Situation-Aware Model Refinement for Semantic Image Segmentation

Lukas Habermayr[1,2], Markus Hofbauer[1], Joao-Vitor Zacchi[2], and Christopher B. Kuhn[1]

Abstract— **The quality of semantic image segmentation models can be affected by external factors such as weather or daytime. Those factors can lead to safety-critical mistakes. In this work, we propose a systematic approach to detect and alleviate such weaknesses of semantic segmentation models. We systematically evaluate a semantic segmentation model under different external factors and analyze which factors have the largest impact on the performance. Then, we collect new training data under the most harmful external factors and fine-tune the model. We use the CARLA simulator to obtain driving data under various environment settings. We deploy a state-of-the-art semantic segmentation model in two distinct driving environments. Then, we use the proposed process to detect which external factors affect model performance the most. We collect new training data under those factors and fine-tune the model. The proposed approach outperforms collecting the same amount of random additional data by up to 10.6%. Our results show the benefit of using an iterative refinement approach as opposed to merely collecting larger data sets. Finally, we use the knowledge about which factors affect performance the most to train a simple decision tree classifier to predict the model's performance given the current external factors. Problematic environments can be detected at an average accuracy of 87.5%.**

I. INTRODUCTION

Robust visual perception is an important challenge in robotics. In the field of autonomous driving, it is essential for the autonomous vehicle to know where it can drive and whether safety-critical objects such as pedestrians are present. This task can be approached with semantic image segmentation, where the class membership of each individual pixel of an image is predicted. Over the last years, models developed to perform this task have significantly improved and achieved remarkable performance on challenging benchmark data sets, while still being sufficiently lightweight for use in mobile systems [1].

While most research is focused on optimizing the model architecture to improve performance, less attention is given to the validation of a model once it is trained. Despite the accuracy rates of up to 90 % achieved by state-of-the-art architectures [2] in driving environments being remarkable, errors remain inevitable after the first training of a model. A major reason for safety-critical mistakes of computer vision algorithms is insufficient awareness of the operational environment [3]. The operational environment is defined by external factors such as weather, infrastructure, and other road participants.

[1]Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany {lukas.habermayr, markus.hofbauer, christopher.kuhn}@tum.de

[2]Fraunhofer IKS, Hansastraße 32, 80686 Munich, Germany {joao-vitor.zacchi}@iks.fraunhofer.de

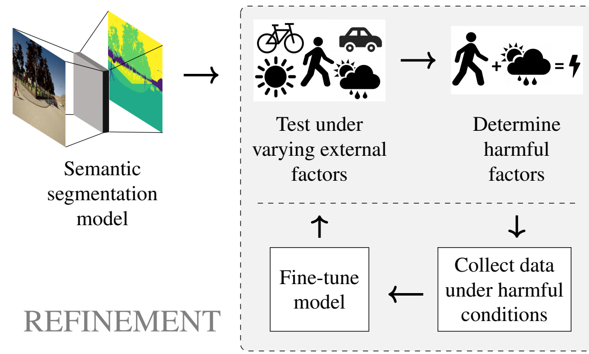


Fig. 1: Overview of the proposed situation-aware refinement approach for semantic segmentation.

In this work, we propose a systematic approach for reducing classification errors made due to external factors. We follow the idea of model refinement as introduced by Jha et al. [4] and apply it to semantic image segmentation. Rather than focusing on the model architecture, we keep the given architecture as it is and propose to iteratively refine the training data set instead. An overview of the proposed approach is shown in Figure 1.

We first train a state-of-the-art semantic segmentation model on driving data collected in the CARLA simulator [5] as a baseline. We choose CARLA for all our experiments to easily generate scenarios with the desired different external factors. Then, we record the model performance in new driving environments under varying external factors such as different weather or traffic conditions. Next, we calculate the correlations between model errors and external factors. After determining which external factors affect the performance the most, we obtain additional training data in CARLA under those specific external factors. Then, we refine our model by fine-tuning it with the extended training data set. In experiments in two distinct driving environments in CARLA, the proposed refinement approach significantly improves the segmentation results. As a reference approach, we also collect the same amount of additional data using random external factors. The systematic refinement approach outperforms using random data by 7.8 % to 10.6 %.

