Identificação de Vagões com Inteligência Artificial

Rayson Laroca[1], Alessander Cidral Boslooper*[,2], David Menotti[1]

[1] *Departamento de Informática, Universidade Federal do Paraná (UFPR)*
[2] *Especialização em Engenharia Industrial 4.0, Universidade Federal do Paraná (UFPR)*

raysonlaroca@gmail.com, alexcidral@hotmail.com, menottid@gmail.com

Resumo Neste trabalho, apresentamos uma solução robusta e eficiente para contagem e identificação de vagões ferroviários utilizando câmeras e inteligência artificial. Esta solução tem custo-benefício atraente e pode substituir soluções baseadas em identificação por radiofrequência (RFID). Primeiro, nós realizamos a coleta em campo de vídeos com resolução *Full HD* à 30 *frames* por segundo. Em seguida, empregamos uma metodologia baseada em *deep learning*, que é brevemente apresentada aqui por ser objeto, em andamento, de Registro de *Software* na UFPR, que obteve resultados impressionantes, isto é, contagem de vagões perfeita (100%) e taxa de identificação acima de 99,7% em cenários reais; no entanto, 11,6% dos vagões perfeitamente contados foram rejeitados pelo sistema por estarem com seus códigos de identificação danificados. Os próximos passos do nosso estudo são embarcar a solução em campo e enviar a composição férrea em formato texto para um servidor na nuvem, utilizando conceitos de internet das coisas (IoT, do inglês *Internet of Things*) com protocolo *Message Queuing Telemetry Transport* (MTTQ). Os resultados obtidos foram surpreendentes visto que o sistema proposto requer razoável poder de processamento (podendo ser embarcado em tecnologia de ponta) e por termos utilizado um número relativamente pequeno de imagens para o aprendizado das redes de reconhecimento de caracteres.

Palavras-Chaves: Identificação de Vagões, *Deep Learning*, Visão Computacional.

1. INTRODUÇÃO

No transporte ferroviário, é cada vez mais comum o desenvolvimento de tecnologias ou sistemas que permitam a contagem e identificação automática de vagões durante a passagem de um trem por um ponto de controle, como uma estação ou depósito [9], [10], [13], [15], [16]. Dada a sua importância, este é um tema que vem sendo abordado na literatura científica desde os anos 90 [1], [2].

Atualmente, a maioria dos sistemas ferroviários utiliza a tecnologia de identificação por radiofrequência (RFID) [10]. Esse método fornece resultados rápidos e precisos, mas a instalação de *hardware* extra em cada vagão aumenta consideravelmente os custos [14]. Por exemplo, no contexto de identificação de *containers* (que é muito similar ao de identificação de vagões), Verma *et al*. [13] afirmaram que embora *containers* modernos tenham espaços reservados para a instalação de leitores RFID, tais leitores não são utilizados pelas maiores transportadoras devido aos altos custos (tanto de instalação quanto de manutenção).

Por outro lado, a contagem e identificação automática de vagões utilizando visão computacional possui características como baixo custo e instalação conveniente, isto é, alta viabilidade e alta manutenibilidade [9], [10]. Tal abordagem, baseada em imagens, é possível pelo fato de todos os vagões serem identificados por um código alfanumérico que contém informações sobre o peso, tipo e subtipo do vagão [12], [17]. Este código é pintado na parte frontal, traseira e também nas laterais de cada vagão, com a sua posição exata em cada parte variando de acordo com o tipo e características do vagão.

Entretanto, a identificação dos vagões com base em imagens é uma tarefa desafiadora, visto que não existem parâmetros definidos para a pintura do código no vagão, como as cores de primeiro e segundo plano; tipo, tamanho da fonte, e distância dos caracteres, etc. [16] (diferentemente de medidores de energia elétrica [5] ou placas de veículos [6], [7], onde tais parâmetros são bem definidos). Além disso, o código de identificação pode ficar sujo e danificado devido à viagens prolongadas ao ar livre [10], e a superfície ondulada do vagão pode tornar a projeção do texto inclinada e serrilhada [13].

Outro ponto a ser levado em consideração é que a abordagem computacional deve ser muito eficiente [3], [4], isto é, ser capaz tanto de executar em dispositivos embarcados (por exemplo, *Raspberry Pi*) quanto de coletar as informações de todos os trens que passam pelo ponto onde a câmera está instalada.

