DOI: 10.48011/asba.v2i1.1672

Inspeção de Componentes de Vagões Ferroviários Baseado em Visão Computacional e Aprendizagem Profunda

Leonardo Borges de Castro * **Flávio Garcia Pereira** *

* *Programa de Pós-Graduação em Engenharia de Controle e Automação, Instituto Federal de Ciência e Tecnologia do Espírito Santo, ES, (e-mail: castro.b.leonardo@gmail.com, flavio.garcia@ifes.edu.br).*

Abstract: : The health of railroads assets is essential to ensure the safety and efficiency of these logistics operations. Nowadays, inspection activities in railroads operations are made from maintenance teams by on-site observations, inside workshops or at main line operations, exposing the team to operational risks. In some cases, it is not possible to implement the inspection process due to the big amount of assets that need to be inspected in short time available, compromising the quality of the inspections. In the current context of Industry 4.0, with the advancements in hardware to image acquisition and video analytics technologies, this work will propose to study a method to detect multiples objects in wagons, which will open space for future image processing applications using machine learning models to identify defects in wagon components. The experiments were developed using the deep learning methods through convolutional neural networks based on regions, called R-CNN, and application of transfer learning techniques to accelerate the model training process. The results generated by this experiment to detect 8 objects in a side view of the wagons, showed 59.37 % in the mean average precision indicator (mAP).

Resumo: Acompanhar a saúde dos ativos das ferrovias de carga são essenciais para a garantia da segurança e eficiência destas operações logísticas. Atualmente atividades de inspeções nas operações ferroviárias são realizadas por pessoas de forma visual e presencial, sendo que muitas destas inspeções ficam inviáveis por exigir emprego de grande quantidade de recurso ou expor as pessoas a riscos, comprometendo assim a qualidade dos processos de manutenção e operação. Com o avanço de tecnologias de aquisição de imagens e métodos de visão computacional no contexto atual da Indústria 4.0, este trabalho irá propor o estudo de um método de localização e classificação de componentes de vagões ferroviários, o que irá abrir espaço para futuras aplicações de processamento de imagens utilizando modelos de aprendizagem de máquina para identificação de defeitos em componentes de vagões. Assim, foi desenvolvida uma aplicação utilizando o método de aprendizagem profunda através de redes neurais convolucionais baseadas em regiões, chamadas de R-CNN, e também a aplicação de técnicas de transferência de aprendizagem para acelerar o processo de treinamento do modelo. Os resultados gerados por este experimento apresentou 59,37% no indicador de média da precisão média (mAP) de detecção de 8 objetos em uma vista lateral dos vagões.

Keywords: Railroad, Wagons, Computer Vision, Deep Learning, Region-based Convolutional Neural Networks.

Palavras-chaves: Transporte Ferroviário, Vagões, Visão Computacional, Aprendizagem Profunda, Redes Neurais Convolucionais baseadas em Região.

1. INTRODUÇÃO

O minério de ferro faz parte do dia a dia da humanidade e a sua mineração se tornou uma atividade essencial para a vida moderna. Ele está presente em boa parte dos bens de consumo, na construção civil, indústria automotiva e aviação. Mas para chegar até os consumidores finais e transformar o minério de ferro, é necessário um longo percurso, passando pela sua extração, britagem, peneiramento, transportes, estocagem e embarque nos portos quando destinado à exportação, conforme ilustrado no infográfico apresentado por VALE (2020).

Segundo dados de BRASIL (2020), na secção de grupos de produtos, as exportações de minério de ferro no Brasil entre os anos de 2017 a 2019 representaram, em média, 54,13% das exportações brasileiras em peso líquido, ou 9,11% em valor, demonstrando a importância deste produto na balança comercial brasileira. Um dos grades diferenciais das mineradoras de ferro está na sua infraestrutura logística para transporte deste produto, utilizando desde

uma grande estrutura ferroviária, ou até mesmo a navegação para distribuição do produto pelo mundo. Segundo dados da ANTT (2020), o minério de ferro representou 76,02% dos produtos transportados por meio ferroviário no Brasil, entre os anos de 2017 e 2019.

Nesse contexto, e associado ao cenário da quarta revolução industrial apresentada pela Indústria 4.0, os setores de logística começam a se beneficiar da utilização de tecnologias de visão computacional em seus processos de operação e manutenção, tornando as atividades de inspeção mais confiáveis, atrativas financeiramente e facilitando o processo de tomada de decisão nas estruturas ferroviária, conforme estudo de viabilidade de tecnologias de visão computacional em ferrovias apresentadas por Nayebi et al. (2019).

No cenário ferroviário ainda existe o fator geográfico de distribuição dos ativos por longas distâncias, o que torna inviável muitas inspeções por parte do homem, e também, por ser um setor com grande quantidade de ativos móveis, tais como, vagões e locomotivas, que demandam a aplicação de grande quantidade de recursos para garantia de inspeções, muitas vezes exigida para garantir a segurança operacional da ferrovia e minimizar os impactos em comunidades e meio ambiente.

Atualmente, a maioria das ferrovias já possui tecnologias implantadas que fazem inspeção automatizada de seus ativos, por exemplo, aplicações em vagões ferroviários, em que são coletadas informações de integridade das rodas e conjuntos dos truques. Estes equipamentos são instalados ao lado da via, e fazem medições do perfil das rodas dos vagões, condições das sapatas de freios, além de coletar informações da geometria dos truques, utilizando processamento digital de imagens.

