This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3410160

0000000000000000000000000000000000000000000000000 Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Assessment of Vehicle Category Classification Method based on Optical Curtains and Convolutional Neural Networks

Joaquin Barreyro[1], Leopoldo R. Yoshioka[1], Claudio L. Marte[1], Caio G. Piccirillo[2], Max M. D. Santos[3], Senior Member, IEEE, and João F. Justo[1]

[1]Universidade de São Paulo, Escola Politécnica, 05508-010, São Paulo, SP, BRAZIL
[2]Universidade Federal do ABC, 09210-580, Santo André, SP, BRAZIL
[3]Federal University of Technology – Paraná, 84017-220, Ponta Grossa, PR, BRAZIL

Corresponding author: Claudio L. Marte (e-mail: claudio.marte@usp.br).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, “This work was supported in part by the U.S. Department of Commerce under Grant BS123456.”

ABSTRACT Automatic Vehicle Category Identification (AVC) is essential for Electronic Toll Collection (ETC) Systems. Currently, various sensors and classification algorithms are used to achieve high accuracy in vehicle category classification. This work proposes a new classification methodology, based on non-intrusive sensors, which uses binary vehicle profile images. We used a convolutional neural network (CNN) based on AlexNet, which could be embedded in real-time computational systems. The CNN was trained using transfer learning and data augmentation techniques. The proposed methodology was tested using binary images of optical curtains installed in toll plazas, located on highways in the State of São Paulo, Brazil. Four experiments were carried out to evaluate the CNN classification accuracy under different operational and environmental conditions. In a controlled experiment, we used about 11,000 images, including original images and augmented images, to train the CNN with 11 vehicle categories. The method achieved an accuracy of 98.02%, very close to one of the current systems based on intrusive sensors. Additionally, we evaluated the performance of the methodology in another experiment, reproducing real operating conditions, involving about 194,000 images. The proposed methodology reached a global accuracy of 96.48%. The reduction in accuracy was due to classification failures resulting from interferences that occurred in the images, due to variations in environmental conditions, such as rain and night operation. The applicability of the proposed methodology can be extended to other contexts, such as the free-flow toll system, through the use of other types of sensors, such as LIDAR or cameras.

INDEX TERMS Intelligent transportation systems, Optical sensors, Road transportation, Tariff.

I. INTRODUCTION

Road networks represent a significant global transportation system. For example, only the United States has more than four million miles of paved highways, as road transport accounts for 65% of the total transportation services [1]. In China, road transport has become a dominant modal, with an estimation that by 2030 there will be about 400 million vehicles circulating there [2]. In South America, Brazil has about 2 million kilometers of road networks, representing about 95% of passengers and 65% of transportation of goods [3].

As Tong *et al.* [4] pointed out, Intelligent Transport Systems (ITS) have become essential in assuring reliable operation and management of highways, highlighting the emergence of artificial intelligence (AI) technologies in that area. According to those authors, Automatic Vehicle Classification (AVC) is essential in Electronic Toll Collection (ETC) systems because the toll rate depends on the vehicle category.

The set of vehicle classification criteria is defined by the transport administration body of each country. As shown in Ref. [5], vehicle characteristics such as height, number of axles, and dual wheels are used for classification. The classification system established by the Federal Highway Administration Research and Technology (FHWA) [6] is adopted by many researchers as a reference for their studies. In Brazil, the criterion for classifying vehicle categories is determined by the number of axles and double wheels' presence (or not) [7].

In most AVC equipment, axle counting is done through several sensors installed on the ETC lane. According to Balid *et al.* [8], sensors can be intrusive or nonintrusive. Intrusive sensors, such as piezoelectric bars, pneumatic tubes, or inductive loops, require civil work for installation. The combination of magnetic loop and piezoelectric sensors are the most common sensing methods used for vehicle detection and axle counting [8]. On the other hand, nonintrusive sensors such as infrared, microwave radar, laser scanning (LiDAR), and cameras do not require civil work [9].

Intrusive sensors are expensive and require that traffic be interrupted during installation and maintenance. On the other hand, nonintrusive sensors have weaknesses related to sensitivity to external disturbances, such as pedestrian crossing, weather conditions, and occlusion.

Regarding nonintrusive techniques, studies in the literature have been mostly focused on images [10]. Moreover, most studies refer to a situation where cameras, in general for surveillance purposes, are installed to provide a broad and open view of a highway, thus allowing to obtain images with complete framing of the vehicles [11]. However, the cameras installed in the automatic lanes (ETC), as they are positioned close to the road, often do not allow capturing images with complete framing of the vehicle profile, especially in the longer ones, which makes the use of camera images, installed on ETC lanes, inappropriate for vehicle classification.

To our best knowledge, the use of optical curtains for vehicle classification purposes has not been previously reported in the literature. However, it should be noted that it allows capturing the complete vehicle profile, despite having a lower resolution than a camera, which is supposed to be sufficient to identify the number of vehicle axles. Therefore, this work explores the assumption that binary images of the vehicle profile, obtained by optical curtains, typically used to detect the separation between vehicles, allow performing the vehicle category classification with accuracy compatible with the intrusive sensor's methods. Additionally, this work aims to verify if, using a simple CNN architecture, such as AlexNet, it is possible to achieve very high accuracy in real-time for toll collection purposes.

The remainder of this paper is organized as follows. Section 2 presents the related work that supports this investigation. Section 3 details the proposed methodology, describing the data acquisition process and CNN's design. Sections 4 and 5 describe the set of experiments, results, and discussions. Finally, Section 6 presents the conclusions.

II. RELATED WORK

Several strategies in the literature address the problem of vehicle classification. In real ITS applications, piezoelectric transducers and inductive loops are the most common sensors used in classification. Dong *et al.* [12] presented a method to extract vehicle characteristics from the signal generated by the inductive loop using the Gradient Tree Boosting algorithm. González *et al*. [13] used piezoelectric transducers installed on the pavement surface. Amodio *et al.* [14] proposed a method that uses a set of magnetometers to obtain the magnetic signature of the vehicle. They applied the Dynamic Time Warping (DTW) algorithm to map the behavior of the signal.

Many image-based classification methods are reported in the literature. Gothankar *et al.* [15] applied lateral vehicle images on a CNN to recognize vehicles in the image. They counted the vehicle's wheels using two approaches. First, based on CNN, and second using Hough's circular transform. Finally, they proposed a decision algorithm to unify both criteria. Sasongko [16] used the vehicle side-view image, already in a segmented format, and applied the ResNet CNN algorithm to classify vehicles into five categories based on the number of axles.

Liu *et al.* [17] proposed a method that used road surveillance cameras. They used a CNN to identify the vehicles within the image and classified them using a subgroup of the FHWA categories (four categories) based on their size. Moreover, Zhang *et al.* [18] used the Mask R-CNN method applied to the same type of images. They compared three methods to identify axles: 1) Canny-based ellipse detector, 2) Template Matching, and 3) Mask R-CNN. Shvai *et al.* [10] used a CNN to classify frontal images and then combined the classification results with optical sensor information using a classifier based on the Gradient Tree Boosting technique. Finally, Hussain *et al.* [19] used a frontal image of the vehicle to classify, while in our work, the classification is made by counting the axles, using the vehicle's profile binary image.

Several investigations have applied LiDAR technology. Nezafat *et al.* [20] used profile images obtained with LiDAR and applied three different types of CNNs (VggNet, ResNet, and AlexNet) to classify and compare the results using a similar technique. Wu *et al.* [21] obtained six signal characteristics. In addition, they compared four classification methods: Random Forest, Bayesian network, K-NN, and SVM. However, off-the-shelf automotive LIDARs are limited by having a point cloud refresh rate of around 20 to 100Hz, which is insufficient to capture wheel detail at the resolution required for vehicles with typical speeds of around 40 km/h on ETC lanes. On the other hand, LIDAR based on MEMS technology that can reach 500Hz or more [22]. LiDAR could be an interesting solution for the Flee-Flow toll system. For ETC lanes, which are segregated, the optical curtain seems to be a better solution with the currently available technologies.

