# How Images Defects in Street Scenes Affect the Performance of Semantic Segmentation Algorithms

Hoda Imam[1], Bassem A. Abdullah[2], Hossam E. Abd El Munim*[3]
Computer and Systems Engineering Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt[1,2,3]
Poolia IT Cloud, Stockholm, Sweden[1]

***Abstract*—Semantic segmentation methods are used in autonomous car development to label pixels of road images (e.g. street, building, pedestrian, car, and so on). DeepLabv3+ and PSPNet are two of the best performance semantic segmentation methods according to Cityscapes benchmark. Although these methods achieved a very high performance with clear road images, yet these two methods are not tested under severe imaging conditions. In this work, we provided new Cityscapes datasets with severe imaging conditions: foggy, rainy, blurred, and noisy datasets. We evaluated the performance of DeepLabv3+ and PSPNet using our datasets. Our work demonstrated that although these models have high performance with clear images, they show very weak performance among the different imaging challenges. We proved that the road semantic segmentation methods must be evaluated using different kinds of severe imaging conditions to ensure the robustness of these methods in autonomous driving.**

***Keywords*—Semantic segmentation; deep learning; cityscapes; DeepLabv3+; PSPNet**

## I. Introduction

Autonomous vehicles are vehicles that can move with little or no human interaction. It collects all the environment surrounding information to simulate human behavior in driving safely. Autonomous vehicles rely on sensors, actuators, driving algorithms, machine learning technologies, and powerful micro-controllers with GPUs to execute the self-driving software.

Self-driving software uses semantic segmentation algorithms that take road scene images as input and give a label to each pixel in the input images. These labels describe the object class that these pixels present (road, traffic light, vehicle, human, etc.). Fig. 1 shows an example of input and ground truth images used in semantic segmentation algorithms. Semantic segmentation is very powerful as it helps self-driving software with understanding scene images at the pixel level.

In recent years, after the emergence of convolutional neural networks (CNNs), segmentation made huge progress. Many semantic segmentation methodologies depending on CNN have been developed in [1-7]. These methodologies were trained and evaluated using large scale datasets [8-11].

These networks are designed and tested to work efficiently with clear images. Also, all the images in the large scale datasets [8-11] are clear images. Yet, semantic segmentation methodologies don't take into consideration the different types of defects in images coming from video cameras.

Defects in images could be a result of bad weather or electronic noise. These defects in images decrease the performance and the accuracy of semantic segmentation methodologies and thus lead to a wrong driving decision taken by the vehicle's self-driving system.

Overall, the state-of-the-art methods take into consideration only the performance of these methods on clear images, as these methods are limited by the existing datasets. These methods ignore the performance with unclear images. Semantic segmentation methods should take into consideration these challenges and handle these severe imaging conditions. Although certain works studied object detection methodologies with challenges as foggy [12], rainy [13, 14], blurred [15], and noisy [16-18] images, yet only a few works [12, 19] studied these challenges with semantic segmentation methodologies. Here, we are studying road semantic segmentation methodologies with different challenges.

In this work, we address different kinds of severe imaging conditions: fog, rain, blurring, and noise. We study the performance of semantic segmentation with these four imaging defects. As collecting real datasets with these conditions is very hard, we decided to use Cityscapes dataset [11] and introduce fog, rain, blurring, and noise on the clear images of the dataset.

Even that author in [12] addressed the performance of semantic segmentation methods [1, 2] with fog. These two methods have very low performance on the Cityscapes benchmark. The mIoU of these two methods is 73.6% and 67.1% respectively on Cityscapes test set.

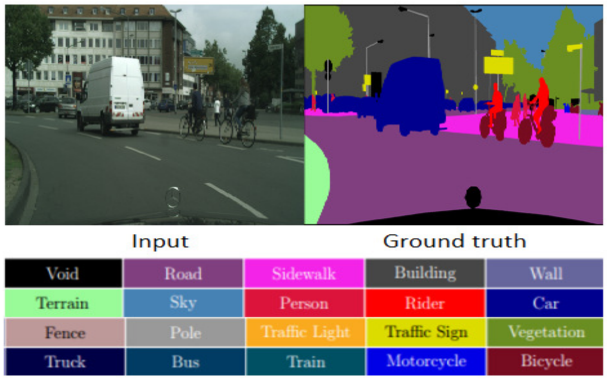


Fig. 1. Example of the Input Image used with Semantic Segmentation Methods and the Ground Truth Image that these Methods Seek to Achieve as an Output.