Em aplicações similares de reconhecimento de caracteres ópticos (OCR, do inglês *Optical Character Recognition*), técnicas baseadas em *deep learning* vêm alcançando resultados surpreendentes e vencendo diversas competições e desafios de aprendizado de máquina [8], dada a habilidade de tais técnicas de aprender representações ou extrair características automaticamente. Portanto, neste artigo, propomos um sistema robusto e eficiente baseado em visão computacional e *deep learning* para a contagem e identificação automática de vagões, eliminando os altos custos relacionados à utilização da tecnologia RFID.

Nos experimentos realizados, a abordagem proposta alcançou resultados impressionantes, considerando tanto a acurácia obtida (**100%** na contagem e **99,7%** na identificação dos vagões) quanto o tempo necessário para executar o algoritmo (em uma placa de vídeo *high-end*, o sistema proposto é capaz de processar **16** imagens por segundo). Portanto, esta solução pode ser utilizada em dispositivos embarcados.

O restante deste artigo está organizado da seguinte forma. O conjunto de dados utilizado para o aprendizado (treinamento, validação e teste) do nosso sistema é descrito na Seção 2. A metodologia proposta é brevemente apresentada na Seção 3, pois encontra-se em fase de Registro de *Software* na UFPR. Relatamos e discutimos os experimentos na Seção 4. Por fim, conclusões e trabalhos futuros são apresentados na Seção 5.

2. Conjunto de Dados

Para o treinamento e avaliação do nosso sistema, coletamos 14.935 imagens (extraídas de vídeos gravados à 30 *frames* por segundo - FPS) de aproximadamente 1.000 vagões diferentes. Estas imagens foram coletadas em cinco dias diferentes em quatro localidades distintas entre novembro de 2019 e fevereiro de 2020. Assim, como pode ser visto na Fig. 1, capturamos imagens de vagões de diferentes tipos e em diferentes condições. Note que, para este trabalho, utilizamos apenas imagens diurnas.

Fig. 1 Amostras do conjunto de dados.

Observe que (i) a região de interesse (isto é, a região de cada vagão contendo o código) ocupa uma porção muito pequena da imagem; (ii) existem vários blocos de texto no vagão (por exemplo, nome de empresas e *slogans*); e que (iii) a distância da câmera aos vagões pode variar consideravelmente. Estes fatores dificultam a localização dos códigos alfanuméricos nas imagens.

As imagens foram adquiridas com quatro câmeras diferentes em resolução *Full HD*. Como tais câmeras pertencem à distintas faixas de preço, presumivelmente as imagens têm diferentes níveis de qualidade. Segundo

Kong *et al*. [3], um sistema para esta tarefa deve funcionar com câmeras não calibradas (isto é, com ângulo e posição arbitrários) para que ele possa ser implantado em qualquer local próximo a qualquer linha de trem. Quando possível, coletamos imagens de um mesmo trem com duas câmeras diferentes, uma em cada lado do trilho, assim como Liya & Jilin [11]. Dessa forma, caso a região do código esteja danificada em apenas um dos lados do vagão, podemos tratar o problema, pois temos redundância de informação.

A Fig. 2 ilustra alguns desafios característicos presentes nas imagens coletadas por nós, e também a grande variabilidade existente nos códigos alfanuméricos em diferentes vagões. Como pode ser visto, a região contendo os caracteres pode estar ou parecer corrompida pela presença de ruídos causados por desgaste, materiais sujos, iluminação externa variada e assim por diante. Além disso, o espaçamento entre os caracteres pode variar consideravelmente, isto é, a distância entre os caracteres não é uniforme, o que dificulta a segmentação de caracteres através de abordagens baseadas em heurísticas.

Fig. 2 Amostras dos códigos alfanuméricos que identificam as locomotivas e os vagões.

3. Metodologia

Esta seção descreve a abordagem proposta e é dividida em duas subseções, uma relacionada à tarefa de contagem de vagões e outra à tarefa de identificação.

Abordamos as duas tarefas utilizando Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*), o tipo mais popular de redes profundas (*deep networks*). Destacamos que nosso sistema opera de forma totalmente automática, ou seja, não há qualquer tipo de intervenção humana durante a execução do método proposto.