There are investigations using other types of sensors. Alexandre *et al.* [23] used acoustic sensor data to classify vehicles using a hybrid genetic algorithm and an extreme neural network. Sliwa *et al.* [24] used a wireless sensor network (WSN) to collect radio frequency (RF) fingerprints from vehicles. They extracted 92 features from 9 signal channels. They compared the results with different Machine

Learning (ML) approaches, such as Support Vector Machine (SVM), Deep Boltzmann Tree, Random Forest, and Proximity Forest.

Although several approaches concerning vehicle classification techniques have emerged, we have found that the main advancements in this area are based on the use of intrusive magnetic sensors, camera images and infrared, combined with different neural network architectures. For instance, Maiga et al. (2023) [34] proposed the utilization of a low-complexity CNN model, achieving an accuracy of around 95.8%. Also, they show that more sophisticated CNN networks like ResNet and DenseNet achieved slightly better results, around 97.4%. On the other hand, Tan et al. (2024) [35] recently conducted a comprehensive survey, demonstrating that classification accuracy depends on various factors, such as the type of sensor used and the chosen classification algorithm, however, the determinant factor that impacts accuracy is the number of categories in evaluation.

As presented previously, many vehicle classification methods have been found in the literature based on different sensors, applying a wide range of algorithms. However, most studies used a few categories without considering dual wheels or suspended axles. Moreover, few studies have considered real situations on an ETC lane, such as different types of lighting, daytime, and nighttime, the incidence of sunlight, shadows, and adverse weather conditions, such as rain or fog.

We identified a research gap related to using nonintrusive sensors for automatic lanes (ETC) applications. Therefore, in this work, we proposed a new methodology that uses signals generated by optical curtains, which already exist on ETC lanes but are currently used only for vehicle separation purposes.

III. METHODOLOGY

This section presents the proposed methodology that uses a nonintrusive sensor based on an optical curtain and a classification algorithm based on CNNs based on AlexNet. Here, we present a) the concepts and technologies of vehicle classification in Brazil, optical curtain, and fundamentals of CNNs, b) the vehicle classification process by applying CNN on binary images of the side profile of vehicles, resulting from scanning through an optical curtain.

TABLE I
VEHICLE CATEGORIES SPECIFICATION DEFINED BY ARTESP

Category	Axle	Car type	Double wheels
1	2	car/pickup	no
2	2	truck/bus	yes
3	3	truck/bus	yes
4	4	truck	yes
5	5	truck	yes
60	6	truck	yes
61	7	truck	yes
62	8	truck	yes
63	9	truck	yes
7	3	car+ trailer (1 axle)	no
8	4	car+ trailer (2 axles)	no
9	2	motorcycle	no

A. CONCEPTS AND TECHNOLOGIES INVOLVED

1) VEHICLE CLASSIFICATION CRITERIA IN BRAZIL

There are several vehicle classification criteria globally. This study uses the Sao Paulo State Transportation Agency (ARTESP) specification. Table I shows the 12 categories defined by ARTESP, associated with the number of axles, considering the lifted axles' presence (or not), and double wheels. Although a motorcycle (category 9) is specified in Table I, it is not considered in this work.

2) BINARY IMAGE GENERATION OF VEHICLE PROFILE USING OPTICAL CURTAIN

In ETC applications, the optical curtain is generally used to separate vehicles moving near each other. Fig. 1 shows an optical curtain in a toll plaza, which consists of two LED columns installed transversally at the vehicle lanes. The LEDs are uniformly spaced with each other. One of the columns corresponds to a set of emitter LEDs and the other the receiver LEDs. When there is no obstruction, the emitted light beams are captured by the receiver on the opposite side, indicating logic state "1". However, when a vehicle passes through this optical curtain, its body and wheels block the light beam, indicating the logic state "0". Therefore, a binary image of the vehicle profile can be constructed by sampling the light beams from the LEDs.

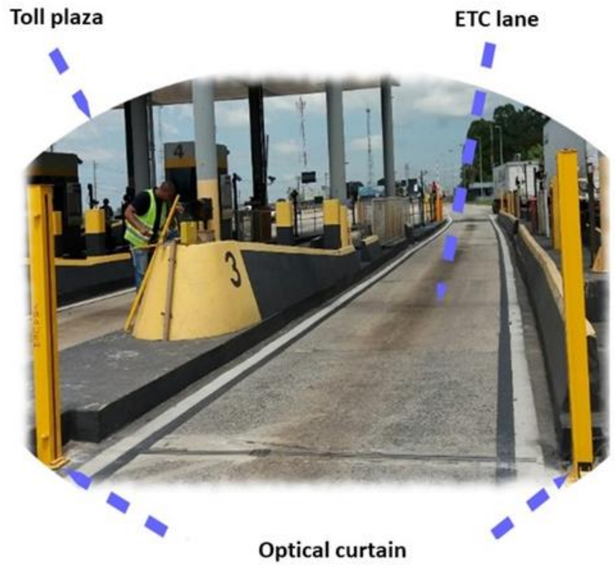

FIGURE 1. Example of an optical curtain used in an automatic lane in a toll plaza.

Fig. 2 details the vehicle's binary image generation process and the geometric parameters involved. The height (h) is generally higher than most automobiles and lower than some heavy commercial vehicles, limiting the amount of information. The curtain is composed of N equidistant and vertically distributed sensors. In this way, the vertical resolution of the curtain is h/N. The curtain periodically sends the status of the N sensors to the data acquisition equipment. The complete vehicle profile is composed of M vectors with N vertical elements, where M is a function of the sampling frequency (fs), the speed (v), and the vehicle length (l). Therefore, the binary image consists of an N×M matrix, where each element represents a binary pixel, which assumes a value of 0 or 1.

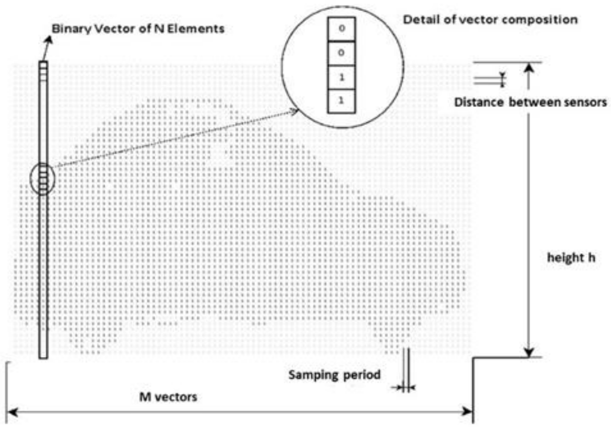

FIGURE 2. Vehicle profile binary image.

Since we are capturing the vehicle profile (side view), the vehicle speed is essential. The vehicle speed (v) influences the matrix width (M). Vehicles at lower speeds generate elongated images, while vehicles at higher speeds generate compressed images. This effect is illustrated in Fig. 3a, where the same vehicle is represented for three different speeds (v, 2v, and v/2).

Optical curtains for ETC applications available on the market typically have a sampling frequency of 240 Hz [25]. Therefore, with a vehicle traveling at 36 km/h, or 1000 cm/s, the horizontal resolution is 1000cm/240, approximately 4 cm. So, columns spaced every 4 cm represent the vehicle profile. For example, if the vehicle travels at 72 km/h, the columns are spaced every 8cm. In Brazil, the maximum speed in segregated automatic lanes at toll plazas is 40 km/h. Thus, to comply with current legislation, it is reasonable to assume that the speed variation range of ETC lanes with lane segregation is between v/2 to 2v.

The width of the matrix (M) depends on the speed and length of the vehicle. However, matrices with variable dimensions are inconvenient for further computational processing, as most ML algorithms require fixed dimensions of the input arrays. For example, many CNN architectures default the input matrix dimension to 224×224×3 (assuming a color image consisting of three channels - RGB). Therefore, in this example, it is necessary to pre-process the binary images resulting from the optical curtain to compensate for the variability of speed and length of the vehicle so that the matrix, which represents the binary image, fits the size of 224×224.

To this end, we adopted the strategy of filtering the original image, preserving only the consecutive vectors with at least one different value. Therefore, even if the vehicle's passage generates M vectors, only L vectors, not repeated, are considered, resulting in a matrix of dimension N×L. We thus have an image compression, where the most uniform regions of the vehicle are suppressed. On the other hand, the region of the wheel of interest for classification is adequately preserved due to the circular shape. Fig. 3b shows the result of eliminating repeated vectors. As a result, the width of the images has become closer to each other.

