



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

FCT - Faculdade de Ciências e Tecnologia
DMC - Departamento de Matemática e Computação
Bacharelado em Ciência da Computação

Trabalho de Conclusão de Curso
(Modalidade Trabalho Acadêmico)

Revisão bibliográfica

Aprendizado profundo na detecção de anomalias em dados ferroviários

Orientador: Prof. Dr. Cassio Machiaveli Oishi

Autor: Miguel de Campos Rodrigues Moret

Presidente Prudente, 20 de novembro de 2025

Aprendizado profundo na detecção de anomalias em dados ferroviários

Revisão bibliográfica apresentado ao curso de Ciência da Computação do Departamento de Matemática e Computação da Universidade Estadual Júlio de Mesquita Filho - Faculdade de Ciências e Tecnologia como requisito para a aprovação na disciplina de Projeto Científico I.

Aprendizado profundo na detecção de anomalias em dados ferroviários

Trabalho aprovado, Presidente Prudente, 2025:

Prof. Dr. Cassio Machiaveli Oishi

Orientador

Prof. Dr. Nome Avaliador

Convidado 1

Presidente Prudente – São Paulo

2025

RESUMO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla tempor velit enim, vitae imperdiet ligula finibus eget. Vestibulum tristique aliquet lectus sit amet fermentum. Praesent sapien eros, porttitor vitae dictum eget, posuere sed nunc. Vestibulum nec purus augue. Aliquam erat volutpat. Aenean imperdiet sapien scelerisque, sagittis mi id, cursus nisl. Integer vel arcu vitae nibh porta ultricies. Sed sem nunc, ornare lobortis maximus ornare, convallis vitae lectus. Cras vitae nunc dictum, auctor urna non, posuere felis. Donec vulputate enim magna, vitae euismod turpis pellentesque ac.

PALAVRAS-CHAVES

Palavra, Palavra, Palavra, Palavra, Palavra, Palavra, Palavra.

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla tempor velit enim, vitae imperdiet ligula finibus eget. Vestibulum tristique aliquet lectus sit amet fermentum. Praesent sapien eros, porttitor vitae dictum eget, posuere sed nunc. Vestibulum nec purus augue. Aliquam erat volutpat. Aenean imperdiet sapien scelerisque, sagittis mi id, cursus nisl. Integer vel arcu vitae nibh porta ultricies. Sed sem nunc, ornare lobortis maximus ornare, convallis vitae lectus. Cras vitae nunc dictum, auctor urna non, posuere felis. Donec vulputate enim magna, vitae euismod turpis pellentesque ac.

KEYWORDS

Word, Word, Word, Word, Word, Word, Word, Word.

LISTA DE FIGURAS

1	Imagem mostrando as múltiplas camadas de um modelo de aprendizado profundo	9
2	Exemplificação de transferencia de aprendizado	10
3	Exemplificação de uma rede neural convolucional	11
4	Exemplificação do funcionamento de um transformador visual	12
5	Exemplificação de imagens presentes no dataset	13

SUMÁRIO

1	Introdução	8
1.1	Justificativa	8
1.2	Objetivos	8
2	Trabalhos Relacionados	8
2.1	Relação com o meu trabalho	8
3	Aprofundamento teórico	8
3.1	Aprendizado profundo	8
3.2	Transferencia de aprendizado	9
3.3	Rede neural convolucional	10
3.4	Transformadores visuais	11
4	Metodologia	13
4.1	Dataset	13
4.2	Modelos	13
4.3	Resultados Esperados	13
4.4	Avaliação dos Resultados	13
5	Referências	14

1 Introdução

1.1 Justificativa

Diante desses avanços, este trabalho concentra-se na investigação de métodos de aprendizado profundo aplicados à detecção automática de anomalias em trilhos ferroviários a partir de imagens. O objetivo é avaliar diferentes modelos, analisar seu desempenho e discutir suas potencialidades como ferramentas de apoio à manutenção preventiva da infraestrutura ferroviária. Ao reunir e analisar a literatura existente, busca-se destacar tendências, limitações e oportunidades de pesquisa, contribuindo para o desenvolvimento de soluções mais eficientes e acessíveis para o setor.

1.2 Objetivos

Sendo assim, este trabalho tem como objetivo investigar o uso de técnicas de aprendizado profundo para a detecção de anomalias em dados ferroviários, buscando desenvolver e avaliar modelos capazes de identificar automaticamente falhas estruturais em trilhos por meio de imagens. Para isso, serão conduzidas etapas de pré-processamento e análise exploratória dos dados, implementação de diferentes arquiteturas de aprendizado profundo, avaliação de desempenho por meio de métricas apropriadas e comparação dos resultados com estudos relacionados da literatura.