Finally, we use the correlation between external factors and model performance for predicting the model performance. For this, we train a simple binary decision tree classifier to classify the environment as challenging or not, using only the external factors as input. This allows to predict in which scenes the segmentation model performs the lowest at an accuracy of 87.5 %.

In summary, our contributions are:

- A comprehensive analysis of the influence of external factors on the performance of semantic segmentation models in a simulation environment.
- A systematic and situation-aware approach for the refinement of training data sets for semantic segmentation models that significantly outperforms using random additional data.
- An approach for leveraging information about the correlation between external factors and model performance to detect challenging environments.

The rest of this paper is structured as follows. In Section II, related work is discussed. We introduce our approach for the refinement of semantic segmentation models in Section III. In Section IV, we evaluate the proposed refinement approach in two driving scenarios and present a straightforward model for predicting the performance of a segmentation model based on external factors. Section V concludes the paper.

II. Related Work

The task of semantic segmentation has been approached in a variety of ways over the last years [1], [2], [6], [7]. We select the DeepLabV3+ architecture [1] for semantic segmentation in this work due to its competitive performance at reasonable computational costs. However, the proposed approach is independent of the specific architecture and can be applied to refine any given model.

Regarding the use of such deep learning methods in safety-critical applications, Jha et al. [4] argued that traditional safety monitors are not suitable for a resilient safety architecture. They pointed out that the assumption of independent component failures does not hold true for such methods. They proposed a refinement stage based on the deviations of a deep learning based model from the ground truth, e.g., the classification errors made by the model. However, they did not further specify how such a refinement could look like. The proposed approach is a model-independent way of implementing such a refinement.

In order to obtain the data necessary for the proposed refinement process, we rely on simulation data. Open source autonomous driving simulators such as CARLA [5], LGSVL [8], AirSim [9], or Deepdrive [10] are powerful tools for acquiring the data needed for refinement. We use CARLA since there is an active research community and it has been used for a wide range of tasks in autonomous driving [11]–[15]. Khan et al. [14] used CARLA to analyze the impact of individual weather factors on semantic segmentation. They concluded to add more rainy images to the training data set in order to increase model robustness. We build on their work to propose a general framework that considers any kind of external factor and combination thereof.

The idea of observing mistakes made by a semantic segmentation model in order to prevent future mistakes has been used in other applications before. Kuhn et al. [16] recorded pixel-wise misclassifications made by a semantic segmentation model and used them to train a pixel-accurate failure prediction model. This allows to detect or predict [17] areas with a high predicted failure rate, which can then be reclassified to correct some of the model's initial mistakes [18]. Instead of correcting a model's mistakes, the proposed approach aims to prevent the model from making such mistakes in the first place by collecting additional data that allows the model to alleviate its weaknesses.

Context information, such as the external factors investigated in this work, is an important part of autonomous driving. Trapp et al. [19] proposed a concept where context information such as weather conditions and the system state are used together to predict the performance of a system. Colwell et al. [20] aimed to improve safety of an autonomous vehicle by restricting its functionalities whenever the system is operating outside of its predefined bounds. Instead of restricting the system, our work aims to refine the system in order to allow it to handle as many scenarios as possible. In works such as [21], [22] or [23], future disengagements of an autonomous system were predicted by monitoring the input and output of the car. In this work, we show that external factors can also be used to predict system performance using a straightforward decision tree classifier.

III. Situation-Aware Refinement

In this section, we present the proposed situation-aware refinement process. First, we conceptually elaborate the underlying idea and then discuss each step in more detail.

A. Concept of Refinement

The idea of a refinement stage for deep-learning-based computer vision modules was first introduced by Jha et al. [4]. The idea is to constantly monitor and improve the perception models used for tasks such as semantic segmentation. After training, a model is deployed and continuously interprets new sensor data. The model's predictions then need to be compared to the actual ground truth. A large deviation of the prediction from the ground truth indicates a challenging input. The knowledge of which inputs led to the largest deviation can be used to refine the model. As a result, systematic faults can be located and eliminated during the development phase, making the model more robust and safe. This process needs to be conducted in an environment where deviations from the ground truth cannot cause any harm and where the ground truth is always available. Simulation environments fulfill both requirements and allow to simulate any weather and driving condition. Jha et al. [4] did not propose a specific implementation of such a refinement process. In this work, we present our approach for realizing a situation-aware refinement for semantic segmentation.

