YoloP-based Pre-processing for Driving Scenario Detection

Marianna Cossu[1], Riccardo Berta[1], Luca Forneris[1], Matteo Fresta[1], Luca Lazzaroni[1], Jean-Louis Sauvaget[2], Francesco Bellotti[1]

[1] Department of Electrical, Electronic and Telecommunication Engineering (DITEN)- University of Genoa, Via Opera Pia 11a, 16145 Genova, Italy.
{marianna.cossu, luca.forneris, matteo.fresta, luca.lazzaroni}@edu.unige.it, {riccardo.berta, francesco.bellotti}@unige.it

[2] STELLANTIS N.V. in Taurusavenue 1, 2132 LS, Hoofddorp, The Netherlands.
{jeanlouis.sauvaget}@stellantis.com

Abstract. Recognition of driving scenarios is getting ever more relevant in research, especially for assessing performance of advanced driving assistance systems (ADAS) and automated driving functions. However, the complexity of traffic situations makes this task challenging. In order to improve the detection rate achieved through state-of-the-art deep learning models, we have investigated the use of the YoloP fully convolutional neural network architecture as a pre-processing step to extract high-level features for a residual 3D convolutional neural network We observed thar this approach reduces computational complexity, resulting in optimized model performance, also in terms of generalization from training on a synthetic dataset to testing in a real-world one.

Keywords: Driving scenarios, synthetic datasets, automated driving, YoloP, deep learning, pre-processing, video classification, time-series, convolutional neural network, three-dimensional convolution.

1 Introduction

Development of advanced driving assistance systems (AFAS) and automated driving functions (ADFs) needs a precise analysis of their behaviour in different operational design domains (ODD). To this end, original equipment manufacturers (OEMs) and suppliers are developing rule-based and machine learning-based systems to detect different types of driving/traffic scenarios (e.g., [1, 2]), in which the various systems and functions should be tested. Moreover, detecting driving scenarios stands as a fundamental objective in expanding the ODD of ADFs. This capability empowers the systems to anticipate and proactively adapt to dynamic external conditions. However, detection of scenarios through deep learning models is very challenging especially when, for cost reasons, the input is provided by a single camera, and research in the field has just started. In order to improve the performance of such a detector, we have focused on high-level image pre-processing.

In general, pre-processing plays a key role in training machine learning (ML) models. In the domain of computer vision, raw data pre-processing can encompass several steps, such as input scaling, dimensionality reduction [3], normalization, data cleaning [4], feature extraction and selection. In this work we investigate the use of the state-of-the-art YoloP fully convolutional neural network architecture as a high-level feature extractor to enhance performance of a state-of-the-art residual 3D convolutional neural network that we have trained to classify 1 second driving scenario single camera recordings. In this context, this article focuses on three main research questions (RQs): (1) What is the effect of the proposed high-level pre-processing on the detector's training time? (2) What is the effect on network performance, in terms of the typical classification metrics? (3) What is the effect on the model's generalization ability?

The remainder of the paper is organized as follows: Section 2 presents the dataset used during the experiments; Section 3 illustrates the pre-processing approaches implemented in the paper; Section 4 presents the experiment executed and the results obtained, while the final section provides the concluding remarks.

2 Datasets

We trained our model to classify 5 driving scenarios (Fig. 1), namely cut-in executed in front of the ego vehicle (1) and behind it (2); cut-out in front of the ego (3) and behind it (4); and ego lane change (5).

Fig. 1. Examples of the synthetic dataset's classes

We used two different types of video-clip datasets: a synthetic one, exploiting a CarLA-based scenario generator [5, 6]; and a real-world one obtained after a manual labeling of the PREVENTION Dataset [7]. Manual labeling was necessary to extend Prevention to cover all our 5 types of labels. We trained and validated our system on the synthetic dataset only, while we tested it on both the datasets.

The implemented detector (described in the next section) receives as input sample a 1 sec. time-window, consisting of three frames (3 frames × 3 channels × 224 px × 224

px). Each sample is extracted from a single scenario class instance (which typically last more than 1 sec., e.g., 1-10 sec.s), with a stride of 0.2 seconds for training and validation, 0.1 seconds for testing. Split of training, validation and testing set is done at level of single scenario instance. Samples for training are shuffled before their utilization.

