

# Visão Computacional e Aprendizado Profundo Semissupervisionado para Detecção e Medição de Trincas em Estruturas de Aço

## 1 Introdução

A nucleação e a propagação de trincas em estruturas de aço submetidas a carregamentos cíclicos representam o modo de falha predominante em máquinas, componentes mecânicos e construções metálicas, constituindo cerca de 90% das falhas em serviço [1]. Nesses casos, estima-se que 20-40% da vida útil dos componentes seja consumida na fase denominada crescimento estável da trinca [2], estágio em que o aumento do dano possui maior previsibilidade e dimensões detectáveis, tornando imperativo o uso de técnicas para identificar e monitorar esse tipo de defeito.

Para isso, os processos de SHM (*Structural Health Monitoring*) desempenham um papel fundamental, pois contemplam atividades de acompanhamento constante da integridade estrutural. No entanto, o SHM ainda é realizado, em grande parte, de forma manual, através de inspeções visuais ou a partir de ensaios não destrutivos. Esses métodos oferecem baixa escalabilidade, além de estarem suscetíveis a erros humanos e à dificuldade no acesso de alguns locais, comprometendo a precisão e a eficácia desses procedimentos.

O uso de técnicas de visão computacional como uma alternativa para o SHM na detecção de trincas vem se destacando nos últimos anos, principalmente no acompanhamento de estruturas de concreto como pontes, viadutos, represas e pavimentos rodoviários [3]. Porém, o monitoramento de estruturas de aço apresenta obstáculos específicos, como apontado por Li, Zhu e Huang [4]:

- **Variabilidade ambiental:** Condições precárias de iluminação, reflexos e texturas metálicas diversas (presença de corrosão ou camadas de tinta);
- **Variabilidade geométrica:** Geometrias complexas, presença de soldas e/ou ligações parafusadas;
- **Escassez de dados:** *Datasets* públicos e rotulados para trincas em aço são limitados em comparação a estruturas de concreto.

Além da identificação das trincas, é de suma importância a medição de suas propriedades geométricas, como comprimento, largura e orientação, uma vez que essas informações constituem os *inputs* de modelos empíricos para predição de vida remanescente [5]. Esses modelos são extremamente sensíveis aos parâmetros geométricos da trinca, o que pode aumentar as incertezas no cálculo e comprometer a confiabilidade da predição. Uma alternativa para mitigar esses erros consiste na implementação de um sistema de monitoramento automatizado, utilizando redes neurais convolucionais profundas (*Deep Convolutional Neural Networks*, D-CNNs) de segmentação semântica, complementadas por etapas de pós-processamento baseadas em operações morfológicas para obtenção das características geométricas das trincas a partir dos mapas semânticos [3].

Este projeto tem como objetivo desenvolver um modelo de visão computacional e aprendizado profundo para detecção e quantificação de trincas em estruturas metálicas. A proposta envolve a combinação de dados públicos e rotulados provenientes de outros tipos de superfícies e dados não rotulados de estruturas de aço, utilizando técnicas de

aprendizado semissupervisionado. Espera-se que o modelo seja versátil e eficiente, permitindo sua implementação em câmeras industriais, UAVs (*Unmanned Aerial Vehicles*) ou dispositivos portáteis.

## 2 Referencial Teórico

A partir de 2012, a utilização de D-CNNs, combinadas com GPUs modernas e grandes volumes de dados, revolucionou a classificação de imagens em larga escala, como demonstrado pela AlexNet [6]. Em 2014, a arquitetura R-CNN (*Regions with Convolutional Neural Networks*) representou um novo avanço ao aplicar redes convolucionais para cada região de interesse dentro de uma imagem. No entanto, esta abordagem apresentava limitações para detecção de objetos em tempo real devido ao seu alto custo computacional [7].

