

Automatic Counting and Identification of Train Wagons Based on Computer Vision and Deep Learning

Rayson Laroça¹, Alessandro Cidral Boslooper², David Menotti¹

¹Department of Informatics, Federal University of Paraná, Curitiba, Brazil

²Boslooper Railway Technology, Curitiba, Brazil

¹{rblsantos,menotti}@inf.ufpr.br ²alexcidral@hotmail.com

Abstract. *In this work, we present a robust and efficient solution for counting and identifying train wagons using computer vision and deep learning. The proposed solution is cost-effective and can easily replace solutions based on radio-frequency identification (RFID), which are known to have high installation and maintenance costs. According to our experiments, our two-stage methodology achieves impressive results on real-world scenarios, i.e., 100% accuracy in the counting stage and 99.7% recognition rate in the identification one. Moreover, the system is able to automatically reject some of the train wagons successfully counted, as they have damaged identification codes. The results achieved were surprising considering that the proposed system requires low processing power (i.e., it can run in low-end setups) and that we used a relatively small number of images to train our Convolutional Neural Network (CNN) for character recognition. The proposed method is registered, under number BR512020000808-9, with the National Institute of Industrial Property (Brazil).*

1. Introduction

In rail transport, it is increasingly common the development of systems that enable the automatic counting and identification of wagons (also known as railway cars) during the passage of a train through a station or its entry into a depot [2–6]. Given its importance, this is a topic that has been addressed in the literature since the mid-1990s [7, 8].

Currently, most systems rely on radio-frequency identification (RFID) technology for these tasks [5, 9]. This method provides fast and accurate results, however, installing extra hardware on each wagon increases both installation and maintenance costs considerably [9]. These high costs were also highlighted in the context of container identification (which is very similar to the identification of wagons) by Verma *et al.* [2], who stated that although modern containers have spaces reserved for the installation of RFID readers, such readers are not used by any major ocean container carriers due to high costs.

On the other hand, dealing with the automatic counting and identification of train wagons using computer vision has characteristics such as low cost and convenient installation, that is, convenient installation and high maintainability [4, 5]. Such an image-based approach is possible due to the fact that all wagons are identified by an alphanumeric code that contains information about the weight, type and subtype of the wagon [10, 11]. This code is painted on the front, rear and also on the sides of each wagon, with its exact position in each part varying according to the type and characteristics of the wagon.

*An article about the proposed system has been published in the October 2020 issue of **Railway Gazette International** [1], the leading business journal for the worldwide rail industry.*

The counting task is challenging due to the fact that each wagon that passes through the camera must be correctly located and tracked so that it is not counted more than once. The identification task is also difficult since there are no defined parameters for painting the code on the wagon, such as foreground and background colors; type, font size, and character distance, etc. [6] (unlike energy meters [12, 13] or license plates [14, 15], in which such parameters are well defined). Also, the identification code can become dirty and damaged due to prolonged travel in the outdoors [5], and the corrugated surface of the wagon can make the projection of the 2D text slanted and jagged [2].

Another point to be taken into account is that the computational approach must be very efficient [16, 17]; in other words, the system designed for these tasks should be able to run on low-end setups, including embedded devices, as well as to collect information from every single train passing through the control point where the camera is installed.

Deep-learning based techniques have been achieving surprising results and overcoming various competitions and machine learning challenges [18–20], given the ability of such techniques to learn representations or extract features automatically. Therefore, in this work, we propose a robust and efficient system based on computer vision and deep learning for automatic counting and identification of train wagons that eliminates the high costs related to the use of RFID technology.

In our experimental evaluation, conducted using Full HD videos, our system achieved impressive results, i.e., 100% accuracy in the counting stage and a mean recognition rate of 99.7% in the identification task. The proposed approach automatically rejected 11.6% of the train wagons successfully counted, as they have damaged/illegible identification codes. In a high-end GPU (an NVIDIA Titan Xp), our system is capable of processing 16 frames per second (FPS), which means it could be deployed on embedded devices located on site and still perform its functions in a matter of seconds or minutes.

The remainder of this paper is organized as follows. The dataset used in our experiments is described in Section 2. The proposed methodology is briefly presented in Section 3. We report and discuss the experiments in Section 4. Finally, conclusions and future work are presented in Section 5.