2 Trabalhos Relacionados

2.1 Relação com o meu trabalho

3 Aprofundamento teórico

3.1 Aprendizado profundo

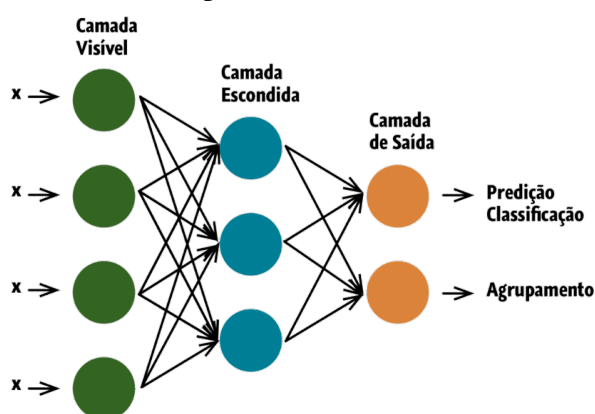
O aprendizado profundo é um subcampo do aprendizado de máquina que se baseia no uso de redes neurais artificiais com múltiplas camadas, capazes de modelar relações complexas em dados. Ao contrário de algoritmos tradicionais de aprendizado de máquina, que frequentemente dependem de extração manual de características, o aprendizado profundo consegue aprender automaticamente representações hierárquicas dos dados.

No contexto de imagens, por exemplo, redes profundas podem aprender a reconhecer pa-

drões visuais em diferentes níveis de abstração: camadas iniciais detectam bordas e contornos, camadas intermediárias capturam formas e texturas, e camadas mais profundas identificam objetos completos ou padrões de defeitos. Isso faz com que o modelo identifique até mesmo defeitos sutis nas imagens.

As redes profundas são compostas por várias unidades chamadas neurônios artificiais, organizadas em camadas. Cada neurônio realiza operações matemáticas sobre os dados de entrada, aplicando pesos e funções de ativação não lineares, o que possibilita à rede modelar relações complexas entre variáveis. O treinamento de uma rede profunda envolve a otimização desses pesos, geralmente por meio do algoritmo de retropropagação, que ajusta os parâmetros para minimizar o erro entre a saída prevista e a saída real.

Figura 1: Imagem mostrando as múltiplas camadas de um modelo de aprendizado profundo



Fonte: Vinicius Soares de Silveira

3.2 Transferencia de aprendizado

Um dos desafios em aprendizado profundo é a necessidade de grandes quantidades de dados rotulados para treinar redes neurais com bom desempenho. Em tarefas específicas, como a detecção de defeitos em trilhos ferroviários, os datasets podem ser pequenos, tornando difícil o treinamento de redes do zero. Nesse contexto a transferência de aprendizado acaba se tornando uma solução eficiente.

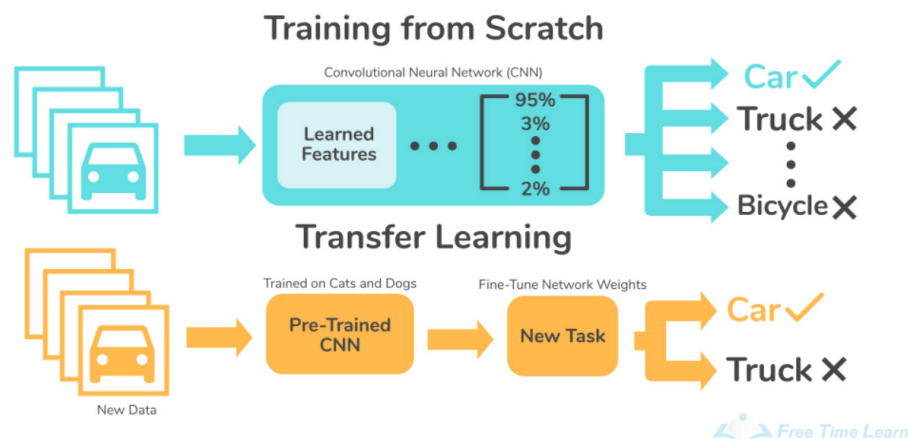
A transferência de aprendizado consiste em aproveitar o conhecimento adquirido por um modelo previamente treinado em um grande dataset genérico, como ImageNet ou COCO, e aplicá-lo a uma nova tarefa com dados limitados. O modelo pré-treinado já aprendeu a identificar padrões visuais fundamentais, como bordas, formas e texturas, que podem ser úteis em

diversos domínios visuais.