*Corresponding Author

In this work, we are not only generating new evaluation datasets but also studying the performance of two powerful methods in semantic segmentation against imaging defects challenges. We study the performance of DeepLabv3+ and PSPNet [5, 4] which are rated as two of the top methods in semantic segmentation. DeepLabv3+ and PSPNet score mIoU of 82.1% and 81.2% respectively on Cityscapes test set.

This work is an expansion to our previous work [20], which studied the performance of semantic segmentation methods with fog and blur challenges. In this paper, we added rain and noise to the challenges used in the performance evaluation of semantic segmentation methods.

In summary, this work contributions are:

(i) Addressing the performance degradation in semantic segmentation methods with severe imaging conditions.

(ii) Creating rainy, foggy, blurred, and noisy datasets for evaluation purposes. We made use of an algorithm provided by [12] to add fog in Cityscapes dataset.

(iii) Using our newly created datasets in performance evaluation of two top semantic segmentation methods(DeepLabv3+ and PSPNet).

This paper is organized as follows: Section 2 reviews shortly the methods of semantic segmentation used in performance measurement. Section 3 describes the challenging evaluation datasets. Section 4 shows the experiments and the performance evaluation results. Finally, Section 5 makes a brief conclusion.

## II. Methods

In this section, we will describe briefly the semantic segmentation methods used in our methods performance search. DeepLabv3+ and PSPNet are two of the best-performing methods according to Cityscapes benchmark. These are two state-of-the-art road semantic segmentation methods used to label pixels of road images (e.g., street, building, pedestrian, car, and so on).

### A. *DeebLabv3+*

DeebLabv3+, the extension of DeebLabv3, is a very powerful semantic segmentation model invented by Google. DeebLabv3+ is mainly composed of two phases:

**Encoder:** In this phase, the model extracts the main features from the input image. It detects the presence of the objects and their location. DeepLabv3+ uses Atrous Spatial Pyramid Pooling (ASPP), which investigates convolutional features by applying atrous convolution at multiple scales.

**Decoder:** In this phase, the model refines the segmentation results along the object boundaries. It applies 1 x 1 convolutions on the low-level features and concatenates it with the upsampled encoded features. It then applies 3 x 3 convolutions and upsamples the features to output the prediction image with the same size of the input image.

DeebLabv3+ scored a performance of 89.0% using the test set of PASCAL VOC 2012 benchmark [10] and 82.1% using the test set of Cityscapes benchmark. Fig. 2 shows the network structure of DeepLabv3+.