2. The Dataset

For the training and evaluation of our system, we collected 14,935 images (extracted from videos recorded at 30 FPS) from approximately 1,000 different wagons. These images were acquired on five different days in four different locations between November 2019 and February 2020. Thus, as can be seen in Figure 1, we captured images of different types of wagons and under different conditions. It should be noted that, in this work, we use only daytime images for two main reasons: (i) an infrared camera is required to acquire images at night so that the identification codes are legible; and (ii) special authorization is required to record videos at certain control points outside business hours.

Note that (i) the region of interest (i.e., the region containing the identification code) occupies a very small portion of the image; (ii) there are several blocks of text in the wagon (for example, company names and slogans); and that (iii) the distance from the camera to the wagons can vary significantly. These factors make it difficult to accurately locate the codes in the images.

The images were acquired with four different cameras in Full HD resolution (i.e., $1,920 \times 1,080$ pixels). As these cameras belong to different price ranges, the images pre-



Figure 1. Sample images of the dataset used in this work.

sumably have different levels of quality. According to Kong *et al.* [16], a system for this task must work with non-calibrated cameras (that is, with an arbitrary angle and position) so that it can be deployed anywhere near any train track. When possible, we collected images of the same train with two different cameras, one on each side of the track, as Liya & Jilin [21]. In this way, if the code region is damaged on only one side of the wagon, we can still correctly identify the wagon since we have information redundancy.

Figure 2 shows some characteristic challenges present in the images collected by us, as well as the great variability that exists in the identification codes in different wagons. As can be seen, the region containing the characters may be or appear corrupted by the presence of noise caused by wear, dirty materials, varied external lighting, etc. In addition, the spacing between characters can vary considerably, that is, the distance between characters is not uniform, which makes it difficult to segment characters using heuristic-based or attention-based approaches.



Figure 2. Samples of alphanumeric codes that identify locomotives and wagons.

3. Proposed Approach

This section describes the proposed approach and is divided into two subsections, which are related to the *wagon counting* and *wagon identification* tasks, respectively.

We leverage the high capability of Convolutional Neural Networks (CNNs), the most popular type of deep networks, to handle both tasks. Note that our system operates fully automatically; in other words, there is no human intervention during the operation of the proposed method.

In this work, in order not to increase the overall cost of the system, we do not correct possible distortions caused by the camera lenses (according to our experiments, such distortions do not impair the performance of the proposed system), and also do not apply preprocessing techniques to the images before feeding them to the networks. This same procedure was adopted by Laroca *et al.* [15] and Kumar *et al.* [17].

3.1. Wagon Counting

Some approaches in the literature (for example, the one proposed by Liya & Jilin [21]) use different types of sensors to indicate the presence of trains at a certain control point. Our solution, on the other hand, is based entirely on computer vision. In this way, implementation and maintenance costs are substantially lower [2, 4, 5, 9].

The wagon counting process is based first on the location of regions of interest in the images recorded and then on the analysis and processing of these regions. More details about this process are not disclosed in this work, as the methodology is registered (under number BR512020000808-9) with the National Institute of Industrial Property (Brazil).

3.2. Wagon Identification

In Brazil, the standard coding for the identification of a train wagon is defined by the ABNT NBR 11691:2019 standard [11]. According to this standard, the identification code is composed of exactly 3 letters and 7 digits (e.g., HFE-094063-1). However, there may exist a letter in the eleventh position (e.g., FHD-643258-1L) that indicates the country's region in which the wagon is registered – we treat this last letter as optional because it was used for many years in previous standards. The last of the 7 digits is a check digit, i.e., it is generated by an algorithm based on the previous digits. Locomotives, on the other hand, are identified by 3 or 4 digit codes (for example, 672 and 8330).

In order to identify a wagon, it is sufficient to recognize only the digits since the letters identify characteristics related to the type, subtype, and maximum permissible weight of the wagon [7, 10, 11]. Nevertheless, in this work, we train a network capable of locating and classifying both letters and digits, so that our system can be easily adapted to identify wagons from other regions/countries where letter recognition is also required. The network was chosen based on promising results reported in related Optical Character Recognition (OCR) tasks, such as image-based automatic meter reading [22] and license plate recognition [23]. Through heuristic rules, we adapt the results (or predictions) produced by the network according to the patterns described in the previous paragraph.

