Benchmarking Framework for Bad Weather Distortion in Semantic Segmentation in the Context of Autonomous Driving

Iris KREMER (337635), Nicolas d'ARGENLIEU (276507)
CS-503 Project Progress Report 1

I. Proposal Redefinition

A. Project Statement

As our previous proposal was not focused enough on Computer Vision and did not bring significant contributions to the domain, we shifted towards a more relevant objective. Following discussions with the TAs, we choose to develop a **Benchmarking Framework for Bad Weather Distortion in Semantic Segmentation in the Context of Autonomous Driving**. Our goal is to develop an accessible framework used to uniformly assess the performances of any bad weather de-noising model, e.g. de-raining or de-fogging. We plan on creating a pipeline composed of the following :

1) Two datasets:
    1a. Real-world natural images of clear-weather and bad-weather road situations,
    1b. Synthetic bad-weather on road situations.
2) A candidate de-noising model,
3) A SOTA Semantic Segmentation model,
4) A scoring module.

The pipeline will apply the **candidate de-noising model** to samples of both datasets. It will then pass the raw and resulting de-noised images through a **SOTA semantic segmentation model**. Finally, the **scoring module** is going to assess the performance of the de-noising model through the relative variation of semantic segmentation results.

The datasets are **BDD100K** [1] and **Cityscapes** [2] from which we are going to take real bad-weather images and clear-weather images to synthesize the bad-weather situations.

Our objective is to produce an automated scoring system outputting *a minima* simple metrics as done by [3] i.e. accuracy and IoU. We also get inspiration from the training method of [4] and the metrics for the de-fogging benchmark of [5]. Moreover we use the tools proposed by [6] for Computer Vision benchmarks. The goal is that the scoring pipeline accepts any models compliant with the specified definition standard.

B. Milestones Overview

The milestones would have been set as follows (*without taking into account the proposal's redefinition delay*) :

- **Milestone 1**
  - Select the various components of the pipeline: **noise-adding models** to create synthetic dataset (e.g. rain-adding model), sample candidate **de-noising model** (e.g. de-raining model), SOTA **semantic segmentation model** and **performance metrics**.
  - Run the de-noising and SOTA models on sample images.
  - Identify relevant categorization of images in both datasets for future evaluation tests.
  - Create distorted synthetic images with noise-adding model to constitute synthetic dataset.
- **Milestone 2**
  - Create basic modular structure in Docker.
  - Insert datasets and related methods into pipeline.
  - Insert candidate de-noising model into pipeline and candidate model integration methods.
  - Insert SOTA model and create automated methods.
  - Develop metric module and insert it into pipeline.
  - Link all aforementioned components.

The 2^{nd} milestone will be followed by a series of tests and documentation creation. However, as we had to redefine our project and research additional literature, we are running late in the proposed planning.

II. Milestone Progress

For Milestone 1, we found that we can reuse synthetic datasets provided as results of the following research papers:

- Rain-distorted dataset from [7] as created in [8].
- Fog-distorted dataset from [9] as created in [10],

Rain-distortion modeling was chosen for its novel method and state-of-the-art results as stated in [7]. This method also allows to control the intensity of the rain. The rain-distorted dataset also provided fog-distorted samples. Similarly, the fog-distortion modeling was selected due to its integration to the official **Cityscapes** dataset as stated in [9]. The fog-distorted dataset also provided rain-distorted images, useful as another evaluation for our framework.

Moreover we selected the different models i.e. de-noising model and SOTA model as follows:

- De-raining model Syn2Real from [11],
- SOTA Semantic Segmentation model HRNet with OCR as implemented in [12] from [13] and [14].

The candidate model was chosen because it is trained on synthetically generated data, but claims to generalize well to real-world data, so we will test this claim with our framework. Finally, the Semantic Segmentation model was chosen according to best performances related in [3].

However, the pre-trained SOTA model only included cross-task embeddings and therefore needed further training to fit semantic segmentation. Because of its incompatibility with the Windows OS, a next step is to train it on the SCITAS cluster (i.e. Linux distribution). Moreover, the different models use different versions of pytorch and will necessitate that we develop our overall framework to handle this.