B. Situation-Aware Refinement

We propose a situation-aware refinement process and apply it to the task of semantic segmentation. We refer to a situation as the local driving scenario, e.g., urban or highway, plus the external factors, such as number of road users and weather conditions. We use the CARLA simulator [5]

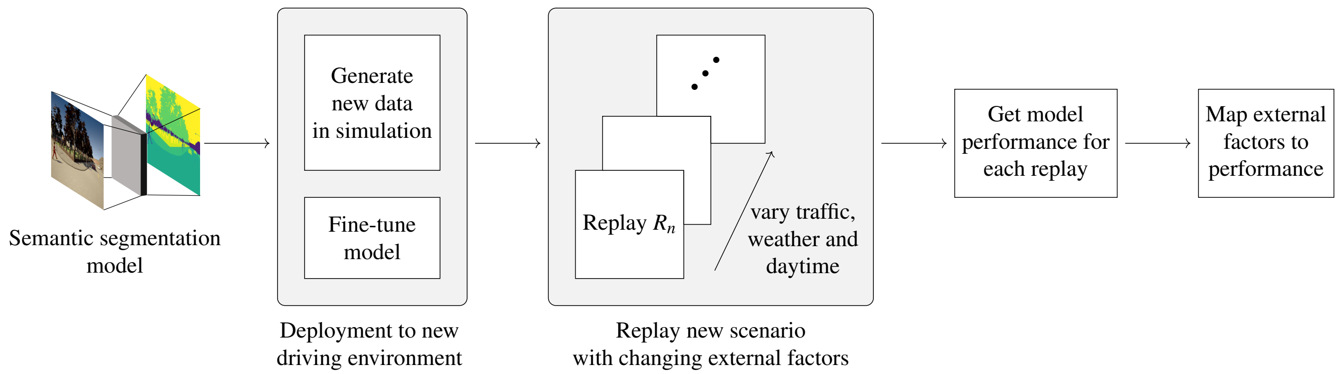


Fig. 2: Summary of the workflow to determine which external factors need to be addressed by collecting further data in the situation-aware refinement. We use a DeepLabV3+ model pre-trained on CARLA's default map and fine-tune on images captured from a different deployment map. We then replay two new scenarios from that map with altering external factors to map different external factors to the model performance.

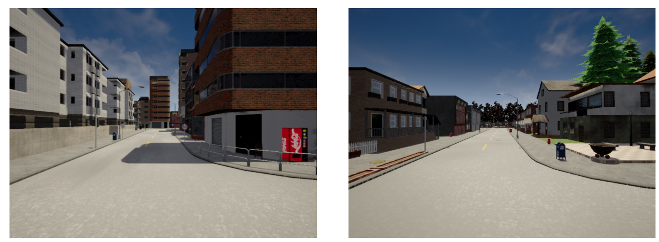
to generate data for varying scenarios and external factors. This allows a detailed analysis of the situation's influence on model performance in the refinement stage. Given a trained semantic segmentation model, we evaluate it multiple times in the same driving scenario by varying the external factors. Then, we evaluate the impact on performance when factors such as weather or traffic are altered. We do this by calculating the correlations between model performance and external factors. After determining which external factors affect the model the most, we record additional training data in CARLA under those specific external factors. We then combine the new data with the initial training data set and fine-tune the model to reduce the weaknesses determined in the previous analysis. Next, we discuss the details of generating the data required for this refinement.

C. Data Generation

The process for generating suitable data for the proposed situation-aware refinement is summarized in Figure 2. In detail, the workflow consists of the following parts:

1) Primary Model: The proposed approach requires a semantic segmentation model that can then be refined. For this paper, we use the state-of-the-art DeepLabV3+ [1] model which has shown competitive performance. We pre-train the model on 5000 images collected in CARLA's default map under dynamic weather and traffic conditions.