Neste trabalho, visando não aumentar o custo computacional do sistema, não corrigimos possíveis distorções causadas pelas lentes das câmeras, e também não aplicamos técnicas de pré-processamento nas imagens antes de alimentá-las às redes (assim como Kumar *et al*. [4] e Laroca *et al*. [5], [6], [7]).

3.1. Contagem

Algumas abordagens existentes na literatura (por exemplo, [11]) utilizam diferentes tipos de sensores para indicar a presença de trens ferroviários em um determinado ponto de controle. A nossa solução, por outro lado, é totalmente baseada em visão computacional. Desta maneira, os custos de implantação e manutenção são substancialmente menores.

O processo de contagem é baseado primeiro na localização de regiões de interesse nas imagens obtidas e então na análise e processamento destas regiões. Mais detalhes sobre este processo não são divulgados neste trabalho, pois a metodologia é objeto de Registro de *Software* em andamento.

3.2. Identificação

O padrão de codificação de identificação de um vagão ferroviário é definido pela norma ABNT NBR 11691:2019 [17], sendo composto (nesta ordem) por 3 letras e 7 dígitos (por exemplo, HFE-094063-1). Ainda pode existir uma letra opcional (por exemplo, FHD-643258-1L) que indica a Regional onde o vagão está lotado (nós tratamos essa última letra como opcional pelo fato dela ter sido utilizada durante muitos anos, porém não mais em normas recentes). O último dos 7 dígitos é um dígito verificador (*check digit*), isto é, ele é gerado por um algoritmo com base nos dígitos anteriores. As locomotivas, por outro lado, são identificadas por códigos de 3 a 4 dígitos (por exemplo, 672 e 8330). Para identificar um vagão, é suficiente reconhecer apenas os dígitos, visto que as letras identificam características relacionadas ao tipo, subtipo e peso máximo admissível do vagão [1], [12], [17]. Porém, neste trabalho, treinamos uma rede capaz de localizar e classificar tanto letras quanto dígitos, de modo que o nosso sistema possa ser facilmente adaptado para a identificação de

vagões de outras regiões/países onde as letras sejam importantes para esta tarefa.

Escolhemos a arquitetura/rede com base em resultados promissores reportados em outras tarefas similares, como leitura de medidores de energia elétrica [5] e reconhecimento de placas de veículos [6], [7]. Através de regras heurísticas, nós adaptamos os resultados (ou predições) produzidos pela rede de acordo com o padrão descrito no parágrafo anterior.

Considerando que (i) a região de interesse pode ter sido incorretamente localizada no estágio anterior e que (ii) o código do vagão pode estar danificado e, consequentemente, ilegível, rejeitamos as predições retornadas pela rede em casos onde vários caracteres tenham sido preditos com valores baixos de confiança e também onde o dígito verificador predito não corresponde ao esperado [12].

Assim que todos os vagões de um trem passarem pelo ponto de controle, o sistema gera um mosaico que resume todas as informações do trem. A geração de mosaicos é importante porque resulta em um grande panorama do trem em que todos os vagões são visíveis como um único bloco [4].

Em trabalhos futuros, conforme as composições vão passando pela câmera, o sistema irá enviar as informações essenciais das composições para um servidor (banco de dados) na nuvem. Se houver algum problema, detectado automaticamente pelo próprio sistema, o operador pode verificar o mosaico com as imagens dos vagões e, se necessário, realizar algum ajuste nas predições ou até mesmo informar outros setores da organização a respeito de problemas (como pichação) na pintura do código de identificação em um determinado vagão.

4. Experimentos

Nesta seção, relatamos os experimentos realizados para verificar o desempenho do método proposto.

4.1. *Setup*

Nos experimentos, utilizamos um PC com processador *AMD Ryzen Threadripper 1920X* e uma placa de vídeo *NVIDIA Titan Xp*.

Todas as 14.935 imagens coletadas por nós foram utilizadas no treinamento e avaliação do algoritmo proposto para contagem de vagões. Nos experimentos relacionados à tarefa de identificação, apenas 1.000 destas imagens foram utilizadas, dado o grande esforço manual em rotular a posição de cada caractere do código de identificação do vagão (este processo de anotação é necessário para o treinamento de nossa rede de OCR).