III. Discussion & Next Steps

The main problem we encountered until now is **time**. We could not stick to our planning due to the delay of changing the project and conduction the appropriate research to properly define our contribution. To remedy this, we are going to proceed with an *essentials-first* approach, i.e. implementing a functional pipeline before designing proper experiments in order to anticipate eventual coding problems.

Moreover, we also identify some potential problems the most likely of which being:

- **Dataset's processing size** - The datasets we use consist of GigaBits of data. We need to anticipate computing times when conducting initial tests with the datasets. A solution in the case of intractable durations is to reduce the size of the data with which we will design our experiments.
- **Pytorch versions differences** - Our models use different versions of `pytorch` and other python libraries. It may be possible that we are unable to handle the integration of the different versions with the backbone code of the framework. In this case, a possible solution would be segment the pipeline into a sequence of smaller pipelines to be executed one after another by the user.
- **Model modification complexity** - As we are building our pipeline, we may modify and adapt the code of the models that we use (de-noising candidate model, SOTA model). This can lead to lengthy modifications that we are unable to achieve without requiring external help. To anticipate this, we are applying our previously mentioned *essentials-first* approach in order to encounter these problems at an earlier stage in the development.

However at this stage of our project, we don't see any critical issues towards our original goal. Therefore, our next steps are to produce the following deliverables:

- Semantic Segmentation results,
- Following groups of images:
  - Real clear-weather images,
  - Real bad-weather images (rainy and foggy),
  - Synthetic bad-weather images (rainy and foggy) with levels of distortion intensity.
- De-noising model results,
- Dockerfile for the pipeline's backbone result.

IV. Author Contribution Statement

Nicolas d'Argenlieu (N.A.) and Iris Kremer (I.K.) conceived of the new project proposal together and each performed part of the literature research for this task. I.K. selected the BDD100K dataset and N.A. selected the Cityscapes dataset as well as the synthetic datasets. N.A. chose the SOTA semantic segmentation model and investigated how to use it in our project. I.K. tested several and chose the candidate de-noising model. N.A. wrote the majority of the milestone 1 report, I.K. completed its last parts.

References

[1] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," *CoRR*, vol. abs/1604.01685, 2016. [Online]. Available: http://arxiv.org/abs/1604.01685

[3] Thomas E. Huang, "Semantic Segmentation Models of BDD100K," 2020. [Online]. Available: https://github.com/SysCV/bdd100k-models/tree/main/sem_seg

[4] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 558–567.

[5] M. Bijelic, P. Kysela, T. Gruber, W. Ritter, and K. Dietmayer, "Recovering the unseen: Benchmarking the generalization of enhancement methods to real world data in heavy fog." in *CVPR Workshops*, 2019, pp. 11–21.

[6] MMSegmentation Contributors, "OpenMMLab Semantic Segmentation Toolbox and Benchmark," 7 2020. [Online]. Available: https://github.com/open-mmlab/mmsegmentation

[7] M. Tremblay, S. S. Halder, R. de Charette, and J.-F. Lalonde, "Rain rendering for evaluating and improving robustness to bad weather," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 341–360, 2021.

[8] Raoul de Charette, J-F. Lalonde, "Rain Rendering for Evaluating and Improving Robustness to Bad Weather," 7 2020. [Online]. Available: https://github.com/cv-rits/rain-rendering/

[9] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, Sep 2018. [Online]. Available: https://doi.org/10.1007/s11263-018-1072-8

[10] Christos Sakaridis, "Fog Simulation on Real Scenes for Partially Synthetic Foggy Data," 10 2018. [Online]. Available: https://github.com/sakaridis/fog_simulation-SFSU_synthetic

[11] R. Yasarla, V. A. Sindagi, and V. M. Patel, "Syn2real transfer learning for image deraining using gaussian processes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2726–2736.

[12] Xie Jingyi, Ke Sun, Jingdong Wang, "High-resolution networks and Segmentation Transformer for Semantic Segmentation," 7 2020. [Online]. Available: https://github.com/HRNet/HRNet-Semantic-Segmentation

[13] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-resolution representations for labeling pixels and regions," *CoRR*, vol. abs/1904.04514, 2019. [Online]. Available: http://arxiv.org/abs/1904.04514

[14] Y. Yuan, X. Chen, X. Chen, and J. Wang, "Segmentation transformer: Object-contextual representations for semantic segmentation," in *European Conference on Computer Vision (ECCV)*, vol. 1, 2021.