Integrating Computer Vision into CACTO: A System for Fraud Prevention and Secure Transport Monitoring

Kaiky Maia
fkms@academico.ufpb.br
Universidade Federal da Paraíba
João Pessoa, Paraíba

Ludmila Gomes
lvgg@academico.ufpb.br
Universidade Federal da Paraíba
João Pessoa, Paraíba

Allan Vasconcelos
allan.vasconcelos@academico.ufpb.br
Universidade Federal da Paraíba
João Pessoa, Paraíba

Thaís G. do Rêgo
thais@ci.ufpb.br
Universidade Federal da Paraíba
João Pessoa, Paraíba

Yuri Malheiros
yuri@ci.ufpb.br
Universidade Federal da Paraíba
João Pessoa, Paraíba

Telmo Silva Filho
telmo.silvafilho@bristol.ac.uk
University of Bristol
Bristol, UK

Marcelo Iury
marceloiury@gmail.com
Universidade Federal da Paraíba
João Pessoa, Paraíba

ABSTRACT

The rapid growth of freight transport in Brazil has increased the demand for intelligent tax auditing tools capable of handling high-volume, real-time data. This paper presents the integration of three computer vision modules into CACTO, a large-scale monitoring platform used by the State Treasury Secretariat of Paraíba (SEFAZ-PB). The modules address key inspection tasks: vehicle re-identification and multi-camera tracking, direction recognition and route monitoring, and open-truck cargo detection. By combining distributed architectures with scalable machine learning pipelines, the system enhances proactive and evidence-based auditing, supporting fiscal enforcement and reducing vulnerabilities to tax evasion.

KEYWORDS

CACTO, computer vision, vehicle re-identification, route monitoring, cargo detection, tax auditing, fraud prevention, big data

1 INTRODUCTION

The exponential growth of e-commerce and logistics has intensified freight traffic on highways, generating massive, high-speed data flows from hundreds of road sensors and vehicles, for example via truck-mounted cameras. Real-time monitoring of this data demands software architectures that are robust, scalable, and low-latency. Without such capabilities, fiscal authorities face vulnerabilities to tax evasion and lose the ability to perform proactive inspection.

To address this problem, the State Treasury Department of Paraíba (SEFAZ-PB) developed CACTO (Control and Accompaniment of Goods in Transit On-line), a distributed monitoring platform that integrates microservices, data pipelines, and machine learning components to automate inspection tasks and support fiscal enforcement. By leveraging real-time analytics, the system enhances the capacity of tax authorities to detect irregularities and maintain transparency.

This article highlights three artificial intelligence (AI) modules: vehicle re-identification and multi-camera tracking, direction recognition and route monitoring, and open-truck cargo detection. We focus on the design choices and architectural patterns that enable these capabilities in a big data environment. AI plays a central role in preserving vehicle identity even with obscured plates, classifying orientation and inferring direction, and identifying undeclared cargo by cross-referencing images with manifests. We conclude by discussing methodologies, datasets, performance results, and the system's contributions, along with its potential for future expansion.

It is important to note that broader effectiveness indicators, such as infrastructure metrics and operational impact, could not be obtained at this stage. Collecting and analyzing such data remain open directions for future work, offering opportunities to further validate scalability and long-term fiscal contributions. While related studies exist in areas such as vehicle re-identification [2], our focus lies in integrating multiple modules into a single solution tailored to this fiscal auditing context.

2 CACTO

CACTO is a web-based system developed to modernize tax auditing in the State of Paraíba, Brazil. Its core purpose is to monitor cargo vehicles in real time and detect potential tax irregularities. The platform integrates OCR camera data with electronic tax documents—NF-e (Nota Fiscal Eletrônica, Electronic Invoice), CT-e (Conhecimento de Transporte Eletrônico, Electronic Transport Document), and MDF-e (Manifesto Eletrônico de Documentos Fiscais, Electronic Manifest of Fiscal Documents)—enabling cross-referencing that generates prioritized alerts and guides auditing actions with greater precision.

Operating in a big data environment characterized by high volume, variety, and velocity, CACTO continuously processes thousands of records. Its infrastructure ensures real-time processing,

In: XXIV Workshop de Ferramentas e Aplicações (WFA 2025). Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (WFA'2025). Rio de Janeiro/RJ, Brasil. Porto Alegre: Brazilian Computer Society, 2025.