Para eliminar qualquer viés no processo de treinamento do nosso sistema, nós adotamos um protocolo de avaliação *leave-one-day-out*, onde utilizamos para treinamento imagens de todos os dias, exceto um; as imagens deste dia são utilizadas para testar o sistema. Este processo é realizado para cada dia diferente em que coletamos imagens. Na seção de resultados, nós reportamos os valores médios de acurácia (para a tarefa de contagem) e reconhecimento (para identificação) obtidos nessas 5 avaliações (visto que foram 5 dias em que coletamos imagens).

4.2. Resultados

4.2.1. Contagem

Através da detecção da região de interesse (aqui, o código de identificação dos vagões), nos vídeos coletados por nós, o algoritmo proposto obteve **100%** de acurácia na tarefa de contagem de vagões, independentemente da câmera utilizada e das condições da imagem e dos vagões. Para alcançar tal resultado, nós exploramos regras heurísticas ao processar as predições obtidas em *frames* consecutivos, garantindo que a contagem de vagões funcione até mesmo caso a região de interesse não tenha sido localizada em alguns *frames* de um determinado vagão.

Alguns resultados de detecção da região de interesse (ROI) são mostrados na Fig. 3. Como pode ser visto, predições bem localizadas foram obtidas em vagões de diferentes tipos, cores e condições.

Fig. 3 Exemplos dos resultados obtidos na detecção da região de interesse.

É importante ressaltar que o tamanho da composição não influencia nos resultados obtidos, isto é, o sistema proposto é capaz de processar eficientemente vídeos de diferentes tamanhos. Nos vídeos coletados por nós, a menor composição férrea é composta por 34 vagões, enquanto a maior é composta por 135 (uma diferença de 4 vezes).

4.2.2. Identificação

Os resultados de identificação são mostrados na Tabela 1. Como pode ser visto na primeira parte, nossa abordagem alcançou uma taxa de acurácia média de 97,3%, variando de 95,6% à 99,2% com base em quais vídeos foram utilizados para o treinamento e para a avaliação do sistema. Tais resultados são surpreendentes, visto que utilizamos menos de 1.000 imagens para o treinamento das redes de reconhecimento de caracteres (conforme detalhado na subseção anterior). Conforme mostrado na Fig. 4, o método proposto reconheceu com sucesso diferentes códigos de identificação independentemente da cor e tipo dos respectivos vagões.

Tabela 1. Resultados de identificação.

Dia	Taxa de Acurácia	Taxa de Rejeição	Taxa de Reconhecimento
# 1	95,6%	15,0%	99,0%
# 2	98,9%	11,4%	100,0%
# 3	94,3%	19,3%	99,5%
# 4	99,2%	4,8%	100,0%
# 5	98,5%	7,4%	100,5%
Média	**97,3%**	**11,6%**	**99,7%**

Fig. 4 Exemplos de códigos alfanuméricos reconhecidos corretamente pelo sistema proposto.

Em relação aos erros, percebemos que eles ocorreram majoritariamente em códigos de identificação com problemas na pintura. Alguns exemplos são mostrados na Fig. 5.

Fig. 5 Exemplos de códigos de identificação danificados, isto é, com problemas na pintura.

Como pode ser visto de maneira ampliada na Fig. 6, geralmente menos caracteres são preditos pela rede de reconhecimento quando parte do código está danificada. Na imagem da esquerda, o último caractere não está visível, enquanto na imagem da direita os dois primeiros caracteres estão obstruídos por uma pichação. Note que tais casos são facilmente rejeitados considerando o padrão dos códigos alfanuméricos de identificação de vagões definido pela ABNT [17].

Fig. 6 Exemplos de predições obtidas em códigos de identificação danificados.

Levando isso em consideração, rotulamos os vagões com problemas na pintura do código de identificação como *'danificados'* e desconsideramos tais erros em uma segunda avaliação. Os resultados podem ser vistos na

segunda parte da Tabela 1. Considerando uma taxa de rejeição de 11,6%, nosso algoritmo alcançou uma impressionante taxa de reconhecimento de **99.7%**, o que é muito próximo da perfeição. Em outras palavras, com base nestes resultados, o sistema proposto é muito robusto e capaz de reconhecer corretamente praticamente todos os códigos legíveis, independentemente de fatores externos e da distância da câmera.