A ideia básica deste trabalho é avaliar técnicas de aprendizagem de máquina para localização e classificação de componentes em imagens coletadas de vagões ferroviários, o que irá permitir a formação de base de dados futuras para classificação de falhas nos componentes de vagões. Será implementada e avaliada a aplicação de técnicas de aprendizagem profunda, utilizando rede neurais convolucionais, mais especificamente as redes neurais convolucionais baseadas em regiões (em inglês, *Region-based Convolutional Neural Networks - R-CNNs*).

O método de detecção de objetos proposto neste artigo tem a vantagem de permitir que em uma única imagem seja possível localizar diversos componentes a serem inspecionados, evitando, assim, o posicionamento de muitas câmeras para inspecionar os componentes individualmente, como é feito na maior parte das tecnologias disponíveis para inspeção de vagões hoje em dia. Outra vantagem é a possibilidade de identificar a sequência de vagões em um fluxo de imagens, sem a necessidade de equipamentos adicionais para detectar a presença dos vagões e sincronizar um disparo de captura das imagens, método que atualmente é utilizado para reduzir os custos computacionais envolvidos neste processamento. A ilustração da Figura 1 permite uma visualização ampla de uma possível solução para utilização em processos de manutenção de componentes, sendo o foco deste trabalho na fase 5, a qual foi chamada de "Detectar e Classificar Múltiplos Objetos".

Dentre os demais processos apresentados na Figura 1 está a fase 1, a qual corresponde à identificação do componente através de um código (TAG) utilizando tecnologias de identificação por radiofrequência (RFID), reconhecimento automático de caracteres (do inglês, *Optical Character Recognition - OCR*) ou mesmo integração com sistemas de produção que possam relacionar os ativos ao momento de captura das imagens, representado na fase 2. A fase 3 corresponde ao armazenamento das entradas para processamento na fase 4, a qual irá realizar a classificação de cada imagem capturada e associá-la a um identificador do ativo informado pela fase 1. Em seguida, na fase 5 é feito o processo de detecção de objetos no enquadramento do ativo. Já na fase 6 é aplicado novos modelos de classificação, mas agora para identificação de anomalias, de forma a permitir a segregação de defeitos nos componentes para futura análise de um especialista em manutenção (fase 7). Diante da análises dos especialistas é possível realimentar os modelos de classificação (fases 4, 5 e 6) ou mesmo emissão de laudos para reparo nos ativos.

Para este trabalho, as imagens serão coletadas através de um enquadramento lateral dos vagões. O desenvolvimento será codificado em Python e suas principais bibliotecas de aprendizagem de máquina. Para o treinamento do modelo serão utilizados recursos de máquinas disponível em nuvem utilizando unidades de processamento gráfico (em inglês, *Graphics Processing Unit - GPU*).

Este artigo está dividido em seis secções, sendo que na Secção 2 será apresentado o vagão ferroviário e o método de manutenção baseado em inspeção. Já a Secção 3 aborda a visão computacional em aplicações ferroviárias. Na Secção 4 serão apresentados métodos de detecção de objetos utilizando aprendizagem profunda e a técnica de transferência de aprendizagem. A Secção 5 e 6, mostram os resultados e as conclusões, respectivamente.

2. O VAGÃO FERROVIÁRIO

Entre os materiais que se movimentam sobre a linha férrea, temos os que tracionam o trem, que são as locomotivas, também chamadas de material de tração, e os que são rebocados, que são os vagões ou carros, ou também chamados de material rodante Brina (1982). Existem diversos tipos de vagões que geralmente circulam nas ferrovias, mas este trabalho será delimitado a analisar os vagões tipo gôndolas, que são destinados ao transporte de minério. Na Figura 2 é apresentada uma ilustração de um vagão gôndola GDE *Greenbrier Maxion* ideal para transporte de minério, com lotação nominal de 93,5 ton de minério de ferro.

Ao analisar os dados das três principais ferrovias que fazem transporte de minério de ferro no brasil, disponibilizados por ANTT (2020), a frota de vagões destas ferrovias somaram um total de 60.536 vagões em dezembro de 2019, evidenciando a grande quantidade de ativos móveis que demandam recursos de manutenção para garantia de sua integridade e segurança. Parte da estratégia de manutenção destes ativos podem ser baseadas em inspeções visuais e medições, as quais os técnicos devem verificar diversos itens e seus modos de falha em poucos minutos, a cada ciclo de operação do vagão, ou até mesmo manter observações destes vagões em movimento ao longo das ferrovias. Muitos destes itens inspecionados nos vagões são críticos quanto

