

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET  
DEPARTAMENTO DE COMPUTAÇÃO– DC  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

**Stevan Henrique Ramon de Góes**

**YOLO-CBAM: Arquitetura YOLOv8  
leve com módulo de atenção para  
segmentação de defeitos em pás de  
aerogeradores**

**Stevan Henrique Ramon de Góes**

**YOLO-CBAM: Arquitetura YOLOv8  
leve com módulo de atenção para  
segmentação de defeitos em pás de  
aerogeradores**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Sistemas de Computação

Orientador: Hermes Senger

São Carlos

2025

*Dedico este trabalho a todos que nunca deixaram  
de fazer perguntas, mesmo quando as respostas  
exigiam mais esforço do que o esperado.*

---

# Agradecimentos

---

Primeiramente, agradeço a Deus, fonte de força, sabedoria e perseverança, por me sustentar nos momentos de dificuldade e por guiar cada passo desta caminhada acadêmica.

À minha família, expresso minha mais profunda gratidão pelo apoio incondicional, pela paciência nos momentos de ausência e pelo incentivo constante, mesmo quando os desafios pareciam maiores que as forças. Nada disso teria sido possível sem a confiança e o amor de vocês.

Aos professores que contribuíram direta ou indiretamente para a minha formação, agradeço pelo conhecimento compartilhado, pelas orientações e pelo estímulo ao pensamento crítico e científico.

Aos colegas, amigos e todos que de alguma forma fizeram parte desta trajetória, seja com apoio técnico, palavras de incentivo ou simples compreensão, deixo aqui meu sincero agradecimento.

Agradeço ao MitDev USP pela concessão da bolsa ao longo deste período, que contribuiu para viabilizar o desenvolvimento deste trabalho.

Por fim, agradeço à instituição e a todos que possibilitaram a realização deste trabalho, contribuindo para meu crescimento acadêmico, profissional e pessoal.

*“A imaginação é mais importante que o conhecimento,  
pois o conhecimento é limitado, enquanto a imaginação envolve o mundo.”  
(Albert Einstein)*

---

# Resumo

---

A crescente demanda global por energia limpa e sustentável intensificou a adoção de tecnologias renováveis, com a energia eólica destacando-se como uma das fontes mais promissoras. Garantir a integridade estrutural e a eficiência operacional das turbinas eólicas é essencial, visto que suas pás estão constantemente expostas a condições ambientais adversas que causam erosão, fissuras e desgaste superficial. As inspeções visuais convencionais, no entanto, ainda são limitadas pela subjetividade humana, altos custos operacionais e baixa escalabilidade.

Em resposta a essas limitações, este trabalho apresenta uma abordagem automatizada de inspeção visual baseada em aprendizado profundo, propondo um modelo de segmentação leve derivado da arquitetura YOLOv8 e aprimorado com o Módulo de Atenção de Bloco Convolutivo (CBAM). A inclusão de mecanismos de atenção permite que a rede capture melhor as características espaciais e de canal, melhorando sua capacidade de detectar defeitos sutis e de baixo contraste, mantendo a eficiência computacional para implantação no mundo real.

O modelo proposto, denominado YOLOv8+CBAM, foi treinado em amostras recortadas e anotadas do conjunto de dados público Blade30. Os resultados experimentais demonstram melhorias significativas no desempenho, com a Interseção sobre União média (mIoU) aumentando de 0,57 para 0,61 em comparação com o modelo YOLOv8. O método apresenta alta sensibilidade a pequenos defeitos, robustez contra desequilíbrio de classes e inferência eficiente, adequada para ambientes com recursos limitados.

De modo geral, este estudo contribui para o avanço da detecção automatizada de defeitos em pás de turbinas eólicas, integrando imagens obtidas por drones e técnicas de aprendizado profundo em um pipeline de segmentação unificado. Extensões futuras podem explorar a quantificação da severidade dos defeitos e a integração com sistemas de mosaico de imagens para dar suporte a fluxos de trabalho de inspeção e manutenção totalmente autônomos.

**Palavras-chave:** Energia eólica. Inspeção automatizada. Aprendizado profundo. Segmentação de imagens. Pás de turbinas eólicas.



---

# Abstract

---

The increasing global demand for clean and sustainable energy has intensified the adoption of renewable technologies, with wind power standing out as one of the most promising sources. Ensuring the structural integrity and operational efficiency of wind turbines is essential, as their blades are constantly exposed to harsh environmental conditions that cause erosion, cracks, and surface wear. Conventional visual inspections, however, remain limited by human subjectivity, high operational costs, and low scalability.

In response to these limitations, this work presents an automated visual inspection approach based on deep learning, proposing a lightweight segmentation model derived from the YOLOv8 architecture and enhanced with the Convolutional Block Attention Module (CBAM). The inclusion of attention mechanisms enables the network to better capture spatial and channel-wise features, improving its ability to detect subtle and low-contrast defects while maintaining computational efficiency for real-world deployment.

The proposed model, named YOLOv8+CBAM, was trained on cropped and annotated samples from the public Blade30 dataset. Experimental results demonstrate significant performance improvements, with the mean Intersection over Union (mIoU) increasing from 0.57 to 0.61 compared to the baseline YOLOv8. The method shows high sensitivity to small defects, robustness against class imbalance, and efficient inference suitable for resource-constrained environments.

Overall, this study contributes to advancing automated defect detection in wind turbine blades by integrating UAV-based imaging and deep learning techniques into a unified segmentation pipeline. Future extensions may explore defect severity quantification and integration with image mosaicking systems to support fully autonomous inspection and maintenance workflows.

**Keywords:** Wind energy. Automated inspection. Deep learning. Image segmentation. Wind turbine blades.

---

# Lista de ilustrações

---

Figura 1 – Detecção de erosão na ponta de um aerogerador [Reddy et al. 2019]. . . . .	23
Figura 2 – Metodologia . . . . .	32
Figura 3 – Exemplo de <i>bounding box</i> alinhada aos eixos. Esse tipo de caixa delimitadora mantém seus lados paralelos aos eixos da imagem, sendo amplamente utilizado em abordagens tradicionais de detecção. . . . .	33
Figura 4 – Exemplo de <i>bounding box</i> rotacionada. Nesse caso, a caixa delimitadora acompanha a orientação do objeto e foi reduzida para permanecer dentro dos limites da imagem. . . . .	33
Figura 5 – Exemplo das Técnicas de brilho, contraste, saturação, matiz. Fonte: [v7labs 2024]	34
Figura 6 – Exemplo da Técnica Mosaico. Fonte: Roboflow (2020). Mosaic Data Augmentation - Deep Dive [YouTube] [Roboflow 2020]. . . . .	35
Figura 7 – Comparação da acurácia e de parâmetros entre o RTMDet, destacado em vermelho, e outros detectores de objetos considerados estado da arte na detecção em tempo real [Lyu et al. 2022]. . . . .	36
Figura 8 – Macro arquitetura da RTMDet [Lyu et al. 2022]. . . . .	36
Figura 9 – Inferências da rede <i>baseline</i> para um exemplo da classe <i>defeito</i> . As regiões detectadas como defeito são indicadas em vermelho. . . . .	39
Figura 10 – Inferências da rede <i>baseline</i> para outro exemplo avaliado. A classe <i>defeito</i> é destacada em vermelho. . . . .	40
Figura 11 – Inferências da rede <i>baseline</i> para outro cenário avaliado. A classe <i>ok</i> é apresentada em azul. . . . .	40
Figura 12 – <b>Mask R-CNN</b> para segmentação de instâncias [He et al. 2017]. . . . .	44
Figura 13 – Representação simplificada das duas etapas do módulo CBAM. . . . .	49
Figura 14 – Diagrama ilustrativo dos módulos de atenção utilizados neste trabalho. A atenção de canal (acima) utiliza <i>average pooling</i> e <i>max pooling</i> para realçar canais relevantes. A atenção espacial (abaixo) aplica as mesmas operações ao longo dos canais e utiliza uma convolução para gerar o mapa espacial refinado. Este diagrama é adaptado com base em [Woo et al. 2018]. . . . .	51

Figura 15 – Imagens com vários defeitos (parte superior) e contaminações (parte inferior) presentes no conjunto de dados Blade30. Vários defeitos, entre eles, rachaduras na borda, rachaduras na superfície vermelha e na superfície branca causadas por trovões, erosão na borda principal. Contaminações diversas como manchas de óleo, manchas de areia aderida, etc [Yang et al. 2023]. . . . .	54
Figura 16 – Imagem do conjunto Blade30 ilustrando uma pá alinhada aos eixos da imagem.	55
Figura 17 – Imagem do conjunto Blade30 mostrando outro exemplo em que a pá apresenta bom alinhamento em relação aos eixos da imagem. . . . .	55
Figura 18 – Imagem do conjunto Blade30 em que a pá não está alinhada aos eixos da imagem, evidenciando maior variação de orientação. . . . .	56
Figura 19 – Imagem do conjunto Blade30 ilustrando outro caso em que a pá apresenta orientação distinta em relação aos eixos da imagem. . . . .	56
Figura 20 – Exemplo de anotações do dataset convertidas para o formato PNG a partir do arquivo JSON original, evidenciando imprecisões inerentes ao processo de anotação manual, no qual as regiões de falha não apresentam contornos geométricos bem definidos. . . . .	57
Figura 21 – Exemplos de imagens originais em alta resolução do conjunto de dados Blade30, utilizadas na inspeção visual de pás eólicas. . . . .	57
Figura 22 – Exemplos de regiões recortadas ( <i>crops</i> ) de $256 \times 256$ pixels contendo defeitos estruturais visíveis. As imagens são alinhadas com suas respectivas máscaras binárias. . . . .	58
Figura 23 – Comparação entre variantes do modelo YOLOv8 e YOLOv8M+CBAM com base no desempenho de segmentação (30 épocas, <i>batch size</i> = 4). . . . .	63
Figura 24 – Exemplos de resultados visuais obtidos pelo modelo <i>YOLOv8+CBAM</i> . Da esquerda para a direita: recorte da imagem original, máscara anotada ( <i>ground truth</i> ), máscara segmentada após binarização e mapa de probabilidade produzido pela rede. . . . .	64
Figura 25 – Exemplos adicionais de resultados visuais obtidos pelo modelo <i>YOLOv8+CBAM</i> . Da esquerda para a direita: recorte da imagem original, máscara anotada ( <i>ground truth</i> ), máscara segmentada após binarização e mapa de probabilidade produzido pela rede. . . . .	65

---

# Lista de tabelas

---

Tabela 1 – Comparação entre trabalhos relacionados à detecção e segmentação de defeitos em pás de turbinas eólicas . . . . .	30
Tabela 2 – Tabela de rótulos e valores adotados para o pré-processamento. . . . .	32
Tabela 3 – Hiperparâmetros de treinamento da rede base de comparação . . . . .	36
Tabela 4 – Efeito das técnicas de <i>data augmentation</i> no conjunto Blade30 segundo a métrica COCO mAP . . . . .	38
Tabela 5 – Impactos observados após integração do CBAM. . . . .	49
Tabela 6 – Estatísticas do conjunto de dados Blade30. . . . .	53
Tabela 7 – Principais inconsistências encontradas nas máscaras do Blade30. . . . .	58
Tabela 8 – Hiperparâmetros e configurações utilizados no treinamento dos modelos de segmentação . . . . .	59
Tabela 9 – Comparação de desempenho entre variantes do YOLOv8 e YOLOv8 + CBAM na segmentação de defeitos . . . . .	62
Tabela 10 – Desempenho dos modelos YOLOv8 na tarefa de segmentação (COCO) [Ultralytics 2024]. . . . .	63
Tabela 11 – Resumo qualitativo das observações sobre as previsões. . . . .	64

---

# Sumário

---

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>13</b>
<b>2</b>	<b>OBJETIVOS DA PESQUISA . . . . .</b>	<b>17</b>
<b>3</b>	<b>REVISÃO DE LITERATURA E ABORDAGEM METODOLÓGICA DA REVISÃO . . . . .</b>	<b>19</b>
<b>3.1</b>	<b>Metodologia para a pesquisa bibliográfica . . . . .</b>	<b>19</b>
<b>3.2</b>	<b>Referencial teórico . . . . .</b>	<b>20</b>
<b>4</b>	<b>ESTUDOS PRELIMINARES . . . . .</b>	<b>31</b>
4.0.1	Pré-processamento . . . . .	31
4.0.2	Métodos . . . . .	32
4.0.3	RTMDet . . . . .	35
4.0.4	Detalhes de implementação . . . . .	35
4.0.5	<i>Aumento de dados</i> . . . . .	37
4.0.6	Métrica de avaliação escolhida para o estudo inicial . . . . .	37
4.0.7	Resultados do estudo inicial . . . . .	37
4.0.8	Conclusões sobre o estudo inicial . . . . .	39
<b>5</b>	<b>METODOLOGIA . . . . .</b>	<b>43</b>
<b>5.1</b>	<b>Redes de segmentação . . . . .</b>	<b>43</b>
<b>5.2</b>	<b>Modelo base: YOLOv8 . . . . .</b>	<b>44</b>
5.2.1	Justificativa do modelo base . . . . .	46
<b>5.3</b>	<b>Módulos de atenção . . . . .</b>	<b>47</b>
<b>5.4</b>	<b>Método proposto: integração com CBAM . . . . .</b>	<b>49</b>
5.4.1	Módulo de atenção de canal . . . . .	49
5.4.2	Módulo de atenção espacial . . . . .	50
<b>6</b>	<b>EXPERIMENTOS . . . . .</b>	<b>53</b>
<b>6.1</b>	<b>Conjuntos de dados: Blade30 . . . . .</b>	<b>53</b>

6.1.1	Estrutura do conjunto de dados . . . . .	54
6.1.2	Versão recortada: BLADE30 . . . . .	57
6.1.3	Desafios enfrentados . . . . .	58
6.1.4	Utilização do dataset no trabalho . . . . .	58
<b>6.2</b>	<b>Infraestrutura experimental . . . . .</b>	<b>59</b>
<b>6.3</b>	<b>Aplicação de aumento de dados . . . . .</b>	<b>60</b>
<b>7</b>	<b>MÉTRICAS DE AVALIAÇÃO . . . . .</b>	<b>61</b>
7.0.1	Comparação de desempenho . . . . .	62
7.0.2	Comparação visual . . . . .	62
7.0.3	Resultados dos experimentos . . . . .	63
<b>8</b>	<b>AMEAÇAS À VALIDADE . . . . .</b>	<b>67</b>
8.0.1	Validade interna . . . . .	67
8.0.2	Validade externa . . . . .	67
8.0.3	Validade de construto . . . . .	68
8.0.4	Perspectivas de Trabalhos Futuros e Resultados Promissores . . . . .	68
<b>Conclusão</b>	<b>. . . . .</b>	<b>71</b>
<b>Referências</b>	<b>. . . . .</b>	<b>75</b>

---

# Capítulo 1

## Introdução

---

A energia eólica consolidou-se como um dos pilares fundamentais na transição global em direção a sistemas energéticos mais sustentáveis e ambientalmente responsáveis [Zhang et al. 2018].

Dentre os diversos componentes críticos que compõem a infraestrutura de geração de energia por meio dos ventos, as pás das turbinas eólicas (do inglês, *Wind Turbine Blades* – WTBs) desempenham um papel essencial na eficiência da conversão energética, sendo diretamente responsáveis pela captação e transferência da energia cinética do vento para o gerador mecânico [García Márquez et al. 2012].

No entanto, em função de suas dimensões consideráveis, exposição prolongada a condições ambientais extremas (como ventos intensos, radiação solar, salinidade e umidade), bem como de seu projeto aerodinâmico altamente complexo, as WTBs tornam-se particularmente vulneráveis a uma ampla gama de defeitos estruturais e superficiais [Liu et al. 2015].

Entre os tipos mais recorrentes de danos observados, destacam-se erosões de material, trincas, avarias provocadas por impacto, raios, além da adesão de partículas ou materiais estranhos à superfície das pás [Liu et al. 2015]. Tais falhas comprometem significativamente o desempenho aerodinâmico das turbinas, reduzem sua integridade estrutural e afetam diretamente a segurança operacional do sistema como um todo [García Márquez et al. 2012].

Diante desse cenário, torna-se indispensável o desenvolvimento e a aplicação de técnicas de inspeção que sejam, simultaneamente, precisas, confiáveis e escaláveis, a fim de garantir a durabilidade das estruturas ao longo do tempo e otimizar a relação custo-benefício da manutenção [García Márquez et al. 2012] em parques eólicos.

Tradicionalmente, as inspeções visuais manuais ainda são amplamente utilizadas para avaliar as condições das pás. Esses procedimentos, normalmente conduzidos por técnicos especializados por meio de acesso por cordas ou plataformas elevadas, oferecem certa flexibilidade operacional em situações específicas [Ribrant e Bertling 2007].

No entanto, apresentam diversas limitações intrínsecas: são altamente subjetivos, demandam muito tempo, implicam riscos à segurança dos inspetores e são pouco eficazes para inspeções em

larga escala, como aquelas requeridas em grandes parques eólicos [Ribrant e Bertling 2007].

Adicionalmente, o fator humano pode ocasionar falhas na detecção de defeitos sutis ou em estágio inicial, especialmente sob condições climáticas adversas ou em ambientes com iluminação deficiente.

Com o objetivo de superar essas limitações operacionais e garantir maior eficiência nos processos de inspeção, a indústria tem adotado, de forma crescente, veículos aéreos não tripulados (VANTs ou drones) [Marin et al. 2009] [57], equipados com câmeras de alta resolução. Esses dispositivos possibilitam a realização de inspeções visuais rápidas, sem contato direto com as estruturas, e sob diferentes ângulos de observação [Tchakoua et al. 2014].

Apesar dos avanços na coleta de imagens, o volume massivo de dados visuais gerado por drones representa um novo desafio: a interpretação consistente, eficiente e automática dessas imagens para identificar anomalias. Essa dificuldade evidencia a necessidade de sistemas robustos de análise de imagens, automatizados e capazes de detectar, localizar e caracterizar múltiplos tipos de defeitos com alta precisão e confiabilidade [Chandrasekhar et al. 2021].

Nos últimos anos, avanços significativos em aprendizado profundo, em especial no uso de redes neurais convolucionais (*Convolutional Neural Networks* – CNNs) e de arquiteturas codificadoras-decodificadoras como a YOLOv8 [Zhang et al. 2024], demonstraram grande potencial em tarefas de segmentação semântica aplicadas a diferentes domínios.

Estudos recentes demonstram que o uso de mecanismos de atenção em redes neurais profundas tem elevado significativamente o desempenho de tarefas de segmentação semântica, especialmente em cenários com alta complexidade visual, presença de ruídos e defeitos sutis [44].

Dentre os mecanismos mais eficazes destaca-se o *CBAM* [Woo et al. 2018], que combina atenção espacial e por canal para realçar regiões informativas e suprimir padrões irrelevantes [Zhao et al. 2025]. Essa estratégia permite a detecção precisa de regiões defeituosas mesmo quando estas se apresentam com baixa resolução, baixo contraste ou em formatos irregulares.

Com base nessas inovações, este trabalho propõe um modelo leve de segmentação semântica, denominado **YOLOv8+CBAM**, voltado à detecção automática de defeitos estruturais em pás de turbinas eólicas a partir de imagens reais adquiridas por VANTs (Veículos Aéreos Não Tripulados).

A arquitetura baseia-se na estrutura da YOLOv8 [Qadri et al. 2023, Zhang et al. 2024], reconhecida por seu desempenho em tempo real e precisão em detecção. O modelo é modificado para incluir módulos CBAM ao longo do *backbone*, permitindo um refinamento progressivo das representações extraídas em cada estágio da rede.

Esses módulos aplicam atenção por canal e atenção espacial de forma sequencial, melhorando a sensibilidade do modelo a padrões visuais sutis, como fissuras, erosões e rachaduras que ocorrem em regiões específicas das pás.

O modelo foi treinado e validado em imagens reais do conjunto de dados *Blade30*, composto por recortes de regiões com defeitos anotados e suas respectivas máscaras binárias. Mesmo com alta desproporção entre pixels de fundo e de defeito, a abordagem alcançou desempenho consistente em métricas como *mIoU* e *mAP@50*, superando variantes da YOLOv8 sem mecanismos de atenção.

Entre os principais benefícios da YOLOv8+CBAM destacam-se:

- **Alta precisão** na segmentação de alvos pequenos e esparsos, com forte alinhamento espacial

---

entre a predição e a máscara de referência;

- **Baixo custo computacional**, adequado para ambientes com recursos limitados;
- **Robustez** a variações de iluminação, perspectiva e ruído, comuns em ambientes de inspeção reais [Yang et al. 2023].

Além disso, a arquitetura proposta é modular, o que facilita experimentações com diferentes estratégias de pré-processamento, aumento de dados e técnicas de regularização.

A visualização dos resultados qualitativos evidencia a capacidade do modelo em segmentar precisamente regiões danificadas, mesmo em casos de baixo contraste ou ruído acentuado. Essas informações estruturadas podem ser empregadas para análises preditivas, planejamento de manutenção e monitoramento da integridade estrutural ao longo do tempo [Xue et al. 2020].

Em suma, o modelo YOLOv8+CBAM representa uma solução eficaz e escalável para inspeção automatizada de pás de turbinas eólicas, integrando o estado da arte em segmentação semântica com requisitos práticos de aplicação em campo.