We also consider that the images inserted into the classifier must have a fixed size. Thus, we adjust the image width L to a fixed value. Finally, we use a nearest-neighbor interpolation that makes the image size compatible with the CNN.

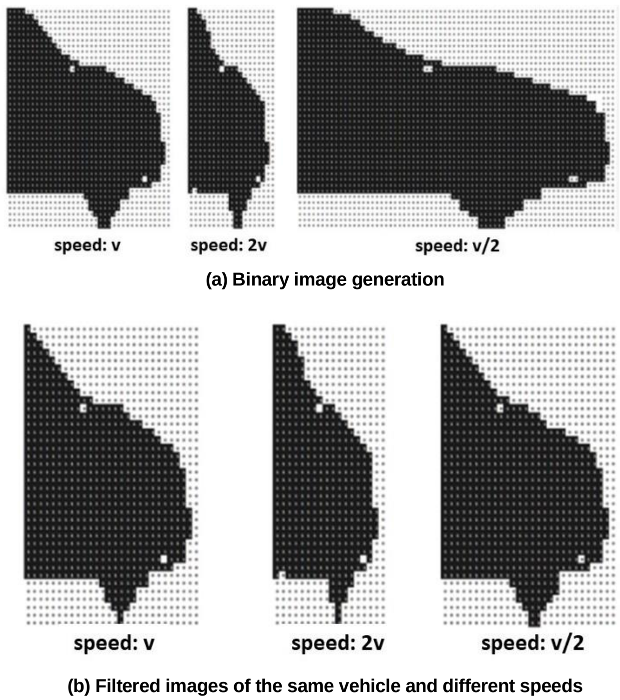

FIGURE 3. Comparison between images with or without a speed filter.

We emphasize that, as we use binary images, the CNN input layer could have been reduced from 224×224×3 to 224×224×1, as there is no color information. However, this dimensionality reduction may require retraining the entire network from scratch, and not being able to take advantage of resources, such as Transfer Learning.

3) MACHINE LEARNING TECHNIQUES

We assume that the vehicle profile image data generated by the off-the-shelf (OTS) optical curtain for the ETC application contains enough information to allow vehicle category classification with high accuracy. Therefore, it becomes valuable to determine if the application of ML algorithms

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3410160

allows for achieving a vehicle classification performance compatible with intrusive sensor-based systems. Several ML algorithms are described in the literature, such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), K-means, Random Forest, and genetic algorithms.

Here, we chose a CNN because, according to the literature, it is the most suitable for performing supervised classification from the dataset of the images. Several CNN architectures have been presented, such as VggNet, ResNet, and AlexNet [20]. Among those architectures, AlexNet has the advantage of having a pre-trained network available for 1000 categories of objects using the ImageNet image database [26]. In addition, it is one of the least complex, requiring fewer computational resources [27], which is relevant for real-time applications in ETC lanes. Thus, this work focuses on verifying vehicle classification performance using the AlexNet architecture.

AlexNet was proposed in 2011 by Krizhevsky et al. [26]. As shown in Fig. 4, it consists of eight layers, five convolutional layers, and three fully connected ones. Our strategy is to specialize AlexNet's for vehicle classification through the Transfer Learning technique, which consists of taking advantage of previous training efforts. Specialization is done by modifying the last three fully connected layers, adapting to the 11 categories of vehicles to be classified. The modified network requires fewer image samples for training to obtain satisfactory results. That is because the five convolutional layers, responsible for extracting the features of the image, were preserved. Shin et al. [28] showed the effectiveness of the Transfer Learning technique when compared to the training process starting from scratch.

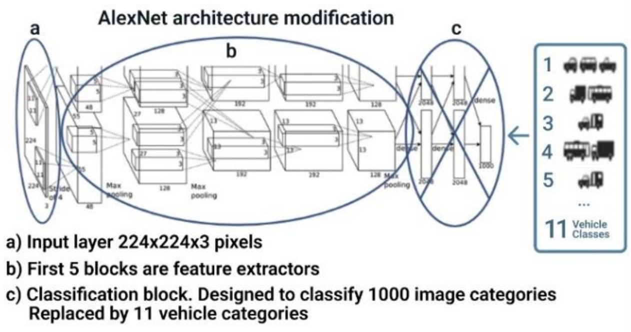

FIGURE 4. CNN AlexNet architecture using Transfer Learning. Adapted from [25].

Despite defining several vehicle categories, most of the image samples collected occur in a few categories, such as cars and trucks with two or three axles (i.e., categories 1, 2, and 3 - see Table III). Vehicle categories with many axles or special vehicles are uncommon. Therefore, sometimes it becomes necessary to add image samples artificially produced for some vehicle categories to the training dataset. We can apply the transformation to available images to generate additional training data. The data augmentation by transformations includes image translation, rotation, mirroring, random cuts, and noise addition. Perez and Wang [29] presented a detailed discussion about applying the Data Augmentation technique. In this work, we use Data Augmentation to complement data samples of the less frequency vehicle categories, for example, categories 7, 8, 61, and 62 (Table III).

The available dataset, including data augmentation, must cover all vehicle categories. About 70% of the data is used for training, while the remaining is for validation. The training process of CNN consists of adjusting the neurons' weights and giving a specific input, and the network can calculate the desired output. The stochastic gradient descent with momentum (SGDM) is one of the most practical training algorithms [30]. When CNN must process several images, ample memory space is generally required. The mini-batch concept has been developed to address this problem. Instead of using a single image in each iteration, a set of images can determine the error measurement that adjusts the neurons' weights. However, the best mini-batch size is still an open issue. Masters and Luschi [31] have used smaller mini-batches than other authors.

B. PROPOSED VEHICLE CLASSIFICATION METHODOLOGY

We now present a methodology to classify the categories of vehicles. The novelty is using binary images of vehicle profiles obtained from optical curtains. This methodology is based on AlexNet CNN architecture, modifying the last layers to classify 11 vehicle categories according to specifications in Table I. The choice of AlexNet appears to be appropriate as it has a simpler structure compared to the most recent architectures, allowing a more efficient implementation in terms of computational resources, which is essential for real-time applications, such as in the case of ETC lanes at toll plazas.

Fig. 5 illustrates an overview of the proposed methodology. The vehicle classification process is divided into two steps: first, it creates the vehicle profile image from the optical curtain, and then the CNN algorithm is used to classify the vehicles.

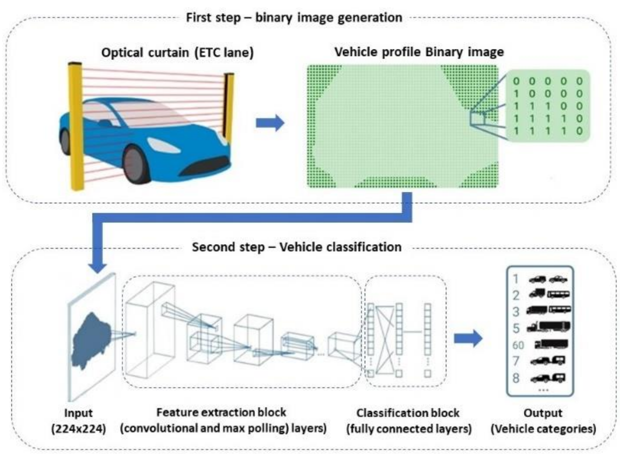

FIGURE 5. Proposed vehicle category classification methodology.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3410160

1) VEHICLE CLASSIFICATION APPLYING CNN

To classify the vehicle category, we adapt the CNN to the categories of interest. The following steps must be executed: 1) choose a specific architecture; 2) modify the last layers; 3) build a dataset to train the network; 4) complement the data set with augmented images in categories where there are not enough samples; 5) prepare a set of real images for validation; 6) determine the set of hyperparameters to train the network, and validate the results. If the results are unsatisfactory, the hyperparameters (i.e., number of Epochs, Mini Batch Size, Learn Rate, number of hidden neurons, and Regularization parameters) must be adjusted; and finally, 7) retrain and compare the results. Fig. 4 shows the blocks of the AlexNet architecture. In particular, the last block (classification layers) is replaced by the layers suitable for the context of this work.