Considering that (i) the region of interest may have been incorrectly located in the previous stage and that (ii) the wagon identification code may be damaged and, consequently, illegible, we reject the predictions returned by the network in cases where several characters have been predicted with low confidence values and also in cases where the predicted check digit does not correspond with the expected one [10, 11].

As soon as all wagons of a train pass through a control point (usually a train station or a train depot), the system generates a mosaic that summarizes all information (i.e., the identification code of each wagon and its respective position in the composition) about that train. According to Kumar *et al.* [17], the generation of mosaics is important because it results in a large panorama of the train in which all its wagons are visible as a single block.

The above information is sent to a server/database in the cloud. If there are any problems, automatically detected by the system itself, the operator can check the mosaic with the images of the train wagons and, if necessary, make some adjustments to the predictions or even inform other sectors of the railway company about problems in the painting of the identification code (e.g., graffiti) on a given wagon.

4. Experiments

In this section, we describe the experiments carried out to verify the robustness and effectiveness of the proposed method.

4.1. Setup

All experiments were performed on a computer with an AMD Ryzen Threadripper 1920X 3.5GHz CPU, 64 GB of RAM and an NVIDIA Titan Xp GPU.

In the wagon counting stage, all 14,935 images collected by us were used either for training or for evaluating the proposed algorithm. In the experiments related to the wagon identification task, only 1,000 of these images were employed, given the great effort required to manually annotate the bounding box (x, y, w, h) of each character in the identification code. Such an annotation process is necessary for training our OCR network, which is based on the YOLO object detector [24].

To eliminate any bias in the training process of our system, we adopted a *leave-one-day-out* evaluation protocol, in which we employ images from all days except one for training, while the images from the excluded day are used for testing. This procedure is performed for each different day on which we collected images. In the results section, we report the mean values of accuracy (for the counting task) and recognition rate (for identification) achieved in these 5 evaluations (we collected images on 5 different days).

4.2. Results

This section reports the results achieved and is divided into two subsections, which are related to the *wagon counting* and *wagon identification* tasks, respectively.

4.2.1. Wagon Counting

Through the detection and tracking of the regions of interest (here, the identification codes) in the videos acquired by us, the proposed algorithm achieved 100% accuracy in this stage, regardless of the camera used and the image/wagon conditions. To achieve such an impressive result, we explored heuristic rules when processing predictions regarding consecutive frames, ensuring that the number of wagons is correct even if the identification code of a given wagon has not been located.

Some detection results are shown in Figure 3. As can be seen, well-located predictions were obtained in wagons of different types, colors and conditions. It is worth

noting that the number of wagons does not affect the results obtained by the proposed system, as it is capable of efficiently and robustly processing videos of different sizes. In the videos collected by us, the smallest train composition is composed of 34 wagons, while the largest is composed of 135 (≈ 4 times larger).



Figure 3. Examples of the detection results obtained for the wagon counting task.

4.2.2. Wagon Identification

The results obtained in this stage are shown in Table 1 (in the second column). As can be seen, our approach reached an average recognition rate of 97.3%, ranging from 94.3% to 99.2%, based on which videos were used for training and testing. Such results are surprising considering that we used less than 1,000 images for training the OCR network, as detailed in Section 4.1. According to Figure 4, the proposed method successfully recognized identification codes, regardless of the color and type of the respective wagons.

Table 1. Wagon identification results.

Day	Recognition Rate (Without Rejection)	Rejection Rate	Recognition Rate (With Rejection)
# 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%	100.0%
Average	97.3%	11.6%	99.7%

Regarding the errors, we noticed that they occurred mostly in identification codes with problems in painting. Some representative examples are shown in Figure 5.

As can be seen in Figure 6, generally fewer characters are predicted by the OCR network when part of the identification code is damaged or illegible. For example, in the middle image of Figure 6 the last character is not visible, while in the rightmost image the first two characters are occluded by graffiti. It should be noted that such cases are easily rejected taking into account the coding standard defined in [11].

In this sense, we manually labeled all wagons with problems in the identification code as ‘damaged/illegible’ and disregard errors in their codes in a second assessment.



Figure 4. Examples of train wagons correctly identified by the proposed system.



Figure 5. Examples of damaged/illegible identification codes.