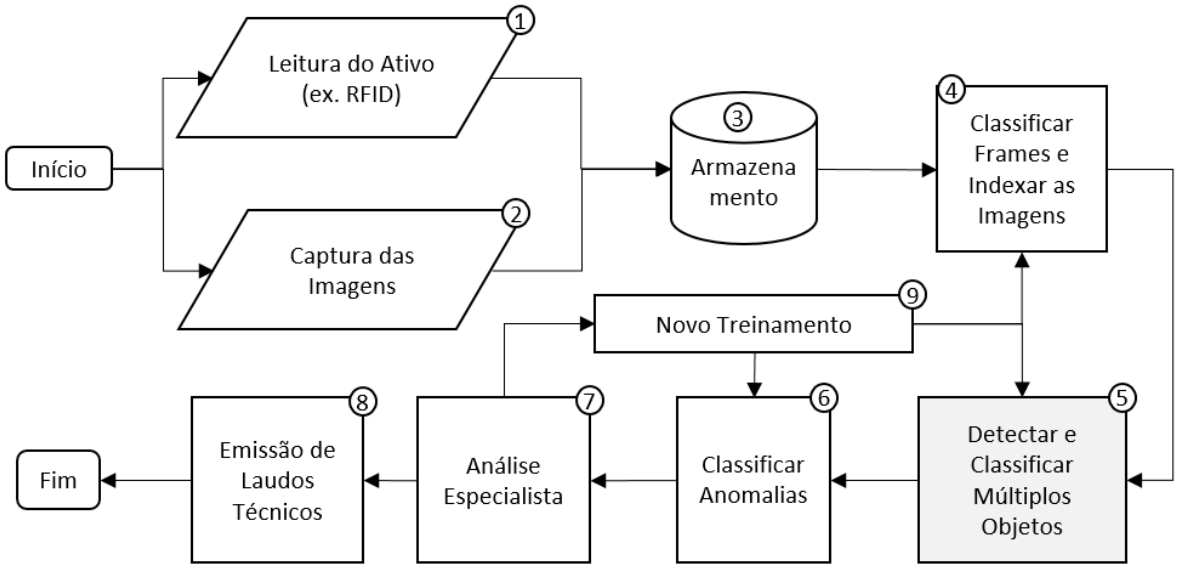

Figura 1. Visão Computacional em Aplicações de Manutenção

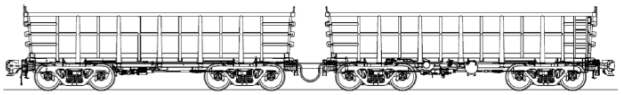

Figura 2. Vagão Gôndola GDE *Greenbrier Maxion*

à segurança operacional dos trens, podendo levar a graves acidentes ferroviários, em caso de falhas.

Diante da criticidade destas atividades e da exposição dos técnicos de manutenção aos riscos do ambiente operacional com vagões em movimento diariamente, e também pela grande quantidade de ativos a serem inspecionados, o que torna essas atividades muitas vezes árdua e tediosa, com um elevado grau de especialização técnica para avaliar os diversos componentes dos vagões e seus modos de falhas, ou mesmo necessidade de parada dos vagões em oficina para inspeção em locais de difícil visualização, os recursos de visão de máquina nestas atividades podem trazer ganhos significativos, tornando o ambiente mais seguro, produtivo e rastreável, o que proporciona maior segurança e eficiência nas operações ferroviárias.

3. VISÃO COMPUTACIONAL NAS FERROVIAS

Visão computacional consiste na aquisição das imagens utilizando câmeras digitais, organização destas imagens e análise utilizando algoritmos para produzir informações úteis, tais como localização e tipos de defeitos (Schlake and et al., 2009).

As operações ferroviárias ao redor do mundo já se beneficiam de tecnologias de visão computacional, ou também chamada de *Machine Vision Detectors* (MVDs) conforme definido por Nayebi et al. (2019). O que é possível reforçar pela publicação de Meyer (2017), o qual cita 4 ferrovias australianas que adquiriram 23 sistemas de visão computacional em dois anos, comprovando o quanto esta tecnologia vem sendo aceita no setor ferroviário.

Muitas aplicações desenvolvidas para o setor ferroviário utilizam equipamentos instalados ao lado da via férrea, também chamados de Sistema *Wayside*, o qual faz o monitoramento do material rodante em movimento. Segundo Meyer (2017), estes equipamentos evoluíram nas ferrovias ao longo dos últimos anos podendo ser representados pelas seguintes gerações, sendo que os sistemas baseados em visão computacional se encontram na terceira e quarta geração.

- Primeira Geração: dados discretos de rolamentos ou rodas quentes;
- Segunda Geração: dados 1D dos detectores acústico de rolamento e detectores de impacto das rodas nos trilhos;
- Terceira Geração: dados 2D dos medidores de perfil de rodas e espessura de sapata e discos de freios;
- Quarta Geração: dados 3D, como o mapeamento 3D dos trilhos.

Sistemas de visão computacional erram! Esta tecnologia vem avançando significativamente ao longo dos últimos anos, mas ainda é esperado erros de detecção, através de falsos positivos ou falsos negativos. E nestas aplicações que analisam grande quantidade de ativos em pouco tempo, mesmo uma taxa de erro baixa irá gerar grande quantidade de alarmes falsos. Por outro lado, os sistemas de visão habilitam inspeções em diversos ângulos e focos, que muitas vezes não são analisados com frequência por inspeções físicas, como por exemplo, uma visão inferior completa dos vagões (Nayebi et al., 2019).