2) Model Deployment: Next, we deploy the model in a new CARLA map to simulate real life deployment. For this, we first fine-tune the semantic segmentation model on 400 images recorded in CARLA's *Town 2* map under dynamic weather and traffic conditions. Then, we deploy the fine-tuned segmentation model in two distinct scenarios in *Town 2*. The two deployment scenarios we used in this work are an *Urban Canyon* scenario and a *Residential Area* environment. Figure 3 shows the first frame of each scenario captured in clear conditions and without road users. The vehicle's trajectory is a straight line and does not contain any turns. A log file of the scenarios is stored in order to replay it with changing external factors as indicated in Figure 2.



Urban Canyon　　Residential Area

Fig. 3: First frame of each of the two deployment scenarios.

3) Scenario Replay: Next, we perform the key step for analyzing the correlation between external factors and model performance. We replay the trajectories driven in the two deployment scenarios in ten different combinations of external factors. The external factors of each replay are given in Table I. The trajectory logging of CARLA allows to replay exactly the same trajectory while only the external factors are varied. This assures that any performance change is due to the external factors and not due to a different driving environment.

Replay R	Replay Description
0	Clear, Noon, No Road Users
1	Clear, Noon, Road Users
2	Clear, Sunset, Road Users
3	Hard Rain, Noon, Road Users
4	25 % Fog, Morning, Road Users
5	60 % Fog, Morning, Road Users
6	Clear, Night, No Road Users
7	25 % Fog, Night, Road Users
8	60 % Fog, Night, Road Users
9	Wet and Cloudy, Noon, Road Users

TABLE I: Description of the external factors of each replay. We recorded each replay for both scenarios.

4) Performance Metric: We evaluate the performance of the segmentation model for a given replay R using the Mean Average Intersection-over-Union (MAIoU) calculated as

$$MAIoU_R = \frac{1}{N} \cdot \sum_{n=1}^{N} AIoU_n, \tag{1}$$

where AIoU_n denotes the average Intersection-over-Union of the n-th image and N is the total number of images in the replay R. We then calculate the correlation between the MAIoU of each replay and all external factors.

D. Situation-Aware Refinement

The situation-aware refinement is based on the results of the correlation of model performance and external factors from the previous step. A high negative correlation between the MAIoU and an external factor indicates a weakness of the model under that factor. Next, we therefore record additional training data under only the external factors with the highest negative correlation to model performance. We always replay the same trajectories driven through the two deployment scenarios. We thus avoid new semantic content and only focus on the changing external factors.

With this approach, the training data set is gradually extended with new samples. In order to avoid over-fitting to the new data, we only add every third image from the new samples. Then, the semantic segmentation model is fine-tuned with the extended data set.

IV. EXPERIMENTAL RESULTS

In this section, we present the results of the experiments performed to evaluate the proposed approach.

A. Experimental Setup

We first train a DeepLabV3+ [1] model on 5000 images from CARLA default map. Then, we deploy the model to the CARLA map *Town 2*. For this, we collect 400 images from *Town 2*. We fine-tune the pre-trained model with the 400 additional images. We then deploy this baseline model to the two scenarios shown in Figure 3 and test it under the ten combinations of different external factors listed in Table I.

B. Correlation Analysis

Next, we evaluate model performance for both scenarios and for each of the ten replays. Then, we calculate the correlations between performance and external factors.

1) Urban Canyon Scenario: Figure 4 visualizes the correlation between the external factors available in CARLA and the MAIoU for each replay. Precipitation and precipitation deposits, i.e., water puddles on the road, have the highest negative correlation to the MAIoU. Kahn et al. [14] already analyzed the influence of rain on image segmentation models extensively. We therefore instead focus on fog density, number of vehicles and number of pedestrians since those factors have the next-highest negative influence on the MAIoU.