ISSN 2596-1683

traceability, and scalability, supporting both reactive monitoring and preventive action through automated alerts.

The system also leverages artificial intelligence to track vehicles across multiple cameras, analyze orientations and routes, and detect undeclared cargo, all by processing large volumes of image data with deep learning models through distributed and scalable pipelines. These capabilities enable continuous oversight, predictive route analysis, and targeted inspections, while dashboards and user interfaces provide auditors with actionable insights for decision-making.

3 RELATED WORK

Regarding related works, searches were conducted in relevant scientific databases (such as IEEE Xplore, ACM Digital Library, and Scopus) using terms associated with tax auditing with computer vision, vehicle re-identification, route monitoring, and cargo detection. However, no projects were identified that present a scope and application comparable to the present work, particularly in the context of integrating multiple artificial intelligence modules into a large-scale tax enforcement platform. Therefore, this study is considered to make an original contribution to the field by applying and integrating computer vision techniques in a real-world fiscal auditing environment.

4 PROPOSED MODULES

The following subsections detail the operational components of CACTO, including dataset preparation, vehicle re-identification, direction recognition, and cargo detection. Screenshots of the system interface are included to illustrate the practical deployment of each module.

4.1 Training and Evaluation Datasets

CACTO integrates image datasets collected from OCR-equipped roadside cameras in Paraíba, automatically stored in MinIO, an open-source, high-performance object storage system compatible with Amazon S3, together with metadata (timestamp, geolocation, camera ID, license plate). Within the tool, cargo frames are tagged as loaded, empty, or undefined (occluded/ambiguous), while orientation samples are categorized as frontal or rear views, with vehicle class labels (truck, car, motorcycle). For vehicle re-identification, the system uses pre-training on the VeRI dataset, which contains 37,778 training images and 11,579 test images [4–6], followed by fine-tuning on 10,000 CACTO-specific images. This process reduces domain mismatch and improves recognition under Brazilian traffic conditions. All three computer vision models were trained using an NVIDIA GeForce RTX 2060 GPU, ensuring feasible training times.

4.2 Vehicle Re-identification and Multi-Camera Tracking

The baseline for our vehicle re-identification pipeline is FastReID [2], a PyTorch toolbox offering diverse backbones (ResNet, OSNet, transformers) and training strategies (hard-mining triplet loss, label smoothing, adaptive learning rate). Pre-trained on large benchmarks, it delivers strong mAP (mean Average Precision) and Rank-1 accuracy and supports easy integration and comparison of custom modules. FastReID's highly modular design allows us to plug in custom feature extractors or attention modules with minimal code changes, which is critical when experimenting with domain-tailored modifications [9]. Additionally, its implementation of various backbone architectures and support for multi-domain pre-training enables straightforward selection of the model variant that offers the best trade-off between inference speed and discriminative power [7].

Building on these strengths, FastReID applies the same feature extraction pipeline to both gallery and query images, generating embeddings—vector representations of image features—that are compared through similarity measures to rank candidate matches. To reduce computational cost, we precompute embeddings for the gallery and only append new images at ingestion (Figure 1). Queries are then matched against this cached gallery, ensuring consistent low-latency retrieval even in challenging cases such as missing license plates (Figure 2).

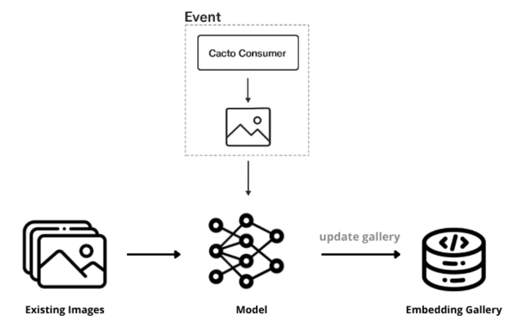

Figure 1: Embeddings for existing gallery images are precomputed, and only new images are appended at ingestion to reduce per-query computation.

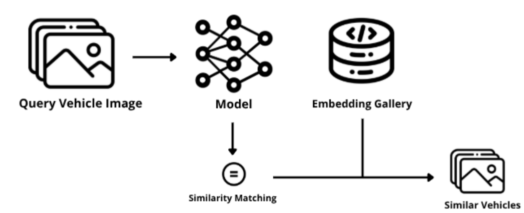

Figure 2: Query images are encoded and matched against cached gallery embeddings, allowing low-latency retrieval even for challenging cases such as missing license plates.