A viabilidade técnica e financeira de aplicações de visão computacional no setor ferroviário pode ser comprovada em diversos estudos já publicados, e também, pela existência de diversos fabricantes especializados nestas aplicações pelo mundo nos últimos anos. Conforme o estudo apresentado por Nayebi et al. (2019), sistemas que monitoram o flange das rodas dos vagões, componentes dos acoplamentos entre vagões, mangotes do sistema de freio, componentes inferiores dos vagões e inspeção do

produto transportados, trazem ganhos financeiros significativos, além de maior segurança operacional. Ou mesmo os estudos de Hart and et al. (2004) e Hart and et al. (2008), que demonstram o potencial da tecnologia para inspeção de componentes dos truques dos vagões e utilização de câmeras no espectro infravermelho para inspeção dos componentes inferiores dos vagões. Schlake and et al. (2009) demostra o desenvolvimento de uma aplicação de visão computacional para inspeção da viga central dos vagões através do programa *Technology Driven Train Inspection* (TDTI) do *Transportation Technology Center, Inc* (TTCI). Nesta aplicação é formada uma visão panorâmica da parte inferior dos vagões. E Mittal and Rao (2017) que apresentam as redes neurais profundas para classificação e localização de componentes da via férrea, tais como, sinais, aparelhos de mudança de via, placas quilométricas, grampos de fixação da via férrea, defeitos nos trilhos, dentre outros. Mittal and Rao (2017) apresenta ainda o conceito de transferência de aprendizagem utilizando modelos que já foram fortemente treinados em outras bases de imagens, tais como *ImageNet* e COCO, permitindo que estes algoritmos sejam mais assertivos na classificação dos elementos.

4. REDES NEURAIS CONVOLUCIONAIS BASEADAS EM REGIÃO (R-CNN)

Segundo LeCun and et al. (2015), métodos de aprendizagem profunda, ou também chamados de *Deep Learning*, vêm se concretizando como o estado da arte em processamento de imagens, vídeo e áudio, assim como as redes recorrente nas aplicações de dados sequenciais para textos e falas. Estes métodos, por envolverem um custo computacional relevante, vem sendo viabilizado cada vez mais no cenário atual de acessibilidade à equipamentos com alto poder de processamento.

4.1 Redes Neurais Convolucionais

A Rede Neural Convolucional (do inglês, *Convolutional Neural Network - CNN*), apesar de ser muito popular hoje em dia para estas aplicações de processamento de imagem, seus primeiros desenhos nasceram em 1980, através do conceito de convolução desenvolvido por Fukushima (1980) em uma arquitetura de rede neural, com a criação da rede *Neocognitron*, mas esta ainda sem utilizar métodos de *back-propagation* para treinamento. Em seguida outros autores evoluíram nestes conceitos, sendo que LeCun and et al. (1989) inclui a técnica de *back-propagation* em uma aplicação de reconhecimento de dígitos de caixa postal escritos a mão livre, e em seus trabalhos seguintes cria a arquitetura mais próxima do que utilizamos atualmente nas redes neurais convolucionais, conforme pode ser observado no estudo sobre aprendizagem profunda realizado por Schmidhuber (2015).

Ele cita ainda que a rede neural convolucional é um tipo especial de rede profunda, com uma facilidade maior de ser treinada e mais generalista se comparando com as típicas redes neurais de camadas totalmente conectadas entre si. Este método se diferencia por quatro principais características: conexões locais, pesos compartilhados, *pooling* e o uso de multicamadas LeCun and et al. (2015).

Este método vem apresentando bom desempenho nas atividades de classificação de elementos nas imagens, pois conseguem isolar características através da aplicação de uma variedade de filtros (convoluções) em suas diversas camadas. Uma vez que estas redes redes convolucionais são treinadas segundo o método de *back-propagation*, elas conseguem aprender qual camada, região ou filtro mais representam o elemento a ser segmentado na imagem.

Uma das grandes vantagens deste método é não demandar pré-processamento, mas precisam de uma grande quantidade de imagens para treinamento, sendo estes nas suas mais diversas condições de luz e posicionamentos, tornando estas redes mais robustas em tarefas de classificação.

4.2 Redes Neurais Convolucionais Baseadas em Região

Uma variação das CNNs, são as redes neurais convolucionais baseadas em regiões (R-CNN), proposta por Girshick and et al. (2014). Estas redes, além de permitirem a classificação dos elementos, incluem a possibilidade de detecção de múltiplos objetos nas imagens, pois este método é iniciado dividindo a imagem em regiões utilizando método de busca seletiva, em seguida é utilizado as redes CNNs em cada uma destas regiões, adicionando na etapa final a técnica de máquina de vetor de suporte (SVMs), para classificação dos objetos e regressão das posições destes objetos. Todas essas etapas tornam o tempo de processamento para treinamento e testes um ponto crítico deste método.

4.3 Fast R-CNN e Faster R-CNN

O próprio Ross Girshick, juntamente com outros pesquisadores, evoluiu o método da R-CNN, adicionando novas técnicas para tornar o modelo mais rápido. Desta forma, nasceram os métodos *Fast R-CNN*, conforme descrito por Girshick (2015), o qual é criado o conceito de *region of interest pooling (RoIP)*, em seguida o método *Faster R-CNN* descrito por Ren and et al. (2017), que adiciona o conceito de *region propose network (RPN)*.