É importante destacar que problemas na pintura dos vagões podem ser facilmente resolvidos pelas companhias ferroviárias, tanto do ponto de vista operacional quanto econômico, através de uma nova pintura e manutenção periódica em tais regiões. Ainda, através do mecanismo de rejeição, o sistema proposto identifica quais vagões possuem códigos danificados (com base na posição destes vagões na composição férrea).

A Fig. 7 ilustra um mosaico gerado pelo sistema após o processamento de todas as imagens de um trem que passou pelo ponto onde a câmera está instalada. Observe que as regiões rejeitadas pelo sistema (aquelas mostradas em vermelho) são referentes aos códigos com problemas de pintura. Em algumas situações, não foi possível detectar nenhuma região de interesse; deste modo, indicamos tais casos no mosaico com um provável *frame* de onde estaria o código alfanumérico (linha 2, coluna 3 da Fig. 7). Em azul, são exibidos códigos considerados como *'danificados'*, porém reconhecidos corretamente pelo sistema proposto.

Fig. 7 Mosaico gerado pelo sistema após um trem passar pelo ponto onde a câmera está instalada.

Destacamos que a manutenção da conservação dos códigos de identificação é essencial para obtenção de resultados ótimos, como é o caso resultante da composição ilustrada na Fig. 8.

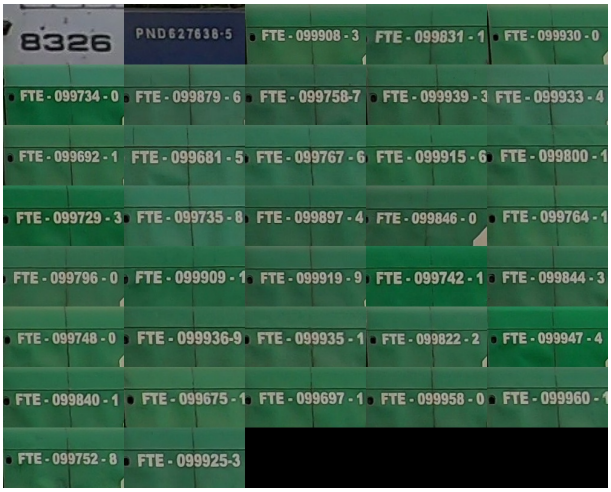

Fig. 8. Composição com códigos de identificação bem conservados resulta em contagem e identificação perfeitas (100%).

Na Fig. 9, ilustramos uma composição que foi filmada por duas câmeras, uma em cada lado do trilho. Nota-se que temos 8 erros no reconhecimento de 6 vagões. Combinando os os resultados, podemos baixar o erro para apenas 2 vagões, visto que em um dos lados do vagão o código de identificação é passível de reconhecimento.

Fig. 8 Mosaicos de uma mesma composição avaliada por duas câmeras, uma em cada lado do trilho.

Por fim, subimos um vídeo no *YouTube* para ilustrar o funcionamento do sistema proposto.

5. Conclusões e Trabalhos Futuros

Neste trabalho, apresentamos um sistema robusto e eficiente baseado em visão computacional e *deep learning* para contagem e identificação automática de vagões ferroviários. Comparado com métodos baseados em RFID, a abordagem proposta (baseada em processamento de imagens) é muito mais vantajosa economicamente, isto é, possui menor custo de instalação e de manutenção [9], [10], [13], [14].

A robustez do nosso sistema é notável, visto que ele alcançou a contagem perfeita em 16 composições (vídeos) e uma taxa de reconhecimento acima de **99,7%** (considerando uma taxa de rejeição de 11.6%), mesmo tendo sido treinado em relativamente poucas imagens e sendo capaz de processar **16** imagens por segundo. Acreditamos que nosso sistema se torne ainda mais robusto quando utilizarmos mais vídeos (e consequentemente imagens) para o aprendizado do sistema.

É importante destacar que o *software* desenvolvido neste trabalho está em processo de registro na Agência de Inovação da Universidade Federal do Paraná (UFPR).

