DEVELOPMENT OF AN ANNOTATED DATABASE FOR ASSESING THE PERFORMANCE OF DEEP LEARNING-BASED VEHICLE DETECTION AND TRACKING MODELS

Tudor BARBU[1], Silviu-Ioan BEJINARIU[2], Ramona LUCA[3]

Rezumat. *În această lucrare este descrisă dezvoltarea unei colecții de imagini ce poate fi utilizată pentru evaluarea performanței algoritmilor de detecție și urmărire a vehiculelor. Baza de date cu imagini conținând vehicule a fost creată folosind multe videoclipuri înregistrate de trafic care apoi au fost adnotate automat prin aplicarea unor detectoare de obiecte bazate pe rețele neuronale convoluționale (CNN). Setul de imagini a fost împărțit în seturi de date pentru antrenare, validare și testare și apoi folosit cu succes pentru a antrena, valida și testa detectoare de vehicule bazate pe învățare profundă. Sunt descrise, de asemenea, mai multe simulări de detecție și urmărire a vehiculelor. Este prezentată o soluție de clasificare a vehiculelor folosind învățarea bazată pe transfer, împreună cu rezultatele obținute pentru detecție și contorizare.*

Abstract. *The development of a voluminous database aimed at performance evaluation of the vehicle detection and tracking algorithms is described here. The vehicle database has been created using many recorded traffic videos and annotated automatically by applying some convolutional neural network (CNN) – based object detectors. It has been split into training, validation and testing datasets and then successfully used to train, validate and test deep learning-based vehicle detectors. Some multiple vehicle detection and tracking simulations are also described. A transfer learning-based vehicle classification solution using this database and those detection and counting results is also provided here.*

Keywords: annotated vehicle database, multiple vehicle detection and tracking, transfer learning, training and validation datasets, CNN-based vehicle classification

DOI 10.56082/annalsarsciinfo.2024.2.22

1. Introduction

Object detection, counting and tracking is a major and still challenging computer vision process consisting of locating objects of a certain class in video frames and determining their trajectories. It has a wide range of application areas: video monitoring, law enforcement, security systems, video indexing and retrieval,

[1] Habilitated PhD, Senior Researcher I, Institute of Computer Science of the Romanian Academy – Iasi Branch, Iasi, Romania, Corresponding member of The Academy of the Romanian Scientists, e-mail: tudor.barbu@iit.academiaromana-is.ro.
[2] PhD, Senior Researcher II, Institute of Computer Science of the Romanian Academy – Iasi Branch, Iasi, Romania, e-mail: silviu.bejinariu@iit.academiaromana-is.ro.
[3] PhD, Senior Researcher, Institute of Computer Science of the Romanian Academy – Iasi Branch, Iasi, Romania, e-mail: ramona.luca@iit.academiaromana-is.ro.

robotic vision, medical imaging, autonomous systems and augmented reality. But such applications are mostly developed for outdoor scenes and therefore the quality of the videos is affected by camera motion, deformation, motion blur, variations in brightness and illumination, object occlusions. To achieve its goal a detection and tracking system must be as little as possible sensitive to them.

Many methods for object detection are presented in the literature. In [1] the authors proposed a template matching based method for object detection. Temporal and frame differencing methods are proposed in [2]. Other methods are based on dictionary-based models [3], deformable part-based models [4], cascade classifiers [5], invariant descriptors with SVM-based classifiers [6], active contours [7], genetic algorithms, deep learning [8]. Similarly, a wide range of methods are used for tracking objects in video sequences. The most common methods are based on Kalman filters [9], mean-shift procedures [10], optical flow estimation [11], Hidden Markov Models (HMM) [12], adaptive template matching [13], object matching [14], and convolutional neural networks (CNN) [15]. Most frequently, the classes of objects subjected to detection and tracking are: faces (with applications in biometry, emotion recognition, security), humans and body parts (with applications in gait analysis for medical recovery, traffic monitoring, security), vehicles, animals (traffic monitoring, security, defense).

In this paper, the creation of a voluminous vehicle database, which has been created for the development the above-mentioned artificial intelligence (AI) techniques, is described. The paper is organized as follows. In the second section is described the proposed vehicles database. Some experiments related to vehicles detection and counting are described in the third section. The fourth section is related to vehicle tracking and the last section concludes the paper.