---

## Capítulo 2

# Objetivos da pesquisa

---

Este trabalho parte da premissa de que a integração de arquiteturas de segmentação semântica baseadas em redes neurais profundas, aliadas a mecanismos de atenção espacial e de canal, representa uma solução eficaz e eficiente para a detecção de defeitos estruturais em pás de turbinas eólicas.

Considerando o alto risco operacional associado a falhas não detectadas nesses componentes e os desafios impostos por imagens de alta resolução obtidas por drones, este estudo propõe uma abordagem automatizada capaz de segmentar, com precisão, regiões danificadas nas pás a partir de imagens reais em contexto industrial.

A hipótese central deste trabalho é que a adaptação da arquitetura YOLOv8, com a incorporação de módulos CBAM (*Convolutional Block Attention Module*), aliada a uma estratégia de inferência baseada em divisão de imagens em fatias (*slice-based inference*), é capaz de superar os desafios da segmentação de defeitos pequenos e esparsos, com baixo custo computacional e mantendo desempenho robusto em dados reais.

Dessa forma, o objetivo geral desta pesquisa consiste em desenvolver e validar um modelo leve de segmentação semântica que maximize a acurácia na detecção de falhas estruturais em pás de turbinas eólicas, mesmo sob condições de variabilidade visual e limitações operacionais de hardware.

Os objetivos específicos incluem:

- ❑ Adaptar a arquitetura YOLOv8 para tarefas de segmentação semântica, com foco na identificação de defeitos em imagens de alta resolução capturadas por drones;
- ❑ Integrar módulos de atenção CBAM ao backbone do modelo, com o intuito de reforçar a sensibilidade da rede a padrões visuais sutis e regiões de interesse pouco contrastadas;
- ❑ Implementar uma estratégia de inferência baseada em fatias, otimizando o processamento de imagens de grandes dimensões e melhorando a segmentação de pequenos defeitos;

- ❑ Avaliar o desempenho da proposta em um cenário realista de inspeção de pás, utilizando métricas como mIoU, precisão, revocação e mAP@50;
- ❑ Demonstrar, por meio de resultados qualitativos e quantitativos, a aplicabilidade prática da solução proposta para inspeção automatizada em contextos industriais.

Além dos objetivos acima, este trabalho busca contribuir com a comunidade científica por meio das seguintes entregas:

- ❑ Proposição de um modelo leve e preciso para segmentação de defeitos em ativos de energia renovável, com potencial para implantação em dispositivos embarcados ou drones;
- ❑ Apresentação de uma metodologia experimental replicável, com pipeline completo de pré-processamento, treinamento e inferência, facilitando a reprodutibilidade científica;
- ❑ Geração de evidências que demonstram o impacto da incorporação de mecanismos de atenção na capacidade da rede neural em detectar padrões de defeito pequenos, esparsos e com baixo contraste;
- ❑ Utilização de um conjunto de dados real e desafiador (Blade30) [Yang et al. 2023], evidenciando o potencial de generalização da arquitetura proposta para diferentes tipos de falhas estruturais;
- ❑ Estabelecimento de uma base sólida para futuras pesquisas na área de inspeção visual automatizada de pás de turbinas eólicas, especialmente em aplicações de manutenção preditiva com visão computacional.

Dessa forma, torna-se essencial o desenvolvimento de um método de segmentação robusto, leve e sensível a padrões visuais sutis, capaz de atuar de forma confiável em ambientes reais de inspeção.

Essa necessidade motiva o presente trabalho, que investiga o uso de arquiteturas de aprendizado profundo aprimoradas pelo mecanismo de atenção para aprimorar a detecção de defeitos em pás de aerogeradores sob condições visuais desafiadoras.

Em síntese, este trabalho não se limita a “detectar defeitos em pás”, mas busca aperfeiçoar a segmentação semântica de pequenos defeitos estruturais em imagens reais de pás de turbinas eólicas (WTBs) capturadas por drones, enfrentando condições adversas de iluminação, texturas complexas e forte desbalanceamento entre classes.

Trata-se de um *gap* técnico concreto, ainda pouco explorado na literatura, uma vez que a maioria dos estudos concentra-se em detecção por bounding boxes e não em segmentação pixel a pixel.

A proposta de integrar o módulo de atenção CBAM à arquitetura YOLOv8 tem como objetivo aprimorar a extração e a ponderação das características relevantes, mantendo a leveza e a viabilidade para futuras aplicações embarcadas.

Essa abordagem se diferencia por unir eficiência computacional, precisão e aplicabilidade prática, contribuindo de forma relevante para os sistemas de inspeção automatizada e manutenção preditiva no setor eólico.

---

## Capítulo 3

# Revisão de literatura e abordagem metodológica da revisão

---

Esta seção descreve a metodologia utilizada na condução da revisão bibliográfica e apresenta os principais estudos que fundamentam esta pesquisa.

### 3.1 Metodologia para a pesquisa bibliográfica

Para embasar teoricamente o desenvolvimento desta pesquisa, foi realizada uma ampla busca por artigos científicos recentes e diretamente relacionados ao tema proposto. Esta etapa foi conduzida utilizando ferramentas especializadas como o Parsifal, uma plataforma reconhecida por facilitar revisões sistemáticas da literatura, especialmente em áreas de tecnologia e ciências aplicadas.

A seleção das fontes foi feita priorizando bases de dados amplamente reconhecidas pela comunidade acadêmica, como o IEEE Xplore e a Scopus. Estas bases foram escolhidas devido à sua relevância, abrangência e rigor acadêmico, garantindo a qualidade e atualidade das referências utilizadas.

Os critérios de busca incluíram palavras-chave diretamente relacionadas ao objeto de estudo, combinadas por meio de operadores booleanos para refinar os resultados e identificar publicações altamente pertinentes. Além disso, foram aplicados filtros como intervalo temporal, considerando preferencialmente artigos publicados nos últimos cinco anos, bem como o idioma, restringindo a pesquisa a artigos em inglês e português.

A análise inicial dos artigos identificados foi realizada com base nos títulos e resumos, permitindo uma triagem eficiente e a exclusão de publicações que não atendiam aos critérios de inclusão definidos previamente. Posteriormente, os artigos selecionados foram revisados integralmente para assegurar que os estudos incluídos na dissertação fossem relevantes e contribuíssem de forma significativa para o avanço da discussão proposta.

Por meio desta metodologia sistemática, foi possível construir um embasamento teórico sólido, alinhado às melhores práticas acadêmicas e às tendências mais recentes da literatura científica na área abordada.

## 3.2 Referencial teórico

Muitas pesquisas documentadas na literatura acerca da identificação de danos em pás de turbinas eólicas (WTB), empregando técnicas de aprendizado de máquina, visam enfatizar o impacto de abordagens diversas aplicadas por pesquisadores e profissionais no tratamento deste desafio específico de visão computacional.

A detecção de defeitos em superfícies de pás de turbinas eólicas (WTB) usando técnicas de visão computacional é um subconjunto da detecção de objetos. Nesse contexto, os objetos são categorizados como diversos tipos de danos, defeitos ou contaminações de superfície nas WTBs. Compreender a evolução dos algoritmos de detecção de objetos ao longo do tempo é crucial para abordar esse problema de forma eficaz.

Portanto, esta seção está organizada em duas partes. A primeira parte apresenta uma visão geral concisa da literatura na área de reconhecimento de objetos, enquanto a segunda parte foca em estudos que utilizam aprendizado profundo para processar imagens e automatizar a identificação de danos.

## Avanços nos algoritmos de identificação de objetos

Antes da adoção generalizada de técnicas de aprendizado profundo para detecção de objetos, os algoritmos tradicionais eram predominantemente utilizados. Esses algoritmos exigiam a extração de características como etapa de pré-processamento antes da classificação. Um dos métodos pioneiros, a detecção de objetos de Viola-Jones [Castrillón et al. 2011], introduziu o conceito de imagens integrais e empregou características de Haar para detecção.

Em 2004, a Transformada de Características Invariantes à Escala (SIFT, do inglês *Scale Invariant Feature Transform*) [Hossein-Nejad e Nasri 2017] avançou a área, permitindo a extração de características de imagens que eram então comparadas com um banco de dados de características para identificar objetos.

Subsequentemente, a abordagem do Histograma de Gradientes Orientados (HOG, do inglês *Histogram of Oriented Gradient*) [Nabizadeh e Kubat 2015] foi introduzida em 2006. Ela utilizava uma técnica inovadora de extração de características combinada com um classificador de máquina de vetores de suporte (SVM, do inglês *Support Vector Machine*) para detecção de objetos. Pesquisas também exploraram a extração de características no domínio da frequência como forma de melhorar a precisão da detecção [Nur-A-Alam et al. 2021] [Byeon et al. 2024].

Em contraste, os métodos baseados em aprendizado profundo revolucionaram a detecção de objetos ao eliminar a necessidade de extração manual de características. Redes neurais convolucionais (CNNs) tornaram-se a base para realizar cálculos diretamente em imagens [Li et al. 2022].

## Segmentação e classificação de defeitos em imagens

A segmentação e a classificação de defeitos em imagens são tarefas centrais em diversos domínios da engenharia e da indústria, como manufatura, infraestrutura civil e sistemas aeroespaciais. Esses processos envolvem a análise automática de imagens para localizar, reconhecer e categorizar imperfeições que possam comprometer a integridade ou o desempenho de componentes físicos.

Com o avanço das técnicas de visão computacional e aprendizado profundo, tornou-se possível aplicar essas abordagens em estruturas cada vez mais complexas e de difícil acesso. Um exemplo relevante é o das pás de turbinas eólicas, cuja inspeção visual é fundamental para garantir a operação segura e eficiente dos aerogeradores.

Devido à grande extensão e curvatura dessas pás, estratégias específicas de aquisição e processamento de imagens são necessárias para permitir uma análise eficaz de sua superfície.

Como as superfícies de WTB são extensas, capturar toda a superfície de uma pá em alta definição em uma única imagem de drone não é viável. Para lidar com isso, pesquisadores desenvolveram técnicas de composição de imagens (*image stitching*) para criar composições precisas da WTB completa. Um método, descrito em [Yu et al. 2023], realiza diversos cálculos geométricos para alinhar imagens da mesma WTB e gerar uma visão panorâmica.

Além da composição, esforços também foram feitos para melhorar a qualidade das imagens de WTB capturadas sob diferentes condições de iluminação. Por exemplo, o método proposto em [Peng et al. 2022] utiliza mapas de textura e desenho (*cartoon*) para melhorar a iluminação das imagens de WTB, garantindo melhor visibilidade e análise.

Pesquisadores também desenvolveram outras técnicas de pré-processamento para isolar áreas relevantes nas imagens de WTB. Uma abordagem, proposta em [Zhou et al. 2023], utiliza um método chamado “regression crop” para extrair a superfície da WTB das imagens e focar a detecção apenas em uma região específica. Outra técnica, detalhada em [Yu et al. 2023], apresenta uma versão aprimorada da U-Net, capaz de segmentar WTBs em imagens infravermelhas utilizando operações convolucionais separáveis em profundidade de divisão hierárquica.

Uma melhoria adicional na U-Net, projetada especificamente para segmentação de WTBs, foi proposta em [Wang et al. 2022]. Esse método incorpora uma rede de atenção eficiente por canal e uma rede de atenção espacial pontual para refinar o processo de segmentação. Além disso, um ótimo desempenho na atividade de segmentação de imagens de WTB foi alcançado com o Pixel U-Net [44], uma versão avançada da U-Net que substitui operações de max pooling e convolução transposta por operações de pixel unshuffle e pixel shuffle, melhorando tanto a precisão quanto a eficiência.

Além das pesquisas sobre técnicas de pré-processamento para imagens de WTB, pesquisadores exploraram aprendizado profundo (DL) para desenvolver redes especificamente adaptadas para detectar defeitos em WTBs. Um método proposto em [Zhang e Wen 2022] utiliza uma versão aprimorada do YOLOv5 para identificar pequenos defeitos em WTBs de forma eficaz. Outro estudo, apresentado em [Ran et al. 2022], introduz uma variante de YOLO com atenção e equilíbrio de características, que melhora o processo de detecção ao reunir informações mais detalhadas de características. Essa arquitetura aprimorada para o YOLOv5s (onde “s” denota small) também substitui a tradicional perda de interseção sobre união (IoU) pela perda de IoU eficiente (EIoU)

para cálculos mais precisos de caixas delimitadoras.

A análise de defeitos em pás de turbinas eólicas abrange uma variedade de metodologias, desde técnicas convencionais baseadas em processamento de imagens até abordagens modernas que utilizam características criadas manualmente.

Wang e Zhang [Wang e Zhang 2017], por exemplo, usaram recursos Haar-like combinados com classificadores em cascata para identificar rachaduras na superfície das pás, com foco principal em distinguir regiões rachadas das não rachadas. De forma semelhante, Huang e Wang [58] aprimoraram o uso desses recursos, integrando-os ao algoritmo Jaya K-means paralelo para maior precisão na detecção de rachaduras.

Para detectar eficientemente rachaduras de grande escala em turbinas eólicas, Peng et al. [Peng e Liu 2018] propuseram um framework analítico utilizando imagens capturadas por drones. Alternativamente, Ruiz e Magda [Ruiz et al. 2018] adotaram uma abordagem distinta, convertendo sinais operacionais de turbinas eólicas em representações em escala de cinza e utilizando características multicanal de textura para reconhecimento de padrões.

O advento de metodologias baseadas em aprendizado profundo trouxe uma mudança significativa no paradigma de detecção de defeitos em pás de turbinas eólicas. Shihavuddin et al. [Shihavuddin et al. 2021] introduziram uma rede piramidal de características combinada com técnicas de aumento de dados offline, especialmente projetadas para imagens de alta resolução. Este método envolveu o treinamento de diversos detectores Faster-RCNN em conjuntos de dados privados heterogêneos, gerando resultados promissores. Explorações subsequentes investigaram a eficácia de modelos YOLO e EfficientDet, comprovando ainda mais o potencial das estruturas de aprendizado profundo nessa área.

A notoriedade das redes neurais convolucionais em relação aos descritores tradicionais foi destacada, especialmente com o uso de classificadores em conjunto. Melnychenko et al. [32] contribuíram categorizando defeitos em pás de turbinas eólicas utilizando pedaços das imagens para treinamento e inferência. Yu et al. [Yu et al. 2020] enfrentaram desafios associados a imagens desfocadas, implementando um modelo de rede neural convolucional de super-resolução complementado por técnicas de pré-processamento de variância Laplaciana. Avanços notáveis na detecção de defeitos também foram atribuídos a outras arquiteturas de aprendizado profundo.

Wightman et al. [61] trazem uma análise sobre o efeito da escolha de hiperparâmetros e variações na rotina de treinamento em relação ao desempenho final de uma rede de classificação de imagens.

Bello et al. [Bello et al. 2021] buscam, também, uma rede para o problema de classificação de imagens, expondo quais ganhos devem ser atribuídos a arquiteturas mais modernas e quais são frutos de estratégias de treinamento melhores.

Ambos os autores [61] [Bello et al. 2021] concluem que variações na rotina de treinamento têm grande influência no desempenho final de um método, podendo-se obter resultados equivalentes ao estado da arte com arquiteturas de redes mais antigas.

Wang et al. [59] apresenta uma abordagem inovadora em duas etapas para a detecção automática de rachaduras em pás de turbinas eólicas, utilizando imagens capturadas por drones.

Na primeira etapa, um método de janela deslizante paralela combinado com um classificador em cascata identifica as áreas com rachaduras nas imagens e, na segunda etapa, um algoritmo Jaya

K-means paralelo é aplicado para segmentar essas áreas e identificar os contornos das rachaduras com precisão [59].

A abordagem foi validada com imagens de um parque eólico comercial, demonstrando alta eficiência e precisão, mesmo em condições de iluminação variada. Além disso, o método foi implementado com sucesso em computadores pessoais, mostrando sua viabilidade para aplicações em campo e em tempo real. Comparações com outros algoritmos destacam a superioridade do método proposto em termos de rapidez e qualidade de detecção [59].

Yu et al. [Yu et al. 2023] apresenta um modelo aprimorado de U-Net para a segmentação de imagens infravermelhas de pás de turbinas eólicas. O modelo utiliza convolução separável em profundidade e fusão de camadas de convolução e normalização em lote para melhorar a precisão da segmentação e reduzir a complexidade computacional. O modelo é avaliado em um conjunto de imagens capturadas por Veículos Aéreos Não Tripulados (VANT) equipados com câmeras infravermelhas, demonstrando bons resultados em termos de métricas de avaliação de segmentação [Yu et al. 2023].

O estudo feito por Yang et al. [Yang et al. 2021] propõe um modelo de reconhecimento de imagem baseado na Rede Neural Convolutiva AlexNet com transferência de aprendizado e Florestas Aleatórias para detectar danos nas pás das turbinas eólicas. A segmentação de limiar Otsu é empregada para eliminar o fundo da imagem, aumentando a eficácia da detecção. As imagens de pás de turbinas eólicas obtidas por VANTs são utilizadas para validar o modelo, obtendo resultados satisfatórios [Yang et al. 2021].



Figura 1 – Detecção de erosão na ponta de um aerogerador [Reddy et al. 2019].

A pesquisa conduzida por Reddy et al. [Reddy et al. 2019] apresenta um método de detecção de rachaduras e danos em pás de turbinas eólicas por meio da análise de imagens utilizando redes neurais convolucionais. O modelo treinado com imagens capturadas por drones classifica imagens das pás como defeituosas ou não defeituosas, reduzindo o tempo e risco associados à inspeção manual. O método proposto alcança altos níveis de precisão na classificação de falhas, sendo econômico e passível de implantação com o uso do microframework Flask [Reddy et al. 2019]. O autor utiliza a biblioteca Keras para implementar algumas técnicas de *aumento de dados* como cisalhamento, zoom e inversão horizontal para gerar novos dados e tornar o treinamento mais robusto.

O estudo de Dervilis et al. [Dervilis et al. 2014] explora o monitoramento da saúde estrutural em turbinas eólicas, enfatizando a melhoria do seu modelo probabilístico. Discute-se o uso de análise probabilística de componentes principais, redes neurais convencionais autoassociativas e com funções de base radial, introduzindo um índice de novidade e um limite de alarme para detecção

de desvios. A pesquisa [Dervilis et al. 2014] se concentra na detecção de danos nas pás através de análise de vibração, com um estudo de caso envolvendo testes de fadiga em uma pá de 9 metros e uso de componentes principais para identificar danos precocemente. Destaca-se a eficácia das técnicas de aprendizado de máquina no monitoramento estrutural, o uso de respostas de alta frequência na detecção de danos, a importância da seleção de características e o potencial de redes de sensores sem fio [Dervilis et al. 2014].

Em 2024, Dwivedi et al. [Dwivedi et al. 2024] propuseram um método para monitoramento de superfícies utilizando um “Vision Transformer”(ViT) para detectar danos estruturais em painéis solares e pás de turbinas eólicas. O modelo ViT alcançou um nível de desempenho superior em comparação com outros modelos de aprendizado profundo, como MobileNet, VGG16, Xception, EfficientNetB7 e ResNet50, com uma taxa de acurácia de 97%. Esse trabalho [Dwivedi et al. 2024] alcançou alta precisão porque as imagens utilizadas para as turbinas eólicas apresentavam áreas de danos claras e severas, o que ajudou o modelo a diferenciar entre áreas saudáveis e áreas danificadas.

Enquanto isso, Shekaramiz et al. [Altice et al. 2024] apresentaram um modelo personalizado de Rede Neural Convolutiva (RNC) para detectar anomalias nas pás de turbinas eólicas. Comparado aos modelos Xception, ResNet-50, AlexNet e VGG-19, o modelo personalizado de RNC alcançou uma acurácia de teste de 96,67%. Os resultados indicaram que o modelo Xception proposto superou outras arquiteturas, atingindo uma acurácia de 99,92%.

Em 2022 [Ran et al. 2022], a detecção de objetos foi utilizada para identificar e localizar falhas em turbinas eólicas. Xiukang e colegas introduziram uma versão aprimorada do algoritmo YOLO chamada AFB-YOLO, que equilibra atenção e recursos, utilizando o YOLOv5s como base. O AFB-YOLO aprimora a rede de pirâmide de recursos. O modelo alcançou uma precisão média (mAP50) de 83,7%, indicando uma melhoria de 4,0% na precisão de detecção em comparação com o modelo original YOLOv5s.

Jihong Guo et al. [Guo et al. 2021] utilizaram redes neurais profundas para desenvolver um modelo de detecção de danos e diagnóstico de falhas. O modelo combina uma etapa de proposta de regiões baseada no Haar-AdaBoost com um classificador de rede neural convolutiva (RNC). Alcançando uma acurácia de 97% para 5 classes (Normal, Defeitos, Rachaduras, Defeitos de Revestimento, Erosão de Bordas), o modelo proposto supera os modelos SVM e VGG16, além de outros métodos.

