características

relevantes estão sendo utilizadas. Comparação, permitindo comparar diferentes modelos. Otimização, ajudando a entender onde melhorar no treino. E também na confiança, métrica de qualidade baseada na atenção.

A abordagem de camadas de atenção, neste contexto de detecção de patologias em concreto armado, pode representar um avanço significativo na interpretabilidade e na eficácia dos modelos. Incorporando esse mecanismo, os modelos conseguem focar seletivamente nas regiões estruturalmente relevantes das imagens, como padrões de fissuras, simulando o processo de inspeção visual humana. A capacidade de atribuir pesos diferentes a certas características não apenas melhora a precisão na identificação de patologias, mas também possibilita obter uma melhor explicação visual das decisões do modelo.

Referências

ABNT. **NBR 5674: Manutenção de edificações – Requisitos para o sistema de gestão de manutenção**. Rio de Janeiro, 2012. Válida a partir de 25.08.2012. Citado na página 17.

AI, S.; KWON, J. Extreme low-light image enhancement for surveillance cameras using attention u-net. **Sensors**, v. 20, n. 2, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/2/495>>. Citado na página 61.

ALIPOUR, M.; HARRIS, D. K. Increasing the robustness of material-specific deep learning models for crack detection across different materials. **Engineering Structures**, v. 206, p. 110157, 2020. ISSN 0141-0296. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0141029619304389>>. Citado na página 93.

ARAFIN, P.; BILLAH, A. M.; ISSA, A. Deep learning-based concrete defects classification and detection using semantic segmentation. **Structural Health Monitoring**, v. 23, n. 1, p. 383–409, 2024. Disponível em: <<https://doi.org/10.1177/14759217231168212>>. Citado 2 vezes nas páginas 46 e 93.

BAHREINI, F.; HAMMAD, A. Dynamic graph cnn based semantic segmentation of concrete defects and as-inspected modeling. **Automation in Construction**, v. 159, p. 105282, 2024. ISSN 0926-5805. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0926580524000189>>. Citado na página 51.

BAI, Y.; ZHA, B.; SEZEN, H.; YILMAZ, A. Deep cascaded neural networks for automatic detection of structural damage and cracks from images. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, V-2-2020, p. 411–417, 2020. Disponível em: <<https://isprs-annals.copernicus.org/articles/V-2-2020/411/2020/>>. Citado na página 93.

BI, H.; WEN, V.; XU, Z. Comparing one-stage and two-stage learning strategy in object detection. **Applied and Computational Engineering**, v. 5, n. 1, p. 171–177, 2023. ISSN 2755-2721. Disponível em: <<https://doi.org/10.54254/2755-2721/5/20230665>>. Citado na página 28.

CARVALHO, R. C. **Cálculo e Detalhamento de Estruturas Usuais de Concreto Armado – segundo a NBR6118 2003**. [S.l.]: Quarta Edição, EDUFSCAR, ISBN 978-85-7600-356-4, 2022. 415 p. Citado na página 22.

CODEBRIM. **COncrete DEfect BRidge IMage Dataset**. ZENODO, 2019. Disponível em: <https://zenodo.org/record/2620293#.YgLkC9_MKMo>. Citado 21 vezes nas páginas 9, 10, 18, 23, 25, 26, 68, 69, 77, 78, 79, 80, 82, 83, 86, 87, 88, 89, 90, 93 e 94.

DENG, L.; YUAN, H.; LONG, L.; CHUN, P. jo; CHEN, W.; CHU, H. Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images. **Automation in Construction**, v. 162, p. 105410, 2024. ISSN 0926-5805. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0926580524001468>>. Citado na página 48.