This design yields a lean, maintainable system with low runtime overhead under varying loads, reducing CPU processing time from approximately 12 minutes to around 1 minute. We evaluated our pipeline on the VeRi benchmark, where it achieved **93.50% Rank-1**, **97.08% Rank-5**, and **98.51% Rank-10** accuracy, indicating the probability that the correct vehicle appears among the top 1, top 5, or top 10 retrieved results, respectively. The model also reached a **mAP (mean Average Precision) of 71.77**, **mINP (mean Inverse Negative Penalty) of 30.27**, and an overall metric score of **82.64**, confirming the effectiveness of our chosen backbone and training strategy. Figure 3 illustrates the CACTO interface, where an alert image is compared against the gallery and the most similar vehicles are displayed in ranked order.

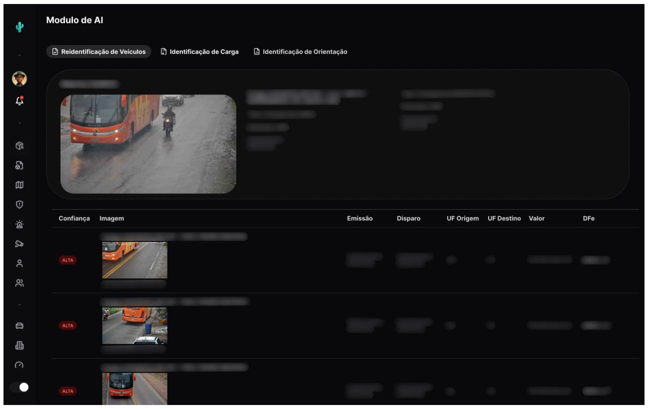

Figure 3: CACTO interface for vehicle re-identification. The top panel shows the alert image (query), while the bottom panel displays the most similar matches ranked by the AI across different cameras.

4.3 Direction Recognition and Route Monitoring

The module determines each vehicle's orientation (front or rear) from frames captured by CACTO's cameras. In a unified detection + OCR pipeline, every vehicle is first localized by an orientation model, isolating it within its visual context before subsequent processing [1].

For each detected vehicle ROI (Region of Interest), a dedicated YOLOv12n [8] plate-detection model extracts the license-plate region. That region is then passed to a YOLO-based OCR model which detects and recognizes alphanumeric characters, sorts them by their X-coordinate, and reconstructs the full plate string.

The reconstructed plate is compared against a target plate, accepting either an exact match or at least five identical characters. If this similarity criterion is met, the system outputs the vehicle's orientation (front or rear) with high confidence.

Low-confidence detections or partially recognized plates are flagged as "undefined" for manual review or enhanced preprocessing. All outputs—orientation, plate text, and frame metadata (timestamp, camera ID, geolocation)—are stored in CACTO, enabling cross-referencing with tracking and cargo data for a comprehensive operational view.

Each image underwent pre-processing that included autoorientation of pixel data (with EXIF-orientation stripping) and resizing to 640×640 pixels with stretching. To improve generalization, a data augmentation pipeline was applied, generating three augmented versions of each source image with random brightness adjustment (-15% to +15%), exposure adjustment (-10% to +10%), Gaussian blur (0–4 pixels), and salt-and-pepper noise affecting 0.1% of pixels [3].

The final YOLOv12n orientation model, after training and evaluation, achieved an overall performance of **83.1% precision**, **76.7% recall**, **76.8% mAP@50**, and **60.9% mAP@50-95** across all categories. The dataset, totaling 4,178 images collected from the state's surveillance camera network, was split into training (**95%**, 3,975 images), validation (**1%**, 53 images), and test (**4%**, 150 images) sets. Detected vehicles are shown alongside their estimated routes in Figure 4. The AI classifies the vehicle's orientation, and this output is then cross-referenced with camera metadata to infer the direction of movement and determine the likely paths. This visualization allows operators to track possible routes in real time, improving decision-making in enforcement workflows.