O estudo de Yadong Xue et al. [Xue et al. 2020] propõe uma abordagem baseada em deep learning para a detecção automática de áreas de vazamento de água em túneis escavados, tendo foco em um problema crítico para a manutenção e segurança estrutural de infraestruturas subterrâneas. A metodologia [Xue et al. 2020] se baseia em uma versão aprimorada do Mask R-CNN, complementada por estratégias de aumento de dados, transferência de aprendizado e uma arquitetura em cascata para melhorar a qualidade das amostras positivas durante o treinamento. A estratégia em cascata ajuda a refinar progressivamente os resultados de segmentação por meio de estágios sucessivos com limiares de decisão crescentes de IoU, mitigando ruídos nas predições e reduzindo o risco de sobreajuste [Xue et al. 2020]. A validação experimental demonstrou que o modelo aprimorado alcançou um AP0.5 de 0.806, superando significativamente o desempenho da versão original do Mask R-CNN (AP0.5 = 0.530), validando a eficácia do modelo para segmentação

de precisão em ambientes complexos como túneis [Xue et al. 2020].

Além disso, considerando que a contagem de pixels de vazamento não é uma métrica confiável para avaliação de segurança, os autores realizaram experimentos de calibração em campo para estabelecer uma correlação entre a área real de vazamento e os pixels segmentados [Xue et al. 2020].

A curva de calibração obteve um erro médio de apenas 2,59%, permanecendo dentro dos limites aceitáveis da engenharia. Para reforçar a robustez do modelo, os autores [Xue et al. 2020] enriqueceram o conjunto de dados por meio de ruído gaussiano, ruído sal-e-pimenta e transformações geométricas. O conjunto final incluiu 3.555 imagens, ampliando a capacidade de generalização do modelo. Com isso, a pesquisa oferece uma solução automatizada, precisa e aplicável à engenharia para quantificar áreas de vazamento em túneis, sendo uma contribuição significativa ao setor de transporte subterrâneo e à manutenção preditiva de infraestrutura [Xue et al. 2020].

O artigo de Suat Ince et al. [Ince et al.] investiga a eficácia de diferentes variantes da arquitetura U-Net na tarefa de segmentação de acidentes vasculares cerebrais isquêmicos em imagens médicas, particularmente aquelas oriundas da técnica de imagem ponderada por difusão (diffusion weighted imaging - DWI), amplamente usada por sua sensibilidade e confiabilidade no diagnóstico precoce. Foram comparadas três arquiteturas: U-Net, U-Net++ e Attention U-Net, utilizando critérios quantitativos rigorosos como interseção sobre união (Intersection over Union - IoU), coeficiente de similaridade de dice (dice similarity coefficient - DSC) e recall [Ince et al.]. O modelo Attention U-Net obteve os melhores resultados, com valores de IoU (0.8223) e DSC (0.9021) [Ince et al.], demonstrando que a introdução de mecanismos de atenção contribui para uma segmentação mais precisa e eficaz [Ince et al.].

Embora o U-Net tenha apresentado recall elevado, teve desempenho inferior em métricas gerais quando comparado às outras variantes [Ince et al.]. A pesquisa de Suat Ince et al. [Ince et al.] destaca a importância do diagnóstico preciso e automatizado, especialmente nas fases iniciais de AVC, nas quais a intervenção rápida pode impactar decisivamente os desfechos clínicos.

O uso de redes profundas baseadas em U-Net se mostrou vantajoso por sua capacidade de extrair características relevantes mesmo diante de ruídos e sobreposição de intensidades entre tecidos [Ince et al.]. A análise mostrou que Attention U-Net é particularmente promissora para aplicações clínicas, devido à sua habilidade em manter precisão e eficiência simultaneamente [Ince et al.]. Os autores concluem que, embora U-Net e U-Net++ ainda sejam adequados em cenários específicos, o uso de módulos de atenção pode se tornar essencial em tarefas de segmentação médica que exijam alta sensibilidade e especificidade.

O trabalho de Chen, Yan, et al. [Chen et al. 2024] apresenta uma estratégia inovadora de detecção de defeitos de soldagem utilizando imagens de ultrassom e uma rede de segmentação em duas etapas, projetada especificamente para inspeção de qualidade em soldas de estruturas ferroviárias. A proposta integra uma rede baseada no framework DeepLabv3+ e um bloco de julgamento regional (*Region Judgment Block* - RJB), permitindo separar áreas válidas das regiões com ruídos que poderiam comprometer a análise [Chen et al. 2024]. O desafio central abordado é a alta sensibilidade dos sinais recebidos, que frequentemente geram falsos positivos em regiões não soldadas [Chen et al. 2024]. A solução apresentada utiliza regras específicas de tipo de

região e um módulo de refinamento que melhora a precisão da segmentação ao eliminar sinais errôneos [Chen et al. 2024].

O modelo de [Chen et al. 2024] é avaliado extensivamente por meio de experimentos que testam diferentes parâmetros, funções de perda, taxas de dilatação e combinações de canais de entrada, validando a robustez da abordagem. Os resultados mostram que o modelo proposto supera métodos anteriores [Chen et al. 2024] ao isolar com maior precisão os defeitos reais, o que favorece análises dimensionais mais eficazes e confiáveis para avaliação de integridade estrutural. O estudo reforça a relevância de combinar inteligência artificial com técnicas não destrutivas avançadas para melhorar a inspeção em indústrias onde a segurança estrutural é crítica [Chen et al. 2024], como na fabricação de trens de alta velocidade.

Esse trabalho de Zhao, Lixiang, et al. [Zhao et al. 2025] propõe a rede de adaptação de domínio com aprendizado contrastivo cruzado-supervisionado (*Cross-Supervised Contrastive Learning Domain Adaptation Network - CSCLDAN*) para realizar segmentação de defeitos em superfícies de aço sob condições de domínio variável. A motivação central é lidar com o desafio da variação de aparência e contexto dos defeitos em diferentes ambientes industriais, o que causa um desvio de domínio (*domain shift*) que compromete o desempenho de redes convolucionais tradicionais [Zhao et al. 2025]. O CSCLDAN combina duas abordagens principais [Zhao et al. 2025]: aprendizado cruzado supervisionado entre uma CNN (UNet) e um Transformer (MCTrans) para captar tanto informações locais quanto globais, e aprendizado contrastivo de consistência autoconsciente (*self-aware consistency contrastive learning*), que mantém a uniformidade das características dos defeitos mesmo em diferentes contextos visuais.

Além disso, a função de perda de alinhamento de características em nível de pixel permite que amostras positivas sejam aproximadas enquanto as negativas sejam separadas, aumentando a discriminação do modelo [Zhao et al. 2025]. Nos experimentos, CSCLDAN superou sete modelos de adaptação de domínio e cinco métodos supervisionados em quatro bases de dados de defeitos em aço, alcançando melhorias significativas em métricas como IoU e F1-score [Zhao et al. 2025]. Embora o modelo seja mais complexo em termos de parâmetros, sua precisão e capacidade de generalização o tornam vantajoso para cenários industriais reais [Zhao et al. 2025].

Os autores [Zhao et al. 2025] reconhecem limitações, como o desempenho sob condições não ideais (ex.: desequilíbrio de classes e escassez de dados), sugerindo como futuras direções o uso de técnicas de few-shot learning e a compactação do modelo. Esta abordagem [Zhao et al. 2025] representa um avanço promissor para aplicações industriais que exigem detecção robusta de defeitos sob múltiplas condições operacionais.

Nos últimos anos, arquiteturas de detecção de objetos como a família YOLO (*You Only Look Once*) [Redmon et al. 2016] ganharam ampla popularidade devido à sua elevada velocidade de inferência e à precisão competitiva, mesmo em ambientes com restrições de tempo real.

Embora inicialmente concebida para detecção de objetos via caixas delimitadoras (*bounding boxes*), a arquitetura YOLO tem sido amplamente adaptada para diferentes domínios, incluindo aplicações industriais.

Por exemplo, em [Fu et al. 2024], uma versão leve do YOLOv7 foi utilizada para o diagnóstico precoce de defeitos em pás de turbinas eólicas (*Wind Turbine Blades — WTB*), superando métodos

tradicionais baseados em classificação.

De maneira semelhante, a abordagem proposta em [Zhang et al. 2025], baseada no YOLOv5, atingiu uma acurácia superior a 94% na categorização de quatro tipos principais de defeitos em WTB, evidenciando a eficácia da arquitetura para sistemas de tomada de decisão em tempo real [Fu et al. 2024, Zhang et al. 2025].

Apesar desses avanços, a maior parte das pesquisas ainda utiliza versões anteriores da arquitetura, sendo que a YOLOv8 permanece pouco explorada nesse contexto, mesmo apresentando melhorias importantes, como a remoção do uso de âncoras (*anchor-free*), um novo cabeçalho de detecção (*detection head*) e melhores capacidades para tarefas de segmentação. Esta lacuna motivou a adoção da YOLOv8 como base da arquitetura proposta neste trabalho.

Paralelamente à evolução das arquiteturas de detecção, os mecanismos de atenção (*attention mechanisms*) têm transformado a maneira como redes convolucionais refinam seus mapas de características (*feature maps*).

Módulos como CBAM (*Convolutional Block Attention Module*), GAM, *Shuffle Attention*, *Coordinate Attention* e ECA (*Efficient Channel Attention*) introduziram estratégias mais refinadas para destacar regiões relevantes da imagem.

O CBAM, por exemplo, aplica atenção de forma sequencial nos domínios de canal e espaço, priorizando regiões informativas [Woo et al. 2018]. O GAM se vale de informações de gradientes para evidenciar áreas discriminativas [Hu et al. 2023]. O *Shuffle Attention* propõe uma operação de embaralhamento para aprimorar o fluxo de informação entre canais e regiões espaciais [Zhang e Yang 2021], enquanto o *Coordinate Attention* capta dependências espaciais aliadas à atenção entre canais. Já o ECA utiliza convoluções unidimensionais com tamanhos de kernel adaptativos para captar interações locais entre canais de maneira eficiente [Wang et al. 2020].

Tais mecanismos motivaram a escolha do CBAM como complemento à arquitetura YOLOv8, visando aumentar a sensibilidade da rede à detecção de defeitos estruturais, os quais são, via de regra, pequenos e esparsos.

Embora existam trabalhos que exploram a combinação entre YOLO e mecanismos de atenção em outros contextos, sua aplicação específica no monitoramento estrutural de pás de turbinas eólicas ainda é limitada.

Por exemplo, o trabalho de [72] utilizou uma versão do YOLOv4 combinada ao MobileNetv1 com técnicas de aprendizado por transferência para reconhecimento de imagens com defeitos.

O SC-YOLO, proposto por [48], incorporou uma estratégia de atenção entre estágios (*cross-stage attention*) com o objetivo de melhorar a detecção de alvos em pequena escala.

Já no estudo mais recente de [Hang et al. 2024], o MIP-YOLO introduziu um módulo de percepção de informações multivariadas (*Multivariate Information* — MIP) aliado ao componente C2TR para melhorar a detecção de fissuras em imagens de WTB, obtendo bons resultados com o custo adicional de aumento na complexidade da arquitetura.

Esses trabalhos demonstram que a sinergia entre mecanismos de atenção e arquiteturas baseadas em YOLO pode ser extremamente promissora em tarefas que exigem alta sensibilidade a detalhes visuais sutis.

O presente trabalho se apoia nesse cenário de avanços ao propor a adaptação da YOLOv8 para

segmentação semântica de defeitos estruturais em pás de turbinas eólicas, com a integração de módulos CBAM ao backbone da rede.

Diferentemente de abordagens anteriores, que focam em classificação ou detecção com bounding boxes, a proposta aqui apresentada concentra-se na segmentação em nível de pixel de defeitos estruturais minuciosos, em um contexto industrial real.

Ao explorar a arquitetura enxuta da YOLOv8 e os mecanismos de atenção do CBAM, esta proposta busca oferecer uma solução precisa e eficiente para inspeção de pás utilizando conjuntos de dados reais, contribuindo de forma significativa para o avanço das técnicas de monitoramento estrutural automatizado.

Coletivamente, esses estudos evidenciam o papel crucial das metodologias de aprendizado profundo na ampliação das capacidades de detecção de defeitos em inspeções de pás de turbinas eólicas. No entanto, o processamento de imagens de alta resolução capturadas por drones apresenta desafios substanciais devido à variabilidade nas escalas dos objetos e aos recursos computacionais necessários.

Para mitigar esses problemas, é comum a prática de dividir imagens em pedaços menores, uma abordagem que reduz a carga computacional, melhora a clareza dos objetos e fortalece o desempenho dos modelos.

Embora exista uma vasta literatura dedicada à detecção de defeitos nas pás, ainda há lacunas significativas nas metodologias e esquemas de classificação utilizados. A ausência de dados públicos com anotações dificulta a avaliação comparativa de desempenho entre detectores de defeitos em superfícies das pás de turbinas eólicas.

Este estudo busca preencher essa lacuna ao propor uma abordagem unificada e estabelecer um marco para pesquisas futuras na área.

Utilizar drones para detectar defeitos nas superfícies de pás de turbinas eólicas tem-se mostrado uma alternativa eficiente e econômica, mas enfrenta desafios como o processamento de imagens de alta resolução, a detecção de pequenos objetos e as mudanças de escala devido à posição do drone.

Diante dos desafios impostos pela inspeção automatizada de ativos em ambientes industriais, este trabalho tem como objetivo principal aperfeiçoar a segmentação de defeitos estruturais em pás de turbinas eólicas, a partir da análise de imagens em alta resolução capturadas por drones.

As principais contribuições desta pesquisa são destacadas a seguir:

- ❑ Proposição de um algoritmo de segmentação semântica de defeitos estruturais, o qual incorpora uma estratégia realista de inferência baseada em divisões por fatias (slice-based inference) para otimizar a detecção de objetos em imagens de grande dimensão espacial;
- ❑ Validação da abordagem proposta em um cenário realista de inspeção, demonstrando sua aplicabilidade prática na segmentação de defeitos estruturais em pás de turbinas eólicas, utilizando imagens de alta resolução capturadas por drones;
- ❑ Adaptação e otimização da arquitetura YOLOv8 para a tarefa de segmentação semântica, diferenciando-se das abordagens convencionais baseadas em detecção por bounding boxes, possibilitando uma análise pixel a pixel mais precisa;

- ❑ Melhorar a segmentação semântica de defeitos pequenos e esparsos em imagens reais de WTBs capturadas por drones, sob forte desbalanceamento entre fundo e defeito e condições de iluminação irregulares;
- ❑ Proposta de uma metodologia experimental replicável, incluindo métricas robustas de avaliação (como mIoU e Dice), para quantificar o desempenho da segmentação de defeitos em pás, incentivando futuras pesquisas a adotarem padrões consistentes.

Tabela 1 – Comparação entre trabalhos relacionados à detecção e segmentação de defeitos em pás de turbinas eólicas

Trabalho	Arquitetura / Método	Tarefa	Conjunto de Dados	Principais Contribuições e Limitações
[Wang e Zhang 2017]	Haar-like + Classificador em Cascata	Classificação / Detecção	Imagens de WTB capturadas em campo	Método clássico e interpretável; baixa robustez a defeitos sutis e forte dependência de pré-processamento manual.
[Peng e Liu 2018]	Processamento geométrico + UAV	Detecção de rachaduras	Imagens de parques eólicos	Eficaz para defeitos extensos; não realiza segmentação pixel a pixel e apresenta baixa sensibilidade a falhas pequenas.
[Shihavuddin et al. 2021]	Faster R-CNN + FPN	Detecção de defeitos	Datasets privados heterogêneos	Bom desempenho em alta resolução; alto custo computacional e pipeline complexo.
[Yu et al. 2020]	CNN + Super-resolução	Classificação	Imagens UAV desfocadas	Mitiga desfoque; não fornece localização precisa nem segmentação.
[Wang et al. 2019] [57]	Janela deslizante + Jaya K-means	Segmentação de rachaduras	Imagens UAV reais	Boa precisão para rachaduras; sensível a ruídos e abordagem multietapas.
[Yu et al. 2023]	YOLOv5 + Atenção	Detecção de defeitos	Imagens de WTB	Melhoria na detecção de pequenos alvos; limitado a bounding boxes.
[Altice et al. 2024]	CNN customizada / Xception	Classificação	Imagens UAV	Alta acurácia; ausência de segmentação e localização precisa.
[Dwivedi et al. 2024]	Vision Transformer	Classificação de danos	Imagens com danos severos	Excelente desempenho para danos evidentes; baixa avaliação em defeitos sutis e sem segmentação.
[Yang et al. 2021]	AlexNet + Random Forest	Classificação	Imagens UAV	Pipeline híbrido eficaz; dependente de segmentação prévia.
[Yu et al. 2023]	U-Net aprimorada	Segmentação semântica	Imagens infravermelhas	Boa segmentação em IR; aplicação restrita a sensores específicos.
<b>Este trabalho</b>	<b>YOLOv8 + CBAM</b>	<b>Segmentação semântica</b>	<b>Blade30 (imagens reais recortadas)</b>	<b>Segmentação pixel a pixel de defeitos pequenos e esparsos, com baixo custo computacional e robustez a desbalanceamento; limitado pela disponibilidade de dados anotados.</b>

---

# Capítulo 4

## Estudos preliminares

---

Nesta seção são apresentados os estudos iniciais conduzidos com o objetivo de analisar o comportamento da arquitetura RTMDet quando submetida a diferentes estratégias de aumento de dados durante o processo de treinamento. O estudo busca compreender como essas técnicas impactam o desempenho do modelo em termos de acurácia e capacidade de generalização, considerando um cenário experimental controlado. A partir dessa análise preliminar, pretende-se identificar configurações de aumento de dados mais adequadas, bem como estabelecer uma base metodológica sólida para os experimentos subsequentes desenvolvidos ao longo deste trabalho.

### 4.0.1 Pré-processamento

Uma vez que o conjunto de dados representa as áreas de interesse como polígonos, uma etapa de pré-processamento foi realizada para transformação do problema para um formato tradicional de detecção de objetos. Os polígonos foram convertidos para os retângulos de menor área que os envolvem (*bounding boxes*).

A grande maioria das imagens do conjunto não possuía anotação, e várias categorias estavam sub-representadas com menos de 10 anotações associadas. Essa situação é um problema para modelos de aprendizado profundo, que apresentam melhor desempenho com grande quantidade de dados [Sun et al. 2017]. Como solução foi desenvolvida uma heurística para conversão do problema original para um problema binário. O número de exemplos por classe aumentaria em troca de uma redução da informação contida nas anotações.

A heurística atribui um valor numérico a cada um dos rótulos de uma anotação. Um rótulo com valor próximo de zero indica um defeito leve, enquanto um valor maior indica defeito grave. A média dos valores dos rótulos representava a "pontuação" da anotação. Uma pontuação de anotação maior do que 1 indicava a classe binária **defect** e uma pontuação menor ou igual a 1 indica a classe binária **ok**.

O processo é ilustrado a seguir:

Rótulos	Valores	Média	Classe Binária
"markings,text"	1, 0	0.5	ok
"markings,crack"	1, 3	2.0	defect

Tabela 2 – Tabela de rótulos e valores adotados para o pré-processamento.

## 4.0.2 Métodos

A metodologia deste estudo compreende em fazer testes controlados para analisar a efetividade de técnicas de *data augmentation*. Um modelo base (*baseline*) de rede neural com uma arquitetura considerada estado da arte, foi treinado e serviu como referência. O modelo base foi comparado a diferentes variações com a mesma arquitetura, mas cada um utilizando um tipo de *data augmentation* durante o treinamento. Uma visão geral do processo é ilustrada na figura 2.

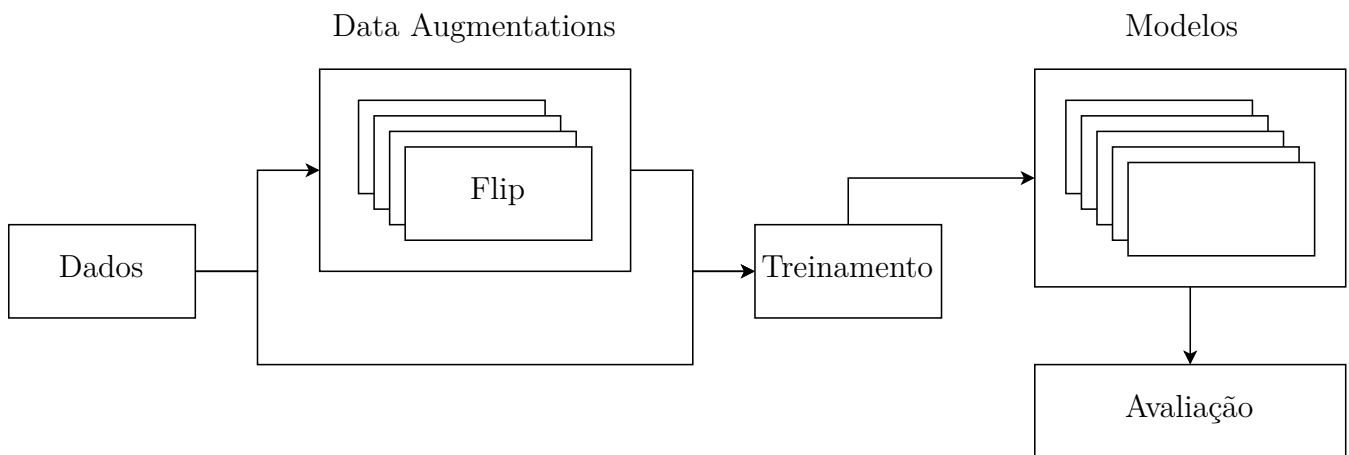


Figura 2 – Metodologia

A arquitetura escolhida foi a RTMDet [Lyu et al. 2022], e para garantir que qualquer divergência nos resultados não seja atribuída a variações no modelo, os hiperparâmetros foram fixados. Em todos os experimentos, o treinamento durou um número fixo de épocas e os resultados reportados foram os melhores valores alcançados de acordo com a métrica COCO mAP [Lin et al. 2015].