2. Vehicle Database Development

The proposed database is intended to be used for development of methods for automatic deep learning-based multiple vehicle detection, recognition and tracking, as part of a computer vision research project in the video traffic monitoring domain: SIMPATIA - Intelligent solutions for monitoring traffic participants using advanced machine vision tools and rigorous mathematical modeling.

The SIMPATIA database is required for training, validation and testing the deep learning-based vehicle detection and tracking frameworks within the mentioned project. It was obtained by extracting the frames from about 130 video sequences (15 GB) recorded in traffic using a video camera or high-resolution phone camera. The recordings were made in different weather conditions (sun, rain, snow) and different times of the day (day, night). Also, the recordings were made at 30 frames/second by placing the camera in different positions: above and on the side

of the road or from the car in traffic. The videos contain more than 550K frames of different sizes: [1920 x 1080], [1280 x 720] and [848 x 480].

In the preprocessing step, the frames were extracted and resized to the same dimensions [848 x 480] pixels. Also, their quality was increased by applying 3D filters to remove the additive white gaussian and quantum noises. From the entire set of frames, almost 60K of them were selected and automatically annotated using an existing deep-learning pretrained model from Ultralitics [16]. In a final step, the automatic annotation was manually validated to eliminate false positive results. The annotated objects represent the most common land vehicles: cars, buses, trucks, trams, bicycles and motorcycles (Fig. 1).

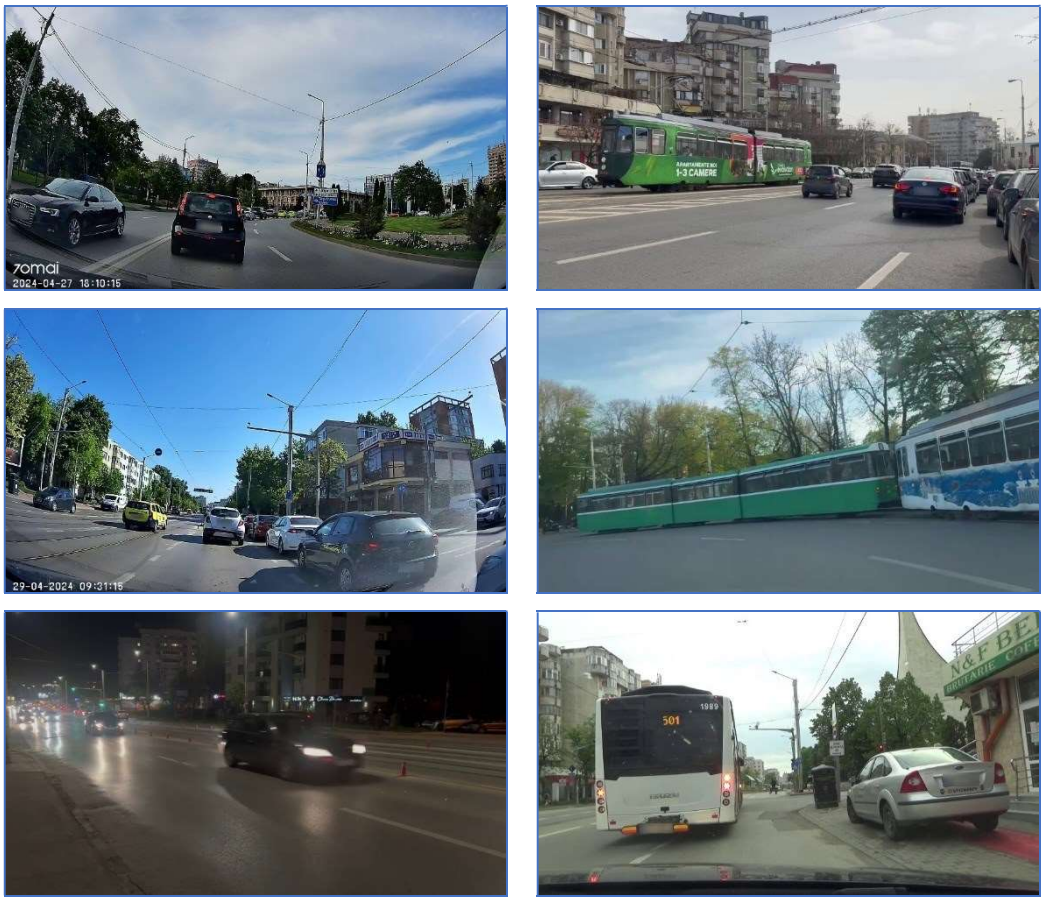

Fig. 1. Sample images from the selected frames

The vehicles database contains 59.032 images in .JPG file format. Because each image contains one or more objects, the total number of objects is grater than 432K. Also, each image file has an associated text file with annotations: the 2D

bounding boxes of the detected objects are described by their position and dimensions [left, top, width, height].

