Evaluating and Increasing Segmentation Robustness in CARLA

Venkatesh Sambandham Konstantin Kirchheim
Frank Ortmeier

Otto Von Guericke University
Magdeburg, Germany

Workshop on AI Safety (WAISE-2023)
September 18, 2023

Presentation Overview

Introduction

- Robustness in Deep Learning models,
  - Ability of a model to maintain performance under uncertain, adversarial and unexpected conditions.
  - Robustness testing plays a pivotal role in assessing generalization performance and ensuring system safety, particularly in critical applications such as autonomous driving and medical diagnosis.
  - Majority of research focuses on testing adversarial robustness.

Prior Works and Motivation

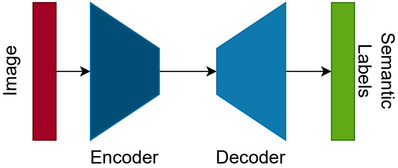
- **Prior Works**
  - Prior works Evaluated the robustness of CNNs for Computer Vision tasks[11, 5, 7].
  - Overlay existing images with synthetic perturbations like noise, rain, snow and fog.
  - Vision Transformers (ViTs), although claimed to have superior performances compared to CNNs, robustness on a computer vision context has not been widely tested.
- **Our Approach**
  - Use CARLA[2] to apply environmental conditions directly into the simulation environment.
  - simulate more realistic scenarios like reflections in water, changing shadows and so on....

Dataset

(a) Training (b) Rain (c) Fog (d) Sun

- **Train-Validation** : 60-min (in-simulation) drive at clear weather conditions and 90° solar angle - 2700 images for training, 750 images for validation.
- **Test**: A Challenging 10-second scenario at varied level of fog, rains and solar angle, 75 variations, 300 images per variation.
- HD images paired with pixel level semantic labels of 23 classes.

Architectures



Encoders

- ResNet [4]
- EfficientNet-B6 [8]
- MiT-B3 (Segformer) [9]

Decoders

- UNet [3]
- FPN [6]
- PSPNet [10]

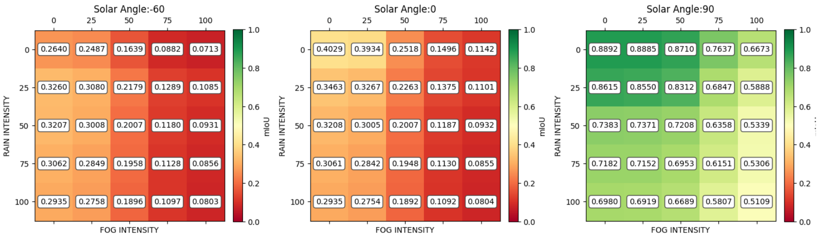
Synthetic Image Perturbations

- Weighted mIoU.
- Conditions
  - Clear Weather
  - solar angle - 90° (Noon)
- Noises
  - Gaussian (G)
  - Pixel Dropout (PD)
- Blurs
  - Gaussian (G)
  - Motion Blur (M)

Decoder	Backbone	Clear (mIoU)	Noise		Blur	
			G	PD	G	M
UNet	ResNet-50	88.42	45.21	23.55	85.39	82.08
	ResNet-101	88.02	30.75	19.82	85.37	82.54
	EfficientNet B6	**88.91**	51.34	17.06	82.91	81.52
	MiT-B3	88.84	57.64	18.92	86.22	83.62
FPN	ResNet-50	87.54	74.37	37.60	84.45	80.17
	ResNet-101	88.42	62.0	31.87	86.09	81.52
	EfficientNet B6	88.81	76.34	59.17	82.82	79.69
	MiT-B3	88.84	**83.75**	61.38	**86.69**	**84.03**
PSPNet	ResNet-50	87.35	67.34	12.86	82.61	78.86
	ResNet-101	87.44	47.41	17.49	81.57	80.45
	EfficientNet B6	87.91	40.22	37.74	78.55	79.31
	MiT-B3	87.74	81.93	**62.09**	85.87	83.16