2) INCLUSION OF VEHICLE LENGTH INFORMATION

Since the optical curtain image is binary, we used only one color channel among the three available. Thus, we can use the remaining channels to include additional information about the vehicle. Therefore, an interesting research question is whether including vehicle length information using an additional channel could improve classification performance, as we know there is a relationship between vehicle length and category. For example, small vehicles tend to be category one, and larger vehicles tend to be trucks with many axles. This artifice is justified because we lost the vehicle length information when normalizing the image to 224×224 pixels, causing a 9-axle truck to have the same image size as a category 1 car.

The strategy is to use a color channel to include the length of the vehicles in the image. To do so, we need to map the maximum and minimum possible length of the vehicle within the color range of the channel. This process is illustrated in Fig. 6, which shows a small car with a yellow background, a medium truck with a light-yellow background, and a large truck with an almost white background.

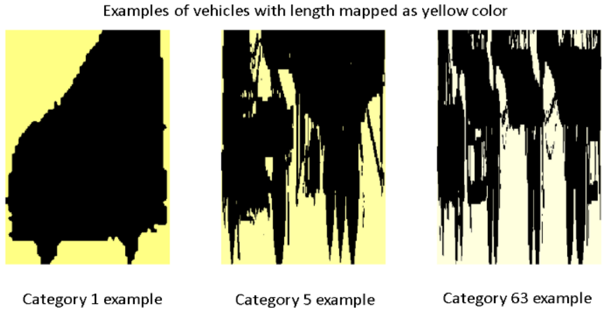

FIGURE 6. Images with a yellow background as a size indicator.

Each pixel is composed of three tones given by the R, G, and B channels. When using binary images and 8 bits per channel quantization level, the three channels are from (0,0,0) to (255,255,255). Here, we took a single channel; it is possible to choose any of the three and use this channel, where the length of the vehicles was mapped. For example, when we choose to map the velocity on channel B, each pixel is: (0,0,B) or (255,255.B). Therefore, the background is a different color depending on the length of the vehicle, and this color is within different shades of yellow.

3) HOW TO COUNT THE NUMBER OF AXLES ON THE GROUND?

We obtain the vehicle classification using CNN, considering all axles, raised or not. We want to know how many axles are on the ground in some contexts. An intuitive approach could be to find colored pixels in the last row of the image (on the ground) and count them. The main difficulties are false positives caused by exhaust fumes, rain, mudguards, or other artifacts. We propose a set of morphological operations to warrant that the pixel we choose corresponds with an axle to solve this problem. Morphological operations are widely used in literature in other fields, supporting our strategy [32].

The two fundamental morphological transformations used in our strategy are erosion and dilation. Both transformations use a structuring element that can be considered a kernel (like on convolutional operations), which is applied to the binary image to extract some features.

The erosion attenuates the surroundings of the image, which has pixels equal to 1 are described in (1):

$$A(x,y) = \min_{(x\prime,y\prime):element\ (x\prime,y\prime)\neq 0} B(x+x',y+y') \quad (1)$$

The dilation shares the same idea but amplifies the surroundings of the image, which has pixels equal to 1, as described in (2):

$$A(x,y) = \max_{(x\prime,y\prime):element\ (x\prime,y\prime)\neq 0} B(x+x',y+y') \quad (2)$$

In both equations, A is the resulting image while B is the input binary image, and it is analyzed within the boundaries of the kernel with elements described as x' and y'.

Fig. 7 illustrates those steps with an example. The original image shows the profile of a vehicle category 5, with 4 axles on the ground highlighted with a blue square and a suspended axle highlighted with a green square. In addition, the vehicle has a mudguard that touches the ground, highlighted with a yellow square, which can be detected as a false ground axle.

- In the first step in removing unwanted objects, we apply the erosion operation on the profile to reduce the noise inside the image and divide it into regions.
- In the second step, we determine the contours of each region and estimate the area.
- In the third step, we eliminate objects with small areas.
- In the fourth step, we apply the image obtained in step 3 as a mask in the original image. Then, we count the axles on the ground. Finally, we look for pixels' groups in contact with the last row.

Some categories do not suspend axles, and we only apply the steps described below in those vehicles. In this way, the method includes a previous step: finding the overall category using the CNN classification. If it belongs to the set of categories able to lift axles, then apply the morphological method. The categories that can lift axles are double wheels and have more than two axles (categories 3, 4, 5, 60, 61, 62, and 63).

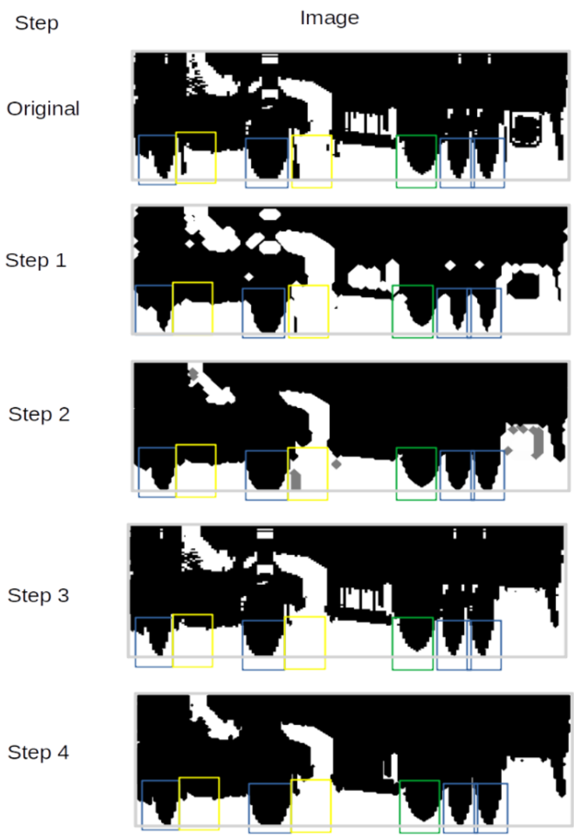

FIGURE 7. Steps to eliminate unwanted objects and count ground axles.

IV. EXPERIMENTAL DESCRIPTION

To validate the proposed classification methodology based on binary images of vehicles using small CNN architecture, we realized four experiments. In each experiment, we used a specific data set of vehicle binary images collected in three toll plazas, in Sao Paulo - Brazil, including several ETC and manual lanes, totaling 212,639 samples, containing different categories of vehicles. In Table II we describe the main objectives of each experiment, highlighting aspects such as the accuracy of classification performance for each category.

In Methodology Section “III.B”, we proposed 3 different approaches: III.B.1, III.B.2, and III.B.3. In section “IV”, we propose four experiments: IV.A, IV.B, IV.C, and IV.D. Each experiment was designed to individually validate each of the approaches presented in III.B.

TABLE II
EXPERIMENT TITLE AND MAIN OBJECTIVE

Experimental Description	Main objective
Experiment 1 - Accuracy Assessment with uniform distribution of vehicles by category	To evaluate the CNN proposed in III.B.1, we test CNN in a controlled way, where the accuracy of the CNN classifier was evaluated individually, using the same amount of data for each of the 11 categories
Experiment 2 - Accuracy Assessment with Vehicle Samples in real situation	To evaluate the CNN proposed in III.B.1, but unlike Experiment 1, now in a real ETC application, using data collected from four consecutive days of operation at three different toll plazas
Experiment 3 - Accuracy Assessment with addition of vehicle length information	To verify and compare the CNN architecture tested in Experiment 1, with new training including the vehicle length information, as proposed in III.B.2
Experiment 4 – Evaluation of Axle Counting Performance and Assessment of Vehicle Speed	a) To evaluate the methodology proposed in III.B.3 (count only the axles in contact with the road.). b) To Compare results between slower lanes (manual) and fast lanes (automatic)

Experimental data were collected using optical curtains installed in manual and ETC lanes of toll plazas, located on highways in the state of São Paulo. The optical curtains used are 180 cm high with 59 uniformly distributed LED sensors and a scanning frequency of 240 Hz. The resulting images consisted of a matrix of 59 rows by M columns, with M being the variable width of the image, which depends on the length and speed of the vehicle. The matrix elements are pixels with binary values. We supplied the datasets (with 212,639 binary images of vehicles), used in experiments of this work, at the Mendeley Open Data Repository [33]:

https://data.mendeley.com/datasets/zsfz3chjzj/2.