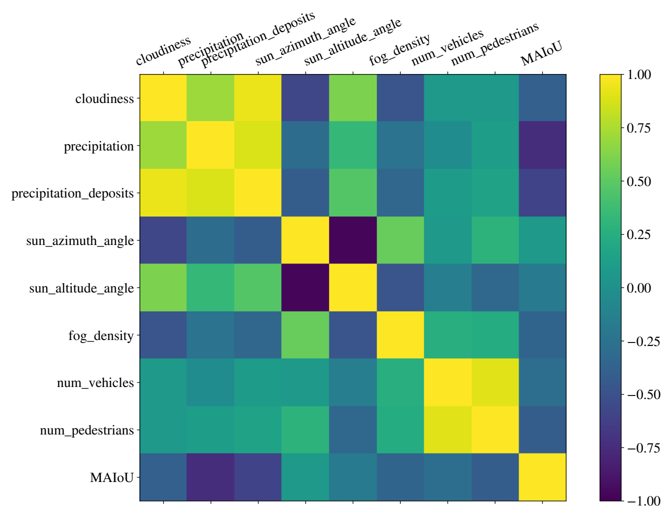


Fig. 4: Correlation of external factors and the MAIoU for the *Urban Canyon* scenario. Yellow indicates a positive correlation and purple indicates a negative correlation.

2) Residential Area: A similar correlation pattern can be observed for the results in the *Residential Area* scenario, shown in Figure 5. Here, fog has the largest negative correlation to the MAIoU. We therefore focus on fog density as well as the number of road users as harmful external factors during the following refinement stage.

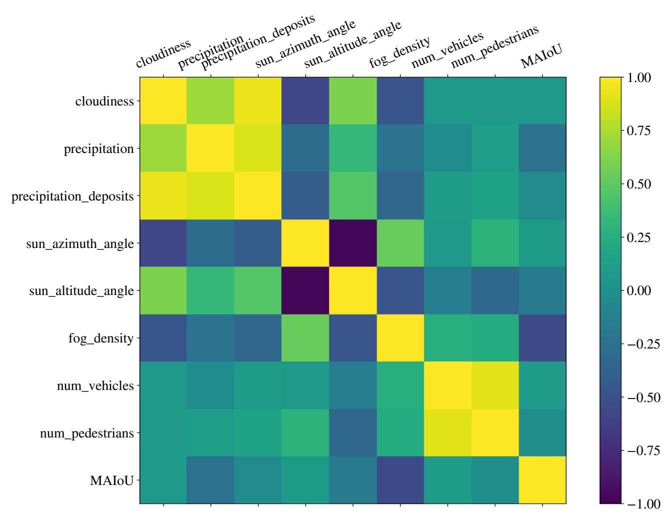


Fig. 5: Correlation of external factors and the MAIoU for the *Residential Area* scenario. Yellow indicates a positive correlation and purple indicates a negative correlation.

C. Refined Models

We next replay and record the same driving sequences in *Town 2* four times. We record once with the highest possible fog density, once with the highest number of road users and once with both maximum fog and the most pedestrians. As a reference, we also record the sequences with randomly varying external factors to investigate the benefits of performing the proposed correlation analysis beforehand. We use an equal number of daytime and night images in each recording. Using the four additional recordings, we then train a total of five refined models.

1) Fog: The *Fog* model is trained on a fine-tuning data set with additional samples recorded with the highest possible fog density in CARLA.

2) Road Users: The *Road Users* model is trained on a fine-tuning data set with additional samples where the number of road users is higher than in the initial data set.

3) Intersection: The *Intersection* model is trained on a fine-tuning data set with an intersection of the two harmful external factors, i.e., both high fog intensity and a high number of road users in each image.

4) Union: The *Union* model is trained on a fine-tuning data set that consists of additional samples with higher fog intensity as well as additional samples with more road users.

5) Random: The reference *Random* model is trained on a fine-tuning data set with additional samples where the external factors were randomly varied during the replay.