Finally, the full dataset was divided into:

- Training dataset – 70% of the data (41.323 frames and their annotations),
- Validation dataset – 10% of the data (5.903 frames and their annotations).
- Testing dataset – 20% of the data (11.806 frames and their annotations).

The dataset is stored in six archives (three archives with images and three archives with annotations) and it is publicly available on SIMPATIA project website: http://iit.academiaromana-is.ro/simpatia/ [17]. The annotated vehicle dataset can be freely used to train, validate and test successfully the deep learning networks for vehicle detection and tracking but, since we first disseminated it in [18], this paper has to be cited when using the database.

3. Training YOLO models using the SIMPATIA dataset

The dataset was used to train two version of the YOLO real-time object detection and image segmentation model: YOLOv5m and YOLOv8m. Both models were trained with the similar parameters: epochs = 50, batch_size = 16, learning_rate = 0.01. The image size was set to 640 and only one class entitled “vehicle” was used for all annotated objects. The training was performed using Ultralitics implementations in Python [16, 19] of the two models on a NVIDIA GeForce RTX 3070, 8GB graphics card. The 50 epochs completed in 10.3 hours for the YOLOv5m model and in 11.9 hours for the YOLOv8m.

The evolution of the Precision = TP/(TP+FP) and Recall = TP/(TP+FN) measures are displayed in Fig. 2 and Fig. 3 respectively (TP, FP, FN are the elements of the confusion matrix). However, the evolution of the two measures is quite fluctuating. this happens because at the moment the types of vehicles are not balanced within the class, there are many more cars and few bicycles and motorcycles.

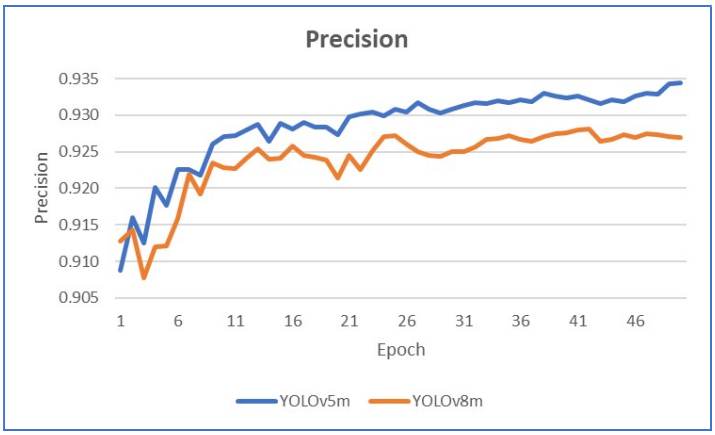

Fig. 2. Evolution of Precision measure during the training process

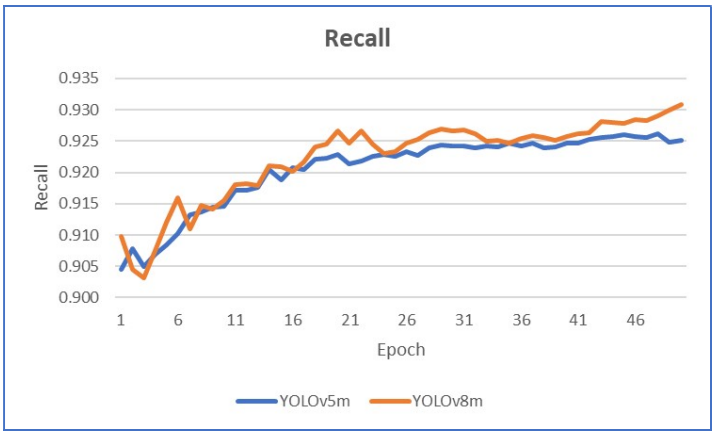

Fig. 3. Evolution of Recall measure during the training process

The training of the two models produced close results, with a plus for YOLOv5m in the case of Precision measure and respectively YOLOv8m in the case of the Recall measure.