Para superar essas limitações, a arquitetura YOLO (*You Only Look Once*) foi desenvolvida por Joseph Redmon et al. [8] em 2015, demonstrando um aumento significativo de desempenho na detecção de objetos ao prever simultaneamente *bounding boxes* e classificações em uma única avaliação completa da imagem. No mesmo ano, a arquitetura U-Net foi introduzida para segmentação de imagens biomédicas [9], estabelecendo um modelo capaz de classificar e extrair informações quantitativas de objetos a partir de um conjunto de dados reduzido, dada a escassez de imagens para treinamento no contexto biomédico. Esse processo de quantificação, chamado de segmentação semântica, atribui a cada pixel de uma imagem um rótulo correspondente à classe do objeto ou região a que pertence (cor, textura, vizinhança etc.). Dessa forma, mapas de segmentação semântica são gerados, onde cada pixel da imagem reflete sua classe semântica.

No contexto da medição de trincas, a segmentação binária (trinca/não trinca) vem sendo utilizada para gerar mapas detalhados pixel-a-pixel dos defeitos [10]. A partir desses mapas, é possível quantificar características como comprimento, largura média, área e orientação da trinca através de operações morfológicas, viabilizando o monitoramento automatizado da estrutura de interesse. Além da obtenção dessas informações, outros tipos de operações podem ser implementados, como filtros para a redução de ruídos no mapa, ou o preenchimento de lacunas artificiais e refinamento de bordas.

Entretanto, observa-se uma lacuna na literatura quanto à aplicação de *pipelines* adaptados às estruturas metálicas, principalmente em ambientes industriais. Variações reais como corrosão, pintura, existência de soldas ou ligações parafusadas e ruídos da aquisição dos dados afetam significativamente a eficácia dos modelos [4]. Além disso, König et al. [3] e Wang e Su [11] destacam que o estado da arte da segmentação semântica de trincas baseia-se fortemente em aprendizado supervisionado, exigindo grande quantidade de dados rotulados manualmente pixel-a-pixel para treinamento dos modelos, o que torna a etapa de pré-processamento trabalhosa e custosa.

Outro desafio encontrado na elaboração de modelos treinados a partir de um conjunto de dados reduzido é o *overfitting*, fenômeno em que o modelo apresenta um excesso de parâmetros para o *dataset* de treinamento, prejudicando sua generalização [12]. Para mitigar esse efeito, técnicas de *data augmentation* são essenciais para aumentar a diversidade do conjunto de dados e diminuir possíveis vieses. Esse processo consiste em realizar transformações nas imagens como: rotação, espelhamento, escalonamento (alterações de *zoom*), cortes aleatórios, além de alterações nas cores, na iluminação e no desfoque.

Tendo em vista o potencial da visão computacional nos processos de SHM e as atuais limitações observadas na sua adaptação para estruturas metálicas, este projeto irá propor uma arquitetura adaptada para detecção e medição de trincas por meio de aprendizado

semisupervisionado, mesclando conjuntos de dados rotulados públicos para estruturas de concreto / pavimentos rodoviários com um *dataset* não rotulado de estruturas de aço. O modelo integrará técnicas de classificação e segmentação semântica com etapas de pós-processamento através de operações morfológicas para aumentar a qualidade dos mapas semânticos e extrair informações geométricas relevantes para determinar a severidade das trincas, incluindo: seu comprimento, largura média e orientação.

### 3 Metodologia

O fluxograma do projeto é apresentado na [Figura 1](#). São propostas quatro etapas principais, que serão executadas em paralelo aos estudos aprofundados para a elaboração do trabalho e à realização das disciplinas indicadas na [Seção 4](#). O detalhamento das atividades previstas para cada uma dessas etapas é apresentado a seguir.