Após estudo do conjunto de dados, considerou-se usar *bounding boxes* rotacionadas (figura 4) por representarem mais fielmente áreas de interesse estreitas e inclinadas. Entretanto essa ideia foi descartada devido ao número limitado de arquiteturas de rede neural que suportam essa configuração.

Neste trabalho foi implementado estratégias de *data augmentation* para melhorar métricas sem a necessidade de adicionar imagens extras ao conjunto de dados.

Seguem abaixo as técnicas de *data augmentation* que foram utilizadas neste trabalho:

- Ajustes de Brilho, Contraste e Matiz: Foi utilizada a *data augmentation* da YOLO HSV para os ajustes de brilho, contraste e matiz, sendo uma categoria de transformação eficaz para aprimorar o desempenho do modelo.
  - Brilho: Alterar o brilho de uma imagem implica-se em adicionar ou subtrair um valor constante de todos os pixels. Isso simula as variações naturais na iluminação, tornando



Figura 3 – Exemplo de *bounding box* alinhada aos eixos. Esse tipo de caixa delimitadora mantém seus lados paralelos aos eixos da imagem, sendo amplamente utilizado em abordagens tradicionais de detecção.



Figura 4 – Exemplo de *bounding box* rotacionada. Nesse caso, a caixa delimitadora acompanha a orientação do objeto e foi reduzida para permanecer dentro dos limites da imagem.

o modelo mais robusto a condições de luminosidade distintas.

- Contraste: A variação no contraste é gerada através do ajuste da diferença entre os valores de intensidade dos pixels na imagem. Isso ajuda o modelo a aprender a distinguir objetos mesmo em situações de iluminação diversas.
  - Matiz: A alteração de matiz afeta a tonalidade da imagem, alterando as cores de forma sutil. Essa variação contribui para uma melhor generalização do modelo, tornando-o menos sensível a diferenças específicas de cor.
- Saturação: A saturação está diretamente relacionada com a intensidade das cores contidas numa imagem. Aumentar ou diminuir a saturação pode criar variações visuais significativas nas imagens.

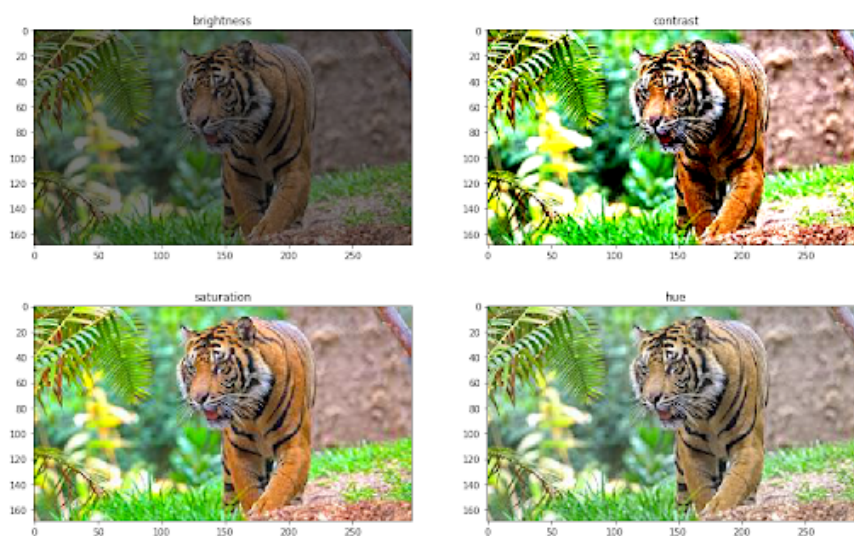


Figura 5 – Exemplo das Técnicas de brilho, contraste, saturação, matiz. Fonte: [v7labs 2024]

- Virar: A imagem pode ser invertida horizontalmente ou verticalmente. Isto produz novas imagens rotacionadas em múltiplos de 90° graus [Khosla e Saini 2020].
- CutOut [DeVries e Taylor 2017]: é uma técnica que compreende em cortar uma parte retangular da imagem original e substituí-la por um valor constante (geralmente zero ou o valor médio da intensidade de pixel na imagem). Essa é uma técnica utilizada para tornar o modelo mais eficaz, forçando-o a focar em diferentes partes da imagem durante o treinamento.
- MixUp [Zhang et al. 2018]: Uma nova amostra é sintetizada sobrepondo uma imagem com outra imagem que é escolhida aleatoriamente a partir dos dados de treinamento [30].
- Mosaico: Ela abrange a criação de uma imagem “mosaico”, sendo o resultado de quatro imagens distintas de mesmo tamanho. Cada uma dessas imagens é escolhida de forma aleatória a partir do conjunto de treinamento original. Esse processo ajuda a diversificar ainda mais o conjunto de dados para o treinamento.
- Combinação: Essa técnica é resultado do conjunto de outras técnicas aplicadas que tiveram as melhores performances [Claro et al. 2020].

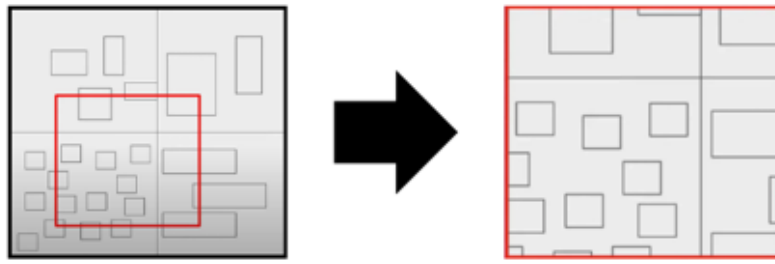


Figura 6 – Exemplo da Técnica Mosaico. Fonte: Roboflow (2020). Mosaic Data Augmentation - Deep Dive [YouTube] [Roboflow 2020].

### 4.0.3 RTMDet

A RTMDet foi selecionada por apresentar desempenho notável no reconhecimento de objetos rotacionados de diferentes tamanhos e na análise de imagens aéreas [Xia et al. 2018].

Na macro arquitetura, são utilizados blocos Cross Stage Partial (CSP) [Wang et al. 2020] com camadas de convolução que utilizam kernels grandes para construir a estrutura da *espinha dorsal*. A *espinha dorsal*, é uma versão modificada da ResNet conhecida como CSPNeXt, tem o propósito de ser leve e eficiente, assim, gerando rápidas inferências sem impactar na precisão de modo negativo. As características de múltiplos níveis, denominadas C3, C4 e C5, são extraídas da *espinha dorsal* e, em seguida, fundidas na etapa do *pescoço*.

O *pescoço* utiliza a Feature Pyramid Network (FPN), que é uma combinação de características com escalas múltiplas, concedendo a detecção eficaz de objetos de tamanhos variados [Lin et al. 2017].

Posteriormente, na etapa final, *cabeça*, representa várias convoluções que preveem simultaneamente a classe e a localização dos objetos.

Esta abordagem de cabeças múltiplas é necessária para manter uma boa precisão na detecção e para a eficiência computacional da rede [Lyu et al. 2022], blocos de detecção com pesos de convolução compartilhados e camadas de normalização em lote são usadas para prever os resultados da classificação e para a detecção das bounding boxes (rotacionadas).

A RTMDet implementa um esquema de ancoragem dinâmica, ajustando automaticamente as âncoras baseando-se na distribuição dos dados de treinamento. Esse esquema melhora a eficiência da detecção, reduzindo falsos positivos e aumentando a precisão [Redmon e Farhadi 2018]. A função de perda é composta por perdas de classificação, localização e correspondência de âncoras, ajustadas para equilibrar esses componentes e melhorar o desempenho global da rede [He et al. 2016].

A rede escolhida como base de comparação para realização dos testes foi a RTMDet [Lyu et al. 2022].

### 4.0.4 Detalhes de implementação

A implementação foi feita utilizando a biblioteca mmdetection [Chen et al. 2019], que permitiu fácil experimentação com vários tamanhos de rede. Os melhores resultados foram alcançados com um modelo RTMDet *large* pré-treinado no dataset COCO [Lin et al. 2015].

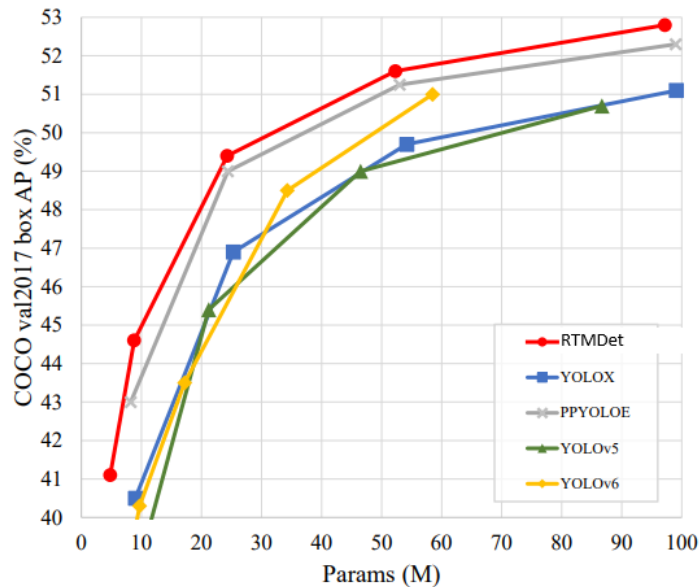


Figura 7 – Comparação da acurácia e de parâmetros entre o RTMDet, destacado em vermelho, e outros detectores de objetos considerados estado da arte na detecção em tempo real [Lyu et al. 2022].

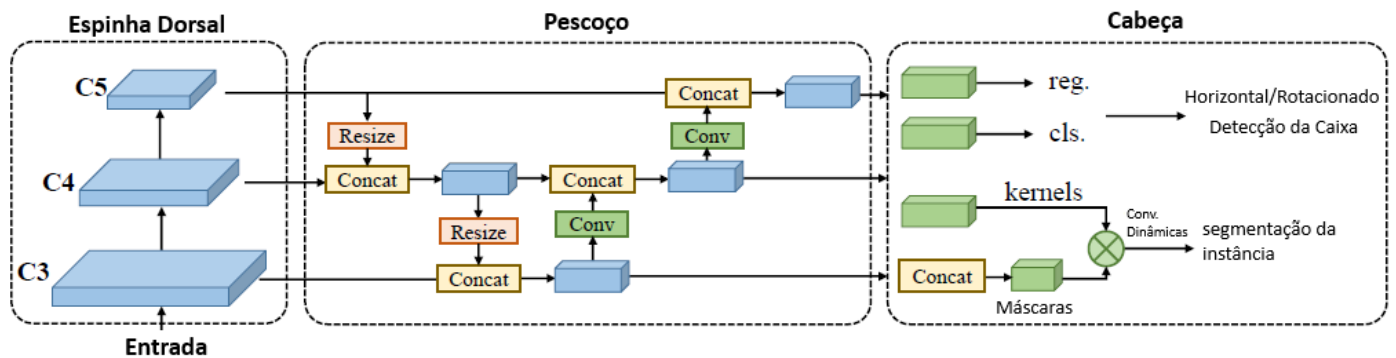


Figura 8 – Macro arquitetura da RTMDet [Lyu et al. 2022].

Hiperparâmetro	Valor
Taxa de aprendizagem inicial	$5 \times 10^{-5}$
Decaimento de TA [Época 40+]	Linear até 1% da inicial
Tamanho do <i>batch</i>	8
Número de épocas	150
Otimizador	AdamW[Loshchilov e Hutter 2019]

Tabela 3 – Hiperparâmetros de treinamento da rede base de comparação

Os dados disponíveis foram particionados, sendo 70% destinados ao treinamento e 30% utilizados na fase de teste. Os hiperparâmetros definidos na tabela 8 foram os responsáveis pelos melhores resultados. O decaimento de taxa de aprendizado (TA) foi aplicado a partir da época 40 até o restante do treinamento, e a TA decaía até 1% do valor inicial. Essa configuração de hiperparâmetros foi fixada em todos os experimentos comparativos. Com essas configurações, o treinamento da rede levava cerca de 3h30m.

### 4.0.5 Aumento de dados

As técnicas de *aumento de dados* foram aplicadas no *pipeline* de treinamento, ou seja, não foi necessária a geração de um conjunto auxiliar com os dados alterados. Para os experimentos referentes a uma técnica, tal técnica era aplicada apenas 50% das vezes. Dessa maneira o modelo era treinado tanto com dados alterados quanto com dados originais.

No contexto de detecção de objetos, as *bounding boxes* devem ser levadas em consideração em qualquer operação que possa alterar as imagens. Esse cuidado foi tomado durante os experimentos. Operações como Virar, CutOut, MixUp e Mosaico, que podem causar alterações nas bounding boxes tinham seus respectivos processos de correção. O caso das técnicas que podem introduzir novas *bounding boxes* em uma imagem (MixUp, Mosaico), também foi considerado e ajustado.

### 4.0.6 Métrica de avaliação escolhida para o estudo inicial

A métrica utilizada para avaliação dos experimentos foi a *Precisão Média* do conjunto de dados COCO [Lin et al. 2015] (COCO mAP). Essa métrica tem várias “versões” que consideram diferentes valores de IoU (Intersection over Union) e objetos de diferentes áreas (pequenas, médias, grandes e todas). Isso fornece uma ideia melhor da performance de um modelo de detecção.

O valor de IoU é calculado entre uma caixa prevista e uma caixa esperada e determina se uma *bounding box* é considerada um acerto. Valores mais altos (próximos a 1) oferecem uma configuração mais difícil, e consequentemente, *bounding boxes* melhores.

A área no contexto do COCO mAP indica o tamanho dos objetos levados em consideração: “pequena” para objetos com área menor que  $32^2$ ; “média” entre  $32^2$  e  $96^2$ ; “grande” apenas áreas maiores que  $96^2$ ; e “todos” indica objetos de qualquer área.

Dentre as “versões” da métrica, a eleita como principal para este trabalho foi a com IoU  $0.50 - 0.95$  para objetos de todas as áreas. Ela foi o critério utilizados para selecionar os modelos que melhor performaram ao longo das 150 épocas de treinamento. A notação  $0.50 - 0.95$  indica que vários limites de valor IoU foram considerados, sendo o resultado final uma média entre eles.

### 4.0.7 Resultados do estudo inicial

Os resultados dos experimentos são sumarizados na tabela 4, que compila o desempenho de todos os modelos treinados no conjunto de teste. Os valores indicados entre parênteses na coluna **Coco mAP (IoU) [Área]** indicam o valor de IoU, e os valores entre colchetes indicam a categoria da área. A combinação mais importante de (IoU)[Área] é a coluna destacada em negrito.

A coluna **Data augmentation** contém as técnicas apresentadas na seção 4.0.2 e o modelo base (*baseline*). A técnica Combinação foi uma composição entre: MixUp, Virar e YOLO HSV (maiores pontuações na categoria  $(0.5 - 0.95)[todas]$ ).

A rede base apresentou desempenho regular nas avaliações gerais e um desempenho fraco tratando-se de objetos com pequena área, coluna  $(0.5 - 0.95)[pequena]$ . Esse resultado é esperado, considerando o baixo volume de dados para uma aplicação de visão computacional e a alta dificuldade do problema de detecção de objetos pequenos. Nenhuma técnica de *data augmentation* foi capaz de sanar satisfatoriamente essa dificuldade.

<i>Data augmentations</i>	COCO mAP (IoU) [Área]					
	(0.5 – 0.95) [todas]	(0.5) [todas]	(0.75) [todas]	(0.5 – 0.95) [pequena]	(0.5 – 0.95) [média]	(0.5 – 0.95) [grande]
<i>baseline</i>	0.434	0.692	0.450	0.000	0.440	0.488
matiz	0.432	0.697	0.443	0.100	0.419	0.480
saturação	0.411	0.642	0.410	0.000	0.445	0.436
brilho	0.429	0.692	0.457	0.100	0.466	0.465
contraste	0.448	0.717	0.457	0.100	0.409	0.521
YOLO HSV	0.448	0.711	0.448	0.000	0.464	0.488
Virar	0.463	0.708	0.483	0.000	0.429	0.532
CutOut	0.448	0.725	0.468	<b>0.200</b>	0.433	0.489
MixUp	<b>0.497</b>	<b>0.728</b>	<b>0.554</b>	<b>0.200</b>	0.454	<b>0.585</b>
Mosaico	0.447	0.696	0.413	<b>0.200</b>	<b>0.484</b>	0.464
Combinação	0.448	0.676	0.467	<b>0.200</b>	0.380	0.532

Tabela 4 – Efeito das técnicas de *data augmentation* no conjunto Blade30 segundo a métrica COCO mAP

Na figura 11 é possível observar a limitação discutida na seção 4.0.2 de *bounding boxes* alinhadas aos eixos: a previsão é correta, mas a maior parte da área prevista não é interessante. A figura 10 mostra previsões em casos de pequenos objetos, e a figura 12 mostra um caso de previsão de região da classe *ok*.

A maior parte das estratégias de **data augmentation** aprimoram o valor da métrica de avaliação, reforçando positivamente a prática. A técnica MixUp gerou um ganho de desempenho de cerca de 15% na versão principal da métrica. Supõe-se que isso foi devido ao modo de funcionamento do MixUp alinhado ao fato de que neste problema não temos dados em abundância: MixUp é uma transformação que pode partir de uma imagem sem anotação e criar uma nova imagem com anotação ao combiná-la com uma imagem com anotação. Além disso não há cenário em que o MixUp elimina *bounding boxes*, coisa que pode acontecer com CutOut e Mosaico.

Outro resultado positivo foi a técnica de Virar, com um ganho próximo de 6%. Suspeita-se que isso foi devido às imagens das pás das turbinas eólicas poderem aparecer em praticamente todas as orientações diferentes, corroborando bem com o objetivo da técnica: tornar o modelo robusto a imagens espelhadas nos mais diferentes sentidos. Uma observação interessante foi a questão dos planos de fundo: aparentemente a existência de algumas imagens de ponta-cabeça (com o céu no lugar do chão) não foi suficiente para atrapalhar o modelo ao ponto de tornar a aplicação de *data augmentation* prejudicial.

As estratégias de ajuste de imagem brilho, saturação e matiz não foram benéficas, alcançando pontuações inferiores ao *baseline*. Isso sugere que não existam muitos casos de, por exemplo, imagens superexpostas ou com saturação excessiva. Contraste e YOLO HSV tiveram contribuição mista, sendo positiva na versão principal da métrica, mas negativa em alguns outros casos.

A técnica chamada Combinação, formada pela composição de outras que tiveram bom desempenho, foi incapaz de superar seus componentes individuais (MixUp, Mosaic, Virar) em quase todos os casos. A única exceção sendo o desempenho superior à da técnica Mosaico na versão (0.75)[*todas*] da métrica. Isso prova que a combinação de estratégias boas não resulta

em uma estratégia boa, mas sugere que essa combinação pode ser usada para compensar alguma deficiência de uma estratégia.



Figura 9 – Inferências da rede *baseline* para um exemplo da classe *defeito*. As regiões detectadas como defeito são indicadas em vermelho.

#### 4.0.8 Conclusões sobre o estudo inicial

De maneira geral, os resultados corroboram com a literatura tradicional ao registrarem um ganho de desempenho na maioria dos casos após a aplicação de técnicas de *data augmentation*. Entretanto, foi observado que é possível escolher técnicas não adequadas para o problema em questão, o que destaca a importância do bom conhecimento do conjunto de dados.

A técnica MixUp produziu resultados excelentes, sugerindo que estratégias capazes de “adicionar anotações” podem beneficiar casos em que elas são escassas, como acontece com o Blade30. As técnicas menos complexas, consistindo em apenas simples transformações nos valores dos pixels das imagens, tiveram resultados mistos que podem ser atribuídos a características específicas do conjunto de dados.

Experimentos com composição de técnicas de *data augmentation* se mostraram capazes de compensar as deficiências de algum componente, mas isso não garantiu uma técnica que funciona bem na maioria dos casos. Ainda sim, a característica de aprimorar deficiências é valiosa e um estudo mais direcionado nessa área pode ser explorado.

Outra futura vertente de pesquisa seria estudar o comportamento das estratégias de *data augmentation* em outros conjuntos de dados diferentes do Blade30. Encontrar e entender situações em que técnicas falham e se sobressaem é de grande valia para pesquisadores e praticantes. Outra



Figura 10 – Inferências da rede *baseline* para outro exemplo avaliado. A classe *defeito* é destacada em vermelho.