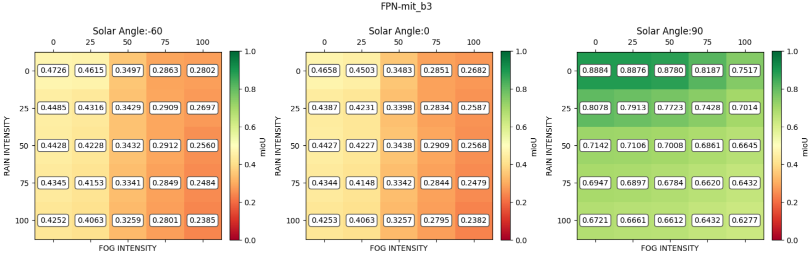
Adverse Weather Conditions

Figure: EfficientNet-B6 + UNet



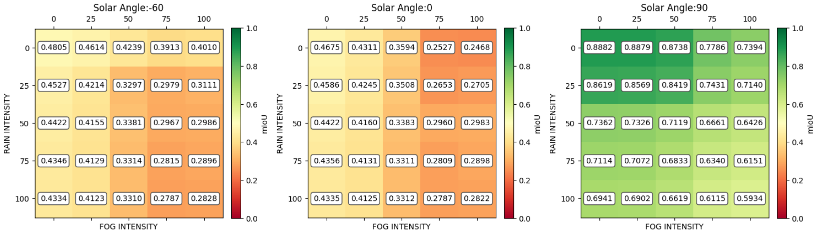
Adverse Weather Conditions

Figure: EfficientNet-B6 + FPN

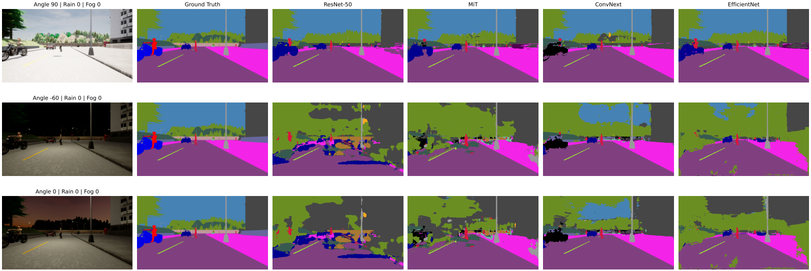


Adverse Weather Conditions

Figure: MiT + FPN



Adverse weather condition



Generate Fine-Tuning data

- **Parameters**
  - Rain - 5 levels
  - Fog - 5 levels
  - Lighting - $0°, 90°, -60°$
- 75 combinations to achieve complete coverage.
- Use 2-projection coverage [1] to identify best scenario.
- 100% coverage in 26 iterations.

Fine-Tuning

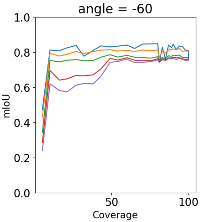
Algorithm 1 Data Generation Algorithm

Require: Initial Dataset $\mathcal{D}_0$, Pre-Trained Model $f(\cdot, \theta_0)$, Test Set $\mathcal{D}_v$

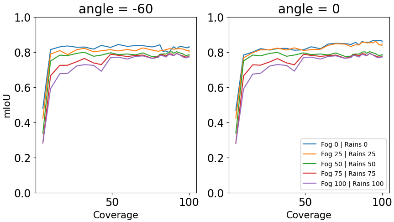
```
while coverage < 1 do
    f(·, θ1) ← train(f(·, θ0), D)
    test(f(·, θ1), Dv)
    coverage ← cov(D)
    s ← scenario that maximizes coverage
    D ← generate(s) ∪ D
end while
```

- 40 images for each scenario.
- 35 for fine-tuning, 5 for validation.
- 5 images were added from previous iterations to avoid drift.

Fine-Tuning results



(a) EfficientNet-B6



(b) MiT-B3

Fine Tuning Results

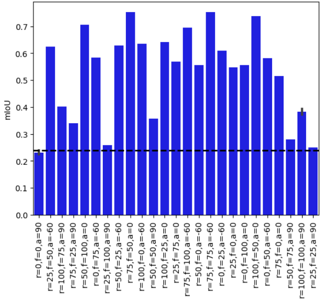


Figure: finetuning on individual scenarios - MiT-B3 model; results under extreme weather conditions $\{$rain(r) = 100, fog (f) = 100, solar angle (a) = $-60°\}$

Fine Tuning Results

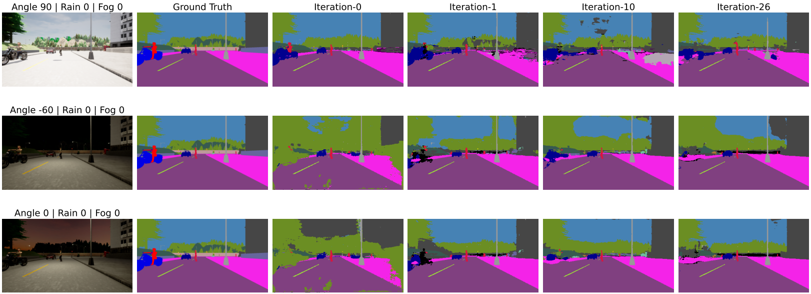


Figure: FPN + EfficientNet-B6 backbone.

