An Application of Engineering Software in Traffic Image Processing: A review based on traditional methods

Lucas Fernandes Ferreira Costa

Reinaldo da Silva Araujo

Ford Motor Company Brasil Ltda

ABSTRACT

The competitive industrial environment motivates automotive companies to continuously search for initiatives that disrupt problem solving and decision making, especially a better understanding of how customers use their vehicles and interacts with such technological experience and diverse use. It is impressive the amount of research related to image recognition and the application of computer vision to get unclear information related to driver intention and the environment during driving events. Nowadays, in-vehicle sensors and online city cameras facilitate almost continuous monitoring of vehicles and traffic, contributing to analytical assessments that are increasingly enhancing the studies about autonomous vehicle and smart cities. This work presents an academic application of computer vision based on engineering software in comparison with a traditional approach using Python. The public data source consists of Brazilian traffic images, and labels will provide tabular information about that environment. A classification model will be built to analyze the conduct of heavy vehicles and will be compared by both methods. Find a way to leverage support tools exploiting Machine Learning and Image Processing seems an innovative and intelligent pathway for customer-centric companies.

INTRODUCTION

This work starts with a bibliometric analysis of recent publications on the topic of traffic image processing. Using the Scopus portal [1], over 14,000 articles were published over the past ten years with the subject as a keyword. To provide an analysis of the most current and widely used techniques, Figure 1 maps the interrelation of these keywords, their temporal trends, and their amount of occurrence in these articles. This approach enables the identification, measurement, and quantification of the scientific production in this subject, facilitating the analysis of the cutting-edge knowledge.

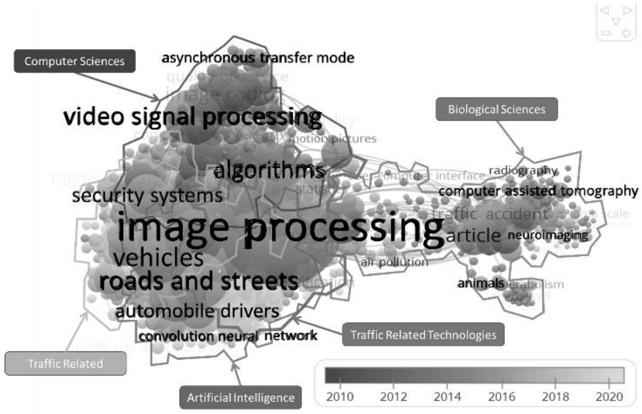

Figure 1. Map of Keywords Interrelationship

From this analysis, it was found that Convolutional Neural Networks (CNNs) are the most modern and have been an extensively discussed approach in this field [2]. The map also revealed five major clusters that serve as the primary sources of information:

- Computer Science
- Artificial Intelligence
- Traffic Related
- Traffic Related Technologies
- Biological Sciences

A systematic review showed that intelligent transportation systems require a deep evaluation of the data collection aspects. In this regard, camera-based systems offer several advantages and unique features that enhance the quality of information, such as their flexibility and efficiency to acquire information from the environment. One important aspect to consider in traffic image processing though is the variability of lighting conditions (sunset, fog, night, etc.) as well as the vehicle segmentation within the image, since a single frame may not fully capture the entire object and may result in inconsistencies requiring specific treatments for the system to function properly [3].

In terms of detection and classification of vehicles in multiple lanes, one must establish some initial image tuning and calibration, so effective decisions can be made from the input data. Masking techniques, whether digital, such as edge detection; as well as background image extraction, such as empty roadways and lanes, are use cases found in the literature The other thing that must be considered in the traffic image processing exercise is the tracking of vehicle movement. Due to the counting process, lane changes and other situations, objects can be positioned on top of each other [4].

When analyzing the problems related to traffic behavior in the current society, there are much knowledge about situations in urban regions related to congestion and its consequences for traffic, but studies are more focused on image processing looking for solutions to the unstable condition of traffic management, as the example of large cities such as São Paulo, Rio de Janeiro, and Texas [5].