Na prática, o funcionamento da transferência de aprendizado pode ser dividido em um processo em duas etapas:

1. **Reaproveitamento do conhecimento existente:** o modelo pré-treinado mantém sua capacidade de reconhecer padrões aprendidos anteriormente, funcionando como uma base de conhecimento.
2. **Adaptação à nova tarefa:** o modelo é ajustado para identificar padrões específicos da nova aplicação. Permitindo que o modelo aprenda o que é relevante para a nova tarefa sem precisar começar do zero, fazendo com que ele aprenda mais, mesmo possuindo poucos dados.

Figura 2: Exemplificação de transferencia de aprendizado



Fonte: Free Time Learn

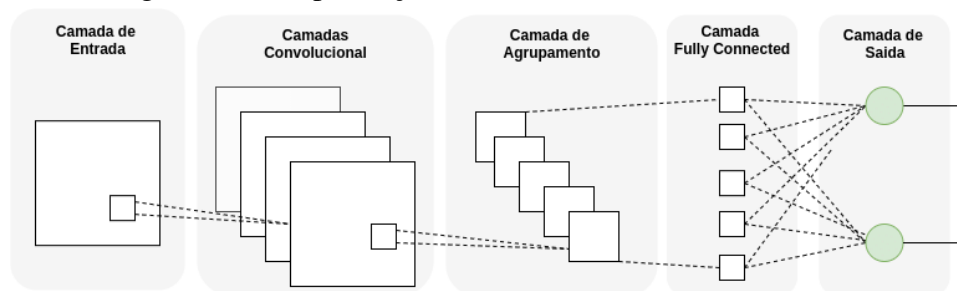
3.3 Rede neural convolucional

As Redes Neurais Convolucionais (CNNs) são uma arquitetura específica de aprendizado profundo projetada para lidar com dados estruturados espacialmente, como imagens. Diferentemente de redes totalmente conectadas, onde cada neurônio se conecta a todos os neurônios da camada anterior, as CNNs utilizam camadas convolucionais que aplicam filtros (kernels) a regiões locais da imagem, extraindo características relevantes com compartilhamento de pesos.

Uma CNN típica é composta por três tipos de camadas principais:

- **Camadas Convolucionais:** Aplicam filtros que detectam padrões locais, como bordas, curvas e texturas. Cada filtro responde a um tipo específico de característica, produzindo mapas de ativação que destacam essas informações.
- **Camadas de Agrupamento:** Reduzem a dimensionalidade dos mapas de ativação, preservando informações importantes e tornando o modelo menos sensível a pequenas variações na posição dos objetos.
- **Camadas Totalmente Conectadas:** Recebem as características extraídas e produzem a saída final.

Figura 3: Exemplificação de uma rede neural convolucional



Fonte: Martin Andreoni

No caso da detecção de anomalias em trilhos ferroviários, as CNNs permitem que o modelo aprenda automaticamente padrões de falhas, como deformações na alma do trilho, fissuras ou ausência de dormentes, sem a necessidade de extração manual de características. A capacidade de capturar hierarquias de features locais torna as CNNs particularmente eficientes para imagens de alta resolução, mesmo em datasets de tamanho moderado.

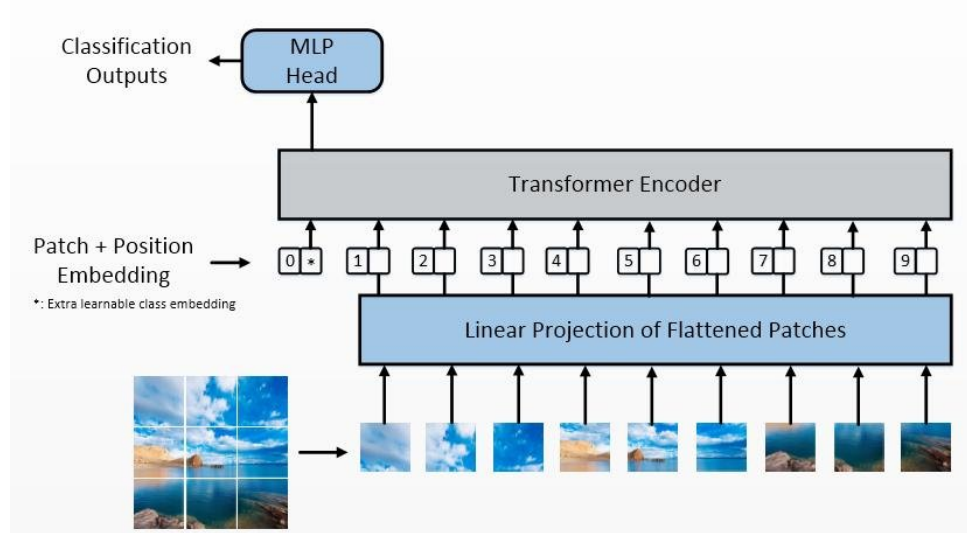
3.4 Transformadores visuais

Os Transformadores Visuais (ViTs) representam uma abordagem mais recente no processamento de imagens, inspirada nos Transformers utilizados originalmente em tarefas de linguagem natural. Enquanto as CNNs focam em padrões locais, os transformadores capturam relações globais na imagem por meio de mecanismos de atenção, permitindo que o modelo considere o contexto completo ao tomar decisões.