Em trabalhos futuros, pretendemos realizar experimentos em cenários mais desafiadores, como imagens obtidas em períodos noturnos, chuvosos, e com neblina. Ainda, planejamos embarcar nossa solução em campo, definindo o melhor *hardware* em termos de custo-benefício e também a melhor posição de cada câmera de modo a evitar sombras, reflexos e até mesmo vandalismo (isto é, as câmeras não devem ficar muito próximas ao solo). Por fim, queremos enviar a composição férrea em formato de texto para um servidor na nuvem utilizando conceitos de internet das coisas (IoT, do inglês *Internet of Things*) com protocolo *Message Queuing Telemetry Transport* (MTTQ).

6. Agradecimentos

Este trabalho foi apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (editais 313423/2017-2 e 428333/2016-8) e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradecemos o apoio da *NVIDIA Corporation* pela doação da placa de vídeo *Titan Xp* utilizada para esta pesquisa.

7. Referências

1. L. De Vena, *Number plate recognition by hierarchical neural networks*, International Joint Conference on Neural Networks, outubro (1993), pág. 2105–2108.
2. J.-H. Jang, J.-H. Ku, K.-S. Hong, J.-H. Kim, *Two-stage recognition of freight train ID number under outdoor environment*, International Conference on Document Analysis and Recognition (ICDAR), agosto (1995), pág. 986–990.

3. Q.-J. Kong, A. Kumar, N. Ahuja, Y. Liu, *Robust segmentation of freight containers in train monitoring videos*, in IEEE Workshop on Applications of Computer Vision, dezembro (2009), pág. 1–6.

4. A. Kumar, N. Ahuja, J. M. Hart, U. K. Visesh, P. J. Narayanan, C. V. Jawahar, *A vision system for monitoring intermodal freight trains*, IEEE Workshop on Applications of Computer Vision, fevereiro (2007), pág. 24–30.

5. R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, D. Menotti, *Convolutional neural networks for automatic meter reading*, Journal of Electronic Imaging, vol. 28 (2019), pág. 1–14.

6. R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, D. Menotti, *An efficient and layout-independent automatic license plate recognition system based on the YOLO detector*, arXiv preprint, vol. arXiv:1909.01754 (2019), pág. 1–14.

7. R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, D. Menotti, *A robust real-time automatic license plate recognition based on the YOLO detector*, International Joint Conference on Neural Networks (IJCNN), julho (2018), pág. 1–10.

8. Y. LeCun, Y. Bengio, G. Hinton, *Deep learning*, Nature, vol. 521 (2015), pág 436-444.

9. C. Li, S. Liu, Q. Xia, H. Wang, H. Chen, *Automatic container code localization and recognition via an efficient code detector and sequence recognition*, IEEE/ASME International Conference on Advanced Intelligent Mechatronics, julho (2019), pág. 532–537.

10. Z. Liu, Z. Wang, Y. Xing, *Wagon number recognition based on the YOLOv3 detector*, IEEE International Conference on Computer and Communication Engineering Technology, agosto (2019), pág. 159–163.

11. D. Liya, L. Jilin, *Intelligent freight train ID recognition system*, IEEE International Conference on Intelligent Transportation Systems, setembro (2002), pág. 417–422.

12. L. J. Pavloski, *Códigos de classificação dos vagões*, http://vfco.brazilia.jor.br/vag/tabela-de-classificacao-dos-vagoes-RFFSA.shtml, 2010.

13. A. Verma, M. Sharma, R. Hebbalaguppe, E. Hassan, L. Vig, *Automatic container code recognition via spatial transformer networks and connected component region proposals*, IEEE International Conference on Machine Learning and Applications (ICMLA), dezembro (2016), pág. 728–733.

14. Y. Yoon, K.-D. Ban, H. Yoon, J. Kim, *Automatic container code recognition from multiple views*, ETRI Journal, vol. 38 (2016), pág. 767–775.

15. W. Zhang, G. Zhou, M. Jiang, *Convolutional neural network for freight train information recognition*, International Conference on Machine Learning and Computing (ICMLC), fevereiro (2017), pág. 167–171.

16. X. Zou, Y. Fu, X. Li, *Image feature recognition of railway truck based on machine learning*, IEEE Information Technology, Networking, Electronic and Automation Control Conference, março (2019), pág. 1549–1555.

17. Associação Brasileira de Normas Técnicas (ABNT), NBR 11691, *Vagão Ferroviário - Classificação, Identificação e Marcação*, 2019.