As considered by the scientific community as effective enough for this type of analysis, the issue of Brazilian traffic was proposed. In this sense, a search was conducted looking for traffic control cameras monitoring public roads within the Brazilian territory [6]. A wide range of options were found that can be used in this comparison of methods, as per Figure 2 below.

Figure 2. Brazilian Traffic Control Camera Images

After acquiring image sources that were publicly available, a behavior analysis was conducted considering the Brazilian Traffic Code (CTB) regarding traffic and conduct rules. Article 29, paragraph IV, reads as follows:

"IV - when a roadway has multiple lanes in the same direction, the right lanes are intended for slower and larger vehicles when there is no dedicated lane for them, while the left lanes are intended for overtaking and faster vehicles" [7].

Computer Vision is the area of study and sub-field of Artificial Intelligence that help computers to see and understand the content of digital images [8]. To improve the performance of the image classification task, i.e., assign label or tag to an image, deep learning models are used on the promise of an automated feature learning, as a specialized method making use of general machine learning algorithm with less digital signal processing expertise to train and operate. As described above in [2], these deep learning methods are achieving state-of-art results on challenging problems.

The image classification will involve tabular data based on cases, containing the quantity of each type of transportation present in the image, as an input for the Artificial Intelligence training. In addition, the conduct of the respective heavy vehicles based on the CTB guidelines mentioned above will be used as target for these cases.

However, this work does not aim to be based on already confirmed premises and conditional exceptions regarding image treatment and preprocessing of the data. The intention is not to obtain an extensive database that guarantees highly refined training. Instead, this study aims to demonstrate the potential and application of these tools in a totally different context, more aligned with the subject of mobility engineering and public roads safety.

The central question of this study is, based on traditional methods, what is the performance of an image analysis engineering software using tabular data in comparison with an artificial intelligence algorithm using Python coding. These applications have been commonly used in industry, such as identifying defects in casting parts [9], or in biomedicine for predictive diseases diagnosis [10], having as initial database images and metadata capable of being integrated and used in training and validation tests using a deep learning model.

METHODOLOGY

Nowadays, it is impressive the amount of data related to driving events and safety on the road. That goes from connected vehicle data until online traffic cameras, and it enables the development of research and application of Artificial Intelligence (AI) algorithms over different contexts inside the Engineering. For this project, an online traffic camera at São Paulo – Brazil was chosen as input variable for the image recognition process. To have a better understanding of the environment and thus, providing more reliable information to train the model, tabular data was used to correlate the state provided by the online traffic camera. The objective of the AI model is to classify the conduct of Heavy Vehicles based on its position on the road.

The dataset creation lies on getting multiple online traffic images of the road SP088 at kilometer 45.5 from the DER Sao Paulo (Departamento de Estradas de Rodagem do Estado de São Paulo) [11]. This traffic camera was chosen upon the others due to its position on the road, having the same direction of the closest roadway. The dataset covered three weekdays in May 2023 and has 300 different images collected. After that, tabular information was used to infer the quantity of light vehicles, motorcycles, and heavy vehicles in that image. As the target of this research, the respected conduct of that heavy vehicle was inferred, having the classes a balanced weight within the population of image cases. The dataset can be understood by the given schema:

- Independent Variables:
    - Online Traffic Camera as Image
    - Number of Vehicles as Values
    - Number of Motorcycles as Values
    - Number of Heavy Vehicles as Values
- Dependent Variable:
    - Heavy Vehicle's Conduct as Classes {Nao se Aplica, Boa Conduta, Ma Conduta}

Figures Figure 3, Figure 4 and Figure 5 highlight the three classes mentioned above regarding the position of the heavy vehicle on the road and thus, its behavior in traffic. It is important to mention that any kind of heavy vehicle present outside the rightest traffic lane was assumed as a misconduct class. Overtaking events were disregarded and considered as misconduct as well.

Figure 3. Not Applicable class (Nao se Aplica)

Figure 4. Good Conduct class (Boa Conduta)

Figure 5. Misconduct class (Ma Conduta)

After the dataset creation, a distribution was made into training and test image datasets in proportion of 80% and 20% respectively, 240 images for training and 60 images for test. This approach was used by Ribeiro *et al.* [12] and followed by this present project.

