BR-RTSD40k: Scaling Brazilian Traffic Sign Recognition via Semi-Supervised Annotation and Transformer-Based Baselines

Arthur Maciel[1], Davi Almeida[1], Gabriel Dantas[1], Victor Gomes[1]
Artur R. Pinto[1], Maria Klara B. Sena[3], Valério de Medeiros Júnior[2]
Frederico Lopes[1], Daniel S. A. Araújo[1]

[1] Federal University of Rio Grande do Norte (UFRN)
Natal, RN, Brazil

[2]Federal Institute of Education, Science and Technology (IFRN)
Natal, RN, Brazil

[3]State University of Rio Grande do Norte (UERN)
Natal, RN, Brazil

{artur.maciel.134, davi.almeida.135, gabriel.lima.712, victor.gomes.090,
artur.revoredo.116}@ufrn.edu.br, mariaklara@alu.uern.br,
valerio.medeiros@ifrn.edu.br, {fred, daniel}@imd.ufrn.br

Abstract. *Traffic sign recognition is a fundamental component of intelligent transportation systems and autonomous driving applications. Although several public traffic sign datasets have been proposed in recent years, there remains a lack of large-scale datasets representing the diversity of Brazilian traffic signage and road conditions. This limitation hinders the development and evaluation of robust models tailored to the Brazilian context. This paper presents the construction of a Brazilian traffic sign dataset, created from street-level imagery collected in real-world driving scenarios and complemented with samples from publicly available sources. To enable efficient large-scale annotation, we propose a semi-supervised labeling pipeline that combines automatic object detection with manual validation and refinement. The resulting dataset comprises 41,016 annotated images, specifically curated to feature the Brazilian R-type (regulatory) signs targeted in this study. To establish a benchmark for future research, we trained an RT-DETR-based object detector and a Convolutional Neural Network (CNN) for sign classification, both of which were rigorously evaluated. Experimental results demonstrate the feasibility of the proposed dataset and provide strong baseline performance for Brazilian traffic sign recognition tasks. To foster reproducible research and accelerate advances in intelligent transportation systems, the complete dataset will be publicly released to the research community.*

Keywords: *Traffic Sign, RTDETR, Dataset, Semi-Supervised Annotation.*

1. Introduction

The rapid growth of urban populations has increased the demand for technologies capable of improving traffic management and public safety. In this context, Intelligent Transportation Systems (ITS) and smart city platforms increasingly rely on computer vision techniques for real-time monitoring and decision-making [Nellore and Hancke 2016,

Dilek and Dener 2023]. Among these applications, traffic sign detection is an important capability for advanced driver assistance systems and autonomous vehicles [Chen et al. 2024].

Brazilian signage alone is far from a simple closed set: the Brazilian Traffic Code (*Código de Trânsito Brasileiro* - CTB) defines over 160 distinct sign types spanning regulatory (R), warning (A), and indication (S) groups, each with its own visual conventions. This diversity makes large-scale, well-annotated data indispensable for training reliable detectors. Nevertheless, most public datasets target European, Chinese, or North American standards, and the Brazilian datasets available so far either lack bounding-box annotations or are too small in scale and class coverage to serve as a comprehensive detection benchmark, leaving a clear gap for the Brazilian context.

Given the breadth of the full CTB taxonomy, this paper narrows its scope to the **R-series regulatory signs**: the group most directly tied to enforcement (speed limits, stop obligations, no-parking restrictions, etc.) and therefore most relevant to driver-assistance, autonomous-driving and surveillance applications. To address this gap, we present the construction of the Brazilian R-Type Traffic Sign Dataset (BR-RTSD40k) dataset, assembled from publicly available imagery captured by diverse consumer-grade cameras (e.g., dashcams and smartphones) and annotated through a semi-supervised pipeline. We then benchmark a Real-Time Detection Transformer (RT-DETR)[Zhao et al. 2024] detector and a Convolutional Neural Network (CNN) classifier on this dataset, providing the first systematic performance baseline for Brazilian regulatory traffic sign detection.