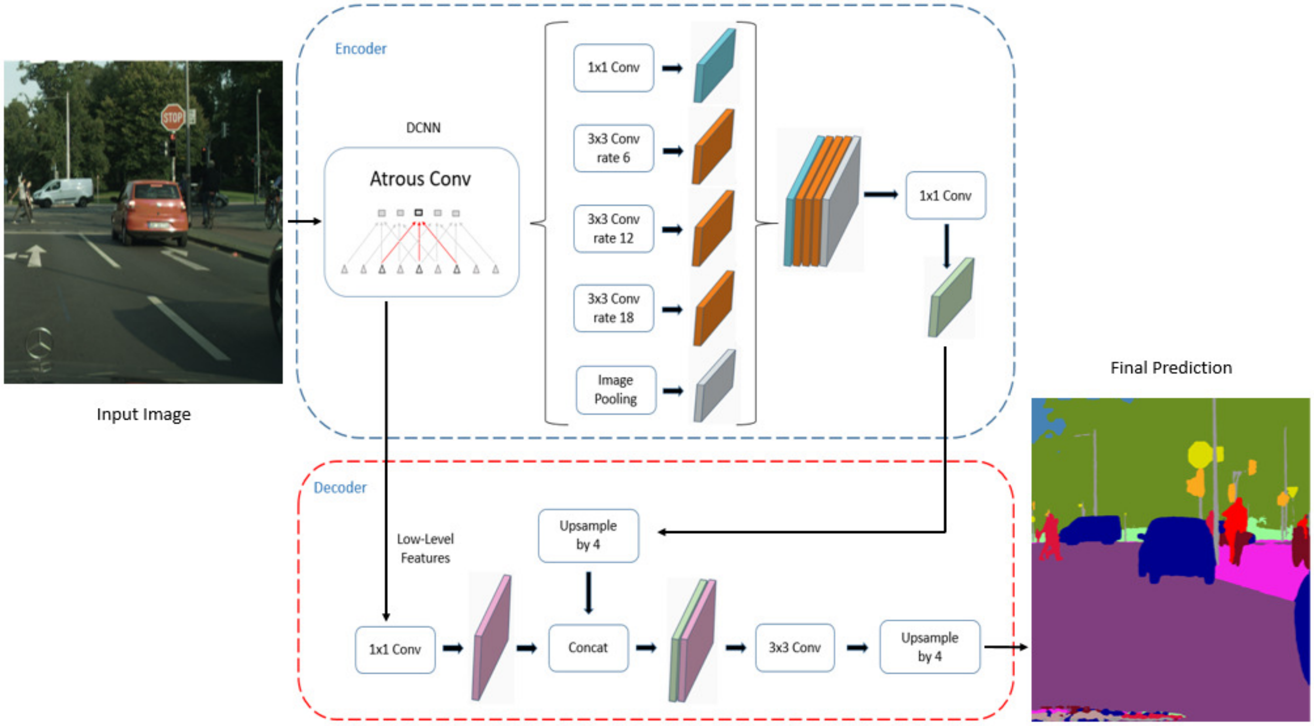


Fig. 2. DeebLabv3+ Method Structure Showing its different Phases. The Predicted Image is the Output of DeebLabv3+ Method using an Input Image from Cityscapes Clear Dataset.

### B. PSPNet

Pyramid Scene Parsing Network (PSPNet) is a semantic segmentation model developed to enhance learning the full context representation of the input scene. PSPNet is mainly composed of four phases:

(i) Creating the feature map of the input image using CNN.
(ii) Applying pyramid pooling mechanism. This pooling mechanism contains four pooling levels presented in a pyramid hierarchy that is proceeded with a 1x1 convolutional layer. Each pyramid level is responsible for analyzing different parts from the input image in different locations.
(iii) Upsampling and concatenating the pyramid levels outputs to give an initial feature maps which contain the local and global information of the input image.
(iv) Applying a convolutional layer to the feature maps to generate the prediction image.

PSPNet scored a performance of 85.4% using the test set of PASCAL VOC 2012 benchmark and 81.2% using the test set of Cityscapes benchmark. Fig. 3 shows the network structure of PSPNet.

## III. Evaluation Dataset

In order to evaluate semantic segmentation methods, we chose to introduce fog, rain, blur, and noise to Cityscapes evaluation set which consists of clear images only. In this section, we will describe in details our proposed challenging datasets and examples from the datasets are shown in Fig. 4.

Due to the difficulty of collecting and annotating images for rainy weather, we choose to generate rain into clear weather images of Cityscapes dataset. In this work, we consider a rain image as a composition of a rain-free image and a rain layer. We formulate the rain image O(i,j) at pixel i,j as the following:

$$O(i,j) = I(i,j) + R(i,j) \tag{1}$$

where I(i,j) denotes the rain-free image and R(i,j) denotes the rain layer. The rain layer is created by the following processes:

**Algorithm 1** Algorithm of adding rain to clear weather images

```
function AddRain($I(i,j), \alpha$)    ▷ $I(i,j)$ clear image, $\alpha$ rain density

    ▷ create black layer withe the same size of the Clear weather image $I(i,j)$
    height, width ← $I(i,j)$.shape
    $B(i,j)$ ← zeros(height,width)

    ▷ Add Gaussian noise with standard deviation equals Rain density $\alpha$
    $N(i,j) \leftarrow B(i,j)$ + Gaussian noise with standard deviation $\alpha$

    ▷ Threshold the output to keep white pixels from (150 to 255) only
    $Z(i,j)$ ← threshold($N(i,j)$ , 150 , 255)

    ▷ Apply diagonal motion filter to the output with kernel size 50
    $R(i,j) \leftarrow Z(i,j) * I50$

    return $R(i,j)$    ▷ $R(i,j)$ rainy image
end function
```