The dataset fed the CNN, where the last three layers were modified to adapt to the vehicle classification problem. Additionally, a new layer with 11 categories replaced the last layer. Next, we present the objectives and datasets used in each experiment.

A. EXPERIMENT 1 - ACCURACY ASSESSMENT WITH UNIFORM DISTRIBUTION OF VEHICLES BY CATEGORY

This controlled experiment evaluated the classifier's accuracy individually for each of the 11 categories. In addition, we aimed to identify specific classification problems in each category. Moreover, we built a dataset with at least 1,000 vehicle samples for each category.

Table III discriminates the composition of the dataset used in experiment 1, in a total of 15,521 images. For each of the

11 vehicle categories, the number of vehicle images used for training and validation is presented. In the case of training images, all images in categories 1, 2, 3, 4, and 60 are original. In contrast, the training images of the other categories include augmented images.

TABLE III
EXPERIMENT 1: NUMBER OF IMAGES USED IN TRAINING AND VALIDATION

Class Type	Training Data			Validation data
	Original	Modified	Total	
1	1034	0	1034	1000
2	1000	0	1000	973
3	1000	0	1000	1006
4	1029	0	1029	510
5	613	399	1012	324
60	1001	0	1001	497
61	208	799	1007	112
62	2	1000	1002	10
63	682	350	1032	828
7	90	999	1089	45
8	29	998	1027	24

We used the Data Augmentation technique since some categories do not have a significant vehicle volume. The criterion for choosing the transformations was keeping vehicles' fundamental characteristics: the number of axles and preserving their morphology. Within the transformations, we highlight: including a horizontal noise line, keeping below 10% of the original image, including a noise column keeping below 10% of the image, shifting the image vertically between 5 and 15%, stretching the image horizontally by duplicating some columns by 10%. The images used for validation were not modified.

Once the network is modified, it must be trained again. It used 10,192 images for training and 5,329 images for validation, distributed among the 11 categories. We used the SGDM as the training algorithm. Table IV presents the hyperparameter configurations that we used in Experiment 1.

TABLE IV
HYPERPARAMETERS USED TO TRAIN THE CNN/ALEXNET

Parameter = Value	Parameter = Value
Momentum = 0.90	L2-Regularization = 1.00e-04
InitialLearnRate = 1.00e-03	Gradient-ThresholdMet = l2norm
LearnRateSchedule = ' none'	Gradient-Threshold = Inf
LearnRateDropFactor = 0.1000	MaxEpochs = 50
LearnRateDropPeriod = 10	MiniBatchSize = 32-64-128

The hyperparameter value was chosen based on the Mathworks documentation [27]. Once verified that the solution converged, the focus shifted to determining two parameters: the number of training epochs and the mini-batch size. It was executed by training the network with three mini-batch sizes (32, 64, and 128). Additionally, the number of epochs varied from 1 to 50. For example, the best global accuracy was obtained with 29 training epochs, a mini-batch size of 32.

B. EXPERIMENT 2 - ACCURACY ASSESSMENT WITH VEHICLE SAMPLES IN REAL SITUATION

The goal was to test the CNN trained in Experiment 1 in a real ETC application, using data collected from four consecutive operation days in three different toll plazas (P1, P2, and P3) to measure the global performance of the proposed methodology. In a practical ETC application, it is sought to have 100% of vehicles classified correctly. The ones, which are not classified correctly and automatically, must be reclassified manually by an operator or technician. This is why global accuracy matters, as its complement (100% - accuracy) indicates how much manual operations are required.

This experiment also aimed to evaluate environmental factors, such as rain and the lane type (manual, ETC) on classification accuracy. However, to get a more detailed analysis, we focused on a specific toll plaza, P2, and observed performance every hour on a specific day. We chose two different 12-hour intervals, with and without rain, and observed, as we expected, that the worst performance was in the interval with the heaviest rain.

We used the same network trained in Experiment 1, but we collected data from three toll plazas labeled P1, P2, and P3, and the data was collected in continuous intervals. The experiment used a total of 194,361 vehicles. As summarized in Table V, we collected 92,841 samples from P1 and 75,869 from P2. Additionally, we collected 25,651 vehicles from P3.

TABLE V
EXPERIMENT 2: NUMBER OF IMAGES AND PERIOD OF OBSERVATION

Station	Period		Number vehicles
	Start	End	
P1	*02/10/2021 00:00hr*	*02/14/2021 23:59hr*	92841
P2	*02/10/2021 00:00hr*	*02/14/2021 23:59hr*	75869
P3	*02/11/2021 00:00hr*	*02/14/2021 23:59hr*	25651

C. EXPERIMENT 3 - ACCURACY ASSESSMENT WITH THE ADDITION OF VEHICLE LENGTH INFORMATION

This experiment aimed to verify the influence of the vehicle length information, suppressed in the normalization process, through the addition of color channels. We used the same dataset of Experiment 1, described in Table III, as well as the same hyperparameters described in Table IV.

D. EXPERIMENT 4 - EVALUATION OF AXLE COUNTING PERFORMANCE AND ASSESSMENT OF VEHICLE SPEED

This experiment aimed to measure the accuracy of the classification system typically used in toll plazas to evaluate the performance of the classification methodology proposed here.

The data were collected, during a continuous period, in five lanes, four manual and one automatic. Table VI shows the

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3410160

composition of vehicles by category, separating them by type of lane (manual and automatic). The dataset includes both binary and camera images to help validate classification. For categories with more than two axles, situations with all axles in contact with the road and suspended axles were considered.

TABLE VI
EXPERIMENT 4: NUMBER OF IMAGES OF VEHICLES USED IN MANUAL AND AUTOMATIC LANES

Class Type	Manual	ETC	Subtotal
1	2539	218	2757
2	346	167	513
3	251	129	380
4	94	98	192
5	72	91	163
60	136	144	280
61	39	41	80
62	0	6	6
63	39	104	143
7	18	0	18
8	1	0	1
Total	3535	998	4533

V. RESULTS AND DISCUSSION

We now present the results obtained in the four experiments described earlier. We used the accuracy index and the F1 score (3) as performance evaluation metrics. We also present a qualitative analysis of the causes of failures that occurred in the classifications of some categories.

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \tag{3}$$

Precision and Recall values are related to false positives and false negatives, respectively. False positive values are vehicles from other categories that were commissioned in the reference category. Likewise, false negative values are vehicles, from the reference category, that were omitted from this reference category, that is, erroneously commissioned in other categories.

Details regarding the utilization of the F1-Score for classification accuracy evaluation are provided in reference [37]. The purpose of adopting this metric is to validate the accuracy obtained from the Confusion Matrix. For instance, a high F1-Score is expected to correspond to a high accuracy index, whereas a low F1-Score is expected to align with a lower accuracy index. This correlation helps ensure the reliability of the accuracy assessment.

A. RESULTS OF EXPERIMENT 1 - ACCURACY ASSESSMENT WITH UNIFORM DISTRIBUTION OF VEHICLES BY CATEGORY

The results of this experiment are summarized in the confusion matrix (Fig. 8), which also brings the results of the F1 score. The results indicated an overall performance of 98.02%. According to the confusion matrix, the highest accuracies occurred in categories 60, 62, and 63 (above 99%), while, according to the F1 score, only category 63 had an outstanding value (0.999). Categories 1 to 4 had accuracies between 98 and 99%, a result similar to that indicated by the F1 score, adding category 60: from 0.983 (cat 60) to 0.987 (cat 1). The effort to improve global accuracy should focus on categories 1 to 4 and 60, as they represent more than 90% of the vehicle volume.