Estas redes de proposta para as regiões (RPN) tem por objetivo encontrar quais os possíveis quadros que mais envolvem objetos na imagem. Para isto são criados os conceitos de âncoras, no qual a imagem original é dividida em vários centros de âncoras, definidos a cada passo. Para cada um destes centros é estabelecido um conjunto de ancoras com diversos tamanhos pré-definidos, conforme ilustrado na Figura 3.

Nestas ancoras, são aplicadas CNNs que irão gerar dois tensores com mapas de atributos para classificação, definindo a probabilidade de cada ancora ser um objeto ou fundo (*background*), e também, mais 4 tensores de mapas de atributos para regressão das posições (Eixo horizontal e Eixo vertical) e tamanho (Largura e Comprimento) das ancoras.

A etapa de treinamento é feita utilizando o método de *backpropagation*, corrigindo os erros de classificação e regressão segundo objetos com classes e posições conhecidas. Para o treinamento de classificação dos objetos, inicialmente é calculada as áreas de intersecção entre a ancora em análise e as ancoras rotuladas para o treinamento (do

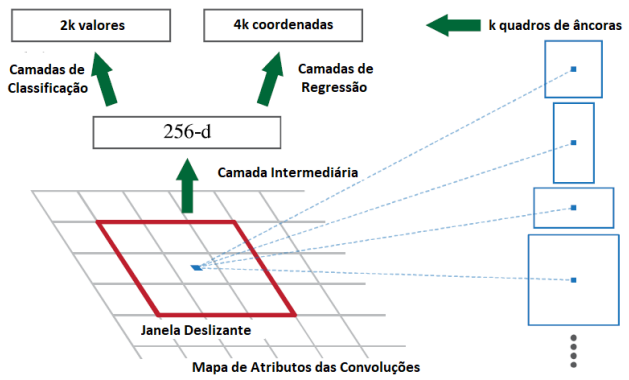

Figura 3. *Region Propose Network (RPN)* - Adaptado de Ren and et al. (2017)

inglês, *Interception over Union - IoU*), caso esta intersecção seja maior que 0,5 é considerado um objeto, caso seja inferior a 0,1 é considerado como fundo da imagem, em seguida é calculado o erro através de entropia cruzada binária, e ajustado os pesos realizando o *backpropagation*. Para as ancoras com probabilidade de serem objetos, é calculado o erro entre a posição e tamanho da ancora para a posição e tamanho das ancoras anotadas, encontrando a perda ou erro. Como resultado temos os pesos ajustados, com as ancoras mais prováveis de serem objetos, e as ancoras que mais se enquadram no objeto. Como podem existir várias ancoras na mesma anotação, é realizado um método de supressão não máximas, mantendo apenas a IoU maior e eliminando as demais.

Os mapas de atributos das CNNs são gerados por redes pré-treinadas, como por exemplo uma VGG ou Resnet, adicionando uma camada extra de convolução 3x3, para gerar um tensor com 512 mapas de atributos.

O processo seguinte é o *region of interest pooling (RoIP)*, o qual irá selecionar as melhores ancoras do processo anterior. Para cada ancora selecionada é feito o processo de *pooling*, reduzindo o mapa de atributos para 512x14x14, e em seguida *max pooling* de 512x7x7.

Na fase seguinte é realizado o processo da R-CNN, com a exceção de não se utilizar uma SVM para a reclassificação e regressão final, mas sim utilizar a própria CNN. A primeira etapa deste processo é criar uma camada de achatamento, chamada de *flatten*, a qual transforma os tensores 512x7x7 anteriores em um vetor. Após esta camada é adicionada duas camadas de neurônios totalmente conectadas, e finalmente a camada de saída com dois neurônios (probabilidade de serem objetos ou fundo) para cada classe identificada, bem como seus 4 neurônios de posição e tamanho de ancora. Ao final deste processo também é feita a supressão de ancoras não máximas. A Figura 4 apresentada por Ren and et al. (2017), ilustra este conceito envolvido na *Faster R-CNN*.

4.4 Outro Método de Detecção de Objetos

Redmon and et al. (2015) descreve outro método de detecção de objetos chamado YOLO (em inglês, *You Only Look Once*), sendo este uma referência em termos de velocidade de predição, chegando ao processamento de 155 quadros por segundo em sua versão *Fast YOLO*, podendo

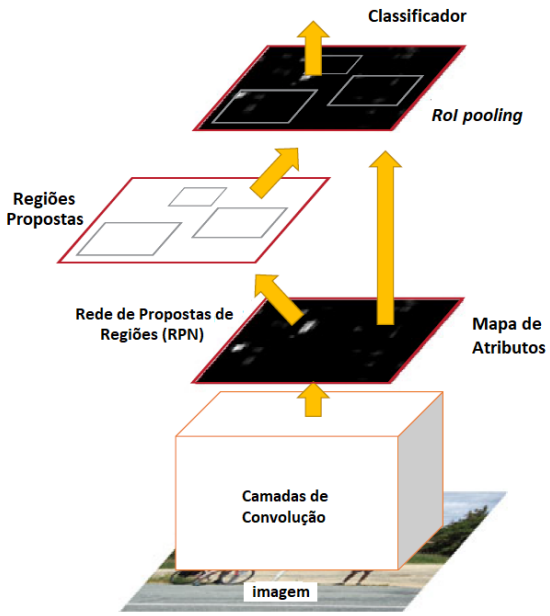

Figura 4. *Faster R-CNN* - Adaptado de Ren and et al. (2017)

assim ser considerado em um grupo de detectores em tempo real. Apesar do seu resultado de precisão média (mAP) ser inferior aos demais métodos estudados por ele, o mesmo apresenta um bom resultado quanto à sensibilidade (*recall*), gerando menor quantidade de falsos positivos. Diferente dos métodos baseado em propostas de regiões (R-CNN), a YOLO utiliza uma única rede neural para fazer uma avaliação da imagem como um todo, quanto a predição da localização e classe do objeto.