- DINIZ, J. d. C. N.; PAIVA, A. C. d.; JUNIOR, G. B.; ALMEIDA, J. D. S. de; SILVA, A. C.; CUNHA, A. M. T. d. S.; CUNHA, S. C. A. P. d. S. A method for detecting pathologies in concrete structures using deep neural networks. **Applied Sciences**, v. 13, n. 9, 2023. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/13/9/5763>>. Citado na página 93.
- FENG, C.; ZHONG, Y.; GAO, Y.; SCOTT, M. R.; HUANG, W. Tood: Task-aligned one-stage object detection. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2021. p. 3510–3519. Citado 3 vezes nas páginas 9, 34 e 35.
- FENG, C.; ZHONG, Y.; GAO, Y.; SCOTT, M. R.; HUANG, W. Tood: Task-aligned one-stage object detection. In: IEEE COMPUTER SOCIETY. **2021 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.], 2021. p. 3490–3499. Citado na página 36.
- FLOTZINGER, J.; RÖSCH, P. J.; OSWALD, N.; BRAML, T. dacl1k: Real-world bridge damage dataset putting open-source data to the test. **Engineering Applications of Artificial Intelligence**, v. 137, p. 109106, 2024. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197624012648>>. Citado 3 vezes nas páginas 47, 56 e 57.
- HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask r-cnn. In: **2017 IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2017. p. 2980–2988. Citado na página 34.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: [s.n.], 2016. p. 770–778. Citado 3 vezes nas páginas 9, 38 e 40.
- HOWARD, A.; SANDLER, M.; CHEN, B.; WANG, W.; CHEN, L.-C.; TAN, M.; CHU, G.; VASUDEVAN, V.; ZHU, Y.; PANG, R.; ADAM, H.; LE, Q. Searching for mobilenetv3. In: **2019 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2019. p. 1314–1324. Citado na página 40.
- HUANG, G.; LIU, Z.; MAATEN, L. van der; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2017. p. 4700–4708. Citado 3 vezes nas páginas 9, 36 e 37.
- JAMSHIDI, M.; EL-BADRY, M.; NOURIAN, N. Improving concrete crack segmentation networks through CutMix data synthesis and temporal data fusion. **Sensors**, v. 23, n. 1, p. 504, 2023. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/23/1/504>>. Citado na página 54.
- JITENDRA, M. S. N. V.; SRINIVASU, P. N.; SRINIVAS, A. S.; NITHYA, A.; KANDULAPATI, S. K. Crack detection on concrete images using classification techniques in machine learning. **Journal of Critical Reviews**, v. 7, n. 9, p. 1236–1241, 2020. Citado na página 93.
- JU, C.; BIBAUT, A.; LAAN, M. van der. The relative performance of ensemble methods with deep convolutional neural networks for image classification. **Journal of Applied Statistics**, v. 45, n. 15, p. 2800–2818, 2018. Citado na página 44.

- KAVEH, H.; ALHAJJ, R. Recent advances in crack detection technologies for structures: a survey of 2022-2023 literature. **Frontiers in Built Environment**, Volume 10 - 2024, n. 8, 2024. ISSN 2297-3362. Disponível em: <<https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuilt.2024.1321634>>. Citado na página 28.
- KIM, J. J.; KIM, A.-R.; LEE, S.-W. Artificial neural network-based automated crack detection and analysis for the inspection of concrete structures. **Applied Sciences**, v. 10, n. 22, p. 8105, 2020. Citado na página 93.
- KRISHNAN, S. S. R.; KARUPPAN, M. K. N.; KHADIDOS, A. O. Comparative analysis of deep learning models for crack detection in buildings. **Scientific Reports**, v. 15, p. 2125, 2025. Citado na página 28.
- LAAN, M. J. van der; POLLEY, E. C.; HUBBARD, A. E. Super learner. **Statistical Applications in Genetics and Molecular Biology**, v. 6, n. 1, 2007. Disponível em: <<https://doi.org/10.2202/1544-6115.1309>>. Citado na página 45.
- LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE Transactions on Neural Networks and Learning Systems**, v. 33, n. 12, p. 6999–7019, 2022. ISSN 2162-237X. Disponível em: <<https://ieeexplore.ieee.org/document/9451544>>. Citado na página 29.
- LIU, A.; HUA, W.; XU, J.; YANG, Z.; FU, J. Concrete crack segmentation based on multi-dimensional structure information fusion-based network. **Construction and Building Materials**, v. 414, 2024. ISSN 0950-0618. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0950061824001235>>. Citado na página 50.
- MISHRA, A.; GANGISETTI, G.; AZAM, Y. E.; KHAZANCHI, D. Weakly supervised crack segmentation using crack attention networks on concrete structures. **Structural Health Monitoring**, v. 23, n. 6, p. 3712–3735, 2024. ISSN 1475-9217. Disponível em: <<https://journals.sagepub.com/doi/10.1177/14759217241228150>>. Citado 2 vezes nas páginas 52 e 56.
- MUNDT, M.; MAJUMDER, S.; MURALI, S.; PANETSOS, P.; RAMESH, V. Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2019. p. 11196–11205. Disponível em: <https://openaccess.thecvf.com/content_CVPR_2019/html/Mundt_Meta-Learning_Convolutional_Neural_Architectures_for_Multi-Target_Concrete_Defect_Classification_With_CVPR_2019_paper.html>. Citado na página 68.
- ORINAITÈ, U.; PALEVIČIUS, P.; PAL, M.; RAGULSKIS, M. A deep learning-based approach for automatic detection of concrete cracks below the waterline. **Vibroengineering PROCEDIA**, v. 44, p. 142–148, August 2022. ISSN 2345-0533. Citado na página 93.
- OZGENEL. **Concrete Crack Images for Classification**. Mendeley Data, 2019. Disponível em: <<https://data.mendeley.com/datasets/5y9wdsg2zt/2>>. Citado 5 vezes nas páginas 9, 10, 67, 73 e 74.
- PAL, M.; PALEVIČIUS, P.; LANDAUSKAS, M.; ORINAITÈ, U.; TIMOFEJEVA, I.; RAGULSKIS, M. An overview of challenges associated with automatic detection of

concrete cracks in the presence of shadows. **Applied Sciences**, v. 11, n. 23, 2021. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/11/23/11396>>. Citado na página 93.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KÖPF, A.; YANG, E.; DEVITO, Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.; STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. Pytorch: An imperative style, high-performance deep learning library. In: WALLACH, H. M.; LAROCHELLE, H.; BEYGELZIMER, A.; BUC, F. d'Alché; FOX, E. B.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems 32 (NeurIPS 2019)**. Curran Associates, Inc., 2019. p. 8024–8035. Disponível em: <<https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>>. Citado na página 67.