2.1 Synthetic dataset

The synthetic dataset contains over 800 simulated scenario instances for each class, with substantial variability. The corresponding number of samples is 73K for the training set, 25K for the validation set, and 6K for the test set. The dataset includes 14 distinct weather and lighting conditions, various traffic densities (ranging from car-free to low, up to high traffic cases), and 21 car models, with different colors. A key focus of our work is on intra-scenario action variability. To this end, each scenario is parametrized (e.g., in terms of duration, speed of the ego vehicle, distance from the leading vehicle), and their distribution is sampled from data extracted from the real world AD-Scene dataset, which accounts for over 1M km manual and automated driving recordings [8]. As a limitation, the simulator currently includes highways with 4/3 lanes, either in a straight configuration or with road curvature degree of 90°, with two different radius values.

2.2 Real-world dataset

To perform tests with a real-world dataset, we selected PREVENTION, which is one the very few publicly available recorded with cameras on highways. PREVENTION labels only front lane-change actions, thus we manually ported part of it to our target classes. During the labelling, we removed the samples that are not represented by the synthetic dataset ODDs (i.e., those with a number of lanes less than 3). However, we did not remove samples with curve radii very different from those of our synthetic dataset, which caused a reduction in performance. The real-world test set includes 10K time windows.

3 Pre-processing

To accomplish our objective, we preprocessed the previous presented dataset using the YOLOPv2 [9], as it reaches better performance compared to other networks sharing the same main goal. [9] allows extracting high-level information, including semantic segmentation of the drivable area and lane lines, and bounding boxes (BBs) of the vehicles.

We performed our experiments applying two different types of pre-processing, which corresponds to the utilization of three versions of input datasets (see Fig. 2):

1. Raw dataset: we directly use the frames extracted from the video-clips. The only pre-processing step involves pixel normalization.
2. Pre1 dataset: it is obtained with the first version of the high-level pre-processing method. This method extracts the most important information for each frame and stores it in the three different color channels. The red channel contains the

vehicle images BBs; the green channel stores the drivable area and lane lines. The gray scale of the original image minus the drivable area is stored in the blue channel.

3. Pre2 dataset: it is obtained with the second pre-processing version. Differently from the first one, it removes all the information that is not extracted by the YoloP, leaving only (in the red channel) the grayscale inside each vehicle's BB.

Fig. 2. An example frame for each pre-process: A) Raw, B) Pre1 and C) Pre2, respectively.

4 Experiments

The experiments were conducted on an Ubuntu PC equipped with an NVIDIA RTX A4000 with 16 GB of VRAM. Results are reported in terms of accuracy, precision, recall, and F1 score [10]. The model used for the classification task is a Residual 3D Convolutional Neural Network (R3D), presented in [2]. The R3D was trained for 6 epochs. Fig. 3 illustrates the training and validation accuracy's trend. From the plots it appears that both the pre-processing methods significantly speed up the training phase and enhance the network's performance.

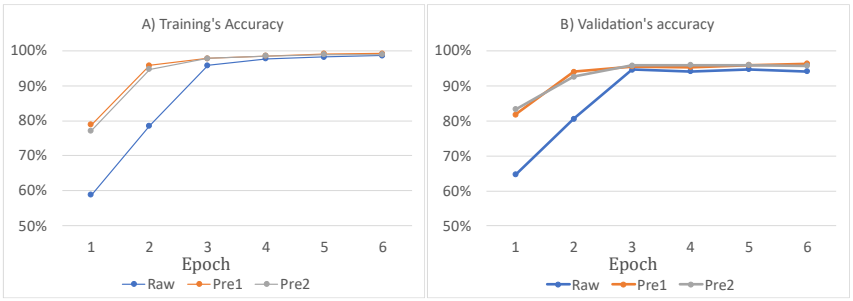

Fig. 3. Comparison in training and validation accuracy among the 3 investigated approaches.

The training times for the R3D model are as follows: 7.17 hours for the Raw dataset, 3.81 hours for Pre1, and 2.92 hours for Pre2, which clearly demonstrates the impact of our pre-processing techniques. Specifically, Pre1 led to a remarkable 50% reduction in training time compared to the Raw dataset, while PreV2 achieved a 60% reduction. This hints at a significant decrease in data dimensionality and complexity achieved by the pre-processing.

The experimental results obtained on the synthetic test set show that all versions of the model achieved high performance values (Table 1). Particularly, both Pre1 and Pre2 outperform the baseline by 3% and 2%, respectively.