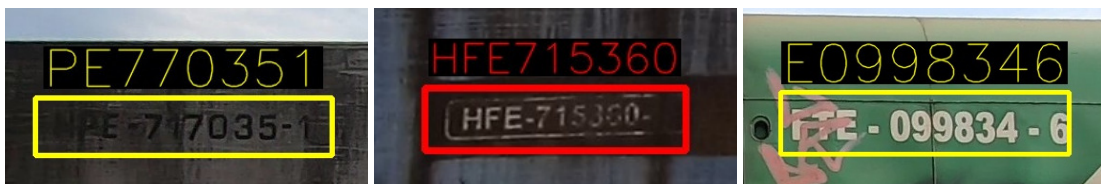


Figure 6. Examples of predictions obtained on damaged identification codes.

The results can be seen in the last column of Table 1. Considering a rejection rate of 11.6%, our algorithm achieved an impressive recognition rate of 99.7%. Based on these results, the proposed system is very robust and capable of correctly recognizing practically all legible codes, regardless of environmental factors and the distance from the camera.

It is important to highlight that problems in the painting of the identification codes can be easily solved by the railway companies, both operationally and economically, with a new painting and periodic maintenance in such regions. Furthermore, through the rejection mechanism, the proposed system automatically identifies which wagons have damaged codes (based on the position of these wagons in the train composition).

Figure 7 shows a mosaic generated by our system after processing all images of a train that passed through the control point where the camera is installed. Note that the regions rejected by the system (those shown in red) usually refer to damaged identifica-

tion codes. In some situations, it was not possible to detect any region of interest (i.e., identification code); thus, we indicate such cases in the mosaic with a probable frame where the wagon code would be located (e.g., row 2, column 3 of Figure 7). In blue, we show ‘damaged’ codes that were still recognized correctly by the system.



Figure 7. Mosaic generated automatically by the system after a train passes the control point where the camera is installed.

We emphasize that the maintenance of the identification codes is essential for the proposed system to obtain optimal results, as illustrated in Figure 8.



Figure 8. Composition with well-maintained identification codes results in perfect wagon counting and identification.

Finally, in Figure 9, we show two mosaics related to a train composition that was recorded by two cameras, one on each side of the track. Note that the system presented 8 errors in the recognition of 6 wagons. By combining the results (through heuristic rules), the number of errors is reduced to just 2 wagons since on one side of the wagon the identification code is recognizable.



Figure 9. Mosaics of the same train composition recorded by two cameras, one on each side of the track.

5. Conclusions and Future Work

In this work, we presented a robust and efficient system based on computer vision and deep learning for automatic counting and identification of train wagons. Compared with RFID-based methods, the proposed approach is economically advantageous since it has lower installation and maintenance costs. Our system is registered, under number BR512020000808-9, with the National Institute of Industrial Property (Brazil).

The robustness of the proposed method is remarkable, as it perfectly counted the number of wagons on all videos used in our experiments, and achieved a mean recognition rate of 99.7% (considering a rejection rate of 11.6%), even though it was trained in relatively few images and being able to process 16 images per second on a high-end GPU. We would like to emphasize that our system is likely to become even more robust if more videos (and, consequently, images) are used for training its networks.

As future work, we intend to carry out experiments in more challenging scenarios, such as images obtained at night, rainy, and foggy periods. In addition, we plan to ship our solution in the field, carefully defining the best hardware in terms of cost-benefit and also the best position of each camera in order to avoid shadows, reflections and vandalism. Finally, we want to explore more advanced data augmentation techniques (e.g., those presented in [25, 26]) in order to achieve even better results without having to manually label more thousands of images for training our system.

Acknowledgments

This work was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) (Social Demand Program) and the National Council for Scientific and Technological Development (CNPq) (grant numbers 428333/2016-8 and 313423/2017-2). The Titan Xp used for this research was donated by the NVIDIA Corporation.