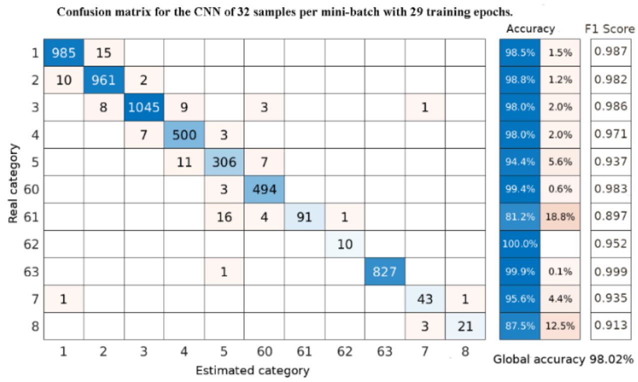

FIGURE 8. Confusion matrix, obtained from Experiment 1.

Analyzing the classification accuracy index, it is observed that, for seven categories, the trained CNN achieved performance above 98%, and in the F1 score, there are more than five categories above 0.98. However, for the four categories, the performance was lower; with the F1 score, six categories were below 0.97.

We now discuss the causes of failures in some categories. In particular, we analyze the gaps between categories 1 and 2, as they concentrate on the most significant volume of vehicles. Moreover, categories 61 and 8 presented the worst performances of the proposed methodology.

For better understanding, we first describe the characteristics of those categories. According to Table I, category 1 refers to a vehicle with two axles and a single wheel (for example, a car or pickup), and category 2 is a vehicle with two axles and a double wheel (for example, a truck or bus). Category 8 is a car with a trailer totaling four axles and single wheels. Moreover, category 61 is a truck with seven axles, of which six are double wheels.

Category 1 had a 1.5% classification failure; 15 vehicles were omitted from this category and commissioned in Category 2. Fig. 9 shows the image of the optical curtain and its picture of a typical case. It is a pickup with a trunk, which does not have dual wheels, but has similar characteristics to many Category 2. An alternative to this type of case would be to review the CNN penultimate layer design. The output of this layer is a number representing the confidence index from 0.0 to 1.0. In the analyzed example, the confidence index calculated for Category 2 was 0.66, and for Category 1 was 0.30. In 9 of the 15 Category 2 commission failures, the confidence index was less than 0.8. However, for Category 1, with 985 vehicles, the confidence index of the classification was higher than 0.9. Based on these results, a future study should be considered, where the impact of the adjustment on the CNN activation function could be explored to improve performance.

FIGURE 9. Example of a classification failure: a category 1 truck detected as a category 2 one.

For categories 61 and 8, which presented the lowest accuracy, both in terms of accuracy index and F1 score, it may be related to the fact that they use a training dataset with many samples of augmented images (80% for category 61 and 97% for category 8). Category 63 has an accuracy of 99,9% in experiment 1 and 100% in experiment 3, and it uses artificial images. However, the percentage of artificial data is only 33%. These outcomes likely stem from limitations in the augmentation process, which may not have adequately accounted for the variability present in real samples of these categories.

B. RESULTS OF EXPERIMENT 2 - ACCURACY ASSESSMENT WITH VEHICLE SAMPLES IN REAL SITUATION

The dataset described in Table V was processed using the CNN trained in Experiment 1. In Table VII, we discriminate the partial accuracy results by type of lane (automatic and manual) and by scenario, as described below:

- Scenario 1: vehicles from three toll plazas, in 96 h;
- Scenario 2: vehicles from toll plaza P2, in 96 h;
- Scenario 3: vehicles from toll plaza P2, in 12 h with rain;
- Scenario 4: vehicles in toll plaza P2, in 12 h without rain.

TABLE VII
EXPERIMENT 2: ACCURACY BY LANE TYPE

		Accuracy		
		Global (%)	Highest (%)	Lowest (%)
Scenario 1: (P1 + P2 + P3)	Total	96.48	98.24	90.13
	Automatic	95.77	97.81	87.21
	Manual	96.95	98.83	90.01
Scenario 2: P2	Total	96.40	98.71	86.81
	Automatic	95.08	100	71.42
	Manual	96.74	99.39	86.81
Scenario 3: P2 with rain	Total	95.69	96.66	92.18
	Automatic	94.07	97.97	88.59
	Manual	96.11	96.98	93.09
Scenario 4: P2 without rain	Total	97.90	98.71	95.97
	Automatic	96.88	98.92	93.55
	Manual	98.20	98.97	96.68

The global accuracy (Scenario 1) was 96.48%. Comparing this result with the one achieved in Experiment 1 (98.02%), we observed a slight reduction. The reduction in accuracy is likely related to the sample images (194,361 vehicles) considered in Experiment 2, which consisted of real operating situations, including a continuous period of 96 h, three toll plazas, two types of lanes (automatic and manual), in addition to different environmental conditions (day, night, sunny or cloudy weather, rain, and fog, among others). We can confirm this assumption through the results of Scenarios 3 and 4. In Scenario 3, with rain, it can be seen that the accuracy index was the lowest (95.69%). In Scenario 4, without rain, the result reached 97.90%, which is very close to the result in Experiment 1 (98.02%). The worsening of accuracy in Scenario 3 can be attributed to the noise caused in the binary image due to water splashes as the vehicle tires pass over water, as illustrated in Fig. 10.

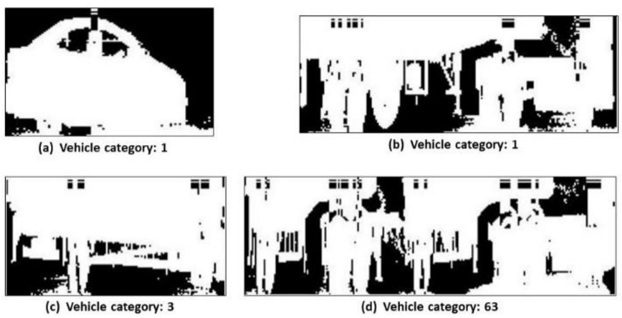

FIGURE 10. Vehicle images with noise caused by rain.

There was better accuracy of manual lanes than automatic ones in all scenarios. In particular, in the occurrence of rain (Scenario 3), the worst result of the automatic lanes is justified by the higher speed (40 km/h) and, as they usually are less protected by the marquees of the toll plazas, increasing the impact of rain on the collected binary images.

Finally, the reduction in global accuracy can also be attributed to the dataset used in Experiment 1 to train and validate the CNN, which did not contemplate the real operational conditions observed in Experiment 2. Thus, to apply the methodology proposed here, we recommend using a complete dataset, including representative images of the real operating conditions of the locality where the toll must operate.

C. RESULTS OF EXPERIMENT 3 - ACCURACY ASSESSMENT WITH ADDITION OF VEHICLE LENGTH INFORMATION

In Experiment 3, we used the same dataset as in Experiment 1, adding vehicle length information through the color channels. The classification results, using CNN with hyperparameters defined in Table IV, were summarized in a confusion matrix shown in Fig. 11. This figure also provides the calculation of the F1 score for each category. The overall result of Experiment 3 was 98.39% accuracy. On the other hand, the Experiment 1 (Fig. 7) achieved 98.02%. Therefore, we observed that the improvement in classification performance, with the inclusion of length information, was marginal.

It is worth mentioning in Fig. 11 that, in six categories (1, 3, 4, 5, 60, and 63), the value indicated by the F1 score is higher than the value of the accuracy indicated by the confusion matrix, which proves that, although improvement in the global performance was marginal, it occurred in more than half of the analyzed categories. These categories (1, 3, 4, 5, 60, and 63) represent more than 86% of the total flow volume.

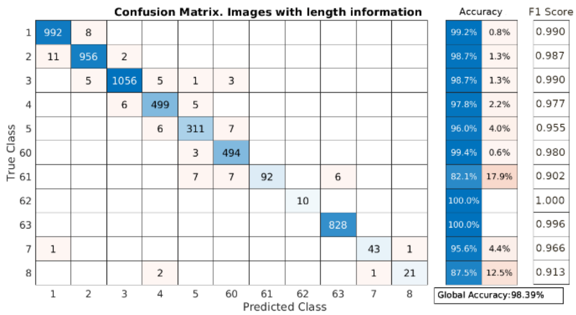

FIGURE 11. The confusion matrix resulted in Experiment 3.

We expected that adding length information to the dataset would help improve CNN's performance significantly. This non-expressive improvement can be attributed to the CNN architecture adopted here, which could not capture the vehicle category characteristic incorporated through the CNN color channel. Another factor is that vehicle length is not always directly related to vehicle category. For example, there are exceptions, such as categories 7 (three axles) and 8 (four axles) with a single wheel.