The main contributions of this work are as follows:

1. We introduce a new Brazilian R-Type Traffic Sign Dataset (BR-RTSD40k) dataset, built from diverse publicly available street-level imagery and complemented by proprietary dashcam footage, providing bounding box and class annotations for R-series regulatory signs, released under a permissive open license at `https://doi.org/10.5281/zenodo.22070392`.
2. We propose a semi-supervised annotation pipeline that combines automatic traffic sign detection with manual validation, substantially reducing the annotation effort required for large-scale dataset construction while maintaining annotation quality.
3. We establish a public benchmark for Brazilian regulatory traffic sign recognition by training and evaluating an RT-DETR object detector and a CNN classifier on the BR-RTSD40k dataset, utilizing standard and appropriate evaluation metrics for both tasks to provide robust baselines for future comparisons.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature; Section 3 details the dataset construction; Section 4 describes the proposed comparative methodology; Section 5 presents and analyzes the results and discusses findings; Section 6 concludes the paper.

2. Related Work

This section contextualizes our contributions. Section 2.1 surveys existing traffic sign datasets to highlight the regional gaps motivating our Brazilian benchmark, while Section 2.2 reviews transformer-based object detectors to justify selecting RT-DETR as our real-time baseline.

2.1. Traffic Sign Detection Datasets

Table 1 summarizes key traffic sign datasets. While GTSRB and GTSDB [Stallkamp et al. 2012, Houben et al. 2013] are standard benchmarks, they offer limited realism regarding object scales. Large-scale datasets like TT100K [Zhu et al. 2016] and MTSD [Ertler et al. 2020] improve environmental diversity but focus on Chinese or global taxonomies and have a restrictive license that limits use to non-commercial research and prohibits the creation of derivative datasets, underrepresenting Brazilian conventions. Locally, Brazilian datasets address this regional specificity only partially: BRTSRD [Silva et al. 2021] provides only pre-cropped images (making it unsuitable for detection), whereas BRTSD [Dalborgo et al. 2023] includes bounding boxes but remains too small to serve as a comprehensive benchmark. This highlights a clear gap, the lack of a large-scale, detection-oriented dataset tailored to Brazilian visual conventions, which motivates the development of our proposed BR-RTSD40k.

Table 1. Comparison of prominent traffic sign datasets and our proposed benchmark.

Dataset	Images	Instances	Classes	Format	Dataset Aug.	Region
GTSRB [Stallkamp et al. 2012]	51,839	51,839[a]	43	Crop	No	Germany
GTSDB [Houben et al. 2013]	900	1,206	43[b]	BBox	No	Germany
TT100K [Zhu et al. 2016]	100,000[c]	∼30,000	221[c]	BBox+Mask	No	China
MTSD [Ertler et al. 2020]	100,000[d]	325,172	313	BBox	No	Global
BRTSRD [Silva et al. 2021]	24,072[e]	24,072[e]	51	Crop	Yes	Brazil
BRTSD [Dalborgo et al. 2023]	5,631	8,065	17[f]	BBox	No	Brazil
BR-RTSD40k (Ours)	**41,016**	**46,055**	**25**	**BBox**	**No**	**Brazil**

Instances denote labeled traffic-sign observations: one sample per cropped image for crop-based datasets and one annotated object for detection datasets. Dataset Aug. indicates whether synthetically transformed samples are included in the released dataset, excluding training-time augmentation.

[a] GTSRB contains 51,839 cropped observations corresponding to 1,728 unique physical traffic signs. [b] GTSDB provides 43 fine-grained class IDs grouped into three competition categories. [c] Only about 10,000 TT100K images contain traffic signs; 221 classes are defined, while 45 have at least 100 examples. [d] MTSD comprises 52,453 fully annotated images and 47,547 real images with partially propagated annotations. [e] BRTSRD contains 24,072 final samples after data augmentation, derived from 3,076 pre-augmentation crops. [f] BRTSD comprises 16 traffic-sign classes and one vegetation-occlusion defect class.