Table 1. Performance reached on the synthetic test set by the R3D trained with and without the proposed pre-processing.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Raw – synthetic	92	92	92	92
Pre1 – synthetic	**95**	**95**	**95**	**95**
Pre2 – synthetic	94	95	94	94

In the second experiment, we utilized the real-world test set to assess the model's generalization ability. Since no real-world samples were included during the training phase and the real-world test set differed in elements such as camera resolution, position and road and lane type, compared to the synthetic dataset, the performance obtained at this stage was notably lower than on the synthetic dataset. However, this experiment highlights the robustness provided by the pre-processing techniques, that led to a substantial improvement in accuracy, up to 10%.

Table 2. Performance reached on the real-world test set by the R3D model trained with and without the proposed pre-processing.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Raw – real-world	63	68	63	60
Pre1 – real-world	66	70	63	63
Pre2 – real-world	**73**	**74**	**73**	**73**

5 Conclusion and future works

In order to improve the performance of a state-of-the-art camera-based driving scenario detector, we have explored the use of the YoloP system as a high-level feature extractor.

Concerning our first RQ, our experimental results demonstrated that the proposed pre-processing allows to cut more than 50% of the computational time needed to perform a full training. Performance measured with the typical classification metrics is increased by 3%, on the original synthetic dataset (RQ2). Finally, we addressed the RQ3 exploiting the PREVENTION dataset. Results show that the proposed pre-processing allowed a significant 10 % increase in accuracy. This indicates a clear improvement in the model's ability to generalize also to real-world scenarios. Since the preliminary results of our experiments are promising, a possible next goal of this research is to expand the ODD of the synthetic training dataset as much as possible (e.g., by introducing more road environment variability such as increasing and decreasing the number of lanes , etc.) and then assess the capability to synthetically generate training datasets that allow training models able to generalize well in the real-world.

Acknowledgements

The authors would like to thank all partners within the Hi-Drive project for their cooperation and valuable contribution. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006664. The sole responsibility of this publication lies with the authors. Neither the European Commission nor CINEA – in its capacity of Granting Authority – can be made responsible for any use that may be made of the information this document contains.

References

1. Izquierdo, R., Quintanar, Á., Lorenzo, J., García-Daza, I., Parra, I., Fernández-Llorca, D., Sotelo, M.Á.: Vehicle Lane Change Prediction on Highways Using Efficient Environment Representation and Deep Learning. IEEE Access. 9, 119454–119465 (2021). https://doi.org/10.1109/ACCESS.2021.3106692.
2. Cossu, M., Villon, J.L.Q., Bellotti, F., Capello, A., De Gloria, A., Lazzaroni, L., Berta, R.: Classifying Simulated Driving Scenarios from Automated Cars. Lect. Notes Electr. Eng. 866 LNEE, 229–235 (2022). https://doi.org/10.1007/978-3-030-95498-7_32.
3. Obaid, H.S., Dheyab, S.A., Sabry, S.S.: The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning. In: 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON). pp. 279–283 (2019). https://doi.org/10.1109/IEMECONX.2019.8877011.
4. Maharana, K., Mondal, S., Nemade, B.: A review: Data pre-processing and data augmentation techniques. Glob. Transit. Proc. 3, 91–99 (2022). https://doi.org/10.1016/j.gltp.2022.04.020.
5. Motta, J., Bellotti, F., Berta, R., Capello, A., Cossu, M., De Gloria, A., Lazzaroni, L., Bonora, S.: Developing a Synthetic Dataset for Driving Scenarios. Lect. Notes Electr. Eng. 866 LNEE, 310–316 (2022). https://doi.org/10.1007/978-3-030-95498-7_43.
6. Cossu, M., Berta, R., Capello, A., De Gloria, A., Lazzaroni, L., Bellotti, F.: Developing a Toolchain for Synthetic Driving Scenario Datasets. Lect. Notes Electr. Eng. 1036 LNEE, 222–228 (2023). https://doi.org/10.1007/978-3-031-30333-3_29.
7. Izquierdo, R., Quintanar, A., Parra, I., Fernández-Llorca, D., Sotelo, M.A.: The PREVENTION dataset: a novel benchmark for PREdiction of VEhicles iNTentIONs. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 3114–3121 (2019). https://doi.org/10.1109/ITSC.2019.8917433.
8. ADScene, towards an industrial scenarios plateform for Driving Assistance Systems design & validation – DSC 2021 EUROPE VR, https://dsc2021.org/adscene-towards-an-industrial-scenarios-plateform-for-driving-assistance-systems-design-validation/, last accessed 2023/05/30.
9. Han, C., Zhao, Q., Zhang, S., Chen, Y., Zhang, Z., Yuan, J.: YOLOPv2: Better, Faster, Stronger for Panoptic Driving Perception, http://arxiv.org/abs/2208.11434, (2022).