D. RESULTS OF EXPERIMENT 4 - EVALUATION OF AXLE COUNTING PERFORMANCE AND ASSESSMENT OF VEHICLE SPEED

Vehicles in Table V were classified using AVC equipment at toll plazas. Table VIII presents the classification results by lane type. The global accuracy of the evaluated system was 98.53%. Additionally, it appears that the accuracy of the axle detection on the ground had a performance of 98.27%, similar to the overall performance of the evaluated system.

TABLE VIII
EXPERIMENT 4: AVC ACCURACY BY LANE TYPE

	Number of vehicles	Accuracy Global cat (%)	Accuracy Ground axle (%)
Total	4533	98.530	98.275
Automatic	998	98.096	98.196
Manual	3535	98.648	98.296

Toll collection applications require high accuracy, close to 99,9%. For this reason, in practice, a reconciliation procedure is adopted to resolve divergence between the result of the classification of the AVC with the toll operator (manual lane) classification or the category information in the electronic tag (automatic lane). When there is a divergence between classifications, manual diligences are performed based on camera images to assign the correct category. Thus, the overall accuracy increases to nearly 99.9% after the reconciliation.

Finally, it appears that the global accuracy of the methodology proposed in this paper reached a performance comparable to that of AVC systems (98,53%), both in the controlled situation (Experiment 1 - 98,02%) and in the real situation (Experiment 2, Scenario 4 - without rain, 97,9%), indicating this methodology is promising.

E. ANALYSIS OF RESULTS

It should be noted that care must be taken when comparing different classification methods found in the literature. Each study is typically conducted under specific premises regarding the type of sensor, number of categories, classification algorithm, environmental conditions, and sample size, among other factors.

As mentioned in Section II, the paper by Tan et al. (2024) [35] presents a comparative study among different works published in the literature. In Table 2 of the aforementioned article, an analysis of seventeen papers published between 2010 and 2019 is presented, considering the aforementioned premises. It is observed that when the number of evaluated categories is greater than 10, the achieved accuracy ranges from 84% to 94%. Conversely, when the number of evaluated categories is less than 5, the classification accuracy significantly increases, ranging from 91% to 99%. It is worth noting that, according to [35], the best results were obtained using a radiofrequency sensor combined with the KNN algorithm, achieving an accuracy of 99.6% for the classification of 2 categories. On the other hand, using an infrared sensor combined with the Bayesian Algorithm, another study presented in Table 2 achieved an accuracy of 99.8% for the classification of 5 categories. It can be verified in the article by Satyanarayana, Majhi and Das (2021) [36] that although this research does not count axles, nor discriminate between suspended axles, the accuracy results achieved for 4 categories, using non-intrusive binary sensors, are within the range indicated above, ranging from 91.3% (using micro-LiDARs) to 98% (extracting video data).

Therefore, it can be concluded that the accuracy values achieved in our research, which were 98.02% for the classification of 11 categories, can be considered very satisfactory when compared to the main results known in the literature. We emphasize that the results obtained here used non-intrusive sensors and a CNN with a low-complexity architecture, aiming for real-time embedded applications, such as ETC lanes.

VI. CONCLUSIONS

In this work, we present a vehicle classification methodology using binary profile images and CNNs, achieving an overall accuracy of 98.02%, which is close to the one of classification systems typically used in toll plazas (98.53%).

In experiments involving nocturnal periods and rain occurrence, a slight performance loss was observed (96.48%). In addition, we observed that the rain introduced noise effects in the images, which impacted the overall performance of the classification. On the other hand, we found that including vehicle length information via the color channel resulted in a marginal improvement in CNN performance.

The experiments have shown that the proposed methodology is promising to be applied in toll plazas. It has the advantage of being a nonintrusive solution, which increases the system's availability. Additionally, the methodology can be used to replace a current classification system or to implement a parallel audit system to the existing system.

Future work could explore the possibility of refining the methods and, for example, taking advantage of other unused features of the binary image, such as the height of the driver's cabin. Another possibility would be to use a statistical approach, considering the adjustment of the confidence index, using probabilistic calculations associated with historical data.

Finally, we emphasize that the applicability of the proposed methodology can be extended to other contexts, such as the free-flow toll system, through the use of other types of sensors, such as LIDAR or cameras.

REFERENCES

[1] N. J. Garber and L. A. Hoel, “Traffic and Highway Engineering”, 5th ed. Boston, MA USA, 2018.

[2] X. Hu, S. Chang, J. Li, & Y. Qin. "Energy for sustainable road transportation in China: Challenges, initiatives and policy implications", *Energy*, vol 35, pp 4289-4301, 2010. doi:10.1016/j.energy .2009.05.024

[3] CNT, "National Confederation of Transport Report 2021", Available at: https://pesquisarodovias.cnt.org.br/downloads/ultimaversao/Pesquisa_CNT_Rodovias_2021_Web.pdf, (accessed Feb. 1, 2022).

[4] W. Tong, A. Hussain, W. X. Bo, and S. Maharjan, "Artificial Intelligence for Vehicle-to-Everything: A Survey," IEEE Access, vol. 7, pp. 10823-10843, 2019, doi: 10.1109/ACCESS.2019.2891073.

[5] Y. -K. Ki and D. -K. Baik, "Vehicle-Classification Algorithm for Single-Loop Detectors Using Neural Networks," IEEE Transactions on Vehicular Technology, vol. 55, no. 6, pp. 1704-1711, Nov. 2006, doi: 10.1109/TVT.2006.883726.

[6] Federal Highway Admin. (FHWA), "Traffic monitoring guide", Washington, DC, USA, Rep., Oct. 2016. [Online]. Available at: https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_fhwa_pl_17_003.pdf

[7] ANTT - Terrestrial Transport National Agency, "Brazil's Highway Concessions," Available at: https://antt-hml.antt.gov.br/web/guest/pedagio (accessed Feb. 1, 2022)

[8] W. Balid, H. Tafish and H. H. Refai, "Intelligent Vehicle Counting and Classification Sensor for Real-Time Traffic Surveillance," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 6, pp. 1784-1794, June 2018, doi: 10.1109/TITS.2017.2741507.

[9] G. Leduc, "Road traffic data: collection methods and applications," JRC Tech. Rep., 2008.

[10] N. Shvai, A. Hasnat, A. Meicler and A. Nakib, "Accurate Classification for Automatic Vehicle-Type Recognition Based on Ensemble Classifiers," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 3, pp. 1288-1297, March 2020, doi: 10.1109/TITS.2019.2906821.

[11] S. K. Rajput, J.C. Patni, S.S. Alshanrani, V. Chaydhari, A. Dumka, R. Singh, M. Rashid, A. Gehlot and A. S. AlGhamdi, “Automatic Vehicle Identification and Classification Model Using YOLOv3 Algotithm for a Toll Management System”, Sustainability, vol.14, no.15 pp. 1-15. doi: 10.3390/su14159163.

[12] H. Dong, X. Wang, C. Zhang, R. He, L. Jia, and Y. Qin, "Improved Robust Vehicle Detection and Identification Based on Single Magnetic Sensor," IEEE Access, vol. 6, pp. 5247-5255, 2018, doi: 10.1109/ACCESS.2018.2791446.

[13] B. González, F. J. Jiménez, and J. De Frutos, “A Virtual Instrument for Road Vehicle Classification Based on Piezoelectric Transducers,” *Sensors*, vol. 20, no. 16, p. 4597, Aug. 2020, doi: 10.3390/s20164597.

[14] A. Amodio, M. Ermidoro, S. M. Savaresi, and F. Previdi, "Automatic Vehicle Model Recognition and Lateral Position Estimation Based on Magnetic Sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 5, pp. 2775-2785, May 2021, doi: 10.1109/TITS.2020.2974808.

[15] N. Gothankar, C. Kambhamettu and P. Moser, "Circular Hough Transform Assisted CNN Based Vehicle Axle Detection and Classification," *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*, pp. 217-221, 2019, doi: 10.1109/ICITE.2019.8880232.