2.2. Transformer-Based Object Detectors

The DEtection TRansformer (DETR)[Carion et al. 2020] reframed detection as direct set prediction, removing hand-designed components such as anchor generation and non-maximum suppression. Its encoder-decoder architecture reasons about global context using a fixed set of learned object queries, producing predictions in parallel via bipartite matching with a set-based global loss. Although DETR matched Faster R-CNN on COCO[Lin et al. 2014], its high computational cost and slow convergence initially hindered real-time use. The Real-Time Detection Transformer (RT-DETR)[Zhao et al. 2024] resolves this tradeoff as the first real-time end-to-end detector, surpassing YOLO[Redmon et al. 2016] architectures while preserving DETR's end-to-end simplicity without extensive post-processing. While YOLO architectures have historically dominated real-time tasks, RT-DETR eliminates the need for complex Non-Maximum Suppression (NMS) tuning while maintaining comparable speed, motivating its selection as our robust detection baseline. Concurrently, a custom CNN was chosen for the cascaded classification stage to provide a lightweight, transparent benchmark specifically tailored to evaluate the newly curated regulatory classes without the overhead of overly parameterized networks.

3. Dataset Construction

This section details the steps involved in building the BR-RTSD40k dataset, covering data collection, Brazilian traffic sign taxonomy, semi-supervised pipeline and quality control procedures that frame the annotation schema.

3.1. Data Collection

The dataset was assembled from three complementary sources to maximize diversity in scene conditions, camera perspectives, sign appearance, and geographic coverage across Brazilian road environments.

Roboflow datasets. Publicly available annotated traffic signs datasets hosted on Roboflow [Projetos 2023] were queried and filtered for Brazilian context, specifically targeting regulatory signs (R-type).

Mapillary API. Street-level imagery was retrieved via the Mapillary API [Mapillary 2026] using geographic bounding boxes covering all Brazilian state capitals, ensuring representative real-world coverage. The acquired frames were then strictly filtered for optimal quality, illumination, and viewing angles.

Proprietary dashcam footage. Original dashcam recordings, captured specifically along Brazilian roads, were incorporated primarily for the qualitative visual validation of the detection results.

3.2. Brazilian Traffic Sign Taxonomy

Regulated by Conselho Nacional de Trânsito (CONTRAN), Brazilian traffic signs span regulatory (R-series), warning (A-series), and indication (S-series) groups. This work is restricted to the R-series (Fig. 1), chosen for their direct enforcement role (e.g., speed limits, stop obligations, no-parking), which makes their accurate detection particularly relevant for driver-assistance and autonomous-driving applications in Brazil. R-series signs share a consistent visual language (predominantly white interior, black characters, and red borders), so inter-class discrimination relies mainly on inner symbols, text, and numerals.



Figure 1. Brazilian Regulatory Signs

Each image was annotated with bounding boxes and class labels from the R-series subset. To reduce extreme class imbalance and avoid unreliable evaluation on underrepresented categories, classes with fewer than 40 instances were removed, and visually related subclasses were merged (R-25a/R-25b into R-25ab; R-25c/R-25d into R-25cd), yielding a final set of 25 classes.

3.3. Semi-Supervised Annotation Pipeline

Over 50,000 raw images were collected from the combined sources. Given this volume, fully manual annotation would be prohibitively costly.As shown in Figure 2, we adopted a model-assisted labeling pipeline that combines automatic detection with human validation, structured in six stages: automatic bounding box generation, annotation refinement, automatic label generation, label validation, active hard negative mining, and dataset consolidation. Details on the model training setup are specified in Section 4. Iterative model retraining steps were executed at intermediate milestones of 3,000, 10,000, and 20,000 validated images to progressively improve the automated proposals, culminating in a final definitive training phase using the complete dataset of 41,016 annotated images.