4.5 Transferência de Aprendizagem

Um dos problemas quando se trabalha com métodos de aprendizagem profunda é a necessidade de preparar uma grande base de dados para o treinamento supervisionado, sendo esta uma tarefa muitas vezes árdua, por demandar recursos para rotular essa grande quantidade de informações nas imagens. Para otimizar este processo, pode ser utilizada a técnica de transferência de aprendizagem.

A transferência de aprendizagem é basicamente o uso de pesos já treinados pela rede neural exposta a outras imagens, diferente das que serão utilizadas no problema em questão. As primeiras camadas da CNN extraem características mais genéricas da imagem, tais como borda, quinas ou mesmo intensidade de pixels, sendo que as últimas camadas farão o ajuste fino para o problema de classificação. Desta forma, a técnica de transferência de aprendizagem é utilizada nestas primeiras camadas da CNN com pesos já treinados em outro domínio, ficando as últimas camadas para serem treinadas no problema em questão Mittal and Rao (2017). Este método também permite que o modelo tenha uma capacidade maior de generalização, ou seja, quando o mesmo for exposto a dados diferente do problema, poderá ter maior facilidade na classificação dos mesmos.

5. EXPERIMENTOS E RESULTADOS

Para solução do problema de detecção de componentes em uma visão lateral dos vagões ferroviários, conforme proposto neste trabalho, será avaliado o método *Faster R-CNN* quanto a sua assertividade em localizar e classificar corretamente os componentes dos vagões. Para isto, foi preparada uma base de dados de 367 imagens com visão lateral dos vagões. Nesta base de imagens foram rotulados os componentes: molas, identificação do vagão (TAG), volante e válvula de freio, escadas, rolamentos e engates fixos e rotativos dos vagões, conforme ilustrado na Figura 5 e Figura 6

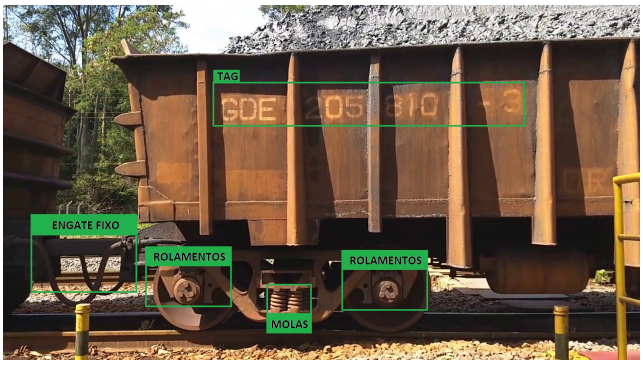

Figura 5. Componentes Rotulados nos Vagões - TAG, Engate Fixo, Rolamentos e Molas

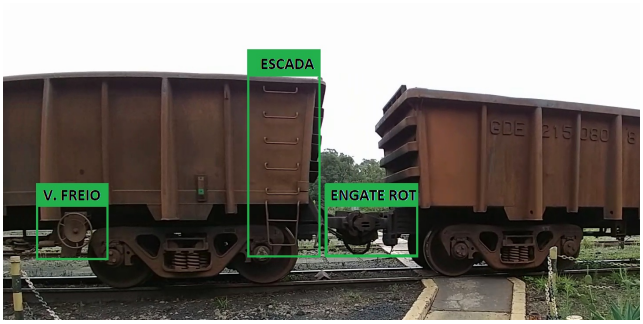

Figura 6. Componentes Rotulados nos Vagões - Escada, Engate Rotativo e Válvula de Freio

Na Tabela 1 estão as distribuições com a quantidade de componentes rotulados para as etapas de treinamentos e testes. É possível observar que a quantidade de componentes variam por imagem, ou seja, em uma mesma imagem não estão presentes todos os componentes rotulados, uma vez que o enquadramento dos vagões podem variar de uma imagem para outra.

Tabela 1. Quantidade de Objetos Rotulados

Objetos Rotulados	Treino	Testes
ENGATE FIXO	90	31
ENGATE ROTATIVO	67	19
ESCADA	115	26
VOLANTE DE FREIO	27	4
MOLAS	370	95
ROLAMENTOS	735	198
TAG	201	56
VÁLVULA	20	3
TOTAL	1625	432

Estes componentes foram rotulados no padrão PASCAL VOC e escolhidos por se destacarem na visão lateral dos vagões, o que permite avaliar a eficiência do método diante das diversas perturbações que possam existir nas imagens, tais como: destaque dos componentes, sombras, proximidade entre os elementos ou obstruções na visualização dos mesmos. Uma vez comprovada a eficácia do método, este poderá ser aplicado para detecção e classificação de outros componentes de interesse na imagem.