[16] A. T. Sasongko and M. Ivan Fanany, "Indonesia Toll Road Vehicle Classification Using Transfer Learning with Pre-trained Resnet Models," *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 373-378, 2019, doi: 10.1109/ISRITI48646.2019.9034590.

[17] C. Liu, D. Q. Huynh, Y. Sun, M. Reynolds and S. Atkinson, "A Vision-Based Pipeline for Vehicle Counting, Speed Estimation, and Classification," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 12, pp. 7547-7560, Dec. 2021, doi: 10.1109/TITS.2020.3004066.

[18] B. Zhang and J. Zhang, "A Traffic Surveillance System for Obtaining Comprehensive Information of the Passing Vehicles Based on Instance Segmentation," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 11, pp. 7040-7055, Nov. 2021, doi: 10.1109/TITS.2020.3001154.

[19] K. F. Hussain, M. Afifi, G. Moussa, “A Comprehensive Study of the Effect of Spatial Resolution and Color of Digital Images on Vehicle Classification,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 3, pp. 1181-1190, June 2018, doi: 10.1109/TITS.2018.2838117.

[20] R. Nezafat, V. Sahin, and M. Cetin, "Transfer learning using deep neural networks for classification of truck body types based on side-fire lidar data." Journal of Big Data Analytics in Transportation vol 1.1, pp 71-82, 2019, doi:10.1007/s42421-019-00005-9.

[21] J. Wu, H. Xu, Y. Zheng, Y. Zhang, B. Lv, and Z. Tian, "Automatic vehicle classification using roadside LiDAR data." Transportation Research Record vol 2673.6, pp 153-164, 2019, doi: 10.1177/0361198119843857.

[22] D. Wang, C. Watkins and H. Xie, “MEMS Mirrors for LiDAR: A Review”, Micromachines, vol. 11, pp. 1-24. 2020, doi: 10.3390/mi11050456.

[23] E. Alexandre, L. Cuadra, S. Salcedo-Sanz, A.Pastor-Sánchez, and C. Casanova-Mateo, "Hybridizing extreme learning machines and genetic algorithms to select acoustic features in-vehicle classification applications." Neurocomputing vol 152, pp 58-68, 2015, doi:10.1016/j.neucom.2014.11.019.

[24] B. Sliwa, N. Piatkowski, and C. Wietfeld, "The Channel as a Traffic Sensor: Vehicle Detection and Classification Based on Radio Fingerprinting," iIEEE Internet of Things Journal, vol. 7, no. 8, pp. 7392-7406, Aug. 2020, doi: 10.1109/JIOT.2020.2983207.

[25] Banner, “A-GAGE MINI-ARRAY Two Piece Measuring Light Curtain: Instruction Manual”, Available at: https://www.bannerengineering.com/us/en/products/sensors/measuring-arrays/two-piece-measuring-light-screen.html#sort=relevancy (accessed Nov. 15, 2022)

[26] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks,"*ACM*

Communications, v. 60, no. 6, pp. 84-90, June 2017, doi.org/10.1145/3065386.

[27] MATLAB. Pretrained deep neural networks. MathWorks Help Center, 01 Jan. 2020. Available at: https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html, (accessed Dez. 1, 2021).

[28] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. M. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics, and Transfer Learning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[29] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning." *arXiv preprint arXiv:1712.04621* (2017).

[30] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala and C. O. Aigbavboa, "A Comparative Analysis of Gradient Descent-Based Optimization Algorithms *on Convolutional Neural Networks," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pp. 92-99, 2018, doi: 10.1109/CTEMS. 2018.8769211.

[31] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 2018, arXiv:1804.07612.

[32] R. M. Haralick, S. R. Sternberg and X. Zhuang, "Image Analysis Using Mathematical Morphology," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 532-550, July 1987, doi: 10.1109/TPAMI.1987.4767941.

[33] J. Barreyro L. R. Yoshioka, and C. L. Marte, "Side Profile Vehicle Binary Images," Mendeley Data, V2, (2021), Available at: https://data.mendeley.com/datasets/zsfz3chjzj/2, (accessed March 17, 2024).

[34] B. Maiga, Y. Dalveren, A. Kara and M. derawi, "Convolutional Neural Network-Based Vehicle Classification in Low-Quality Imaging Conditions for Internet of Things Devices",Sustainability, v.15(23), 2023, doi: /10.3390/su152316292

[35] S. H. Tan, J. H. Chuah, C-O. Chow, J. Kanesan and H. Y. Leong, "Artificial intelligent systems for vehicle classification: A survey", Engineering Applications of Artificial Intelligence, v. 129, 2024, doi: 10.1016/j.engappai.2023.107497

[36] G. S. R. Satyanarayana, S. Majhi and S. K. Das, "A Vehicle Detection Technique Using Binary Images for Heterogeneous and Lane-Less Traffic," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-14, 2021, Art no. 5007514, doi: 10.1109/TIM.2021.3062412.

[37] S. Raschka, "An Overview of General Performance Metrics of Binary Classifier Systems", Performance Metrics. 2014. Available at: https://arxiv.org/pdf/1410.5330.pdf, (accessed March 17, 2024).

JOAQUIN BARREYRO holds a degree in Computer Engineering - from Universidad Nacional del Litoral (2010) and a master's degree in Electrical Engineering from the University of São Paulo (2022). He is currently a software architect - at Tecsidel. In addition, he has experience in the field of Engineering, with an emphasis on Automation, Toll Control, and Intelligent Transportation Systems.

LEOPOLDO RIDEKI YOSHIOKA is B.Sc. (1984) in Electronic Engineering from Instituto Tecnologico de Aeronautica (1984). M.Sc. (1988) and Ph.D. (1991) in Electronic Engineering from Tokyo Institute of Technology (TITech). Assistant Professor (1991-1992). of the Department of Information Processing at TITech. Research and Development Coordinator at COMPSIS Computers and Systems (1997-2010). Currently is a Professor of the Electronic Systems Engineering Department at Escola Politecnica of Universiade de Sao Paulo (USP). Research areas include Intelligent Transportation Systems, Automotive Embedded Systems, and Engineering Innovation.

CLAUDIO LUIZ MARTE Graduated with Full Degree in Electronics from the Federal University of São Carlos [UFSc] (1985) and in Electrical Engineering from the Polytechnic School of USP (1988), master's degree (1994) and doctorate (2000) in Electrical Engineering by EPUSP. He is currently a professor at the Department of Transport Engineering at EPUSP (since 2009). He was director of the Center for Information Technology, Automation and Mobility (CIAM) at the Institute of Technological Research of São Paulo (IPT) from 2004 to 2008. Since 1997, he has been working in the Intelligent Transport Systems, and micro/macro transport and traffic simulation models.

CAIO GARCIA PICCIRILLO holds a Control and Automation Engineering degree from Instituto Mauá de Tecnologia (2017). He is currently a master degree candidate in information engineering with an emphasis on artificial intelligence at the Federal University of ABC. He is currently an R&D Engineer at JKControl Developer at Tecsidel and Assitant Professor at Instituto Mauá de Tecnologia. In addition, he has experience in Computer Science, with an emphasis on Methodology and Computer Techniques.

MAX MAURO DIAS SANTOS received a Ph.D. degree in industrial engineering at Universidade Federal de Santa Catarina, Brazil (2004). He was a Postdoctoral Fellow in electrical engineering at the Universidade de Aveiro, Portugal (2005/2006). His career has spanned academia and industry in the field of automotive, automation, and control. Since 2013, he has been an Associate Professor with the Department of Electronics at the Universidade Tecnológica Federal do Parana - Ponta Grossa, Brazil. His research interests are in the fields of automotive, automation, and autonomous systems. He is a senior member of IEEE and an Associate Editor of IEEE Transactions on Intelligent Vehicles.

JOÃO FRANCISCO JUSTO is B.Sc. (1988) and M.Sc. (1991) in Physics from the University of São Paulo and Ph.D. (1997) in Nuclear Engineering from Massachusetts Institute of Technology. He is currently a Full Professor at the Escola Politécnica of the University of São Paulo. In addition, he was a visiting associate professor at the University of Minnesota (2007-2008). He has experience in the computational modeling of nanomaterials and embedded electronics.