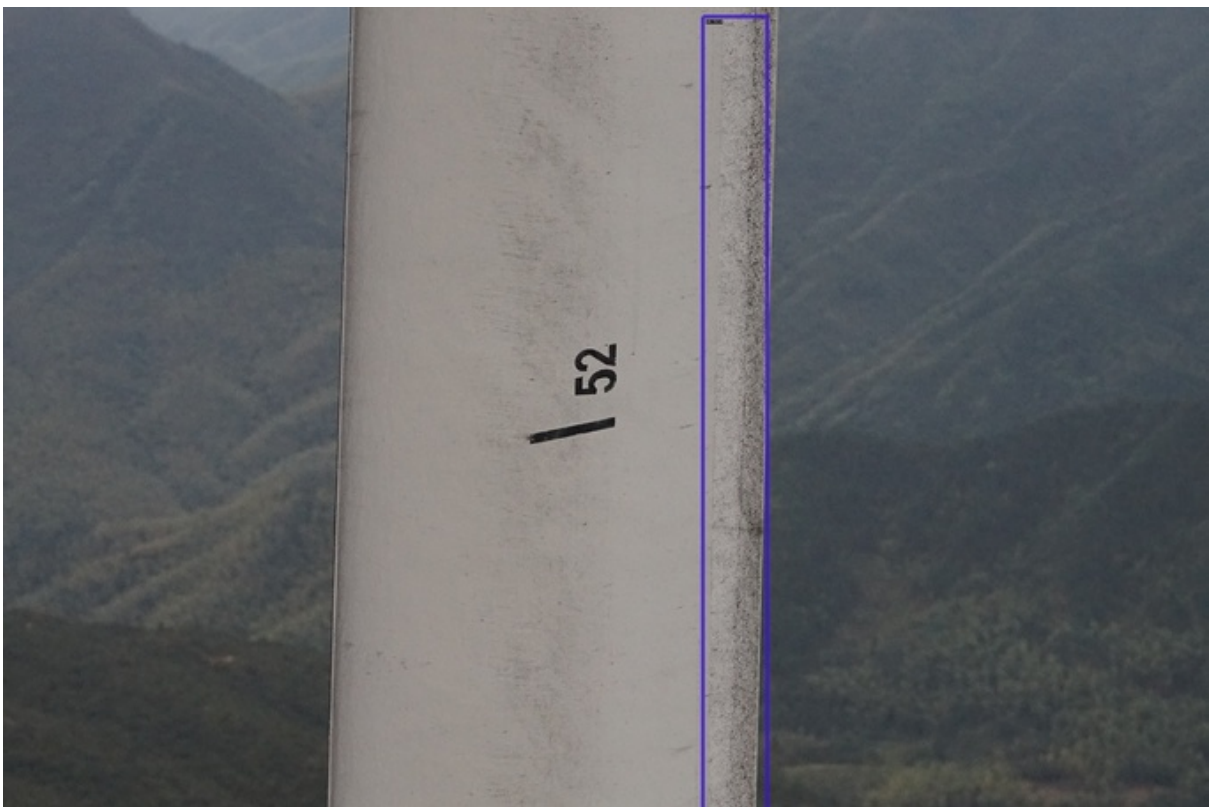


Figura 11 – Inferências da rede *baseline* para outro cenário avaliado. A classe *ok* é apresentada em azul.

possibilidade diferente seria usar o presente trabalho como base de comparação para encontrar soluções melhores para o problema de detecção de defeitos em turbinas eólicas.



---

# Capítulo 5

## Metodologia

---

Neste segmento, discutiremos os detalhes relacionados à execução dos experimentos, especificações do hardware empregado, aspectos da rede neural e métodos utilizados.

### 5.1 Redes de segmentação

A visão computacional teve avanços significativos com o desenvolvimento de técnicas baseadas em redes neurais profundas, possibilitando aplicações robustas em tarefas complexas como detecção, classificação e segmentação de imagens.

Redes de segmentação, em especial, têm-se estabelecido como uma ferramenta indispensável para a extração de informações visuais detalhadas, permitindo a identificação e a delimitação precisa de objetos em imagens [He et al. 2017].

As redes de segmentação consistem em modelos de aprendizado profundo que têm como objetivo atribuir um rótulo específico a cada pixel de uma imagem [He et al. 2017], diferenciando as regiões de interesse.

Uma das principais características das redes de segmentação é a **precisão espacial**, pois essas redes trabalham em nível de pixel, permitindo identificar não apenas a presença de objetos, mas também suas formas e contornos exatos [Chen et al. 2018].

Além disso, essas redes apresentam generalização eficiente [Chen et al. 2018], sendo capazes de operar em cenários complexos e variados, desde ambientes controlados até situações do mundo real com ruídos e variações de escala, iluminação e perspectiva [He et al. 2017].

Modelos como **U-Net** e **DeepLabv3** demonstram excelente desempenho ao combinar características multiescala e técnicas avançadas de decodificação, garantindo a robustez dos resultados [Chen et al. 2018] [46].

Essas redes oferecem escalabilidade, uma vez que podem ser aplicadas em grandes conjuntos de dados com a utilização de um hardware adequado, facilitando a análise em tempo real das imagens.

Além disso, a qualidade dos modelos modernos, como o **Mask R-CNN** [He et al. 2017], garante um equilíbrio entre precisão e eficiência computacional, possibilitando seu uso em dispositivos móveis e embarcados, como drones e câmeras inteligentes [Chen et al. 2018].

No contexto de aprendizado profundo, as redes de segmentação também se beneficiam de técnicas como transferência de aprendizado e otimização de hiperparâmetros, que ajudam a superar desafios relacionados ao treinamento de grandes modelos com conjuntos de dados limitados.

Assim, a combinação de alta precisão, automação e flexibilidade posiciona essas redes como uma das tecnologias mais promissoras para resolver problemas visuais complexos, como o deste trabalho de pesquisa.

Outro ponto relevante é que a técnica de segmentação também facilita o processo de análise subsequente dos defeitos, como a separação das áreas da pá e a criação de máscaras de alta precisão, que posteriormente podem ser utilizadas para categorizar e analisar diferentes tipos de anomalias estruturais.

Com isso, espera-se não apenas identificar a presença de falhas, mas também permitir investigar futuramente detalhes específicos sobre a extensão e o tipo de cada defeito.

## 5.2 Modelo base: YOLOv8

A família de modelos YOLO (*You Only Look Once*) [Redmon et al. 2016] consolidou-se como uma das abordagens mais eficazes para detecção de objetos em tempo real, principalmente por sua arquitetura unificada que realiza o processamento completo da imagem em uma única passagem pela rede. Essa característica torna o modelo extremamente eficiente para inferência, o que é especialmente valioso em aplicações com restrições severas de latência.

A versão denominada YOLOv8, lançada pela Ultralytics, introduziu melhorias arquiteturais significativas e suporte nativo para tarefas de segmentação, o que a torna uma base atrativa para adaptações voltadas à inspeção industrial e outros domínios especializados.

Em sua implementação padrão, a YOLOv8 é composta por três blocos principais: *Backbone*, *Neck* e *Head*. O *Backbone* é responsável pela extração de características hierárquicas a partir das imagens de entrada.

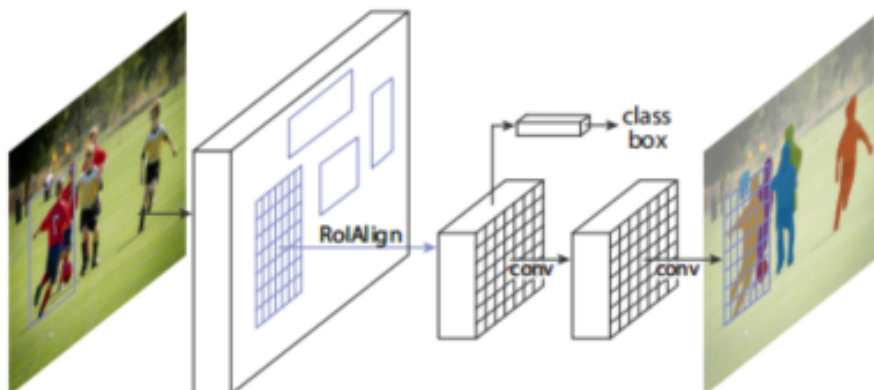


Figura 12 – **Mask R-CNN** para segmentação de instâncias [He et al. 2017].

Essa versão substitui os módulos C3 utilizados na YOLOv5 pelos módulos C2f, que são mais leves e rápidos, proporcionando maior velocidade de inferência sem comprometer a precisão [Reis et al. 2023]. Além disso, o modelo incorpora conexões parciais entre estágios (CSP - *Cross Stage Partial*), originalmente propostas na YOLOv4 [Wang et al. 2020], com o objetivo de reduzir o número de parâmetros da rede e melhorar a reutilização de características, mitigando o problema de desaparecimento de gradientes e facilitando a construção de arquiteturas mais profundas.

O bloco *Neck*, baseado no PANet (*Path Aggregation Network*) [Liu et al. 2018], realiza a fusão de informações multiescalares provenientes de diferentes camadas do *Backbone*. Essa etapa é fundamental para otimizar a localização de objetos em diferentes escalas dentro da imagem.

O módulo SPPF (*Spatial Pyramid Pooling - Fast*) é integrado a esse bloco e contribui para a agregação de informações contextuais a partir de múltiplos campos receptivos, melhorando a capacidade do modelo de lidar com variações no tamanho e forma dos objetos.

Já o *Head* da YOLOv8 é projetado para gerar as previsões finais: caixas delimitadoras (bounding boxes), escores de objetividade e probabilidades das classes. Para isso, utiliza a função de perda *Binary Cross Entropy (BCE)* na classificação, combinada com a *Complete Intersection over Union (CIoU)* e a *Distribution Focal Loss (DFL)* para a regressão das caixas. Essas funções de perda são formuladas da seguinte maneira:

- **Binary Cross Entropy (BCE):** mede o erro entre as probabilidades previstas e os valores reais de cada classe:

$$\text{BCE}(p_s, t_s) = -\frac{1}{N} \sum_{i=1}^N [t_s^i \log(p_s^i) + (1 - t_s^i) \log(1 - p_s^i)] \quad (1)$$

- **Complete IoU (CIoU):** avalia a sobreposição entre as caixas previstas e as reais, penalizando também a distância entre seus centros e a divergência de aspecto:

$$\text{CIoU}(p_b, t_b) = 1 - \text{IoU}(p_b, t_b) + \frac{\rho^2(c_{pb}, c_{tb})}{c^2} + \alpha v \quad (2)$$

- **Distribution Focal Loss (DFL):** melhora a regressão das caixas ao modelar as coordenadas como distribuições discretas, favorecendo localizações mais precisas:

$$\text{DFL}(p_d, t_b) = \sum_{\text{batches}} \sum_{\text{anchors}} \text{softmax}(p_d) \cdot \text{proj} \quad (3)$$

Embora a YOLOv8 tenha sido desenvolvida originalmente para tarefas de detecção de objetos, sua estrutura modular e o suporte a cabeçalhos de segmentação tornam-no um candidato robusto para ser adaptado a tarefas mais complexas, como segmentação semântica ou de instâncias.

Neste trabalho, a YOLOv8 foi empregada como base arquitetural para o desenvolvimento de uma rede de segmentação personalizada voltada especificamente para a detecção de defeitos estruturais em recortes de imagens de pás de turbinas eólicas.

Considerando a necessidade de localizar defeitos pequenos e esparsos com precisão a nível de pixel, a arquitetura tradicional do YOLOv8 para segmentação foi modificada, visando maior leveza e interpretabilidade.

Ao manter os princípios estruturais do modelo original, conseguimos preservar a capacidade de inferência em tempo real, ao mesmo tempo em que introduzimos módulos customizados, como um *Backbone* com atenção baseada em CBAM e uma *Head* de segmentação simplificada, melhor adaptada para análise de imagens de alta resolução.

Essa adaptação mostra como a flexibilidade do YOLOv8 permite que ele vá além das tarefas convencionais de detecção genérica e seja aplicado com sucesso em contextos mais específicos, como a inspeção industrial automatizada e a segmentação de defeitos em componentes estruturais.

### 5.2.1 Justificativa do modelo base

A escolha da arquitetura YOLOv8 como base metodológica deste trabalho não se fundamenta em critérios de novidade, mas sim em critérios de engenharia e adequação ao problema proposto.

O objetivo central desta pesquisa é desenvolver uma abordagem leve, eficiente e aplicável em cenários reais de inspeção automatizada, nos quais a inferência em tempo real é um requisito essencial para operações embarcadas em drones e sistemas.

Diferentemente de arquiteturas amplamente utilizadas em segmentação semântica, como U-Net e DeepLabv3+ e Mask R-CNN, o YOLOv8-Seg apresenta um design unificado que integra, em um único pipeline, as tarefas de detecção e segmentação.

Enquanto modelos clássicos de segmentação produzem máscaras diretamente para toda a imagem, assumindo que os objetos de interesse já estão localizados, o YOLOv8 executa simultaneamente a identificação das regiões relevantes e a geração das respectivas máscaras, dispensando módulos externos de detecção ou heurísticas adicionais para isolar objetos.

Essa característica reduz a fragmentação típica de pipelines que combinam detectores e segmentadores independentes e diminui significativamente a complexidade operacional.

Além disso, por adotar uma abordagem *single-stage*, o YOLOv8 evita as etapas sequenciais e computacionalmente custosas de modelos como o Mask R-CNN, que depende de um gerador de propostas (RPN), de refinamentos sucessivos e de cabeçalhos independentes para classificação, regressão e segmentação.

A arquitetura do YOLOv8 realiza todas essas operações em um único *forward pass*, necessitando apenas de um pós-processamento leve baseado em *Non-Maximum Suppression* (NMS). Na prática, isso resulta em menor latência e menor consumo de memória, permitindo que o modelo opere em ambientes restritos, como dispositivos *edge* ou pipelines industriais de inspeção visual em tempo real.

Outro aspecto relevante é a eficiência do *backbone* e o aproveitamento de recursos multiescala otimizados para lidar com padrões distintos de forma, textura e contexto.

Enquanto U-Net e DeepLabv3+ oferecem excelente desempenho em segmentação densa e homogênea, essas arquiteturas não são projetadas para lidar nativamente com múltiplos objetos, relações espaciais complexas ou cenários em que a detecção é etapa indispensável, como ocorre na inspeção de pás eólicas.

Nesses contextos, o uso de um modelo unificado reduz o acoplamento entre etapas, mitiga erros cumulativos entre detectores e segmentadores e proporciona maior robustez operacional.

Do ponto de vista da implantação, o YOLOv8 também se destaca pela portabilidade. O modelo pode ser exportado para formatos como ONNX, TensorRT e CoreML, o que facilita a integração em sistemas embarcados ou em pipelines híbridos *cloud-edge*. Em contraste, arquiteturas como DeepLabv3+ e Mask R-CNN requerem uma maior capacidade computacional e apresentam uma sobrecarga significativa quando convertidas para mecanismos de inferência otimizados.

Considerando esses fatores, a adoção do YOLOv8-Seg para segmentação em cenários industriais é justificada não apenas por sua acurácia competitiva, mas principalmente pelo equilíbrio entre precisão, velocidade e simplicidade operacional.

A combinação de um pipeline unificado, baixo custo computacional e a capacidade de realizar detecção e segmentação em uma única arquitetura torna o YOLOv8-Seg uma solução particularmente adequada para aplicações práticas que exigem respostas rápidas, escalabilidade e confiabilidade, como ocorre na inspeção automática de defeitos em pás eólicas.

Outro aspecto técnico relevante é o fato de o YOLOv8 adotar um paradigma *anchor-free*, o que elimina a necessidade de caixas delimitadoras pré-definidas e, conseqüentemente, reduz o viés de escala.

Essa propriedade é particularmente importante em cenários de inspeção de pás de aerogeradores, nos quais os defeitos podem variar drasticamente em tamanho, forma e contraste.

Por fim, o YOLOv8 oferece suporte nativo à segmentação semântica, característica que o diferencia das versões anteriores da família YOLO, voltadas apenas para a detecção por *bounding boxes*. Essa capacidade híbrida viabiliza a construção de uma abordagem de segmentação pixel a pixel, preservando a leveza e a velocidade necessárias para aplicações em campo.

A utilização da YOLOv8 neste trabalho não decorre de uma busca por atualidade, mas de coerência com o problema real enfrentado. Diferentemente de arquiteturas mais recentes e pesadas, a YOLOv8 combina eficiência, estabilidade e adaptabilidade a diferentes escalas de defeitos, oferecendo suporte direto à segmentação semântica em tempo real. Além disso, sua estrutura modular favorece a integração de mecanismos de atenção, como o CBAM, sem comprometer o tempo de inferência, algo crítico para inspeções aéreas automatizadas.

Portanto, a escolha pela YOLOv8 é plenamente justificada e tecnicamente sólida, configurando uma contribuição técnica e científica dentro das restrições e necessidades específicas do domínio da inspeção de pás eólicas.

## 5.3 Módulos de atenção

Os mecanismos de atenção têm-se tornado componentes indispensáveis nas arquiteturas modernas de aprendizado profundo, especialmente em tarefas de visão computacional que demandam detecção e segmentação de objetos com alta granularidade.

A principal função desses módulos é permitir que as redes neurais atribuam diferentes níveis de importância a diferentes regiões espaciais ou canais de características, enfatizando dinamicamente as partes mais relevantes da entrada.

Esse foco seletivo facilita a modelagem de dependências contextuais e aprimora a qualidade das representações internas da rede, o que é particularmente crucial em cenários com dados

desbalanceados ou padrões visuais sutis.

No contexto específico da segmentação de defeitos estruturais em pás de turbinas eólicas, os módulos de atenção se mostram uma estratégia promissora para aumentar a sensibilidade da rede a regiões pequenas, esparsas e de baixo contraste, as características frequentemente presentes nesse tipo de aplicação.

O desafio reside em identificar imperfeições que, muitas vezes, ocupam uma fração mínima da imagem e apresentam textura semelhante ao fundo, o que exige do modelo uma capacidade refinada de distinção visual.

Diversos módulos de atenção têm sido propostos na literatura, cada um utilizando abordagens distintas para otimizar o aprendizado de representações mais eficazes. A seguir, são apresentados alguns dos principais mecanismos:

- ❑ **Coordinate Attention:** incorpora informações tanto espaciais quanto de canal por meio de operações de *pooling* unidimensionais ao longo das dimensões espaciais, permitindo à rede codificar melhor a localização dos padrões relevantes.
- ❑ **Shuffle Attention:** realiza uma fusão eficiente entre as dimensões espaciais e de canal utilizando operações de agrupamento e embaralhamento, promovendo uma comunicação mais rica entre diferentes regiões da imagem.
- ❑ **Efficient Channel Attention (ECA):** adota convoluções 1D com tamanhos de kernel adaptativos para capturar interações locais entre canais de forma leve e computacionalmente eficiente.
- ❑ **Convolutional Block Attention Module (CBAM)** [Woo et al. 2018]: o módulo adotado neste trabalho, aplica sequencialmente atenção de canal e espacial de maneira simples e eficiente, destacando seletivamente as informações mais relevantes sem aumentar significativamente o custo computacional.

Apesar de suas diferenças estruturais, esses módulos compartilham o mesmo objetivo: aprimorar as representações intermediárias da rede para potencializar a eficiência do aprendizado e o desempenho em tarefas subsequentes.

Ao melhorar a capacidade da rede de focar nas regiões mais informativas, os mecanismos de atenção tornam-se ferramentas fundamentais para lidar com desafios complexos, como a detecção de padrões mínimos ou ambíguos em imagens reais.

Neste trabalho, optou-se pela utilização do módulo CBAM devido ao seu excelente equilíbrio entre desempenho e custo computacional.

Sua estrutura leve, combinada com sua eficácia na ênfase de regiões críticas tanto no espaço quanto nos canais, o torna especialmente apropriado para modelos de segmentação voltados à implantação em dispositivos com recursos limitados ou aplicações em tempo real.

A adoção do **CBAM** nesta pesquisa fundamenta-se em sua capacidade comprovada de aprimorar a representação de características relevantes com baixo custo computacional. Diferentemente de módulos mais complexos como SE-Block, ECA, GAM ou mecanismos baseados em Transformer, o CBAM atua de forma sequencial e interpretável, aplicando atenção primeiro ao

domínio de canal e, em seguida, ao espaço, o que o torna ideal para integração em arquiteturas leves como o YOLOv8, sem comprometer o tempo de inferência.

Essa abordagem permite refinar o foco da rede em regiões críticas da pá, realçando pequenos defeitos e suprimindo ruídos visuais, especialmente sob condições adversas de iluminação e textura. Assim, a inclusão do CBAM não apenas aprimora a sensibilidade e a robustez da segmentação, mas também mantém a eficiência necessária para aplicações futuras, oferecendo uma contribuição clara e aplicada ao avanço de sistemas de inspeção visual automatizada no domínio eólico.

## 5.4 Método proposto: integração com CBAM

O módulo de atenção CBAM (*Convolutional Block Attention Module*) [Woo et al. 2018] foi integrado em cada bloco convolucional do codificador da arquitetura proposta. Esse módulo é composto por duas subestruturas aplicadas de forma sequencial: atenção de canal e atenção espacial. Cada uma dessas etapas é formulada matematicamente para refinar o mapa de características de maneira direcionada, realçando as regiões mais relevantes, tanto em termos espaciais quanto em profundidade de canal.