Additionally, to ensure a comprehensive state of this research in comparing AI methodologies, it is important to admit a balanced dataset to avoid any kind of further processing or technique to overcome a skewed class proportion. Therefore, for both train and test it has been considered a balanced dataset over all three classes, with the same quantity of cases and weights.

ODYSSEE A-EYE – In one side of the comparison review, an engineering software was used to create a classification model to predict the road conduct classes. ODYSSEE A-Eye is a decision support tool that leverage Machine Learning (ML), image processing and reduced order modelling technology to propose predictions based on a database from multiple sources: Images, qualitative and quantitative data, 3D models, curves, etc. [9]. Several benefits make this application a suitable choice for this project, as having a reduced computing effort and adaptive learning with low code, and the mixture of image recognition and tabular data in the learning database. Existing examples of classifying pathology and severity of a patient based on thorax x-ray images and their physical characteristics, as well as predicting defects of a casting 3D part based upon its image [9], were remarks and have motivated the development of this research. ODYSSEE A-Eye receives the dataset based a customized layout and reads the traffic images as JPG formats and the tabular data as arrays in a CSV file.

Among different types of prediction solvers available, a user script to compare interpolation methods was run to define the best model based on the predicted and the actual values. This kind of process helps the user to choose a suitable method for the problem and avoid multiple iteration to find the best solution. The user script runs several interpolation methods and returns the model with the lower Euclidian norm found. It is important to mention that ODYSSEE A-Eye runs the ML solver under the QUASAR environment that has some implemented models, but also allowing user's customized algorithms in a native language source. For this project, Kriging interpolation method was chosen based on the script recommendation.

In terms of image processing, a QUASAR script is automatically run with multiple pre-defined functions that goes upon color reconstruction and edges detection, for example.

The output is generated in the workspace as a CSV file that contains the predicted classes in a binary vector shape. Due to the lack of a visualization tool in ODYSSE A-Eye, Python codes had to be run to perform the validation analysis.

FASTAI – On the other side of this study, it was used a powerful and accessible tool for building and training deep learning models. FastAI is a user-friendly interface, offering a high-level API's library that reduce the complexity of machine learning process. These cutting-edge libraries are designed to saving time of developing AI models for beginners to experienced practitioners. This objective is achieved due to the remarkable features available that comes from learning architectures until crucial techniques, such as an easy-to-use process for data augmentation. The deep learning models such as DenseNet, VGG, and ResNet that have been trained on notable datasets like ImageNet are ready to use, bringing acceleration in the pace of developing innovation in that field [13].

This application can be accessed through Google Colaboratory system, also known as the Jupyter performing programing tasks notebook in a web browser environment, that allow users to access high performance CPU's such as TPU's and GPU's. This is a powerful resource that support the model development since FastAI integrates popular frameworks available, i.e., PyTorch and TensorFlow [14] that demands an enormous quantity of calculations, reducing the time spent for each iteration from hours to minutes.

The steps taken to run this experiment was decided to be as easy as possible to have sufficient similarity with the other application in this study.

Once a new workbook has been created in Colab, it is necessary to install the FastAI Python's packages and import the libraries with the functionalities needed to begin to develop the computer vision code. The entire data was loaded by using the available resources with any special pre-processing of images or data augmentation and the only preparation was to let the API in charge of creating both datasets for training and test. The proportion used was 20% to be separate randomly as a test dataset.

A Python's dictionary [15] was created by reading a Microsoft Excel spreadsheet with tabular data containing in the rows the filename of each image with the respective values of independent and dependent variables, where the columns were:

- NomeArquivo – filename of JPEG image,
- Vehicles – number of vehicles,
- Motorcycles – number of motorcycles
- HeavyVehicle – number of heavy vehicles,
- HVConductAtTheTraffic – string text containing the behavior perceived in the image (Figure 6)