Some detection results obtained using the trained YOLOv5m model are presented in Fig. 4. It should be mentioned that the time required for vehicle detection on the same hardware mentioned above, is less than 15 ms/frame which makes the trained model feasible for real-time traffic monitoring systems.

4. Vehicle tracking solutions

We have proposed various tracking solutions for the detected vehicles. Some of them represent *tracking by detection* (TBD) techniques, while others constitute motion-based vehicle tracking approaches.

Thus, an effective TBD algorithm has been introduced in [18]. The vehicles detected by a combination of a YOLO-based detector created by us and a variational PDE-based active contour [20] are tracked by a TBD approach that applies an IoU (Intersection over Union) – based metric and several conditions related to the vehicles' sizes, shapes and color distribution content [18].

The motion-based vehicle tracking frameworks considered and implemented by us include a Kalman filtering – based counting approach and an optical flow estimation-based tracking model. The vehicle tracking technique using a Kalman filter that estimates and predicts the position of the moving detected objects allows multiple vehicle objects tracking in real time [9, 21]. In Fig. 4, the vehicles are labelled with a unique identifier as long as they are visible in the video sequence.

Frame 35 Frame 45 Frame 55

Fig. 4. Detection and object tracking results using a Kalman filtering-based approach

The Optical Flow estimation algorithms can detect the movement pattern in consecutive frames caused either by the object or camera movement. We have considered an optical flow-based vehicle tracking solution that applies a version of the PDE variational model Horn-Schunck [22].

In Fig. 5 is presented the result of the considered optical flow algorithm applied to a traffic video sequence. In this example the tracked features are points detected by using the Shi-Tomashi corner detector [23]. The optical flow allows the more accurate reconstruction of the trajectory because it is based on the movement of a significant feature, unlike the detection of objects in which their image can be affected by the viewing angle.

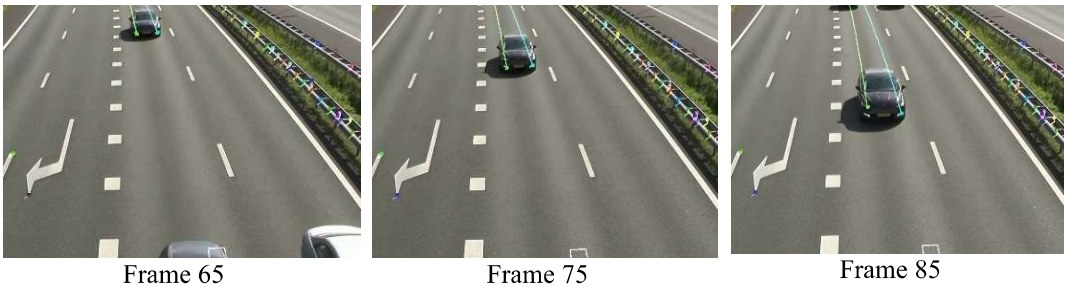

Fig. 5. Determining the vehicle trajectory using the optical flow estimation algorithm

5. Conclusions

We created and annotated a voluminous vehicle database, by using many traffic videos also acquired by us. The database can be used to develop novel deep learning models for the vehicle detection and recognition tasks.

The database was used to train and validate several YOLO-type deep learning models with promising results. The successful results of several multiple vehicle detection and tracking techniques have been described here. This database still needs to be enriched with new images of the less represented classes.

The work presented in this paper is part of the video traffic monitoring project that is acknowledged below and therefore, the enrichment of this database along with the development of new models that will be integrated into a complex video monitoring system will represent the focus of our future research.

Acknowledgment

This research work was supported by a grant of the Romanian Academy, GAR-2023, Project Code 19.

R E F E R E N C E S

[1] A. Banharnsakun and S. Tanathong, *Object detection based on template matching through use of best-so-far ABC*, Computational intelligence and neuroscience, 2014.

[2] T. Barbu, *Multiple Object Detection and Tracking in Sonar Movies using an Improved Temporal Differencing Approach and Texture Analysis*. U.P.B. Scientific Bulletin, Series A, 74, pp. 27–40, 2012.