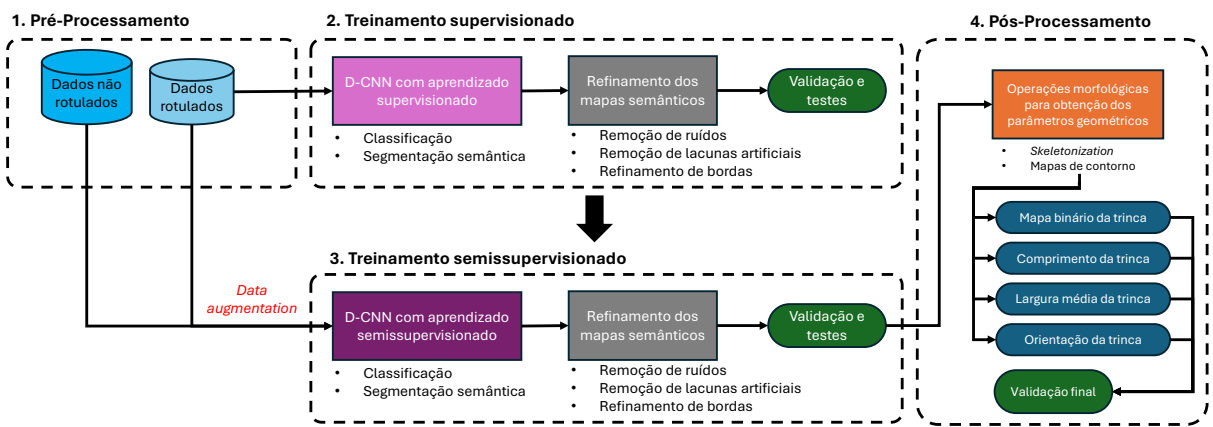


Figura 1: Fluxograma metodológico do projeto.

#### 3.1 Pré-Processamento

A preparação dos dados consistirá em três atividades: 1) aquisição dos dados públicos rotulados, 2) aquisição dos dados não rotulados de estruturas de aço e 3) realização de *data augmentation*. Para a primeira etapa, König et al. [3] indica diversos conjuntos de dados rotulados para estruturas de concreto e pavimentos rodoviários disponíveis online. Para a obtenção de imagens não rotuladas de trincas em estruturas de aço, será conduzida uma busca aprofundada na internet, além do envio de solicitações para empresas de consultoria em engenharia e inspeção de ativos industriais.

Para o conjunto de dados rotulado, serão definidas as partições destinadas aos testes e treinamento. Do conjunto de treinamento, uma fração será alocada para a validação do modelo supervisionado (e seus respectivos mapas semânticos *ground truth*) e o restante para o treinamento em si do modelo. Ainda, serão efetuadas transformações nas imagens via *data augmentation*, como rotações, *flipping*, injeção de ruído e outras técnicas apresentadas por Shorten e Khoshgoftaar [12].

#### 3.2 Treinamento Supervisionado

Nesta etapa, o *dataset* rotulado será utilizado para o treinamento supervisionado da D-CNN. Será realizada uma pesquisa aprofundada para a definição da arquitetura mais adequada à finalidade do projeto. Os mapas semânticos serão refinados através

de técnicas de processamento de imagem, que incluem: remoção de ruídos, remoção de lacunas artificiais e refinamento de bordas, aumentando a qualidade dos resultados. Para a validação deste modelo, serão avaliadas abordagens como *F1 Score* [11], IoU (*Intersect over Union*) e *Dice Score* [13], além de outras métricas consolidadas na literatura.

### 3.3 Treinamento Semissupervisionado

O treinamento semissupervisionado será definido a partir de uma pesquisa aprofundada sobre o estado da arte na segmentação de trincas. Serão avaliadas possibilidades como: *consistency regularisation*, *self-training* com *pseudo-labelling*, interfaces *teacher-student*, além da integração dessas abordagens [13]. As atividades de refinamento dos mapas semânticos e validação do modelo serão similares às da etapa de treinamento supervisionado.

### 3.4 Pós-Processamento

Para extrair os parâmetros geométricos de interesse do mapa semântico das trincas, serão realizadas operações morfológicas de *skeletonization* (diminuição sucessiva de bordas até alcançar uma trinca com um pixel de largura) [10] e elaboração dos mapas de contorno (extrair apenas os contornos da trinca). A partir desses resultados, algoritmos para cálculo do comprimento e orientação da trinca serão implementados. Também será realizada uma pesquisa sobre métodos para aquisição dos fatores de escala, necessários para a obtenção das dimensões físicas reais das trincas (pixel para milímetros, por exemplo). A validação final do modelo será realizada a partir de um *set* pequeno de trincas com suas características físicas anotadas manualmente.