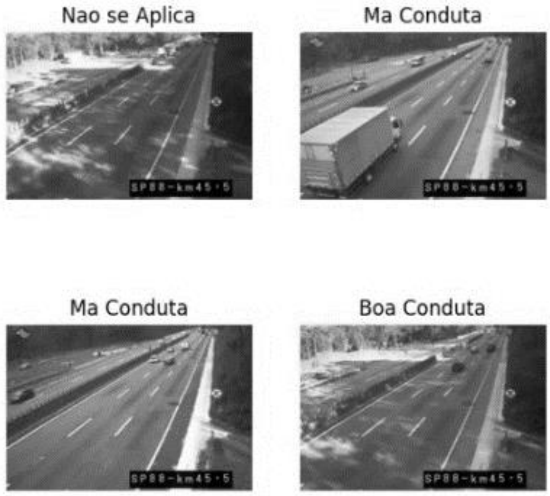

Figure 6. Values of Dependent Variable

With the data correctly loaded and processed, the ResNet was chosen as the Convolutional Neural Network model to be trained, as Wu, Shen, and Hengel confirmed that it has good results in terms of qualitative semantic image segmentation [16]. The transfer learning method [17] is automated and supported by FastAI libraries that operates by freezing certain layers while fine-tuning others to obtain the appropriated deep learning models to make the predictions.

The final step was to evaluate the results directly in the Colab notebook.

RESULTS AND DISCUSSION

In accordance with Foody [18], a confusion matrix was used to validate the results found by both classification models, as it is still at the core of the accuracy assessment literature nowadays. By checking the cross-tabulation of the true class label against the predicted label, it is possible to characterize errors and evaluate interclass confusion that could be solved by a deeper evaluation of the classification model. Lastly, the confusion matrix was displayed with the values normalized to provide a good understanding of the percentage of correct answers for each class.

In addition of the confusion matrix, some classification metrics were used to improve the comparison review. The open-source machine learning library Scikit-Learn [19] has a ready to use classification report that displays the main classification metrics: precision, recall, f1-score, and the accuracy of the model.

The results presented for both models lie on the balanced test dataset containing 20 cases for each of the three classes.

ODYSSE A-EYE – Figure 7 shows the confusion matrix given by the ODYSSE A-EYE results. The model was able to identify the classes Nao se Aplica and Boa Conduta with ease but had an interclass confusion of 50% for the actual class being Ma Conduta and the model predicting Boa Conduta.

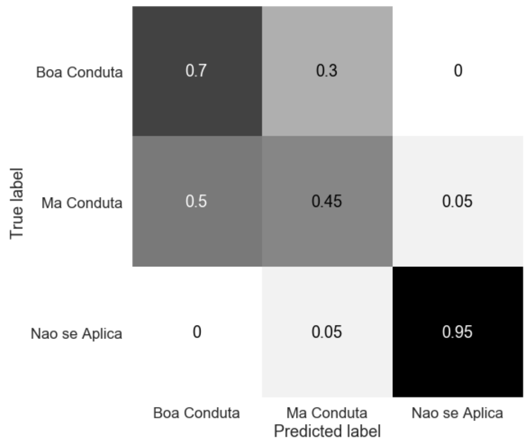

Figure 7. Confusion Matrix ODYSEE A-Eye

Following the classification metrics used, Table 1 highlights the results for this model. The use of these metrics helps to clarify what the model has the best. By the nature of this classification, that is classifying the conduct of heavy vehicles on road, the recall metric takes advantage against the others since it explains how many of the correctly predicted cases turned out to be positive. In other words, False Positives (e.g., misconduct wrongly classified) have a lower concern than False Negatives (e.g., real misconduct being classified as good conduct).

	Precision	Recall	F1-Score
Boa Conduta	0.58	0.70	0.64
Ma Conduta	0.56	0.45	0.50
Nao se Aplica	0.95	0.95	0.95
Model Accuracy			0.70

Table 1. Classification Metrics for ODYSSE A-Eye

FASTAI – Repeating the same evaluation process mentioned above, Figure 8 shows the confusion matrix given by FastAI results. As seen in the ODYSSEE A-Eye model, it is possible to identify that the best class predicted based on the true values was the Nao se Aplica. In most kind of traffic images for this class, the road appeared almost empty, and the grey pattern was purely evident. This explains the easier prediction for that class and, therefore a lower interclass confusion with the ones having heavy vehicle in their respective images.