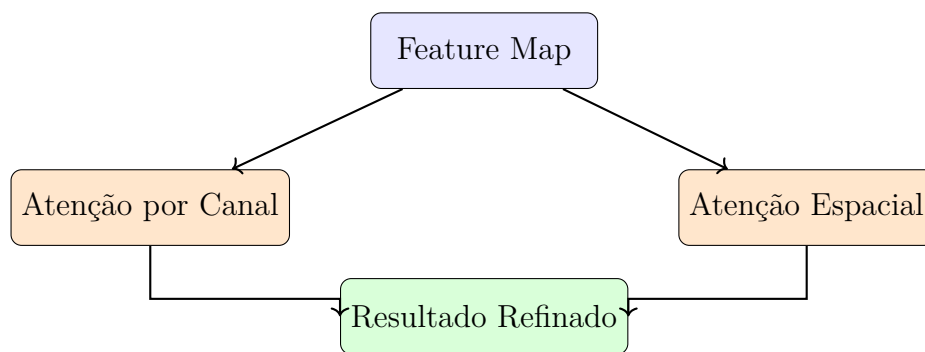


Figura 13 – Representação simplificada das duas etapas do módulo CBAM.

Tabela 5 – Impactos observados após integração do CBAM.

Benefício	Evidência no Experimento
Melhor sensibilidade	Detalhes pequenos foram melhor capturados (ex.: microtrincas).
Separação fundo/defeito	Segmentações mais limpas em cenários de baixo contraste.
Correção de falhas manuais	Modelo superou erros do ground truth em múltiplas amostras.
Estabilidade em ruídos	Resultados mais consistentes em imagens com reflexos.

### 5.4.1 Módulo de atenção de canal

O módulo de atenção de canal tem como principal objetivo destacar os canais mais informativos dentro de um mapa de características intermediário  $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ . Para isso, são aplicadas duas

operações de agregação global ao longo das dimensões espaciais: *average pooling* e *max pooling*, resultando em dois vetores descritivos:

$$\mathbf{F}_{\text{avg}} = \text{AvgPool}(\mathbf{F}), \quad \mathbf{F}_{\text{max}} = \text{MaxPool}(\mathbf{F}) \quad (4)$$

Ambos os vetores são, processados por um mesmo perceptron multicamadas (MLP), cujos pesos são compartilhados, e os resultados são somados:

$$\mathbf{M}_c = \sigma(\text{MLP}(\mathbf{F}_{\text{avg}}) + \text{MLP}(\mathbf{F}_{\text{max}})) \quad (5)$$

Nesta equação,  $\sigma$  representa a função de ativação sigmoide. O resultado é um mapa de atenção de canal  $\mathbf{M}_c \in \mathbb{R}^{C \times 1 \times 1}$ , que é multiplicado ponto a ponto com o mapa de características original, produzindo uma versão refinada:

$$\mathbf{F}' = \mathbf{M}_c \otimes \mathbf{F} \quad (6)$$

### 5.4.2 Módulo de atenção espacial

Após a etapa de atenção de canal, o módulo de atenção espacial tem a função de identificar as regiões mais importantes no espaço da imagem. Com base no mapa refinado  $\mathbf{F}'$ , são aplicadas novamente as operações de *average pooling* e *max pooling*, desta vez ao longo da dimensão de canais:

$$\mathbf{F}'_{\text{avg}} = \text{AvgPool}_{\text{canal}}(\mathbf{F}'), \quad \mathbf{F}'_{\text{max}} = \text{MaxPool}_{\text{canal}}(\mathbf{F}') \quad (7)$$

As saídas dessas duas operações são concatenadas e passam por uma camada convolucional com um kernel de  $7 \times 7$ , resultando no mapa de atenção espacial:

$$\mathbf{M}_s = \sigma(f^{7 \times 7}([\mathbf{F}'_{\text{avg}}; \mathbf{F}'_{\text{max}}])) \quad (8)$$

Por fim, o mapa refinado final é obtido através da multiplicação do mapa de atenção espacial com o mapa de características de entrada desta etapa:

$$\mathbf{F}'' = \mathbf{M}_s \otimes \mathbf{F}' \quad (9)$$

Ao aplicar o CBAM em cada bloco do codificador, a rede torna-se progressivamente mais sensível a regiões e canais relevantes à medida que a profundidade da arquitetura aumenta. Essa propriedade é particularmente valiosa no cenário estudado, onde os defeitos estruturais muitas vezes se apresentam ocultos em áreas visualmente homogêneas, com baixo contraste ou variações mínimas de textura em relação ao fundo.

A integração do CBAM mostrou-se eficaz ao melhorar substancialmente a capacidade do modelo de identificar regiões de baixa visibilidade. Além disso, devido à sua estrutura modular e ao número reduzido de parâmetros adicionais, essa incorporação não acarretou um aumento significativo no tempo de inferência nem no consumo de memória, o que a torna uma solução viável para aplicações embarcadas ou cenários de processamento em tempo real.

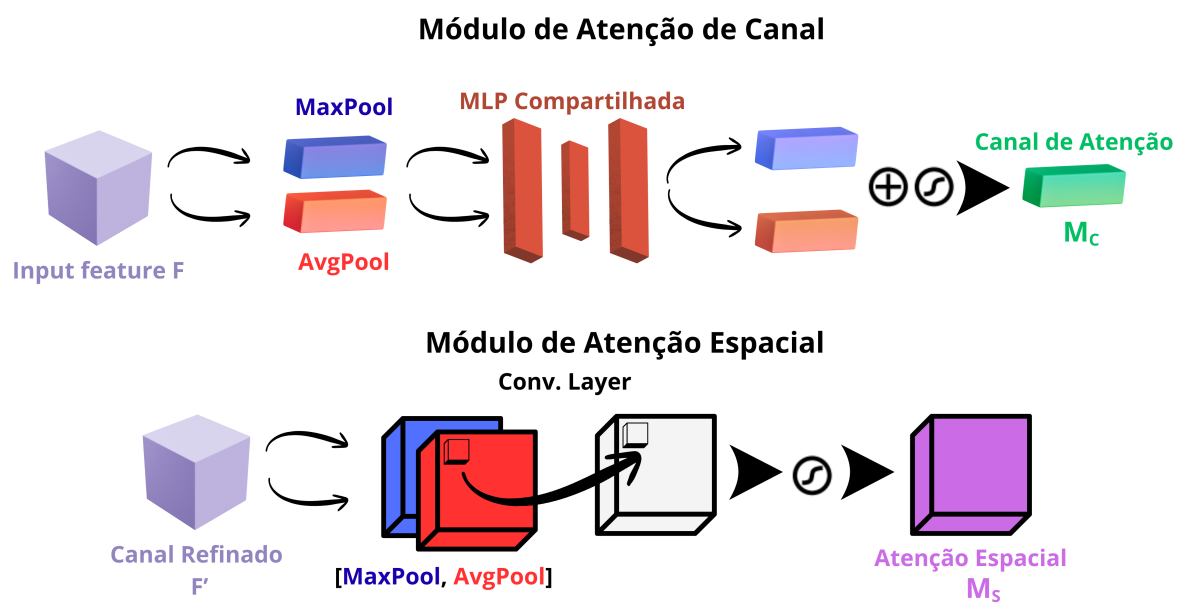


Figura 14 – Diagrama ilustrativo dos módulos de atenção utilizados neste trabalho. A atenção de canal (acima) utiliza *average pooling* e *max pooling* para realçar canais relevantes. A atenção espacial (abaixo) aplica as mesmas operações ao longo dos canais e utiliza uma convolução para gerar o mapa espacial refinado. Este diagrama é adaptado com base em [Woo et al. 2018].



---

# Capítulo 6

## Experimentos

---

Neste segmento, discutiremos os detalhes relacionados à execução dos experimentos, às especificações do hardware empregado, aos aspectos do conjunto de dados utilizado e aos métodos.

### 6.1 Conjuntos de dados: Blade30

O conjunto de dados utilizado é chamado Blade30 [Yang et al. 2023], e consiste em diversas imagens coletadas de diferentes parques eólicos em vários ambientes distintos, como pradarias, terrenos montanhosos e o mar. Os dados são capturados por drones e se originam de 30 pás de turbinas eólicas.

Métricas	Valores
Tamanhos das Pás	19–81,4 m
Cores das Pás	Branco Puro e Vermelho Escuro
Parques Eólicos	Deserto de Gobi, Mar, Céu, Fazenda, Floresta, Terreno Acidentado
Modelos de Drones	DJI M600 e M210
Câmeras	Sony Alpha5100 e Sony Alpha9
Clima	Ensolarado, Nublado, Nevoeiro Leve
Defeitos	No total 156 (12 tipos)
Imagens	Aproximadamente $5400 \times 3600$ de resolução, no total 1302

Tabela 6 – Estatísticas do conjunto de dados Blade30.

A atividade que motivou a criação do conjunto foi a costura de imagens de VANTs, onde o objetivo principal era construir uma imagem contígua de uma pá de turbina eólica a partir de uma sequência de imagens que capturavam diferentes partes.

Os dados são acompanhados de informações como indicações de danos e contaminações presentes nas fotos. As imagens incluem de modo intencional uma variedade de ângulos das pás, fotos com altas resoluções e condições de iluminação diferentes, expondo uma ambientação real.



Figura 15 – Imagens com vários defeitos (parte superior) e contaminações (parte inferior) presentes no conjunto de dados Blade30. Vários defeitos, entre eles, rachaduras na borda, rachaduras na superfície vermelha e na superfície branca causadas por trovões, erosão na borda principal. Contaminações diversas como manchas de óleo, manchas de areia aderida, etc [Yang et al. 2023].

O conjunto de dados [Yang et al. 2023] possui 1302 imagens tiradas por drones de 30 pás de turbinas eólicas. As imagens possuem alta resolução (5456x3632), apresentam formato RGB e as áreas de interesse de algumas imagens são representadas com polígonos associados às fotografias. Uma única área pode conter zero ou múltiplos rótulos, podendo estes variar entre tipos de defeitos (*desgaste, contaminação, rachadura..*) e indicações (*texto, ponta, ok*) formando cerca de 30 categorias.

As turbinas são localizadas em 6 ambientes distintos, envolvendo desertos, mar, fazendas e montanhas, e as fotos foram tiradas em condições de tempo ensolarado, nublado e neblina leve. Em termos práticos, isso se traduz em uma variedade substancial de planos de fundo, o que pode ser um desafio para a detecção.

Por conta da posição natural das pás de um aerogerador, casos em que a pá não se alinha aos eixos vertical e horizontal da imagem são comuns. Isso resulta em casos de pás inclinadas, como pode ser visto na figura 17.

Em suma, para o desenvolvimento deste trabalho, foi utilizado o conjunto de dados público **Blade30**, proposto por Yang et al. [Yang et al. 2023], o qual foi especialmente elaborado para tarefas de inspeção e segmentação de defeitos estruturais em pás de turbinas eólicas.

Este dataset é composto por imagens de alta resolução obtidas em **ambientes reais**, sendo essencial para o treinamento e validação de modelos de segmentação baseados em redes neurais profundas.

### 6.1.1 Estrutura do conjunto de dados

O Blade30 é organizado em 30 subdiretórios, sendo cada um correspondente a uma pá distinta de turbina eólica. Cada subdiretório contém três tipos de arquivos principais:

- ❑ **Imagens originais:** arquivos no formato `.jpg`, com resolução elevada e em cores (RGB), que representam a superfície completa das pás sob diferentes condições de iluminação e textura;
- ❑ **Máscaras binárias:** arquivos no formato `.png` que contêm segmentações por pixel que delimitam a área da pá em cada imagem;



Figura 16 – Imagem do conjunto Blade30 ilustrando uma pá alinhada aos eixos da imagem.



Figura 17 – Imagem do conjunto Blade30 mostrando outro exemplo em que a pá apresenta bom alinhamento em relação aos eixos da imagem.

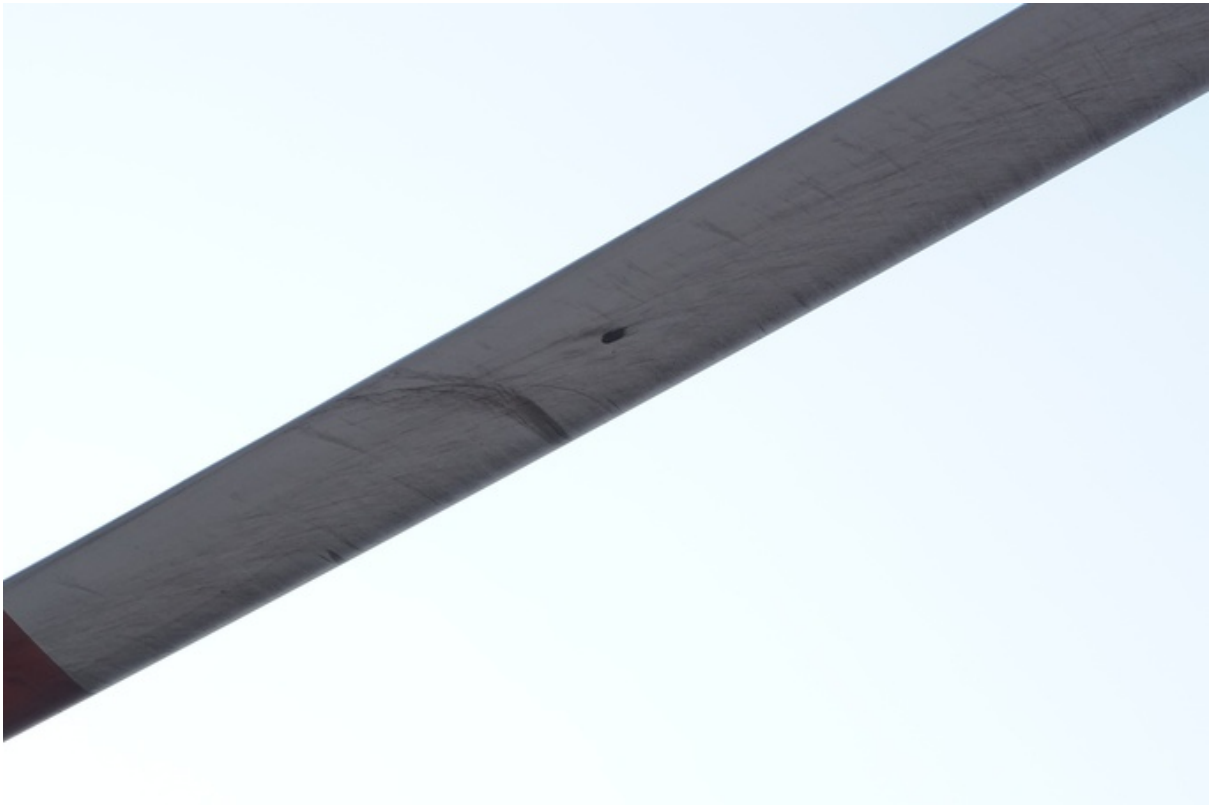


Figura 18 – Imagem do conjunto Blade30 em que a pá não está alinhada aos eixos da imagem, evidenciando maior variação de orientação.



Figura 19 – Imagem do conjunto Blade30 ilustrando outro caso em que a pá apresenta orientação distinta em relação aos eixos da imagem.

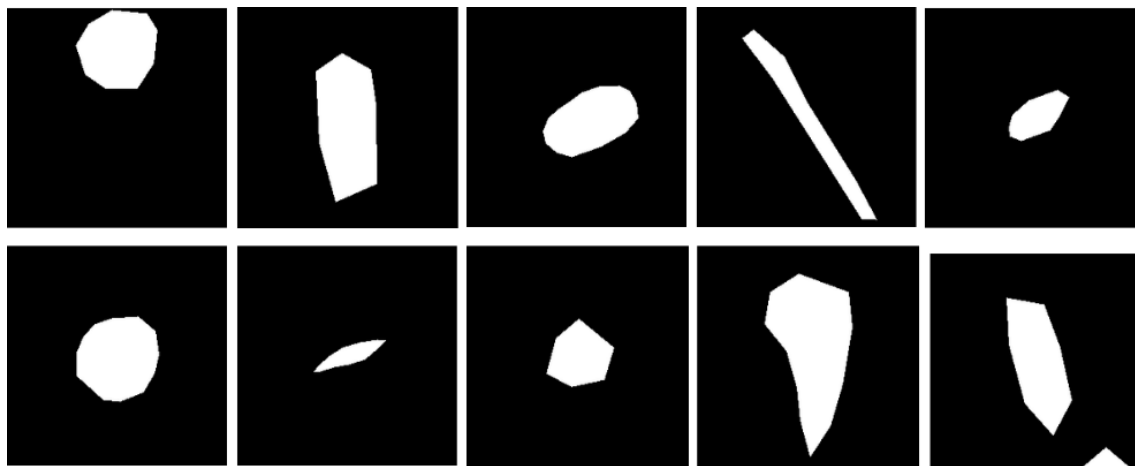


Figura 20 – Exemplo de anotações do dataset convertidas para o formato PNG a partir do arquivo JSON original, evidenciando imprecisões inerentes ao processo de anotação manual, no qual as regiões de falha não apresentam contornos geométricos bem definidos.



Figura 21 – Exemplos de imagens originais em alta resolução do conjunto de dados Blade30, utilizadas na inspeção visual de pás eólicas.

- ❑ **Anotações em JSON:** arquivos no formato COCO (.json), que descrevem por meio de polígonos os defeitos estruturais presentes, como rachaduras, erosões e buracos.

### 6.1.2 Versão recortada: BLADE30

Considerando a alta resolução das imagens originais e a baixa frequência de ocorrência de defeitos, adotou-se uma estratégia de extração de **regiões de interesse (crops)** ao redor dos defeitos anotados.

Esses recortes, com tamanho fixo de  $256 \times 256$  pixels, foram gerados automaticamente a partir das anotações poligonais.

Para cada crop extraído, são fornecidos:

- ❑ A imagem RGB recortada;
- ❑ Uma máscara binária correspondente, alinhada espacialmente;
- ❑ Apenas regiões com anotações de defeitos foram consideradas, sendo descartadas áreas sem anomalias visíveis.

Essa versão refinada do dataset, nomeada **Blade30-Crops**, contém 626 pares imagem-máscara, os quais foram utilizados diretamente no treinamento e validação dos modelos de segmentação.

O foco nas regiões com defeitos permite treinar modelos com maior eficiência, minimizando o viés causado pela grande quantidade de fundo não informativo.

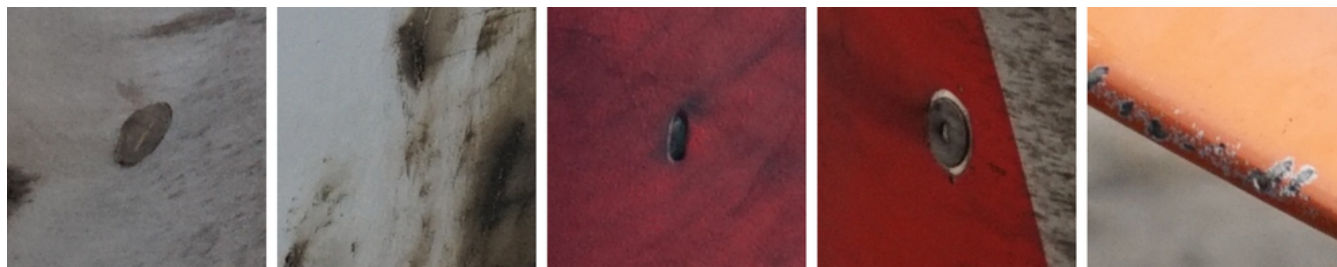


Figura 22 – Exemplos de regiões recortadas (*crops*) de  $256 \times 256$  pixels contendo defeitos estruturais visíveis. As imagens são alinhadas com suas respectivas máscaras binárias.

### 6.1.3 Desafios enfrentados

Apesar de sua utilidade, o Blade30-Crops apresenta alguns desafios:

- ❑ **Desbalanceamento entre classes:** os defeitos ocupam uma área extremamente pequena em comparação ao fundo;
- ❑ **Sutileza visual dos defeitos:** muitos apresentam baixo contraste com a superfície da pá e possuem formas irregulares;
- ❑ **Conversão de anotações:** as anotações em formato COCO (poligonais) precisam ser cuidadosamente transformadas em máscaras binárias, garantindo alinhamento espacial adequado com as imagens de entrada.

Tabela 7 – Principais inconsistências encontradas nas máscaras do Blade30.

Tipo de erro	Descrição
Contornos	Máscaras apresentavam bordas abertas ou imprecisas, especialmente em defeitos pequenos.
Excesso de área	Erosão marcada com área muito maior do que o defeito real.
Omissão de detalhes	Microtrincas ignoradas pelos anotadores humanos.
Variação subjetiva	Anotações com critérios diferentes entre imagens.

Como solução, desenvolveu-se uma pipeline de pré-processamento, incluindo o mapeamento dos polígonos em pixels das anotações, alinhamento espacial das máscaras e imagens e exclusão de amostras vazias ou incorretas.

### 6.1.4 Utilização do dataset no trabalho

Neste trabalho, o subconjunto Blade30-Crops foi utilizado para a tarefa de **segmentação binária** dos defeitos estruturais. O dataset foi dividido em 80% para *treinamento* e 20% para *validação*.

A divisão foi realizada de forma a garantir que imagens oriundas da mesma pá não estivessem em ambos os conjuntos, evitando *vazamento de dados*.