D. Performance Evaluation

We test all five refined models in both the *Urban Canyon* and the *Residential Area* scenario and for all replays listed in Table I. Figure 6 shows the results for the *Urban Canyon*. All refined models perform better than the baseline model. The *Intersection* model achieves an MAIoU of 0.40, outperforming the baseline model by 18.7 % and the *Random* model by 10.6 %. Interestingly, for replay 3 (*Hard Rain, Noon, Road Users*), the models refined with additional fog perform better than the randomly refined model, even though the *Random* model sees more rainy samples during fine-tuning. This indicates that fine-tuning with foggy images can lead to more robustness than fine-tuning with rainy images.

For the *Residential Area* performance shown in Figure 7, all refined models outperform the baseline model as well. The *Intersection* model refined with images with both more fog and more road users performs best again, outperforming the *Random* model by 7.8 %.

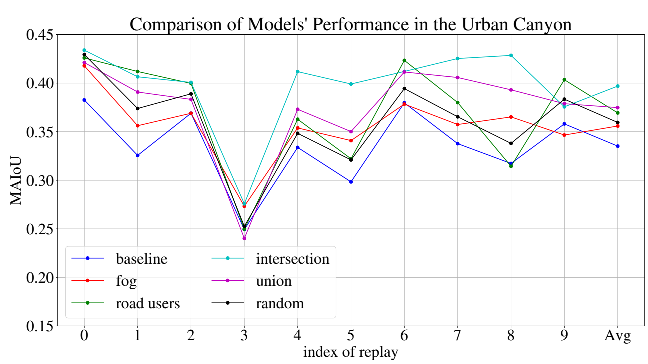


Fig. 6: MAIoU of all refined models for the *Urban Canyon*. *Avg* refers to the average over all replays. On average, the *Intersection* model (turquoise) performs best.

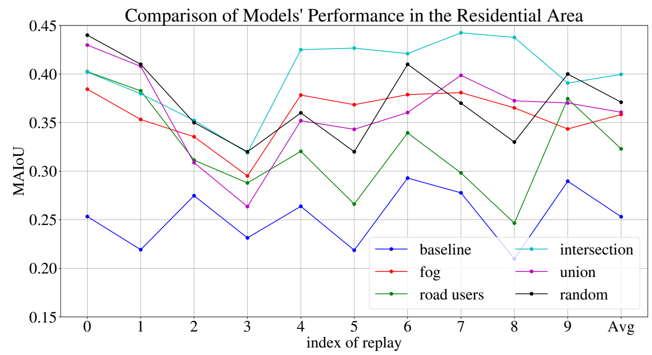


Fig. 7: MAIoU of all refined models for the *Residential Area*. *Avg* refers to the average over all replays. On average, the *Intersection* model (turquoise) performs best.

E. Visualization

In Figure 8, we visualize the segmentation results of the best performing models and compare them to the ground truth labels as well as to the unrefined baseline model. The two left columns show three exemplary images and their corresponding ground truth labels from the *Residential Area* and *Urban Canyon* scenarios under different external factors. The third column depicts the results of the baseline model that was fine-tuned to the new map, but not refined for the two deployment scenarios. It can be seen that the cars are largely misclassified. Details such as road lines or traffic signs are also not detected. The fourth column shows the output of the refined *Intersection* model. It visibly detects cars as well as road lines, parts of the traffic signs and parts of the vegetation. It also detects at least the outlines of the pedestrians. In contrast, the *Random* model largely misses details such as traffic signs and makes significantly more errors even with large objects such as cars. This demonstrates the benefits of using the proposed approach for refinement instead of simply collecting larger training sets.

F. Further Application

In a safety-critical application such as autonomous driving, the data gathered during the refinement process can be further used to predict the model's performance based on the current external factors. Following the concept of dynamic safety management first introduced by Trapp et. al [19], external factors can be used to predict a system's performance. For a proof of concept, we measure the MAIoU of the *Intersection* model for another 38 replays in both scenarios and map it to the external factors. We then divide the replays evenly into two classes according to their MAIoU. Class 1 denotes a performance in the lowest 50 %, while class 2 contains the replays with an MAIoU in the highest 50 %. We then aim to predict whether the performance for a new replay is low or high based only on the external factors as input. For this, we train a binary decision tree classifier with 38 replays, reserving another 10 replays as a test set. The resulting tree classifies the test set at an average accuracy of 87.5 %. The tree achieves an F1-score of 0.86 for class 1