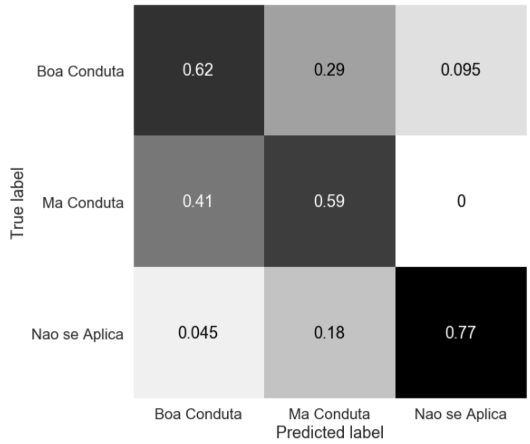

Figure 8. Confusion Matrix FastAI

Table 2 below describes the classification metrics for FastAI. Once again, the worst prediction performance lies on the class Ma Conduta which seems to be the hardest case to classify.

	Precision	Recall	F1-Score
Boa Conduta	0.62	0.62	0.62
Ma Conduta	0.50	0.59	0.54
Nao se Aplica	0.89	0.77	0.83
Model Accuracy			0.67

Table 2. Classification Metrics for FastAI

To illustrate the interclass confusion, FastAI easily enables visualization of those cases, ranking by loss and highlighting the predicted and actual classes, as well as the probability of the classification. Figure 9 shows the top losses cases of interclass confusion, where most have been between Boa Conduta and Ma Conduta, i.e., cases having heavy vehicles present. The text above images states the Predicted/Actual/Loss/Probability of classification.

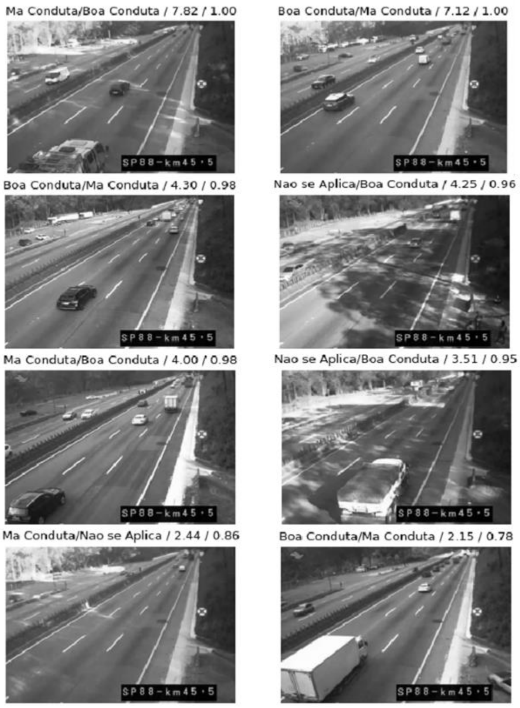

Figure 9. Top 8 Losses Cases for FastAI

By evaluating both models in terms of the recall metric, ODYSSE A-Eye has performed better in classifying the classes Nao se Aplica and Boa Conduta but had a worst performance to predict the class Ma Conduta. FastAI on the other hand shows less variability through the classes prediction and performed better when differentiating Ma Conduta from Boa Conduta. The accuracy of both models was quite comparable, and the better result from ODYSSE A-Eye comes mainly from the impressive correct classification for the class Nao se Aplica. For the nature of this study, a better recall metric for the class Ma Conduta gives FastAI an advantage in terms of image recognition and the conduct classification.

Is evident that for both models a better preprocessing on the images could be a significant benefit for the class prediction. The challenge of dealing with sunlight, images during sunset and events at the end of the road – with lower frame quality on the image, made both models don't perform as they are intended to do. As expected, improve the image segmentation, as well as expand the number of cases in the dataset show potential to enhance their classification results

CONCLUSION

The present work explores the systematic of leverage image recognition and multiclass prediction by different AI methodologies and applications. By analyzing each model and comparing their classification metrics, it was possible to understand the constraints present in the environment of mobility studies using camera-based systems. The AI algorithms have an enormous power to recognize patterns and predict driver behavior, for example, but they depend on a successful image preprocessing to perform into a reasonable Engineering field of application. Therefore, the authors strongly recommend that further studies be conducted using a more prepared dataset, even though both models have shown potential to classify conduct of a vehicle in the traffic.