References

- [1] R. Laroca, A. C. Boslooper, and D. Menotti, “Visual analytics keep track of freight wagons,” *Railway Gazette International*, vol. 176, no. 10, pp. 38–39, 2020.
- [2] A. Verma, M. Sharma, R. Hebbalaguppe, E. Hassan, and L. Vig, “Automatic container code recognition via spatial transformer networks and connected component region proposals,” in *IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2016, pp. 728–733.
- [3] W. Zhang, G. Zhou, and M. Jiang, “Convolutional neural network for freight train information recognition,” in *International Conference on Machine Learning and Computing (ICMLC)*, Feb 2017, p. 167–171.
- [4] C. Li, S. Liu, Q. Xia, H. Wang, and H. Chen, “Automatic container code localization and recognition via an efficient code detector and sequence recognition,” in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2019, pp. 532–537.
- [5] Z. Liu, Z. Wang, and Y. Xing, “Wagon number recognition based on the YOLOv3 detector,” in *IEEE International Conference on Computer and Communication Engineering Technology*, Aug 2019, pp. 159–163.
- [6] X. Zou, Y. Fu, and X. Li, “Image feature recognition of railway truck based on machine learning,” in *IEEE Information Technology, Networking, Electronic and Automation Control Conference*, Mar 2019, pp. 1549–1555.
- [7] L. De Vena, “Number plate recognition by hierarchical neural networks,” in *International Joint Conference on Neural Networks (IJCNN)*, Oct 1993, pp. 2105–2108.
- [8] J.-H. Jang, J.-H. Ku, K.-S. Hong, and J.-H. Kim, “Two-stage recognition of freight train ID number under outdoor environment,” in *International Conference on Document Analysis and Recognition (ICDAR)*, Aug 1995, pp. 986–990.
- [9] Y. Yoon, K.-D. Ban, H. Yoon, and J. Kim, “Automatic container code recognition from multiple views,” *ETRI Journal*, vol. 38, no. 4, pp. 767–775, 2016.
- [10] L. J. Pavloski, “Códigos de classificação dos vagões,” <http://vfco.brazilia.jor.br/vag/tabela-de-classificacao-dos-vagoes-RFFSA.shtml>, 2010, accessed: 2020-06-08.
- [11] Associação Brasileira de Normas Técnicas (ABNT), “Vagão ferroviário - classificação, identificação e marcação,” <https://www.abntcatalogo.com.br/norma.aspx?ID=426576>, 2019, accessed: 2020-06-08.
- [12] G. Salomon, R. Laroca, and D. Menotti, “Deep learning for image-based automatic dial meter reading: Dataset and baselines,” in *International Joint Conference on Neural Networks (IJCNN)*, July 2020, pp. 1–8.
- [13] R. Laroca, A. B. Araujo, L. A. Zanlorensi, E. C. de Almeida, and D. Menotti, “Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach,” *arXiv preprint*, vol. arXiv:2009.10181, pp. 1–24, 2020.
- [14] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, “A robust real-time automatic license plate recognition based on the YOLO detector,” in *International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–10.

- [15] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, “An efficient and layout-independent automatic license plate recognition system based on the YOLO detector,” *arXiv preprint*, vol. arXiv:1909.01754, 2019.
- [16] Q.-J. Kong, A. Kumar, N. Ahuja, and Y. Liu, “Robust segmentation of freight containers in train monitoring videos,” in *IEEE Workshop on Applications of Computer Vision (WACV)*, Dec 2009, pp. 1–6.
- [17] A. Kumar, N. Ahuja, J. M. Hart, U. K. Visesh, P. J. Narayanan, and C. V. Jawahar, “A vision system for monitoring intermodal freight trains,” in *IEEE Workshop on Applications of Computer Vision*, Feb 2007, pp. 24–30.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *International Conference on Neural Information Processing Systems (NIPS)*, Dec 2012, pp. 1097–1105.
- [19] T.-Y. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *European Conference on Computer Vision (ECCV)*, Sept 2014, pp. 740–755.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [21] D. Liya and L. Jilin, “Intelligent freight train ID recognition system,” in *IEEE International Conference on Intelligent Transportation Systems*, Sep. 2002, pp. 417–422.
- [22] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, “Convolutional neural networks for automatic meter reading,” *Journal of Electronic Imaging*, vol. 28, no. 1, p. 013023, 2019.
- [23] I. O. de Oliveira, R. Laroca, D. Menotti, K. V. O. Fonseca, and R. Minetto, “Vehicle re-identification: exploring feature fusion using multi-stream convolutional networks,” *arXiv preprint*, vol. arXiv:1911.05541, pp. 1–11, 2019.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 779–788.
- [25] D. V. Ruiz, B. A. Krinski, and E. Todt, “ANDA: A novel data augmentation technique applied to salient object detection,” in *International Conference on Advanced Robotics*, Dec 2019, pp. 487–492.
- [26] —, “IDA: Improved data augmentation applied to salient object detection,” in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Nov 2020, pp. 1–9.