[3] A. Wu, S. Zhao, C. Deng and W. Liu, *Generalized and Discriminative Few-Shot Object Detection via SVD-Dictionary Enhancement*, Advances in Neural Information Processing Systems", 34, 2021.

[4] T. Mordan, N. Thome, G. Henaff and M. Cord, *End-to-end learning of latent deformable part-based representations for object detection*, International Journal of Computer Vision, 127, pp. 1659-1679, 2019.

[5] M. Oliveira and V. Santos, *Automatic detection of cars in real roads using haar-like features*, Department of Mechanical Engineering, University of Aveiro, 3810, 2008.

[6] T. Barbu, *SVM-based Human Cell Detection Technique using Histograms of Oriented Gradients*, Mathematical Methods for Information Science & Economics: Proc. of AMATHI '12, Montreux, Switzerland, pp. 156-160, Dec. 29-31, 2012.

[7] T. F. Chan and L. A. Vese, *Active contours without edges*, IEEE Transactions on image processing, 10(2), pp. 266-277, 2001.

[8] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu and M. Pietikäinen, *Deep learning for generic object detection: A survey*, International journal of computer vision, 128 (2), pp. 261-318, 2020.

[9] H. A. Patel, and D. G. Thakore, *Moving object tracking using Kalman filter*. International Journal of Computer Science and Mobile Computing, 2(4), pp. 326-332, 2013.

[10] C. Yang, R. Duraiswami and L. Davis, *Efficient mean-shift tracking via a new similarity measure*, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, pp. 176-183, 2005. IEEE.

[11] K. Kale, S. Pawar and P. Dhulekar, *Moving object tracking using optical flow and motion vector estimati*on, 4th International conference on reliability, infocom technologies and optimization (ICRITO)(trends and future directions), pp. 1-6, September 2015. IEEE

[12] Y. Yuan, H., Yang, Y. Fang and W. Lin, *Visual object tracking by structure complexity coefficients*, IEEE Transactions on Multimedia, 17(8), pp. 1125-1136, 2015.

[13] W. Chantara, J. H. Mun, D. W. Shin and Y. S. Ho, *Object tracking using adaptive template matching*, IEIE Transactions on Smart Processing and Computing, 4(1), pp. 1-9, 2015.

[14] V. Srikrishnan, T. Nagaraj and S. Chaudhuri, *Fragment based tracking for scale and orientation adaptation*, 6th Indian Conf. on Computer Vision, Graphics & Image Processing, pp. 328-335, dec. 2008. IEEE.

[15] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri and F. Herrera, *Deep learning in video multi-object tracking: A survey*, Neurocomputing, 381, pp. 61-88, 2020.

[16] YOLOv5 - Ultralytics YOLO Docs, https://docs.ultralytics.com/models/yolov5/, last accesed on 1.11.2024.

[17] SIMPATIA Vehicle Database, http://iit.academiaromana-is.ro/simpatia, last accessed on 1.11.2024.

[18] T. Barbu, S.-I. Bejinariu and R. Luca, *Transfer Learning-based Framework for Automatic Vehicle Detection, Recognition and Tracking*, The 16th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2024, Iasi, Romania, June 27-28 2024, IEEE, doi: 10.1109/ECAI61503.2024.10607565.

[19] YOLOv8 - Ultralytics YOLO Docs, https://docs.ultralytics.com/models/yolov8/, last accesed on 1.11.2024.

[20] T. Barbu, *Robust contour tracking model using a variational level-set algorithm*, Numerical Functional Analysis and Optimization, Vol. 35, Issue 3, pp. 263-274, 2014.

[21] T. Barbu, S. -I. Bejinariu, *CNN-based Moving Vehicle Recognition using GMM-based Foreground Modeling*, *Level-set based Segmentation and Kalman Filter-based Tracking, 2024 International Conference on INnovations in Intelligent SysTems and Applications (INISTA),* Craiova, Romania, pp. 1-6, 4-6 September 2024, IEEE.

[22] B. K. P. Horn and B.G. Schunck, "Determining optical flow." *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[23] J. Shi, C. Tomasi, *Good features to track* , Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn., pp. 593-600, 1994.