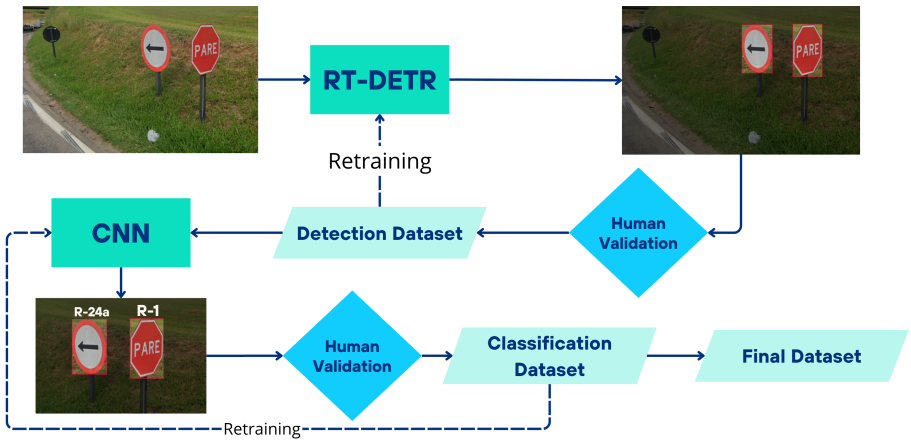


Figure 2. Semi-Supervised Annotation Pipeline

Automatic Bounding Box Generation. A base RT-DETR detector was first trained for 50 epochs on a subset of 3,000 images sourced from Roboflow Dataset, which arrived with pre-existing bounding box annotations exclusively featuring the *road-sign* label.

Annotation Refinement. Human annotators reviewed the pre-filled predictions frame by frame. For each candidate box, the reviewer could: (i) *accept* the box without modification; (ii) *correct* the bounding box geometry; or (iii) *reject* the box entirely as a false positive. Annotators were also instructed to add any traffic signs missed by the automatic detector.

Automatic Label Generation. The resulting bounding boxes from the previous step are cropped from the original images and sent to a Convolutional Neural Network (CNN) classifier trained on Brazilian R-type sign classes. Consistent with real-world traffic scenarios, the collected data naturally follows a long-tailed distribution, where three dominant regulatory classes account for approximately 60% of the instances. To establish a robust baseline classifier and prevent extreme bias toward these majority classes, we applied a hybrid sampling strategy during the CNN training phase. We limited the majority classes to a maximum of 1,500 instances via random undersampling and enriched the minority classes using supplementary data from Roboflow. Rather than forcing a strict class equilibrium, this approach effectively smoothed the natural distribution, mitigating the long-tail effect. The model was trained from scratch for 30 epochs on this refined subset to generate highly accurate initial label proposals.

Label Validation. Human annotators review all generated labels. For each labeled crop, the reviewer could: (i) *accept* the predicted class without modification; (ii) *correct* the label when the prediction was incorrect; or (iii) *discard* the crop when the sign was ambiguous, partially visible, or did not belong to the target taxonomy.

Active Hard Negative Mining. To continuously improve model robustness, a hard negative mining strategy was integrated into the human validation feedback loop. During the established retraining milestones, a curated subset of the rejected false positives was actively recycled into the training sets. For the RT-DETR detection model, images containing rejected bounding boxes were included as background examples (without annotations) to penalize incorrect object proposals. Simultaneously, for the CNN classifier, rejected crops were assigned a dedicated *NO_SIGN* class. This continuous feedback loop progressively enhanced both models' discriminative capabilities against complex background clutter, drastically reducing the false positive rate in subsequent auto-annotation cycles.

Dataset Consolidation. With all bounding boxes and class labels validated, the data was compiled into two distinct subsets. The primary dataset comprises 41,016 full images tailored for the detection task, intentionally retaining its natural, highly imbalanced distribution to serve as a realistic benchmark. In contrast, the CNN supplementary subset consists of 36,326 bounding-box crops. Following the sampling and smoothing strategy detailed in the previous step, this classification collection reduces extreme long-tail disparities and integrates a *NO_SIGN* class for hard negative mining. The final class distribution and instance counts for both subsets are compared in Table 2.

Table 2. Class Distribution for the BR-RTSD40k (BR) and the CNN Supplementary (CNN) Subsets.