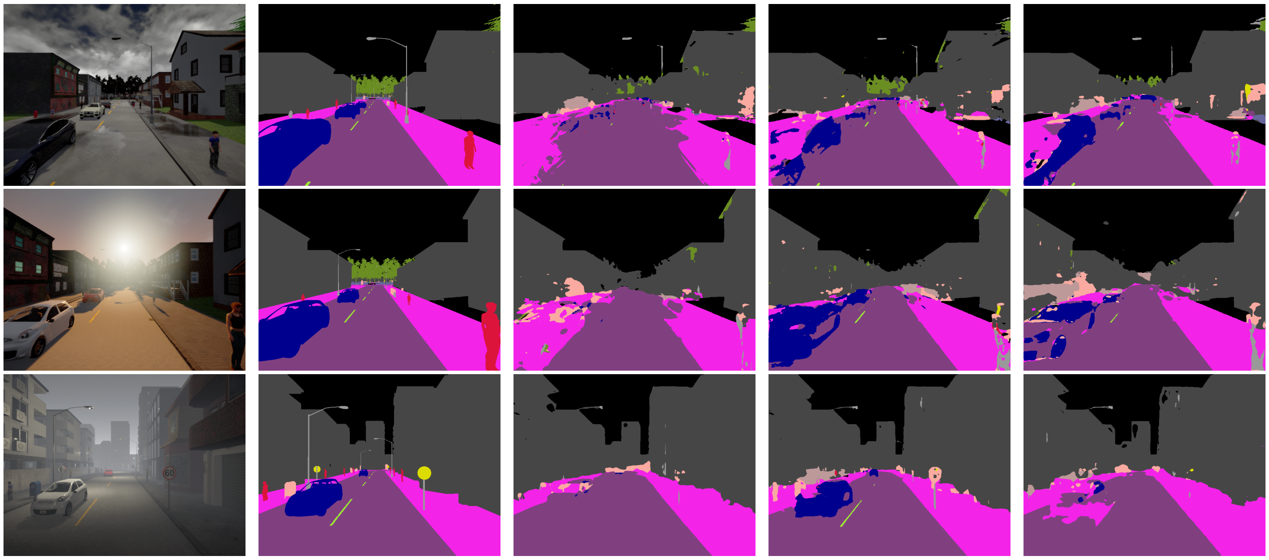


Fig. 8: Visualization of exemplary segmentation results of the baseline, *Intersection*, and *Random* model. The raw input image is shown in the first column. The second column shows the ground truth of the raw images. The third column shows the baseline model's predictions. The fourth and fifth columns show the *Intersection* and *Random* models' segmentation results.

(low performance) and 0.92 for class 2 (high performance). While this classification task is rather simple, it does indicate that considering external factors to predict insufficient model performance is a promising direction.

V. CONCLUSION

In this paper, we proposed a situation-aware refinement approach for semantic image segmentation. We evaluated a semantic segmentation model under varying external factors and calculated the correlation between model performance and factors such as weather or number of road users. We used the information from the correlation analysis to collect new data to fine-tune the model with. We evaluated the proposed approach in two scenarios in the CARLA simulator. The presence of fog and road users was identified to cause a decrease in model performance. The systematic refinement process with data collected under those factors outperformed using random additional data by 7.8 % to 10.6 % on average. This demonstrates the advantage of a systematic refinement process over the conventional approach of simply collecting more, but random data to improve model performance. The knowledge obtained in the refinement process can also be leveraged to predict the general performance of the model based only on the external factors at an accuracy of 87.5 % using a decision tree classifier.

The main limitation of the proposed approach is the use of simulation data. The benefits of the refinement in the real world still need to be investigated. However, collecting just the most useful data is even more important with real-world data due to the expensive manual labeling process. The proposed approach therefore has the potential of reducing costs in addition to improving model performance.