The results presented are valid to verify the feasibility of those methods and the advantages of having an already built AI model, user friendly, that cuts a lot of developing time and code effort, driving the focus to the preparation of the dataset. This systematic review also enables discussions of future developments, where the image recognition and classification method could be used to help the investigation of traffic density or even predict complex scenarios in the mobility field.

REFERENCES

[1] "SCOPUS," 2023. [Online]. Available: https://www.scopus.com/ search/. [Accessed April, 28th 2023].

[2] N. J. VAN ECK and L. WALTMAN, Manual for VOSviewer version 1.6.18, Leiden: University Leiden, 2022.

[3] R. CUCCHIARA, M. PICCARDI and P. MELLO, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE transactions on intelligent transportation systems,* vol. 1, no. 2, pp. 119-130, 2000.

[4] S. GUPTE, O. MASOUD, R. F. MARTIN and N. P. PAPANIKOLOPOULOS, "Detection and Classification of Vehicles," *IEEE Transactions on intelligent transportation systems,* vol. 3, no. 1, pp. 37-47, 2002.

[5] V. JAIN, A. SHARMA and L. SUBRAMANIAN, "Road Traffic Congestion in the Developing World," *Proceedings of the 2nd ACM Symposium on Computing for Development,* pp. 1-10, 2012.

[6] GITHUB, "OSINT Brazuca (Open source intelligence ou Inteligência de Fontes Abertas)," 2023. [Online]. Available: https://github.com/osintbrazuca/osint-brazuca#cameras-online. [Accessed April, 18th 2023].

[7] BRASIL, "Lei nº 9.503, de 23 de setembro de 1997. Institui o Código de Trânsito Brasileiro.," [Online].

Available: http://www.planalto.gov.br/ccivil03/Leis/L9503.htm. [Accessed May, 15th 2023].

[8] J. BROWNLEE, Deep Learning for Computer Vision, 2020.

[9] HEXAGON, "ODYSSEE," 2023. [Online]. Available: https://hexagon.com/products/odyssee. [Accessed May, 20th 2023].

[10] Y. TIAN, "Integrating image and tabular data for deep learning," Towards Data Science, 2020. [Online]. Available: https://towardsdatascience.com/integrating-image-and-tabular-data-for-deep-learning-9281397c7318. [Accessed May, 15th 2023].

[11] DER, "Câmeras Online," 2023. [Online]. Available: http://www.der.sp.gov.br/WebSite/Servicos/ServicosOnline/CamerasOnlineMapa.aspx. [Accessed May, 8th 2023].

[12] V. RIBEIRO, V. GREATI, A. BEZERRA , G. SILVANO, I. SILVA, P. T. ENDO and T. LYNN, "Brazilian Mercosur License Plate Detection: a Deep Learning Approach Relying on Synthetic Imagery," in *Brazilian Symposium on Computing System Engineering (SBESC)*, Natal, 2019.

[13] J. HOWARD and S. GUGGER, "Fastai: A layered API for deep learning," *Information*, vol. 11, no. 2, p. 108, 2020.

[14] J. HOWARD and S. GUGGER, Deep Learning for Coders with fastai and PyTorch: AI Applications Without a PhD, O'Reilly Media, 2020.

[15] Python Software Foundation, "Data Structures - Python 3.11.3 documentation," 2023. [Online]. Available: https://docs.python.org/3/tutorial/datastructures.html#dictionaries. [Accessed May, 5th 2023].

[16] Z. WU, C. SHEN and A. V. D. HENGEL, "Wider or deeper: Revisiting the resnet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119-133, 2019.

[17] K. WEISS, T. M. KHOSHGOFTAAR and D. WANG, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1-40, 2016.

[18] G. M. FOODY, "Status of land cover classification accuracy assessment," *Remote Sensing of Environment*, pp. 185-201, 12 July 2001.

[19] Scikit-Learn, "scikit-learn," [Online]. Available: https://scikit-learn.org/stable/index.html. [Accessed May, 12th 2023].