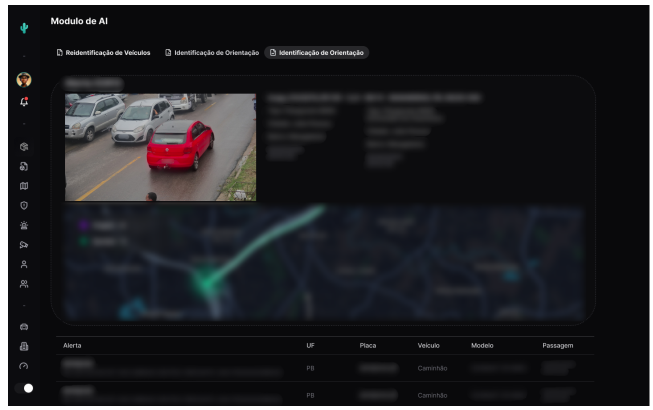

Figure 4: Alert screen in CACTO showing a detected vehicle (top) and the estimated route inferred.

4.4 Open-Truck Cargo Detection

The methodology adopted for detecting cargo in open trucks within the CACTO system is detailed in this subsection. It addresses the challenges related to visual ambiguity in the images used, as well as the processes of image acquisition, class definition and annotation, dataset cleaning and balancing, and YOLO model training.

The automated detection of cargo in open trucks addresses a practical demand from the Paraíba State Treasury to optimize inspection procedures and regulate cargo transportation on state highways. The proposed solution consists of a YOLO-based visual recognition model, capable of quickly and efficiently indicating the presence of cargo in truck beds, assisting agents in the decision-making process. One of the main challenges encountered relates to visual ambiguity in several situations. Trucks with partial loads, non-commercial cargo, or various objects, such as people in the truck bed, complicate accurate classification.

Regarding the image acquisition procedure, as previously described, the annotation process was carried out using the Roboflow platform. The dataset, composed of 3,292 images extracted, was annotated following a standardized procedure that considered the visual challenges, defining four categories: "Loaded", "Empty", "Null" (when no truck is present in the image), and "Undefined" (when the truck bed content cannot be determined). This categorization ensures annotation consistency and improves the effectiveness of model training. The image set was split into training (93%, 3,066 images), validation (3%, 111 images), and test (3%, 115 images). For pre-processing and augmentation, we applied the same procedures described in the orientation module. The final YOLOv12n model, after training and evaluation, achieved an overall accuracy of **86.5% (mAP@50)** and **76.0% (mAP@50-95)** across all categories. Class-wise performance indicates good generalization capacity, with the "Loaded" class reaching **88.9% mAP@50**, "Empty" **82.1%**, and "Undefined" **88.4%**. These results demonstrate the robustness of the solution in handling heterogeneous scenarios and ensuring reliable detection to support SEFAZ operations.

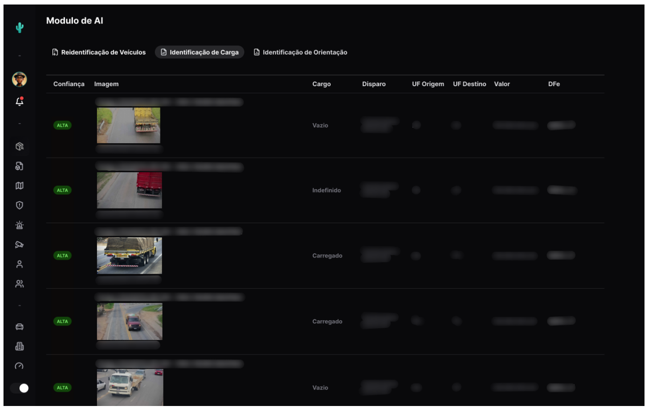

Figure 5: Open-truck cargo detection alerts estimated screen, indicating whether the truck is *Loaded, Empty,* or *Undefined.*

5 CONCLUSION

The development of the CACTO system represents a major step forward in the tax auditing practices adopted by SEFAZ-PB. Conceived as a robust and scalable solution, CACTO integrates modern technologies for data capture, processing, and real-time analysis, enabling both preventive and reactive actions against tax evasion. Its modular architecture, event-driven infrastructure, and data-intensive processing capabilities ensure that fiscal oversight is both reliable and adaptive to the demands of large-scale, high-velocity environments. Moreover, the system is developed under an Internal Use License, ensuring its deployment and operation are exclusively for governmental fiscal purposes.