Class	BR-RTSD40k	CNN	Class	BR-RTSD40k	CNN	Class	BR-RTSD40k	CNN
NO_SIGN	—	4,278	R-6c	2,677	2,247	R-25cd	1,390	1,617
R-1	4,717	2,129	R-7	631	900	R-26	451	626
R-2	850	1,731	R-9	121	231	R-27	324	471
R-3	405	758	R-12	28	69	R-28	200	272
R-4a	1,454	2,064	R-15	742	1,092	R-29	19	48
R-4b	793	1,184	R-19	11,646	5,331	R-32	820	888
R-5	632	1,296	R-24a	1,302	2,200	R-33	390	718
R-6a	9,996	2,633	R-24b	420	627	R-34	568	593
R-6b	1,725	1,796	R-25ab	390	527			

Crucially, this semi-supervised approach proved highly effective as a methodological contribution. Across the 3,000, 10,000, and 20,000 image retraining milestones, the system progressively shifted the human workload from manual drawing to rapid binary verification, drastically reducing cognitive load and enabling parallel annotation. Concurrently, the active hard negative mining visibly minimized automated false positives in later stages. Ultimately, this domain-agnostic framework was instrumental in achieving the dataset's scale efficiently.

4. Proposed Methodology

This section details the training and evaluation setup of both the RT-DETR and the CNN models on BR-RTSD40k dataset, covering neural network architecture, training configuration, hardware and software setup, and evaluation metrics.

4.1. Traffic Sign Detection Model

Dataset Split. The complete dataset of 41,016 annotated images was partitioned into three distinct subsets: a training set (34,864 images), a formal test set for rigorous quantitative metric evaluation (5,742 images), and a qualitative inference set (410 images) reserved exclusively for visual human review of the model's predictions. To strictly prevent temporal data leakage, frames extracted from the same continuous video sequence were grouped together and assigned to a single partition.

Data Augmentation. To enhance model robustness and prevent overfitting, the RT-DETR training phase utilized the baseline augmentation pipeline from PaddleDetection, combining random photometric distortions (brightness, contrast, saturation, and hue), spatial transforms (canvas expansion, IoU-validated cropping, and horizontal flipping), and multi-scale resizing between 480 and 800 px, with the 640x640 px baseline favored.

Training Configuration. The detector was fine-tuned from COCO-pretrained weights using the AdamW optimizer with a cosine-decay schedule and linear warmup. Table 3 summarizes the hyperparameters applied to the model.

Table 3. Key Training Hyperparameters for the RT-DETR Model.

Hyperparameter	Value
Base Model / Backbone	RT-DETR ResNet-50-vd (COCO pretrained, fully fine-tuned)
Input Resolution	640×640 px
Training Epochs	72
Batch Size	12
Optimizer	AdamW (weight decay $= 5 \times 10^{-4}$)
Initial Learning Rate	75×10^{-6}
LR Schedule	Cosine decay with linear warmup (1000 steps)

Evaluation Metrics. To comprehensively assess the object detection model, we adopted the standard Microsoft COCO evaluation protocol [Lin et al. 2014]. We report the primary Average Precision across multiple IoU thresholds (AP@50:95), alongside AP at fixed thresholds of 0.50 and 0.75. To evaluate scale invariance, both AP and Average Recall (AR) are reported across small (area $< 32^2$ px), medium ($32^2 <$ area $< 96^2$ px), and large (area $> 96^2$ px) object scales (AP_S, AP_M, AP_L and AR_S, AR_M, AR_L). Furthermore, we provide Precision, Recall, and the F1-Score, all evaluated at a standard IoU threshold of 0.50.

4.2. Traffic Sign Classification Model

CNN Classification Architecture. For the classification task, we designed a custom 10-layer CNN trained from scratch. This choice provides a controlled classification baseline without relying on external pretraining, allowing the evaluation to focus on the ability of the curated R-type sign crops to support fine-grained class discrimination. The network stacks five sequential blocks that progressively expand the feature channels up to 384, each following a standard motif of 3×3 convolutions, Batch Normalization, ReLU, and Max Pooling for feature extraction and spatial downsampling. The resulting feature maps are then collapsed by an Adaptive Average Pooling layer into a dense vector, which passes through dropout regularization and a final linear classification head.