Optou-se por uma aplicação mínima de técnicas de *aumento de dados*, com o objetivo de preservar as características visuais naturais dos defeitos, como texturas e padrões de desgaste, que são essenciais para uma detecção fidedigna em contextos reais.

## 6.2 Infraestrutura experimental

Todos os experimentos foram realizados em um ambiente computacional local, sem uso de serviços em nuvem. A configuração utilizada foi:

- ❑ **Processador:** Intel Core i5 de 13<sup>a</sup> geração;
- ❑ **Memória RAM:** 16 GB;
- ❑ **GPU:** NVIDIA GeForce RTX 3050 com 6 GB de memória;
- ❑ **Sistema operacional:** Windows 11;
- ❑ **Framework:** PyTorch 2.1 com CUDA 11.8.

O código foi implementado em **Python 3.12**, no ambiente **Jupyter Notebook**, utilizando as seguintes bibliotecas:

- ❑ **Torchvision:** para transformações e carregamento de dados;
- ❑ **Matplotlib:** para visualização dos resultados;
- ❑ **Pillow e OpenCV:** para manipulação das imagens e máscaras;
- ❑ **Scikit-learn:** para divisão do dataset e cálculo das métricas.

O modelo de segmentação foi treinado com o otimizador Adam, taxa de aprendizado de  $1 \times 10^{-3}$  e *batch size* de 4 imagens. Os modelos foram treinados por 30 épocas, com seleção baseada no melhor valor de *mIoU* na validação. O tempo de treinamento variou entre 4 e 27 horas, dependendo da complexidade do modelo.

Para garantir a **reprodutibilidade**, utilizaram-se sementes aleatórias fixas e uma pipeline de pré-processamento consistente em todos os experimentos.

Hiperparâmetro	Valor
Taxa de aprendizagem inicial	$1 \times 10^{-3}$
Otimizador	Adam
Tamanho do <i>batch</i>	4
Número de épocas	30
Critério de seleção	Melhor mIoU na validação
Aumento de dados	Mínimo (sem transformações agressivas)
Framework	PyTorch 2.1 + CUDA 11.8

Tabela 8 – Hiperparâmetros e configurações utilizados no treinamento dos modelos de segmentação

## 6.3 Aplicação de aumento de dados

As técnicas de *aumento de dados* foram incorporadas diretamente ao *pipeline* de treinamento, não sendo necessária a criação de um conjunto auxiliar com imagens previamente transformadas.

Em todos os experimentos, as transformações foram aplicadas de forma probabilística, cada técnica foi executada em aproximadamente 50% das vezes.

Com isso, o modelo foi exposto tanto a amostras originais quanto a versões alteradas, o que contribuiu para a generalização do aprendizado.

No contexto de **segmentação**, em que cada pixel da imagem está associado a uma classe (defeito ou sem defeito), é fundamental garantir o alinhamento exato entre as imagens e suas respectivas máscaras binárias ao aplicar qualquer transformação.

Esse cuidado foi rigorosamente seguido neste trabalho. Transformações como espelhamento horizontal, rotação, recorte, adição de ruído e alteração de brilho/contraste foram sempre aplicadas simultaneamente sobre a imagem e a máscara correspondente, mantendo a consistência espacial entre ambas.

Como não há utilização de *caixas delimitadoras* neste estudo, não foram necessárias correções específicas associadas à geometria de caixas delimitadoras, como ocorre em tarefas de detecção de objetos. No entanto, os cuidados com a integridade das regiões de segmentação foram mantidos com o mesmo rigor.

---

# Capítulo 7

## Métricas de avaliação

---

A fim de avaliar o desempenho do modelo proposto YOLOv8+CBAM e compará-lo com outras variantes da arquitetura base, foram adotadas quatro métricas amplamente utilizadas em tarefas de segmentação semântica binária.

Essas métricas fornecem uma análise abrangente tanto da capacidade do modelo em identificar corretamente os pixels pertencentes à classe de interesse (defeito), quanto de sua precisão na exclusão de falsos positivos.

- ❑ **Precisão (Precision)**: Mede a proporção de pixels corretamente classificados como defeito em relação ao total de pixels que o modelo previu como defeito. Trata-se de uma métrica importante para entender o quanto o modelo evita falsos positivos.
- ❑ **Revocação (Recall)**: Avalia a proporção de pixels corretamente identificados como defeito em relação ao total de pixels que realmente pertencem à classe defeito. Essa métrica indica a capacidade do modelo em detectar todos os casos reais de defeito, sendo essencial em aplicações onde perder um defeito é mais grave do que um falso alarme.
- ❑ **mAP@50 (mean Average Precision com limiar IoU de 0,5)**: Refere-se à média da precisão para diferentes níveis de revocação, considerando um limiar fixo de *Intersection over Union* (IoU) de 0,5. Embora mais comumente aplicada em tarefas de detecção de objetos, essa métrica também foi empregada aqui como forma complementar para avaliar a performance global da segmentação.
- ❑ **Mean Intersection over Union (mIoU)**: Considerada a principal métrica em tarefas de segmentação, o mIoU mede a sobreposição entre as regiões segmentadas pelo modelo e as máscaras reais, fornecendo uma avaliação direta da qualidade espacial das predições. Quanto maior o valor de mIoU, melhor o alinhamento entre a segmentação predita e a máscara de referência.

A métrica de mIoU é definida formalmente pela equação:

$$\text{mIoU} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i + FN_i} \quad (10)$$

em que  $C$  representa o número de classes do problema (neste caso, duas: defeito e fundo), e  $TP_i$ ,  $FP_i$  e  $FN_i$  correspondem, respectivamente, aos verdadeiros positivos, falsos positivos e falsos negativos para a classe  $i$ .

Essa abordagem quantitativa permite comparar diferentes modelos de maneira objetiva, identificando quais variações arquiteturais produzem os melhores resultados em termos de fidelidade espacial e capacidade discriminativa na detecção de defeitos estruturais em imagens reais.

### 7.0.1 Comparação de desempenho

A Tabela 9 apresenta os resultados quantitativos obtidos após o treinamento de quatro variantes do modelo YOLOv8 por 30 épocas, com um *batch size* de 8.

Os modelos avaliados incluem as versões leve (YOLOv8N), média (YOLOv8M), grande (YOLOv8L) e a versão proposta neste trabalho (YOLOv8+CBAM), que incorpora o módulo de atenção CBAM ao *backbone* da arquitetura base.

Observa-se que o modelo YOLOv8+CBAM superou consistentemente as demais versões em todas as métricas avaliadas. Especificamente, essa versão apresentou a maior pontuação de mIoU (0,6069), indicando uma segmentação mais precisa das regiões defeituosas nas imagens analisadas.

Além disso, obteve os melhores valores de precisão, revocação e mAP@50, evidenciando sua superioridade tanto na capacidade de localização espacial dos defeitos quanto na classificação binária dos pixels.

Esse desempenho reforça a hipótese de que a introdução de mecanismos de atenção, mesmo em arquiteturas leves, pode resultar em ganhos significativos na qualidade da segmentação, especialmente em cenários com padrões visuais sutis e regiões de interesse pequenas e esparsas.

Tabela 9 – Comparação de desempenho entre variantes do YOLOv8 e YOLOv8 + CBAM na segmentação de defeitos

Modelo	Precisão	Revocação	mAP@50	mIoU
YOLOv8N	0.6506	0.4868	0.5545	0.4987
YOLOv8N+CBAM	0.7024	0.5213	0.5938	0.5176
YOLOv8M	0.8088	0.6486	0.7686	0.5759
YOLOv8M+CBAM	<b>0.8758</b>	<b>0.6695</b>	<b>0.8027</b>	<b>0.6069</b>
YOLOv8L	0.7852	0.5931	0.6725	0.5432
YOLOv8L+CBAM	0.8297	0.6224	0.7142	0.5661

### 7.0.2 Comparação visual

A Figura 23 apresenta um gráfico de barras comparativo entre as quatro variantes do modelo YOLOv8, utilizando como base as métricas de avaliação discutidas anteriormente. O objetivo é

Tabela 10 – Desempenho dos modelos YOLOv8 na tarefa de segmentação (COCO) [Ultralytics 2024].

Modelo	Tamanho (px)	mAP <sub>box</sub> 50–95	mAP <sub>mask</sub> 50–95	Velocidade (CPU ONNX) [ms]	Velocidade (A100 TensorRT) [ms]	Parâmetros (M)	FLOPs (B)
YOLOv8n-seg	640	36.7	30.5	96.1	1.21	3.4	12.6
YOLOv8s-seg	640	44.6	36.8	155.7	1.47	11.8	42.6
YOLOv8m-seg	640	49.9	40.8	317.0	2.18	27.3	110.2
YOLOv8l-seg	640	52.3	42.6	572.4	2.79	46.0	220.5
YOLOv8x-seg	640	53.4	43.4	712.1	4.02	71.8	344.1

oferecer uma visualização clara e intuitiva das diferenças de desempenho entre os modelos após 30 épocas de treinamento, com *batch size* igual a 4.

Pode-se observar que a versão proposta, YOLOv8 com integração do módulo CBAM, apresenta desempenho superior nas métricas de precisão e mIoU, que são fundamentais para segmentações de alta qualidade.

Isso indica que a adição do mecanismo de atenção não apenas melhora a capacidade de identificar regiões com defeitos, mas também reduz a quantidade de predições incorretas (falsos positivos), mantendo a arquitetura leve e eficiente.

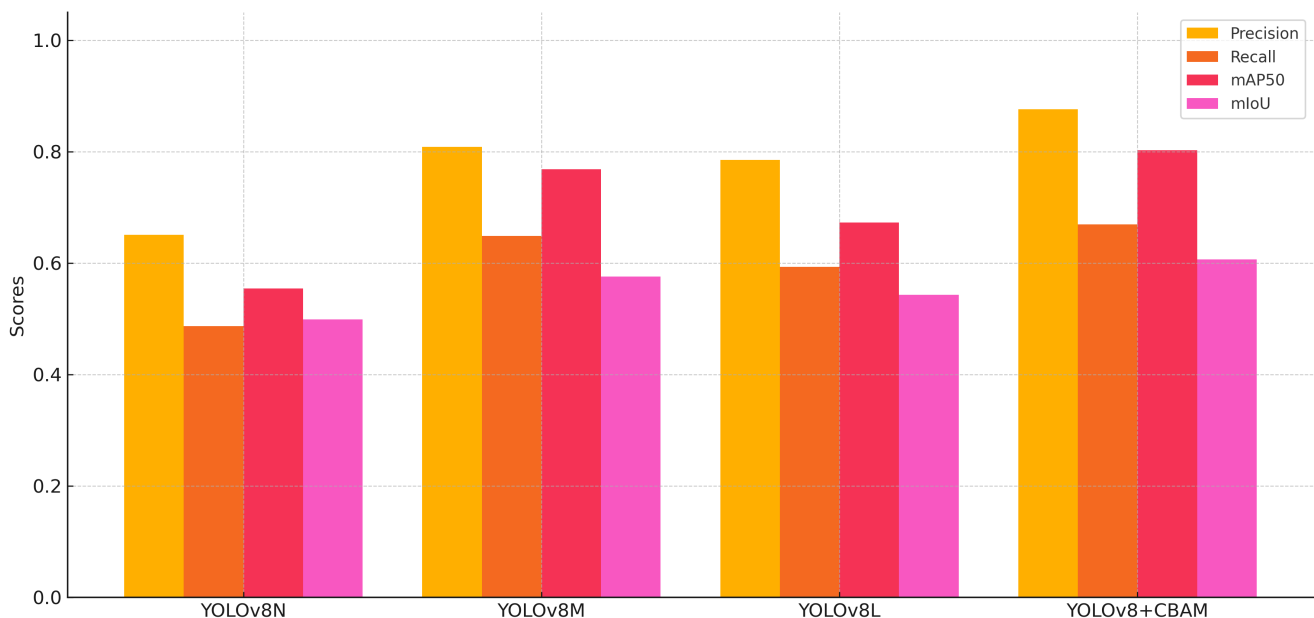


Figura 23 – Comparação entre variantes do modelo YOLOv8 e YOLOv8M+CBAM com base no desempenho de segmentação (30 épocas, *batch size* = 4).

Esses resultados corroboram a hipótese central deste trabalho: a integração de módulos de atenção, como o CBAM, contribui de forma significativa para o aprimoramento da segmentação semântica de defeitos estruturais.

Além disso, essa melhoria é obtida sem comprometer a leveza da rede ou aumentar drasticamente a complexidade computacional, o que a torna especialmente adequada para aplicações em ambientes com restrições de hardware ou exigências de processamento em tempo real.

### 7.0.3 Resultados dos experimentos

Para validar o desempenho do modelo proposto *YOLOv8+CBAM*, foram conduzidos experimentos utilizando regiões recortadas do conjunto de dados Blade30.

O objetivo foi avaliar a capacidade do modelo em segmentar defeitos estruturais em amostras não vistas, simulando cenários reais de inspeção automatizada.

Tabela 11 – Resumo qualitativo das observações sobre as predições.

Cenário	Observações
Defeitos pequenos	O modelo YOLOv8+CBAM apresentou melhor sensibilidade a microtrincas, mesmo onde as máscaras originais apresentavam erros humanos.
Baixo contraste	A atenção espacial do CBAM destacou contornos sutis ignorados pela YOLOv8 pura.
Ruído e reflexos	O modelo proposto filtrou ruído relacionados à luminância, mantendo o contorno estrutural.
Máscaras humanas falhas	Observou-se correção implícita dos erros do ground truth em múltiplos casos.

A Figura 24 apresenta exemplos qualitativos dos resultados obtidos durante a fase de validação. Em cada linha da figura, são mostradas, da esquerda para a direita: a imagem original do recorte, a máscara binária real (*ground truth*), o mapa de probabilidade gerado pela rede e, por fim, a máscara de segmentação binarizada resultante da inferência.

É possível observar que o modelo é capaz de localizar defeitos estruturais com alta precisão, mesmo em situações desafiadoras, como regiões de baixo contraste, bordas suavizadas ou danos de pequena extensão.

Os mapas binarizados foram obtidos a partir da aplicação de um limiar de decisão de 0,3 sobre as probabilidades previstas, valor escolhido por apresentar um bom equilíbrio entre revocação e precisão nas amostras de validação.

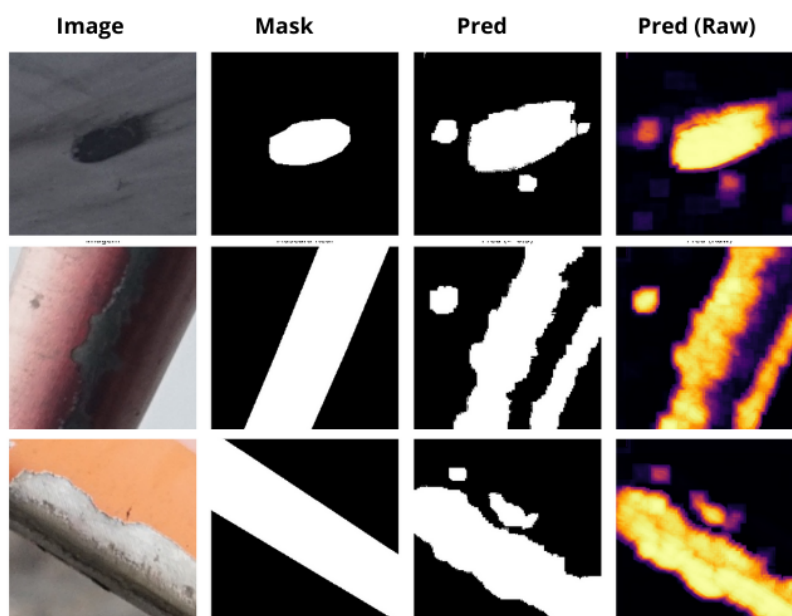


Figura 24 – Exemplos de resultados visuais obtidos pelo modelo *YOLOv8+CBAM*. Da esquerda para a direita: recorte da imagem original, máscara anotada (*ground truth*), máscara segmentada após binarização e mapa de probabilidade produzido pela rede.

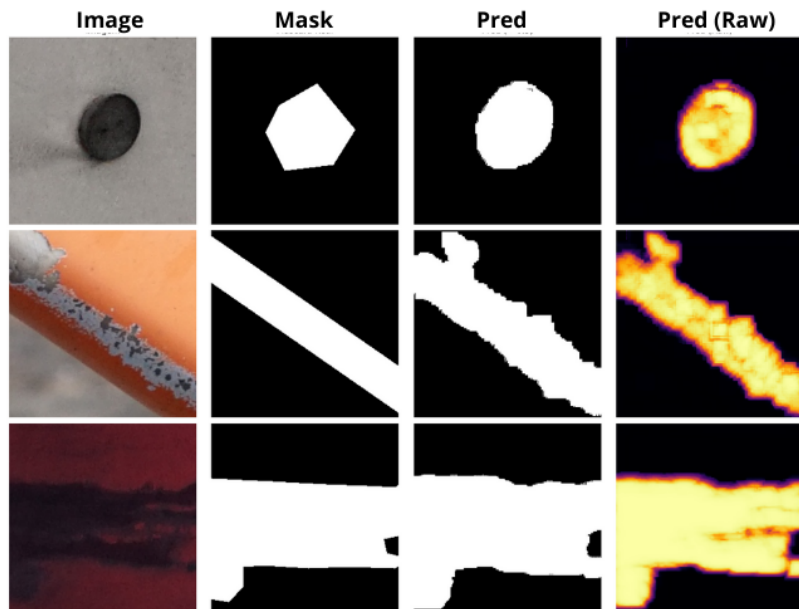


Figura 25 – Exemplos adicionais de resultados visuais obtidos pelo modelo *YOLOv8+CBAM*. Da esquerda para a direita: recorte da imagem original, máscara anotada (*ground truth*), máscara segmentada após binarização e mapa de probabilidade produzido pela rede.

Esses resultados qualitativos complementam as métricas quantitativas discutidas anteriormente e reforçam a robustez do modelo, evidenciando sua capacidade de capturar padrões visuais sutis e segmentar regiões de defeito com alta fidelidade espacial.



---

# Capítulo 8

## Ameaças à validade

---

Como em qualquer estudo experimental envolvendo visão computacional e aprendizado profundo, alguns fatores podem introduzir vieses ou limitar a generalização dos resultados apresentados. A seguir, discutem-se as principais ameaças à validade deste trabalho.

### 8.0.1 Validade interna

A validade interna refere-se à confiança de que os resultados obtidos decorrem exclusivamente do método proposto. Algumas ameaças incluem:

- ❑ **Dependência do pré-processamento:** a estratégia de recortes do Blade30 pode influenciar a maneira como o modelo aprende padrões locais.
- ❑ **Erros nas anotações:** as máscaras do Blade30 apresentam falhas humanas, o que pode introduzir ruído no processo de treinamento.
- ❑ **Variabilidade no treinamento:** diferenças aleatórias de inicialização e seleção de batches podem afetar os resultados.

### 8.0.2 Validade externa

A validade externa diz respeito à capacidade de generalização para outros cenários:

- ❑ **Dataset limitado:** o Blade30 representa apenas um tipo específico de pá eólica; outros materiais ou ambientes podem exigir ajustes.
- ❑ **Condições ambientais variáveis:** iluminação, reflexo, poeira e movimento do drone podem impactar significativamente o desempenho no mundo real.

### 8.0.3 Validade de construto

Relaciona-se à adequação entre o que o estudo pretende medir e o que as métricas realmente medem.

- ❑ O uso de métricas como mIoU e mAP pode não capturar completamente aspectos morfológicos dos defeitos.
- ❑ Máscaras humanas imperfeitas reduzem a fidelidade do ground truth.

Em síntese, embora medidas tenham sido tomadas para mitigar esses riscos, tais limitações devem ser consideradas na interpretação dos resultados e no planejamento de trabalhos futuros.

### 8.0.4 Perspectivas de Trabalhos Futuros e Resultados Promissores

Os resultados obtidos ao longo deste trabalho demonstram que a integração de mecanismos de atenção à arquitetura YOLOv8 constitui uma abordagem eficaz para a segmentação semântica de defeitos estruturais em pás de turbinas eólicas, mesmo em cenários caracterizados por dados escassos, forte desbalanceamento entre classes e presença de defeitos sutis. Ainda assim, diversas possibilidades de extensão e aprofundamento da pesquisa podem ser exploradas, contribuindo para a evolução do método proposto e para sua aplicação em contextos operacionais mais amplos.

Uma primeira direção promissora para trabalhos futuros consiste na quantificação da severidade dos defeitos segmentados. Embora o modelo YOLOv8+CBAM seja capaz de identificar e localizar regiões danificadas em nível de pixel, a estimativa do grau de severidade, extensão ou progressão temporal dos danos pode fornecer informações ainda mais relevantes para estratégias de manutenção preditiva. Métricas derivadas da área segmentada, da morfologia dos defeitos ou da recorrência espacial ao longo do tempo podem ser investigadas como indicadores objetivos de degradação estrutural.

Outra linha de continuidade natural deste trabalho envolve a integração do modelo de segmentação com pipelines de costura e reconstrução de imagens (image mosaicking). Considerando que as inspeções reais por VANTs geram múltiplas imagens parciais da superfície das pás, a combinação entre segmentação semântica e reconstrução panorâmica permitiria a localização precisa dos defeitos ao longo de toda a extensão da pá. Tal integração viabiliza a criação de mapas completos de danos, facilitando análises históricas, comparações entre inspeções sucessivas e apoio à tomada de decisão em manutenção.