## 4 Cronograma

O cronograma evidenciado na Figura 2 destaca as tarefas a serem realizadas durante o Mestrado, divididas entre realização de disciplinas, estudo aprofundado do projeto e a elaboração do projeto em si. A ordem das disciplinas foi planejada de forma a complementar o estudo aprofundado acerca do projeto e a promover um sequenciamento do aprendizado nas áreas de Aprendizado de Máquina, Visão Computacional, Processamento de Imagens e Inteligência Artificial.

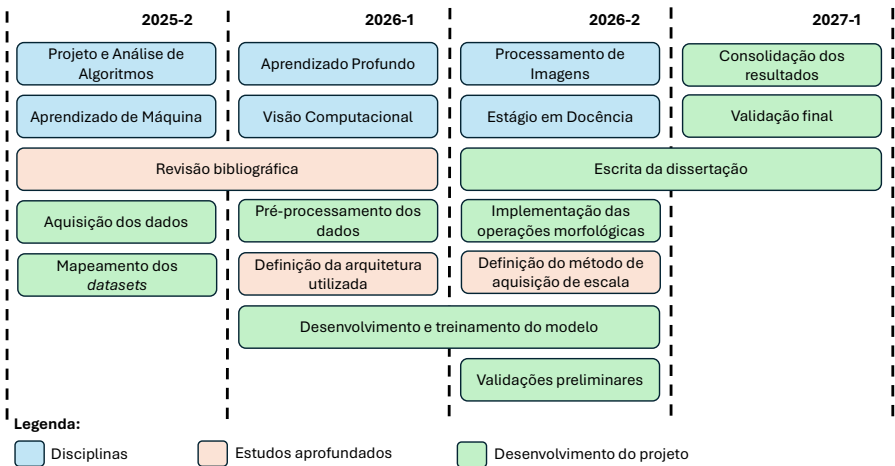


Figura 2: Cronograma do Mestrado.

## Referências

- [1] William F Hosford. *Solid mechanics*. Cambridge University Press, 2010.
- [2] JC Newman Jr. “The merging of fatigue and fracture mechanics concepts: a historical perspective”. Em: *Progress in Aerospace Sciences* 34.5-6 (1998), pp. 347–390.
- [3] Jacob König et al. “What’s cracking? A review and analysis of deep learning methods for structural crack segmentation, detection and quantification”. Em: *arXiv preprint arXiv:2202.03714* (2022).
- [4] Zhihang Li, Huamei Zhu e Mengqi Huang. “A deep learning-based fine crack segmentation network on full-scale steel bridge images with complicated backgrounds”. Em: *IEEE Access* 9 (2021), pp. 114989–114997.
- [5] Z Kala. “Sensitivity analysis of fatigue behaviour of steel structure under in-plane bending”. Em: *Nonlinear Analysis: Modelling and Control* 11.1 (2006), pp. 33–45.
- [6] Alex Krizhevsky, Ilya Sutskever e Geoffrey E Hinton. “ImageNet classification with deep convolutional neural networks”. Em: *Communications of the ACM* 60.6 (2017), pp. 84–90.
- [7] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. Em: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.
- [8] Joseph Redmon et al. “You only look once: Unified, real-time object detection”. Em: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.
- [9] Olaf Ronneberger, Philipp Fischer e Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. Em: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18. Springer. 2015, pp. 234–241.
- [10] Sutanu Bhowmick, Satish Nagarajaiah e Ashok Veeraraghavan. “Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos”. Em: *Sensors* 20.21 (2020), p. 6299.
- [11] Wenjun Wang e Chao Su. “Semi-supervised semantic segmentation network for surface crack detection”. Em: *Automation in Construction* 128 (2021), p. 103786.
- [12] Connor Shorten e Taghi M Khoshgoftaar. “A survey on image data augmentation for deep learning”. Em: *Journal of big data* 6.1 (2019), pp. 1–48.
- [13] Elyas Asadi Shamsabadi et al. “Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling”. Em: *Automation in Construction* 158 (2024), p. 105181.