Dataset Split. The validated sign crops were divided into training and test sets, comprising 29,061 and 7,265 images, respectively, for a total of 36,326 samples, following the long-tail distribution of the dataset. This total differs from the primary detection dataset as a direct result of the previously described hybrid smoothing strategy, which undersampled majority classes and injected supplementary external images to establish a better-balanced subset for the CNN.

CNN Training Configuration. The classifier was trained from scratch under a deliberately simple optimization regime, without learning-rate scheduling or early stopping, to establish a raw, foundational baseline. This configuration was intentionally kept simple to provide a clear reference point for future work investigating more advanced optimization strategies, pretrained architectures and class-imbalance mitigation techniques. Label smoothing was applied to the training loss while standard Cross-Entropy was kept for evaluation, and final weights were selected by peak test-set accuracy. Table 4 summarizes the complete hyperparameter configuration.

Table 4. Training Hyperparameters and Settings for the Custom CNN Classifier.

Hyperparameter	Value
Input Resolution	160×160 px (RGB)
Training Epochs	30
Optimizer	AdamW (weight decay $= 1 \times 10^{-4}$)
Learning Rate	3×10^{-3} (constant)
Regularization	Dropout ($p = 0.30$)
Training Loss	Cross-Entropy (Label Smoothing $= 0.05$)
Batch Size	192

CNN Classification Augmentation. The CNN training incorporated spatial and photometric augmentations to enhance model generalization. Transformations included random rotations with anti-clipping canvas expansion, alongside lighting simulations that randomly adjusted brightness, contrast, and saturation.

Classification Evaluation Metrics. To rigorously evaluate the CNN classifier and address the inherent class imbalances within the traffic sign dataset, we adopted a robust set of statistical metrics. Alongside standard Accuracy, we report the Matthews Correlation Coefficient (MCC) to provide a more reliable measure of overall predictive quality across all true and false prediction categories. Furthermore, to ensure that minority classes are evaluated equitably regardless of their frequency, we utilize macro-averaged metrics, specifically Macro Precision, Macro Recall, and the Macro F1-Score, which compute the unweighted mean of performance across all individual classes.

4.3. Experimental Setup: Hardware and Software

All experiments ran on a single workstation equipped with an NVIDIA GeForce RTX 5090 (32 GB GDDR7), an Intel Core Ultra 9 285K CPU, and 64 GB of DDR5 RAM, under a Python 3.13 environment. Because the detection and classification models have distinct requirements, training was split across two frameworks: the RT-DETR detector

was trained with PaddleDetection[1] for its native, optimized RT-DETR implementation, while the custom CNN classifier was implemented and trained from scratch in PyTorch[2] and TorchVision. Frames from the proprietary dashcam recordings were extracted using NVIDIA DeepStream.

5. Results and Discussion

Table 5 summarizes the quantitative results, which were evaluated exclusively on independent test partitions to guarantee an unbiased assessment of model generalizability.

Table 5. Performance Evaluation Metrics for the RT-DETR Detection and CNN Classification Models.

RT-DETR Object Detection Metrics			
General Metric	**Value**	**Size Metric**	**Value**
AP@50:95	0.854	AP_S (Small)	0.770
AP@50	0.973	AP_M (Medium)	0.894
AP@75	0.970	AP_L (Large)	0.981
Precision@50	0.928	AR_S (Small)	0.830
Recall@50	0.970	AR_M (Medium)	0.928
F1-Score@50	0.948	AR_L (Large)	0.989
CNN Classification Metrics			
Metric	**Value**	**Metric**	**Value**
Accuracy	0.9844	Macro Recall	0.9765
MCC	0.9833	Macro F1-Score	0.9810
Macro Precision	0.9860		

The experimental results validate the BR-RTSD40k dataset as a robust foundation for Brazilian regulatory traffic sign recognition, primarily due to its diversity and the large volume of collected images. Both models demonstrated strong performance: the RT-DETR detector achieved an F1-Score of 0.948, while the CNN classifier reached a Macro F1-Score of 0.981. These high values are partially facilitated by the highly consistent visual nature of R-type signs, which share standardized geometries and color schemes.