Above all, the system's intelligence is consolidated in its three AI-driven modules: the vehicle re-identification and multi-camera tracking service, which ensures persistent monitoring even when license plates are obscured; the trajectory and direction analysis pipeline, which fuses visual orientation with geospatial data to predict travel routes and enable proactive inspections; and the open-truck cargo detection module, which verifies cargo presence against fiscal manifests, closing a critical loophole in freight transparency. Together, these components transform raw surveillance data into actionable intelligence, granting auditors greater agility, evidence-driven decision-making, and the capacity to act preventively against irregularities.

As summarized in Table 1, the performance metrics achieved by these modules confirm the effectiveness of the proposed approach, balancing accuracy and efficiency in real-world operational scenarios. These results provide not only empirical validation of the system's capabilities but also reinforce its potential as a reference model for large-scale, AI-powered fiscal monitoring.

By combining advanced AI with a modernized fiscal infrastructure, CACTO establishes itself not only as a strategic tool in the fight against tax evasion and revenue loss but also as a reference model for data-driven public initiatives in real-time monitoring and fraud prevention.

Table 1: Performance Metrics of the CACTO Computer Vision Modules

Module	Metrics / Dataset
Vehicle Re-identification	**Rank-1:** 93.50; **Rank-5:** 97.08; **Rank-10:** 98.51; **mAP@50:** 71.77; **Dataset:** VeRI (37,778 train, 11,579 test) + 10,000 CACTO images; *Additional:* mINP=30.27, Score=82.64, Runtime$\downarrow$12$\rightarrow$1 min
Vehicle Orientation	**mAP@50:** 76.8; **mAP@50-95:** 60.9; **Precision / Recall:** Prec=83.1 / Rec=76.7; **Dataset:** 4,178 CACTO images
Open-Truck Cargo Detection	**mAP@50:** 86.5; **mAP@50-95:** 76.0; **Precision / Recall:** Prec=79.1 / Rec=81.5; **Dataset:** 3,292 CACTO images; *Additional:* mAP@50 Loaded=88.9, Empty=82.1, Undefined=88.4

REFERENCES

[1] Michael Abebe Berwo, Asad Khan, Yong Fang, Hamza Fahim, Shumaila Javaid, Jabar Mahmood, Zain Ul Abideen, and M. S. Syam. 2023. Deep Learning Techniques for Vehicle Detection and Classification from Images/Videos: A Survey. *Sensors* 23, 10 (2023), 4832. https://doi.org/10.3390/s23104832

[2] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. 2020. FastReID: A Pytorch Toolbox for General Instance Re-identification. arXiv:2006.02631 [cs.CV] https://arxiv.org/abs/2006.02631

[3] Youngmin Kim, Donghwa Kang, and Hyeongboo Baek. 2025. A 2-Stage Model for Vehicle Class and Orientation Detection with Photo-Realistic Image Generation. *arXiv preprint arXiv:2506.01338 [cs.CV]* (2025). https://doi.org/10.48550/arXiv.2506.01338

[4] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. 2016. Large-scale vehicle re-identification in urban surveillance videos. https://doi.org/10.1109/ICME.2016.7553002

[5] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. 2016. A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance, Vol. 9906. 869–884. https://doi.org/10.1007/978-3-319-46475-6_53

[6] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. 2017. PROVID: Progressive and Multimodal Vehicle Reidentification for Large-Scale Urban Surveillance. *IEEE Transactions on Multimedia* PP (09 2017). https://doi.org/10.1109/TMM.2017.2751966

[7] Matheus Oliveira, Lucas Fonseca, and Rodrigo Minetto. 2021. *A Two-Stream Siamese Neural Network for Vehicle Re-Identification with License Plate Recognition.* Master's thesis. Universidade Tecnológica Federal do Paraná (UTFPR). https://repositorio.utfpr.edu.br/jspui/handle/1/26663

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640 [cs.CV] https://arxiv.org/abs/1506.02640

[9] Lucas Nascimento Souza, Carla M. Medeiros, and João P. Silva. 2025. Reidentificação Veicular em Ambientes Reais Utilizando Restrições Espaço-Temporais e Aprendizado Semissupervisionado. In *Anais do Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*. https://sol.sbc.org.br/index.php/sbcup/article/view/36148