O funcionamento básico de um Vision Transformer envolve:

- Divisão da imagem em patches menores.
- Transformação de cada patch em vetores de embedding, que codificam informações visuais.
- Passagem dos embeddings por camadas de autoatenção, que ponderam a importância relativa de cada patch em relação aos outros, integrando contexto global.
- Classificação ou detecção baseada nos embeddings processados, permitindo identificar padrões complexos e sutis, como pequenas fissuras ou deformações.

Figura 4: Exemplificação do funcionamento de um transformador visual



Fonte: Jingxin Liang et al.

Transformadores visuais têm se mostrado altamente eficazes em tarefas onde a variabilidade de cenários é grande ou os datasets são limitados, pois conseguem aproveitar melhor a informação contextual e evitar confusões causadas por objetos próximos ou fundos complexos.

Como observado no trabalho de Shahin et al, modelos híbridos, combinando CNNs e ViTs, também têm sido utilizados para aproveitar a extração local de CNNs e a consciência global dos transformadores, alcançando desempenho superior em detecção de defeitos ferroviários. Isso é feito, treinando CNNs em grandes datasets de imagens e utilizando-as para treinar os ViTs.

4 Metodologia

4.1 Dataset

O dataset inicial que será utilizado neste trabalho é o *Railway Track fault Detection Resized* (224 X 224), disponibilizado na plataforma Kaggle por Gerry. Trata-se de uma versão redimensionada do conjunto original *Railway Track Fault Detection*. Segundo o autor, as imagens possuem alta resolução, o que torna o pré-processamento custoso; por esse motivo o dataset foi recriado com todas as imagens reduzidas para 244x244x3, facilitando seu uso, também foi incluído um arquivo *rails.csv*, que simplifica o carregamento e a organização das amostras.

O conjunto contém 384 imagens de trilhos, sendo metade correspondente a trilhos defeituosos e a outra metade a trilhos em boas condições. A divisão proposta pelo autor distribui as imagens em 6% para teste, 78% para treinamento e 16% para validação.

Os defeitos representados no dataset concentram-se em três categorias: ausência de pinos de fixação, deformações na alma do trilho (parte lateral entre a)

Ele é composto por três tipos de defeitos, falta de pinos, deformações/rachaduras nos trilhos e ausência de dormentes (tábuas transversais aos trilhos)

Figura 5: Exemplificação de imagens presentes no dataset



Fonte: Gerry (adaptado)

Devido ao tamanho reduzido e à qualidade limitada do dataset, permanece em aberto a possibilidade de incorporar novos conjuntos de dados, seja para ampliar a quantidade de amostras disponíveis, seja para garantir maior diversidade visual e melhorar o desempenho dos modelos.

4.2 Modelos

4.3 Resultados Esperados

4.4 Avaliação dos Resultados

Os resultados dos modelos serão avaliados utilizando quatro métricas:

- **Acurácia** (Accuracy)

Mede a proporção de classificações corretas, sendo utilizada para analisar o desempenho dos modelos de forma geral.

$$A = \frac{CC}{CC + CI}$$

- **Precisão** (Precision)

Mede a proporção de positivos verdadeiros entre as predições positivas, sendo utilizada para avaliar o grau de confiabilidade das detecções positivas.

$$P = \frac{VP}{VP + FP}$$

- **Revocação** (Recall)

Mede a capacidade do modelo de recuperar os positivos reais, avaliando o quanto o modelo deixa de detectar casos positivos.

$$R = \frac{VP}{VP + FN}$$

- **F1-score**

Média harmônica entre precisão e revocação, resumindo ambas em uma única métrica.

$$F1 = 2 \frac{P \cdot R}{P + R}$$

Legenda das siglas:

- CC: Classificações Corretas;
- CI: Classificações Incorretas;
- VP: Verdadeiros Positivos;
- FP: Falsos Positivos;
- FN: Falsos Negativos;
- VN: Verdadeiros Negativos;

5 Referências