Furthermore, the class-balancing approach adopted during the CNN training phase proved highly impactful, ensuring strong classification results even for minority classes. Concurrently, the strategic introduction of *NO_SIGN* samples served a vital function; when operating in tandem with the detection model, this strategy significantly suppressed false positives caused by urban background clutter.

To further analyze the detection performance and the handling of background elements, Figure 3 presents the confusion matrix evaluated at a confidence threshold of 0.5. The results demonstrate that the model exhibits a strong capability in distinguishing traffic signs from the background, achieving a high true positive rate of 93% while keeping

[1]https://github.com/paddlepaddle/paddledetection

[2]https://pytorch.org/

false negatives relatively low (7%). Although the model handles background suppression reasonably well overall, the occurrence of false positives indicates a slight margin for improvement, as some background elements are still incorrectly localized as traffic signs. To mitigate this limitation and improve the overall precision of the pipeline, these detections are cascaded into the CNN classifier, which effectively acts as a secondary filter by discarding invalid bounding boxes categorized into its *NO_SIGN* class.

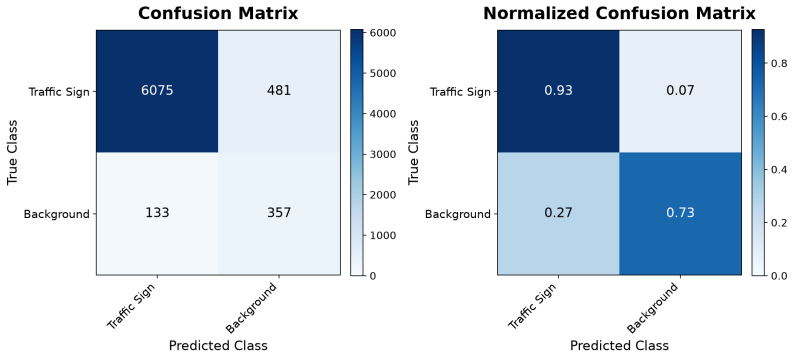


Figure 3. Detection Confusion Matrix (Det. Score = 0.5)

Despite these strong cascaded metrics, a common limitation persists regarding small object detection, where performance drops considerably compared to larger bounding boxes. This reduction in accuracy is primarily a consequence of the hardware utilized for data collection. Standard street-level capture devices (such as dashcams and smartphones) typically utilize wide-angle lenses that capture wide-field views but lack the high spatial resolution required to clearly resolve distant targets, resulting in a very low pixel density for faraway signs. Consequently, these small bounding boxes lack the distinct visual features necessary for the model to perform reliable feature extraction and detection.



Figure 4. Qualitative inference example in a real-world scenario.

Beyond the quantitative metrics, a qualitative assessment was conducted using the isolated inference set to verify the pipeline's behavior under authentic driving conditions. Figure 4 illustrates a practical example where the integrated system successfully localizes and classifies multiple regulatory signs mounted on the same pole. Despite common

urban challenges such as background clutter, variable lighting, and shadows cast by nearby vegetation, both the RT-DETR detector and the CNN classifier maintain high confidence scores and precise localization. This visual evidence further corroborates the robustness of the trained models and their viability for real-world traffic sign recognition applications.

Finally, while the mitigation strategies employed in this study successfully established a strong baseline, the dataset's deliberately preserved long-tailed distribution continues to pose an inherent challenge. By accurately reflecting the natural, real-world frequency of Brazilian traffic signs, BR-RTSD40k serves as a realistic benchmark, encouraging future research to explore algorithmic innovations capable of addressing extreme class imbalance natively. In addition to these algorithmic advancements, future efforts should prioritize a more robust and targeted data collection strategy. Such initiatives would not only help naturally populate the current minority classes but also enable the reintroduction of regulatory signs that were initially discarded due to critically low occurrence rates. Furthermore, expanding the dataset's taxonomy to encompass other categories of the Brazilian traffic code, most notably warning and indication signs, stands as a vital next step toward developing a fully comprehensive national benchmark for advanced driver-assistance systems.