Para este experimento, foi utilizado o código de detecção e segmentação chamado de *Detectron2*, desenvolvido por Wu and et al (2019) através do *Facebook AI Research*. Trata-se de um código aberto desenvolvido em Python 3.6 na biblioteca PyTorch, o qual permite algumas customizações quanto ao tipo de CNN a ser utilizada e quanto à funcionalidades de detecção ou segmentação de objetos, além de permitir inicializar a rede neural com pesos previamente treinados em outras bases de dados, aplicando o conceito de transferência de aprendizagem.

Em termos de processamento de máquina, foi utilizada a plataforma em nuvem *Google Colaboratory (Colab)*, o qual permite acesso livre a processamento em GPU (*Driver* 418.67, Tesla P100 e CUDA: 10.1) e fácil compartilhamento do código com outros colaboradores. A codificação no *Colab* é feita utilizando códigos em Python 3, no formato Jupyter Notebook. O código deste trabalho está disponibilizado no *website*

`https://github.com/leonardoborgeslcastro/Object-Detection`

Aplicando a sequência tradicional de aprendizagem supervisionada de máquina, inicialmente a base de dados de 367 imagens foi dividida em 80% (294) de dados para treinamento e 20% (73) para testes, as quais foram escolhidas de forma aleatória. Na fase seguinte, os dados de treinamento e testes no padrão PASCAL VOC foram registrados no código *Detectron2*, ou seja, transformadas para o formato de entrada nas redes CNNs de treinamento e predição.

A rede escolhida para a aplicação foi a *Faster R-CNN*, que neste caso é configurada por *faster_rcnn_R_C4*. Esta configuração define o modelo *Resnet-50* como arquitetura da CNN para classificação de múltiplos objetos e pesos previamente treinados na base de dados PASCAL VOC 2007 e 2012, permitindo o conceito de transferência de aprendizagem.

A rotina de treinamento foi configurada para o máximo de 30.000 iterações e duas imagens por *batch*, o qual ao final de 2 horas e 48 minutos (0.33 segundos por iteração) apresentou um resultado de perda total de 29,6%, sendo 5,5% no processo de classificação e 21,2% no processo de regressão. Neste cálculo é adicionado também as perdas de classificação e regressão no bloco RPN da *Faster RCNN*, os quais foram 0,6% para classificação e 1,7% para regressão. Para a base de dados de treinamento atual, foi possível observar que o erro não se altera muito a partir de 15 mil iterações, como pode ser observado na Figura 7.

Os testes foram executados nas 73 imagens reservadas para esta etapa, e foi utilizada a métrica de precisão média (AP) para avaliação dos resultados, sendo este um indicador bem consolidado para os problemas de detecção de objetos.

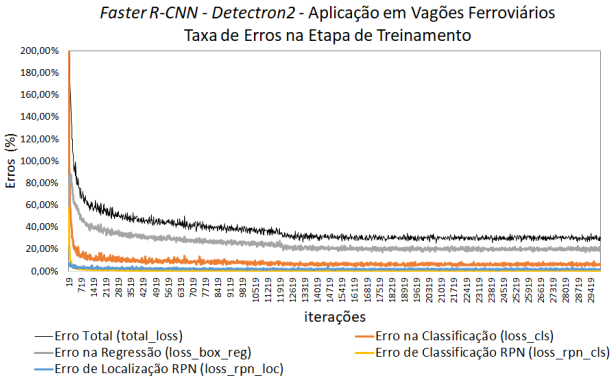

Figura 7. Erros na Etapa de Treinamento no Método *Faster R-CNN - Detectron2*.

O AP é uma ponderação entre sensibilidade (capacidade de predizer corretamente os casos que realmente existem), precisão (corresponde à porcentagem da predição que realmente está correta) e a área de interseção entre o quadro predito da posição do objeto pelo quadro onde o objeto realmente está presente, chamado de IoU. Para o IoU, é definido um valor limite para classificar se o objeto foi localizado corretamente ou não, cujo valor geralmente é 50% ou 75% de interseção, o qual são denominados de AP50 ou AP75, respectivamente. O cálculo do AP é a área sobre o gráfico de sensibilidade pela precisão, e conforme definido por Padilla et al. (2020), na prática, o AP é precisão média para todo o intervalo de valores de sensibilidade.

Para avaliação da detecção de múltiplos objetos, é calculada a média das precisões médias (AP), gerando o indicador mAP. Nos testes realizados no problema deste estudo, o resultado para o AP médio foi de 59,37%, AP50 foi 90,97% e o AP75 foi 71,77%. O resultado de sensibilidade média (em inglês, *Average Recall*) foi de 67,7%. O tempo médio de predição foi de 0,22 segundos por imagem. A Figura 8 ilustra uma amostra de predição da detecção de componentes dos vagões.