RAO, S. N. YOLOv11 architecture explained: Next-level object detection with enhanced speed and accuracy. 2024. Disponível em: <<https://medium.com/@nikhil-rao-20/>>. Citado 2 vezes nas páginas 9 e 31.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 779–788. Citado 2 vezes nas páginas 9 e 32.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2015. v. 28. Citado na página 34.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. 2016. Citado 3 vezes nas páginas 9, 32 e 34.

SAFIUDDIN, M.; KAISH, A. B. M. A.; WOON, C.-O.; RAMAN, S. N. Early-age cracking in concrete: Causes, consequences, remedial measures, and recommendations. **Applied Sciences**, MDPI, v. 8, n. 10, p. 1730, 2018. Citado na página 17.

SHAIKH, T. A.; RASOOL, T.; VERMA, P.; MIR, W. A. A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. **Annals of Operations Research**, Springer, p. 1–77, 2024. Citado na página 43.

SHAMSABADI, E. A.; ERFANI, S. M. H.; XU, C.; COSTA, D. Dias-da. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. **Automation in Construction**, Elsevier, v. 158, p. 105181, 2024. Citado 3 vezes nas páginas 18, 52 e 56.

SHI, M.; HUA, T.; YANG, Z. J.; TAN, C.; WEN, Y. Weakly supervised deep learning-based concrete aggregates automatic segmentation for assessing separation degree. **Journal of Building Engineering**, v. 82, p. 108342, 2024. ISSN 2352-7102. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352710223025251>>. Citado na página 53.

SHI, Z.; JIN, N.; CHEN, D.; AI, D. A comparison study of semantic segmentation networks for crack detection in construction materials. **Construction and Building Materials**, Elsevier, v. 416, p. 135208, 2024. Citado na página 49.

- SOUZA, M. I.; MURTA, M. M. **Patologias, recuperação e reforço estrutural em concreto armado**. 51 p. Dissertação (Trabalho de Conclusão de Curso (Bacharel em Engenharia Civil)) — Instituto Doctum de Educação e Tecnologia, Caratinga, MG, 2012. Disponível em: <<https://dspace.doctum.edu.br/handle/123456789/1190>>. Citado na página 22.
- SOUZA, V. C. M. d.; RIPPER, T. **Patologia, Recuperação e Reforço de Estruturas de Concreto**. São Paulo: Pini, 1998. 262 p. Citado na página 16.
- SZEGEDY C., L. W. J. Y. S. P. R. S. A. D. E. D. V. V. . R. A. Going deeper with convolutions. In: IEEE COMPUTER SOCIETY. **2015 IEEE conference on computer vision and pattern recognition**. [S.l.], 2015. p. 1–9. Citado na página 44.
- TAHERI, S. A review on five key sensors for monitoring of concrete structures. **Construction and Building Materials**, Elsevier, v. 204, p. 492–509, 2019. ISSN 0950-0618. Citado na página 16.
- TIAN, Z.; YANG, F.; YANG, L.; WU, Y.; CHEN, J.; QIAN, P. An optimized yolov11 framework for the efficient multi-category defect detection of concrete surface. **Sensors**, MDPI, v. 25, n. 5, p. 1291, February 2025. ISSN 1424-8220. Citado 3 vezes nas páginas 18, 47 e 56.
- WANG, A.; CHEN, H.; LIU, L.; CHEN, K.; LIN, Z.; HAN, J. et al. Yolov10: Real-time end-to-end object detection. **Advances in Neural Information Processing Systems**, v. 37, p. 107984–108011, 2024. Citado na página 31.
- WIGHT, J. K.; MACGREGOR, J. G. **Reinforced Concrete: Mechanics and Design**. 7th. ed. Hoboken, NJ: Pearson, 2016. ISBN 978-0-13-348596-7. Citado na página 16.
- WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241–259, 1992. Citado na página 44.
- XU, G.; YUE, Q.; LIU, X.; CHEN, H. Investigation on the effect of data quality and quantity of concrete cracks on the performance of deep learning-based image segmentation. **Expert Systems with Applications**, v. 237, p. 121438, 2024. ISSN 0957-4174. Citado na página 49.