6. Conclusion

This study introduced BR-RTSD40k, a comprehensive and robust benchmark dataset designed to advance Brazilian regulatory traffic sign recognition. By implementing a highly scalable model-assisted annotation pipeline that minimized manual drawing effort, we efficiently curated over 41,000 images, yielding a challenging dataset that deliberately preserves the natural long-tailed distribution of urban traffic environments. The proposed baseline framework, which integrates an RT-DETR object detector with a CNN classifier, demonstrated highly effective performance in real-world driving conditions. Achieving an F1-Score of 0.948 in detection and a Macro F1-Score of 0.981 in classification, the cascaded system successfully maintained high localization precision while leveraging the CNN to mitigate background-induced false positives.

Despite these strong baseline metrics, the extreme class imbalance and the hardware limitations associated with small object detection from standard dashcams continue to pose significant challenges. By accurately reflecting the real-world frequency of Brazilian traffic signs, BR-RTSD40k encourages the community to explore novel algorithmic innovations capable of addressing long-tailed distributions natively. Beyond these algorithmic improvements, future efforts must prioritize targeted data collection campaigns. A robust, scale-focused acquisition strategy will naturally enrich current minority classes and enable the reintroduction of rare regulatory signs that were initially discarded. Furthermore, this efficient semi-supervised framework provides a cost-effective, reproducible labeling methodology transferable to other computer vision domains. Ultimately, broadening the taxonomy to encompass other crucial categories of the Brazilian traffic code, including warning and indication signs, stands as a vital next step toward establishing a definitive, fully comprehensive national benchmark for autonomous driving and advanced driver-assistance systems.

References

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-end object detection with transformers. In *ECCV*, page 213–229, Berlin, Heidelberg. Springer-Verlag.

Chen, H., Ali, M. A., Nukman, Y., Razak, B. A., Turaev, S., Chen, Y., Zhang, S., Huang, Z., Wang, Z., and Abdulghafor, R. (2024). Computational methods for automatic traffic signs recognition in autonomous driving on road: A systematic review. *Results in Engineering*, 24:103553.

Dalborgo, V., Murari, T. B., Madureira, V. S., Moraes, J. G. L., Bezerra, V. M. O. S., Santos, F. Q., Silva, A., and Monteiro, R. L. S. (2023). Traffic sign recognition with deep learning: Vegetation occlusion detection in brazilian environments. *Sensors*, 23(13).

Dilek, E. and Dener, M. (2023). Computer vision applications in intelligent transportation systems: A survey. *Sensors*, 23(6).

Ertler, C., Mislej, J., Ollmann, T., Porzi, L., and Kuang, Y. (2020). Traffic sign detection and classification around the world. In *ECCV*.

Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., and Igel, C. (2013). Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In *International Joint Conference on Neural Networks*, number 1288.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer International Publishing.

Mapillary (2026). Mapillary API v4 documentation. `https://www.mapillary.com/developer/api-documentation`. visited on 2026-06-21.

Nellore, K. and Hancke, G. P. (2016). A survey on urban traffic management system using wireless sensor networks. *Sensors*.

Projetos (2023). Brazilian traffic signs dataset. `https://universe.roboflow.com/projetos-abip9/brazilian-traffic-signs-hnifq`. visited on 2026-06-21.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *CVPR*.

Silva, R., Prado, B., Matos, L. N., Santo, F., Zanchettin, C., and Novais, P. (2021). Construction of brazilian regulatory traffic sign recognition dataset. In Tavares, J. M. R. S., Papa, J. P., and González Hidalgo, M., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*.

Stallkamp, J., Schlipsing, M., Salmen, J., and Igel, C. (2012). Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., and Chen, J. (2024). Detrs beat yolos on real-time object detection.

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., and Hu, S. (2016). Traffic-sign detection and classification in the wild. In *CVPR*.