For future work, an interesting direction would be to account for more external factors at a finer granularity. This could allow to more accurately assess the weaknesses of a model, which in turn can allow a more efficient situation-aware refinement for semantic segmentation.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812.788 (MSCA-ETN SAS). This publication reflects only the authors' view, exempting the European Union from any liability. Project website: `http://etn-sas.eu/`.

REFERENCES

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[2] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[3] P. Koopman, U. Ferrell, F. Fratrik, and M. Wagner, "A Safety Standard Approach for Fully Autonomous Vehicles," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2019, pp. 326–332.

[4] S. Jha, J. Rushby, and N. Shankar, "Model-centered assurance for autonomous systems," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2020, pp. 228–243.

[5] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, " CARLA: An Open Urban Driving Simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[7] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” in *International Conference on Learning Representations*, 2016.

[8] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Mozeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, E. Agafonov, T. H. Kim, E. Sterner, K. Ushiroda, M. Reyes, D. Zelenkovsky, and S. Kim, “LGSVL Simulator: A High Fidelity Simulator for Autonomous Driving,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.

[9] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles,” in *Field and Service Robotics*. Springer, 2018, pp. 621–635.

[10] Voyage, “Deepdrive from Voyage - Push the state-of-the-art in self-driving,” https://deepdrive.voyage.auto/, 2019.

[11] M. Hofbauer, C. B. Kuhn, G. Petrovic, and E. Steinbach, “Telecarla: An open source extension of the carla simulator for teleoperated driving research using off-the-shelf components,” in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 335–340.

[12] D. Dworak, F. Ciepiela, J. Derbisz, I. Izzat, M. Komorkiewicz, and M. Wójcik, “Performance of lidar object detection deep learning architectures based on artificially generated point cloud data from carla simulator,” in *2019 24th International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE, 2019, pp. 600–605.

[13] M. Hofbauer, C. B. Kuhn, J. Meng, G. Petrovic, and E. Steinbach, “Multi-view region of interest prediction for autonomous driving using semi-supervised labeling,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.

[14] S. Khan, B. Phan, R. Salay, and K. Czarnecki, “ProcSy: Procedural Synthetic Dataset Generation Towards Influence Factor Studies Of Semantic Segmentation Networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[15] M. Hofbauer, C. B. Kuhn, L. Püttner, G. Petrovic, and E. Steinbach, “Measuring driver situation awareness using region-of-interest prediction and eye tracking,” in *2020 IEEE International Symposium on Multimedia (ISM)*, 2020, pp. 91–95.

[16] C. Kuhn, M. Hofbauer, S. Lee, G. Petrovic, and E. Steinbach, “Introspective Failure Prediction for Semantic Image Segmentation,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Rhodes, Sept. 2020, pp. 1–6.

[17] C. Kuhn, M. Hofbauer, Z. Xu, G. Petrovic, and E. Steinbach, “Pixel-wise failure prediction for semantic video segmentation,” in *IEEE International Conference on Image Processing (ICIP) 2021*, Anchorage, Alaska, 2021.

[18] C. B. Kuhn, M. Hofbauer, G. Petrovic, and E. Steinbach, “Better look twice-improving visual scene perception using a two-stage approach,” in *2020 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2020, pp. 33–40.

[19] M. Trapp, D. Schneider, and G. Weiss, “Towards Safety-Awareness and Dynamic Safety Management,” in *2018 14th European Dependable Computing Conference (EDCC)*, Iasi, Sept. 2018, pp. 107–111.

[20] I. Colwell, B. Phan, S. Saleem, R. Salay, and K. Czarnecki, “An Automated Vehicle Safety Concept Based on Runtime Restriction of the Operational Design Domain,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, June 2018, pp. 1910–1917.

[21] C. Kuhn, M. Hofbauer, G. Petrovic, and E. Steinbach, “Introspective failure prediction for autonomous driving using late fusion of state and camera information,” *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[22] ——, “Introspective Black Box Failure Prediction for Autonomous Driving,” in *31st IEEE Intelligent Vehicles Symposium*, Las Vegas, NV, USA, 2020.

[23] ——, “Trajectory-based failure prediction for autonomous driving,” in *32nd IEEE Intelligent Vehicles Symposium*, Nagoya, Japan, 2021.