Fine Tuning Results

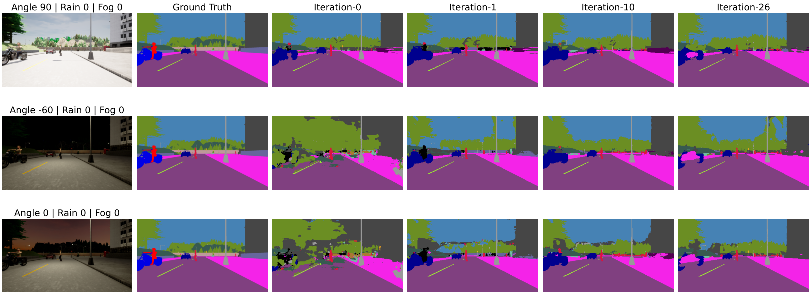


Figure: FPN + MiT-B3 backbone

Fine Tuning Results

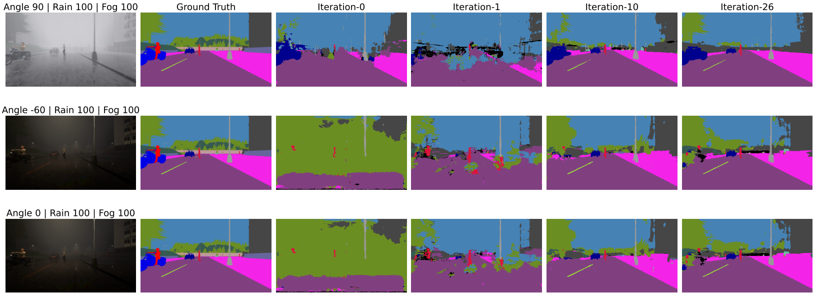


Figure: FPN + EfficientNet-B6 backbone.

Fine Tuning Results

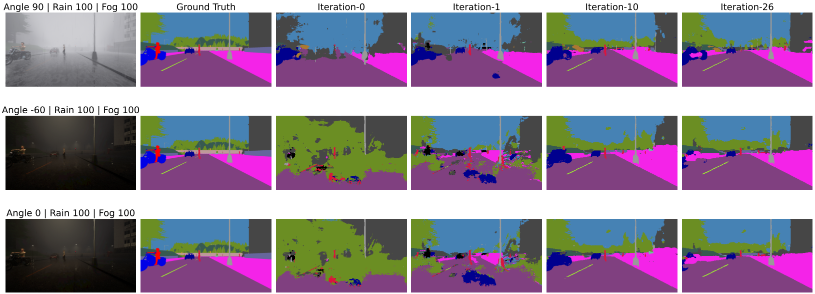


Figure: FPN + MiT-B3 backbone

Conclusions

- **Transformer Robustness**: Our study evaluated segmentation architectures on a dataset generated in a simulation environment, revealing comparable robustness with CNN and Vision Transformer-based backbones.
- **EfficientNet-B6 Efficacy**: Notably, EfficientNet B6 demonstrated competitive performance compared to a Vision Transformer based backbone.
- **Finetuning for enhanced robustness**: Fine-tuning with minimal training data under adverse conditions emerged as a powerful technique for enhancing model robustness.

Future Works

- **Model Variations and Architecture** : Future research should explore the applicability of our findings to fully transformer-based models, CNNs with dilated convolutions, and CNN architectures inspired by the design of Vision Transformers (ViTs).
- **Real World Validation** : Replicating our experiments in field tests using real-world data is crucial to determine whether our conclusions hold in practical scenarios. This step will validate the generalizability of our findings to real conditions, ensuring the reliability of the proposed model configurations.

The End

Questions? Comments?

Cheng, C.H., Huang, C.H., Yasuoka, H.: Quantitative projection coverage for testing ml-enabled autonomous systems. In: Automated Technology for Verification and Analysis: 16th International Symposium, ATVA 2018, Los Angeles, CA, USA, October 7-10, 2018, Proceedings 16. pp. 126–142. Springer (2018)

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An open urban driving simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. pp. 1–16 (2017)

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al.: U-net: deep learning for cell counting, detection, and morphometry. Nature methods **16**(1), 67–70 (2019)

He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

Kamann, C., Rother, C.: Benchmarking the robustness of semantic segmentation models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8828–8838 (2020)

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)

Michaelis, C., Mitzkus, B., Geirhos, R., Rusak, E., Bringmann, O., Ecker, A.S., Bethge, M., Brendel, W.: Benchmarking robustness in object detection:

Autonomous driving when winter is coming. In: Machine Learning for Autonomous Driving Workshop, NeurIPS 2019. vol. 190707484 (2019)

- Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In: Neural Information Processing Systems (NeurIPS) (2021)
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890 (2017)

Zhou, X., Kouzel, M., Alemzadeh, H.: Robustness testing of data and knowledge driven anomaly detection in cyber-physical systems. In: 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W). pp. 44–51 (2022). https://doi.org/10.1109/DSN-W54100.2022.00017