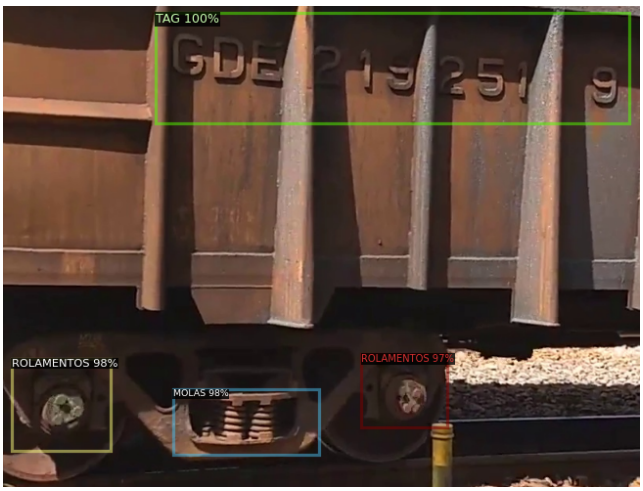

Figura 8. Detecção de Componentes dos Vagões

Na Tabela 2 é possível observar o resultado da precisão média (AP) para cada um dos componentes detectados pelo método na visão lateral dos vagões.

Tabela 2. Precisão Média - Etapa de Testes

Objetos Detectados	AP (%)
ENGATE FIXO	61,71
ENGATE ROTATIVO	54,09
ESCADA	61,60
VOLANTE DE FREIO	49,37
MOLAS	57,21
ROLAMENTOS	61,29
TAG	63,57
VÁLVULA	66,10

6. CONCLUSÕES

Este trabalho apresenta um método de visão computacional e aprendizagem de máquina para detecção de múltiplos objetos em uma visão lateral de vagões ferroviários para transporte de minério de ferro. Com a evolução desta tecnologia será possível segmentar os principais componentes de vagões para posteriormente submetê-los à treinamento de novos modelos de aprendizagem de máquina para detecção de anomalias nestes componentes, informação esta que será muito útil para os mantenedores de vagões, permitindo ganhos de produtividade e segurança em operações ferroviárias.

Neste trabalho foi utilizado o método *Faster R-CNN* com transferência de aprendizagem, chegando a um resultado de 57,44% no indicador de média de precisão média (mAP), indicador este que mede a qualidade dos modelos nas tarefas de detecção de múltiplos objetos em uma imagem. Durante o treinamento do modelo foi possível observar que acima de 15 mil iterações o modelo não evoluiu em melhorias no processo de detecção, o que pode ser evidência de uma necessidade de aumento na base de imagens utilizadas no treinamento, aumentando assim a diversidade das amostras, e também, melhor balanceamento na quantidade de componentes rotulados, que poderá melhorar os resultados finais. Como trabalhos futuros, também poderão ser alterados alguns parâmetros da *Faster R-CNN*, ou mesmo utilizar outras arquiteturas de CNN neste modelo. Para melhoria da base de dados de treinamento poderá ser aplicado técnicas de aumento de dados, também chamadas de *data augmentation*.

Este trabalho deixa ainda como contribuição científica a formação de uma base de imagens rotuladas de componentes em uma vista lateral dos vagões tipo gôndola, bem como a contribuição para o setor ferroviário quanto a aplicação de um método para localizar componentes nos vagões e evoluções para identificar anomalias nestes componentes, sendo este um processo crítico nas operações ferroviárias.

REFERÊNCIAS

ANTT (2020). Anuário do setor ferroviário. `http://www.antt.gov.br/ferrovias/arquivos/Anuario_Estatistico.html`.

BRASIL (2020). Ministério da economia, industria, comércio exterior e serviços. `http://www.mdic.gov.br/index.php/comercio-exterior/estatisticas-de-comercio-exterior/series-historicas`.

Brina, H. (1982). *Estradas de Ferro*. LTC – Livros Técnicos e Científicos Editora S.A, Rio de Janeiro.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 93–202.

Girshick, R. (2015). Fast r-cnn. *Ithaca: Cornell University Library*.

Girshick, R. and et al. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *In CVPR*.

Hart, J. and et al. (2004). A machine vision system for monitoring railcar health: Preliminary results. *Association of American Railroads*.

Hart, J. and et al. (2008). Machine vision using multispectral imaging for undercarriage inspection of railroad equipment. *In Proceedings of the 8th World Congress on Railway Research*.

LeCun, Y. and et al. (1989). Back-propagation applied to handwritten zip code recognition. *Neural Computation*, 541–551.

LeCun, Y. and et al. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Meyer, A. (2017). Implementation of machine vision based wayside condition monitoring systems in australian heavy haul railroads. *Proceeding of the 11th International Heavy Haul Association Conference*, 1074 – 1078.

Mittal, S. and Rao, D. (2017). Vision based railway track monitoring using deep learnig. *arXiv preprint arXiv:1711.06423*.

Nayebi, K., Prebble, C., and Meyer, A. (2019). The impact of machine vision technology in heavy haul operations. *International Heavy Haul STS Conference*.

Padilla, R., Netto, S.L., and da Silva, E.A.B. (2020). Survey on performance metrics for object-detection algorithms.

Redmon, J. and et al. (2015). You only look once: Unified, real-time object detection.

Ren, S. and et al. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *In IEEE Transations on Pattern Analysis and Machine Intelligence*, 39, no. 6, 1137–1149.

Schlake, B. and et al. (2009). Machine vision condition monitoring of heavy-haul railcar structural underframe components. *in Proceedings of the International Heavy Haul Conference*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Biological Cybernetics*, 61, 85–117.

VALE (2020). Minério de ferro e pelotas. `http://www.vale.com/brasil/PT/business/mining/iron-ore-pellets/Documents/carajas/index.html`.

Wu, Y. and et al (2019). Detectron2. `https://github.com/facebookresearch/detectron2`.