Além disso, futuras pesquisas podem investigar a generalização do modelo para diferentes conjuntos de dados, tipos de defeitos e condições de aquisição, incluindo variações mais severas de iluminação, ângulo de captura e resolução espacial. A avaliação do desempenho do YOLOv8+CBAM em novos cenários contribuiria para uma análise mais abrangente de sua robustez e aplicabilidade em ambientes industriais distintos.

Do ponto de vista arquitetural, há espaço para explorar variações na disposição e parametrização dos módulos de atenção, bem como a combinação do CBAM com outros mecanismos de atenção ou estratégias de fusão multiescala. Tais investigações podem aprofundar a compreensão do impacto da

---

atenção espacial e por canal na segmentação de defeitos pequenos e pouco contrastantes, mantendo o compromisso com a leveza computacional do modelo.

Por fim, os resultados alcançados indicam que a abordagem proposta apresenta potencial significativo para aplicações embarcadas e em tempo quase real, especialmente em cenários com restrições de hardware. A adaptação do modelo para execução otimizada em dispositivos embarcados em drones ou estações de processamento em campo representa um passo importante rumo à automação completa do processo de inspeção visual.

Em síntese, o método desenvolvido estabelece uma base sólida para pesquisas futuras na área de inspeção automatizada de pás de turbinas eólicas. Os ganhos observados em desempenho, aliados à eficiência computacional e à precisão da segmentação pixel a pixel, reforçam o caráter promissor da arquitetura YOLOv8+CBAM como componente central de sistemas inteligentes de monitoramento estrutural e manutenção preditiva no setor de energia eólica.



---

# Conclusão

---

Neste trabalho, foi proposto um modelo leve de segmentação semântica denominado *YOLOv8+CBAM*, com o objetivo principal de detectar defeitos estruturais em pás de turbinas eólicas.

A arquitetura desenvolvida baseia-se no modelo YOLOv8, amplamente reconhecido por seu desempenho em tarefas de detecção em tempo real, e foi aprimorada por meio da incorporação do módulo de atenção CBAM (*Convolutional Block Attention Module*) ao seu *backbone*.

Essa integração permitiu que a rede aprendesse a destacar automaticamente regiões espaciais e canais mais relevantes, aumentando sua sensibilidade a padrões sutis e regiões esparsas, características típicas dos defeitos encontrados nas imagens analisadas.

Foram realizados experimentos extensivos utilizando recortes de alta resolução do conjunto de dados Blade30, o que permitiu avaliar a robustez e a eficácia do modelo proposto em um cenário realista de inspeção visual.

Os resultados quantitativos demonstraram que o *YOLOv8+CBAM* alcançou desempenho competitivo em métricas amplamente utilizadas, como mAP@50 e mIoU, superando outras variantes do YOLOv8 em termos de métricas e qualidade de segmentação.

Além disso, as análises qualitativas confirmaram a capacidade do modelo de identificar e delimitar defeitos estruturais mesmo em condições adversas, como imagens com ruído, baixo contraste ou bordas pouco definidas.

Um achado particularmente relevante deste estudo diz respeito à qualidade das anotações originais do conjunto de dados Blade30. Por se tratarem de máscaras geradas manualmente, naturalmente sujeitas a imprecisões humanas, observou-se que algumas delas apresentavam falhas, omissões ou inconsistências nos contornos dos defeitos. De forma surpreendente, o modelo proposto demonstrou capacidade de generalizar além das imperfeições presentes no próprio ground truth, produzindo segmentações mais coesas e precisas que as máscaras originais em diversos casos. Esse comportamento evidencia que o *YOLOv8+CBAM* não apenas aprendeu os padrões relevantes dos defeitos, mas também desenvolveu uma representação robusta o suficiente para corrigir, implicitamente, erros oriundos do processo de anotação humana.

Esse fenômeno ressalta um ponto importante para a área de inspeção automática: modelos de atenção, quando bem calibrados, podem não apenas replicar as anotações humanas, mas superar

limitações intrínsecas ao processo manual, oferecendo um padrão superior de consistência e precisão. Esse achado reforça a relevância de arquiteturas que integram mecanismos de atenção como o CBAM, especialmente em domínios em que a variabilidade visual é alta e as fronteiras dos defeitos não são claramente definidas.

Outro aspecto relevante da proposta é sua eficiência computacional.

A arquitetura foi concebida com foco em cenários de implantação em dispositivos com recursos limitados, como drones e sistemas embarcados, mantendo tempos de inferência reduzidos e consumo de memória moderado, sem comprometer a qualidade da segmentação.

Do ponto de vista prático, esse conjunto de características destaca o potencial do modelo como um componente fundamental em pipelines de inspeção autônoma, nos quais a precisão da segmentação precisa coexistir com restrições de hardware e demanda por processamento em tempo real. A capacidade de operar eficientemente, mesmo sobre dados complexos e ruidosos, reforça sua aplicabilidade em ambientes industriais reais.

Além disso, o desempenho consistente observado ao longo dos experimentos sugere que o *YOLOv8+CBAM* possui boa capacidade de generalização, mesmo quando treinado sob condições adversas e com um conjunto de dados relativamente limitado. Esse ponto é especialmente relevante no contexto da inspeção de pás eólicas, uma vez que a disponibilidade de dados anotados de alta qualidade ainda é um desafio significativo na área.

A metodologia adotada também contribuiu para o avanço do conhecimento na área, demonstrando que abordagens baseadas em recortes de alta resolução podem capturar com maior precisão características estruturais locais dos defeitos, permitindo que a rede aprenda representações mais discriminativas. Esse tipo de abordagem se mostra particularmente útil para cenários nos quais defeitos apresentam bordas finas, extensões irregulares ou contrastes reduzidos em relação ao fundo.

Por outro lado, os resultados obtidos também evidenciam limitações importantes que abrem espaço para investigações futuras. A dependência de recortes artificiais limita, em certa medida, a compreensão contextual global da imagem, o que pode afetar a detecção de defeitos de grandes extensões ou com padrões que dependem de relações espaciais mais amplas. Nesse sentido, abordagens multi-escala surgem como uma alternativa promissora para permitir que o modelo capture tanto informações locais quanto globais de maneira integrada.

Como trabalhos futuros, sugere-se a exploração de estratégias de entrada com múltiplas escalas, o que pode aprimorar ainda mais a detecção de defeitos com diferentes proporções e localizações.

Também se considera promissora a integração de informações temporais provenientes de vídeos, visando aplicações em inspeção contínua com drones em movimento.

Adicionalmente, o uso de aprendizado por transferência a partir de conjuntos de dados sintéticos pode ampliar a generalização do modelo em diferentes cenários industriais.

A expansão da base de dados, seja por meio de novas anotações reais ou pela geração de imagens sintéticas com realismo aumentado, pode impulsionar ainda mais o desempenho da arquitetura proposta.

Por fim, a implantação prática do modelo em sistemas embarcados representa um próximo passo natural, permitindo a realização de inspeções em tempo real com autonomia e precisão elevadas.

A consolidação desse pipeline em um ambiente industrial contribuirá não apenas para a validação externa da abordagem, mas também para o avanço de tecnologias autônomas de inspeção, reduzindo custos operacionais, aumentando a segurança e acelerando a detecção precoce de falhas estruturais em turbinas eólicas.

**Folha de Aprovação**

Defesa de dissertação de mestrado do(a) candidato(a) Stevan Henrique Ramon de Góes, realizada em 15/12/25

**Comissão Julgadora**

Prof(a) Dr(a) Hermes Senger (UFSCar)

Prof(a) Dr(a) Jurandy Gomes de Almeida Junior (UFSCar)

Prof(a) Dr(a) Renato Moraes Silva (USP)

O relatório de defesa assinado pelos membros da comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Computação

---

# Referências

---

ALTICE, B. et al. Anomaly detection on small wind turbine blades using deep learning algorithms. **Energies**, v. 17, n. 5, 2024. ISSN 1996-1073. Disponível em: <<https://www.mdpi.com/1996-1073/17/5/982>>.

BELLO, I. et al. **Revisiting ResNets: Improved Training and Scaling Strategies**. 2021.

BYEON, H. et al. Deep learning model to detect deceptive generative adversarial network generated images using multimedia forensic. **Computers and Electrical Engineering**, v. 113, p. 109024, 2024. ISSN 0045-7906. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790623004482>>.

CASTRILLÓN, M. et al. A comparison of face and facial feature detectors based on the viola-jones general object detection framework. **Machine Vision and Applications**, Springer, v. 22, p. 481–494, 2011.

CHANDRASEKHAR, K. et al. Damage detection in operational wind turbine blades using a new approach based on machine learning. **Renewable Energy**, Elsevier Ltd, v. 168, p. 1249–1264, 5 2021. ISSN 18790682.

CHEN, K. et al. **MMDetection: Open MMLab Detection Toolbox and Benchmark**. 2019.

CHEN, L.-C. et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 40, n. 4, p. 834–848, 2018.

CHEN, Y. et al. Welding defect detection based on phased array images and two-stage segmentation strategy. **Advanced Engineering Informatics**, Elsevier, v. 62, p. 102879, 2024.

CLARO, M. et al. Utilização de técnicas de data augmentation em imagens: Teoria e prática. **Sociedade Brasileira de Computação**, 2020.

DERVILIS, N. et al. On damage diagnosis for a wind turbine blade using pattern recognition. **Journal of Sound and Vibration**, v. 333, n. 6, p. 1833–1850, 2014. ISSN 0022-460X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022460X13009528>>.

DEVRIES, T.; TAYLOR, G. W. **Improved Regularization of Convolutional Neural Networks with Cutout**. 2017.

DWIVEDI, D. et al. Identification of surface defects on solar pv panels and wind turbine blades using attention based deep learning model. **Engineering Applications of Artificial Intelligence**, v. 131, p. 107836, 2024. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197623020201>>.

- FU, Z. et al. Le-yolo: Lightweight and efficient detection model for wind turbine blade defects based on improved yolo. **IEEE Access**, v. 12, p. 135985–135998, 2024.
- García Márquez, F. P. et al. Condition monitoring of wind turbines: Techniques and methods. **Renewable Energy**, v. 46, p. 169–178, 2012. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960148112001899>>.
- GUO, J. et al. Damage identification of wind turbine blades with deep convolutional neural networks. **Renewable Energy**, v. 174, p. 122–133, 2021. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960148121005589>>.
- HANG, X. et al. Study on crack monitoring method of wind turbine blade based on ai model: Integration of classification, detection, segmentation and fault level evaluation. **Renewable Energy**, v. 224, p. 120152, 2024. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0960148124002179>>.
- HE, K. et al. Mask r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2961–2969.
- \_\_\_\_\_. Deep residual learning for image recognition. June 2016.
- HOSSEIN-NEJAD, Z.; NASRI, M. An adaptive image registration method based on sift features and ransac transform. **Computers Electrical Engineering**, v. 62, p. 524–537, 2017. ISSN 0045-7906. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S004579061630917X>>.
- HU, H. et al. Gam: Gradient attention module of optimization for point clouds analysis. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2023. v. 37, n. 1, p. 835–843.
- İNCE, S. et al. U-net-based models for precise brain stroke segmentation. **Chaos Theory and Applications**, Akif AKGÜL, v. 7, n. 1, p. 50–60.
- KHOSLA, C.; SAINI, B. S. Enhancing performance of deep learning models with different data augmentation techniques: A survey. In: **2020 International Conference on Intelligent Engineering and Management (ICIEM)**. [S.l.: s.n.], 2020. p. 79–85.
- LI, Z. et al. 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.
- LIN, T.-Y. et al. Feature pyramid networks for object detection. p. 2117–2125, 2017.
- \_\_\_\_\_. **Microsoft COCO: Common Objects in Context**. 2015.
- LIU, S. et al. Path aggregation network for instance segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 8759–8768.
- LIU, W. et al. The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 44, p. 466–472, 2015.
- LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization**. 2019.
- LYU, C. et al. **RTMDet: An Empirical Study of Designing Real-Time Object Detectors**. 2022.

- MAHARANA, K.; MONDAL, S.; NEMADE, B. A review: Data pre-processing and data augmentation techniques. **Global Transitions Proceedings**, v. 3, n. 1, p. 91–99, 2022. ISSN 2666-285X. International Conference on Intelligent Engineering Approach(ICIEA-2022). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666285X22000565>>.
- MARIN, J. et al. Study of fatigue damage in wind turbine blades. **Engineering failure analysis**, Elsevier, v. 16, n. 2, p. 656–668, 2009.
- MELNYCHENKO, O.; SAVENKO, O.; RADIUK, P. Apple detection with occlusions using modified yolov5-v1. In: **2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)**. [S.l.: s.n.], 2023. v. 1, p. 107–112.
- NABIZADEH, N.; KUBAT, M. Brain tumors detection and segmentation in mr images: Gabor wavelet vs. statistical features. **Computers Electrical Engineering**, v. 45, p. 286–301, 2015. ISSN 0045-7906. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790615000324>>.
- NUR-A-ALAM et al. An intelligent system for automatic fingerprint identification using feature fusion by gabor filter and deep learning. **Computers and Electrical Engineering**, v. 95, p. 107387, 2021. ISSN 0045-7906. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790621003554>>.
- PENG, L.; LIU, J. Detection and analysis of large-scale wt blade surface cracks based on uav-taken images. **IET image processing**, Wiley Online Library, v. 12, n. 11, p. 2059–2064, 2018.
- PENG, Y. et al. Non-uniform illumination image enhancement for surface damage detection of wind turbine blades. **Mechanical Systems and Signal Processing**, v. 170, p. 108797, 2022. ISSN 0888-3270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0888327021011080>>.
- QADRI, S. A. A. et al. Plant disease detection and segmentation using end-to-end yolov8: A comprehensive approach. In: **2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE)**. [S.l.: s.n.], 2023. p. 155–160.
- RAN, X. et al. An improved algorithm for wind turbine blade defect detection. **IEEE Access**, v. 10, p. 122171–122181, 2022.
- REDDY, A. et al. Detection of cracks and damage in wind turbine blades using artificial intelligence-based image analytics. **Measurement**, v. 147, p. 106823, 2019. ISSN 0263-2241. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0263224119306803>>.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 779–788.
- REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. **arXiv preprint arXiv:1804.02767**, 2018.
- REIS, D. et al. Real-time flying object detection with yolov8. **arXiv preprint arXiv:2305.09972**, 2023.
- RIBRANT, J.; BERTLING, L. Survey of failures in wind power systems with focus on swedish wind power plants during 1997-2005. In: **2007 IEEE Power Engineering Society General Meeting**. [S.l.: s.n.], 2007. p. 1–8.

RIZVI, S. Z.; JAMIL, M.; HUANG, W. Enhanced defect detection on wind turbine blades using binary segmentation masks and yolo. **Computers and Electrical Engineering**, v. 120, p. 109615, 2024. ISSN 0045-7906. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790624005421>>.

ROBOFLOW. **Mosaic Data Augmentation - Deep Dive**. 2020. Disponível em: <[https://www.youtube.com/watch?v=V6uj-eGmE7g&ab\\_channel=Roboflow](https://www.youtube.com/watch?v=V6uj-eGmE7g&ab_channel=Roboflow)>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18**. [S.l.], 2015. p. 234–241.

RUIZ, M. et al. Wind turbine fault detection and classification by means of image texture analysis. **Mechanical Systems and Signal Processing**, v. 107, p. 149–167, 2018. ISSN 0888-3270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0888327017306738>>.

SHI, Y.; LI, X.; CHEN, M. Sc-yolo: A object detection model for small traffic signs. **IEEE Access**, IEEE, v. 11, p. 11500–11510, 2023.

SHIHAVUDDIN, A. et al. Image based surface damage detection of renewable energy installations using a unified deep learning approach. **Energy Reports**, v. 7, p. 4566–4576, 2021. ISSN 2352-4847. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352484721005102>>.

SUN, C. et al. **Revisiting Unreasonable Effectiveness of Data in Deep Learning Era**. 2017.

TCHAKOUA, P. et al. Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. **Energies**, Molecular Diversity Preservation International (MDPI), v. 7, n. 4, p. 2595–2630, 2014.

ULTRALYTICS. **YOLOv8 Performance Metrics**. 2024. Acesso em: 12 out. 2025. Disponível em: <<https://docs.ultralytics.com/pt/models/yolov8/#performance-metrics>>.

V7LABS. **v7labs**. 2024. Disponível em: <<https://www.v7labs.com>>.

WANG, C.-Y. et al. Cspnet: A new backbone that can enhance learning capability of cnn. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops**. [S.l.: s.n.], 2020. p. 390–391.

WANG, L. et al. An improved u-net model for segmenting wind turbines from uav-taken images. **IEEE Sensors Letters**, v. 6, n. 7, p. 1–4, 2022.

WANG, L.; ZHANG, Z. Automatic detection of wind turbine blade surface cracks based on uav-taken images. **IEEE Transactions on Industrial Electronics**, v. 64, n. 9, p. 7293–7303, 2017.

WANG, L.; ZHANG, Z.; LUO, X. A two-stage data-driven approach for image-based wind turbine blade crack inspections. **IEEE/ASME transactions on mechatronics**, IEEE, v. 24, n. 3, p. 1271–1281, 2019.

\_\_\_\_\_. A two-stage data-driven approach for image-based wind turbine blade crack inspections. **IEEE/ASME Transactions on Mechatronics**, v. 24, n. 3, p. 1271–1281, 2019.

\_\_\_\_\_. A two-stage data-driven approach for image-based wind turbine blade crack inspections. **IEEE/ASME Transactions on Mechatronics**, v. 24, n. 3, p. 1271–1281, 2019.

- WANG, Q. et al. Eca-net: Efficient channel attention for deep convolutional neural networks. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2020. p. 11534–11542.
- WIGHTMAN, R.; TOUVRON, H.; JÉGOU, H. **ResNet strikes back: An improved training procedure in timm**. 2021.
- WOO, S. et al. Cbam: Convolutional block attention module. In: **Proceedings of the European conference on computer vision (ECCV)**. [S.l.: s.n.], 2018. p. 3–19.
- XIA, G.-S. et al. Dota: A large-scale dataset for object detection in aerial images. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 3974–3983.
- XUE, Y. et al. Deep learning-based automatic recognition of water leakage area in shield tunnel lining. **Tunnelling and Underground Space Technology**, Elsevier, v. 104, p. 103524, 2020.
- YANG, C. et al. Towards accurate image stitching for drone-based wind turbine blade inspection. **Renewable Energy**, Elsevier Ltd, v. 203, p. 267–279, 2 2023. ISSN 18790682.
- YANG, X. et al. Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. **Renewable Energy**, Elsevier Ltd, v. 163, p. 386–397, 1 2021. ISSN 18790682.
- YU, J. et al. An improved u-net model for infrared image segmentation of wind turbine blade. **IEEE Sensors Journal**, v. 23, n. 2, p. 1318–1327, 2023.
- \_\_\_\_\_. An improved u-net model for infrared image segmentation of wind turbine blade. **IEEE Sensors Journal**, v. 23, n. 2, p. 1318–1327, 2023.
- \_\_\_\_\_. An infrared image stitching method for wind turbine blade using uav flight data and u-net. **IEEE Sensors Journal**, v. 23, n. 8, p. 8727–8736, 2023.
- YU, Y. et al. Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor. **Neurocomputing**, v. 376, p. 1–9, 2020. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231219313396>>.
- ZHANG, C. et al. Automated detection and segmentation of tunnel defects and objects using yolov8-cm. **Tunnelling and Underground Space Technology**, v. 150, p. 105857, 2024. ISSN 0886-7798. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S088677982400275X>>.
- ZHANG, C.; YANG, T.; YANG, J. Image recognition of wind turbine blade defects using attention-based mobilenetv1-yolov4 and transfer learning. **Sensors**, MDPI, v. 22, n. 16, p. 6009, 2022.
- ZHANG, D. et al. A data-driven design for fault detection of wind turbines using random forests and xgboost. **IEEE Access**, v. 6, p. 21020–21031, 2018.
- ZHANG, H. et al. **mixup: Beyond Empirical Risk Minimization**. 2018.
- ZHANG, Q.-L.; YANG, Y.-B. Sa-net: Shuffle attention for deep convolutional neural networks. In: **IEEE ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2021. p. 2235–2239.
- ZHANG, R.; WEN, C. Sod-yolo: A small target defect detection algorithm for wind turbine blades based on improved yolov5. **Advanced Theory and Simulations**, v. 5, n. 7, p. 2100631, 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/adts.202100631>>.

ZHANG, Z. et al. Gcb-yolo: A lightweight algorithm for wind turbine blade defect detection. **Wind Energy**, Wiley Online Library, v. 28, n. 6, p. e70029, 2025.

ZHAO, L. et al. Cross-supervised contrastive learning domain adaptation network for steel defect segmentation. **Advanced Engineering Informatics**, Elsevier, v. 64, p. 102964, 2025.

ZHOU, W. et al. Wind turbine actual defects detection based on visible and infrared image fusion. **IEEE Transactions on Instrumentation and Measurement**, v. 72, p. 1–8, 2023.