(i) Creating a black layer B(i,j) with the size of the rain-free image.
(ii) Adding Gaussian noise to the black layer. We used 1D Gaussian distribution. Its standard deviation $\alpha$ determines the rain density.
(iii) Applying motion blur filter to the black layer with the Gaussian noise to create the rain layer. We chose the rain motion to be diagonal. We convolved a 2D filter (50 x 50) across the image. As the direction of 1's across the filter grid gives the direction of the desired motion, we used an identity matrix as a motion blur

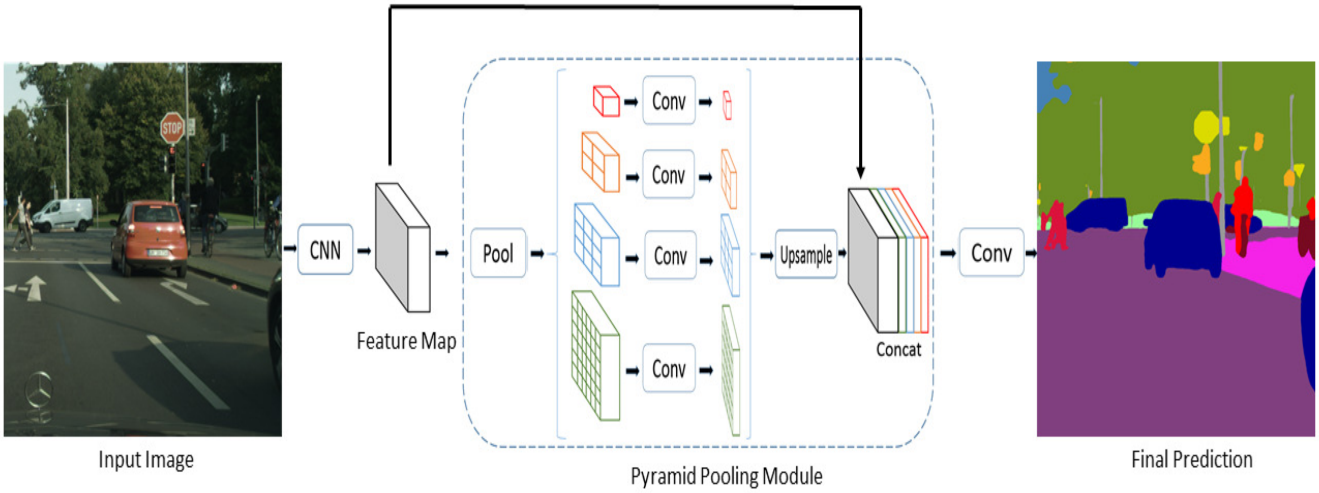


Fig. 3. PSPNet Method Structure Showing its different Phases. The Predicted Image is the Output of PSPNet Method using an Input Image from Cityscapes Clear Dataset.

filter.

Our rainy Cityscapes dataset images created are characterized by the parameter $\alpha$ used to create the rain layer. $\alpha$ determines the rain density. Rain density increases with an increase of $\alpha$ parameter. We created four rainy datasets with $\alpha$ of 15, 20, 25, and 30. Alg. 1 describes the procedures of adding rain to an input clear image.

The author in [12] developed an algorithm to add synthetic fog to the clear weather images of Cityscapes dataset. We chose to use this algorithm to create our evaluation foggy dataset. In this dataset, fog density is defined by the visibility range of the image. We created four foggy datasets with visibility ranges of 600, 300, 150, and 75 meters.

In order to evaluate the performance of semantic segmentation methods, we blurred Cityscapes clear dataset. We convolved the clear images with a Gaussian 2D-kernel that has a standard deviation $\gamma$. The standard deviation $\gamma$ of

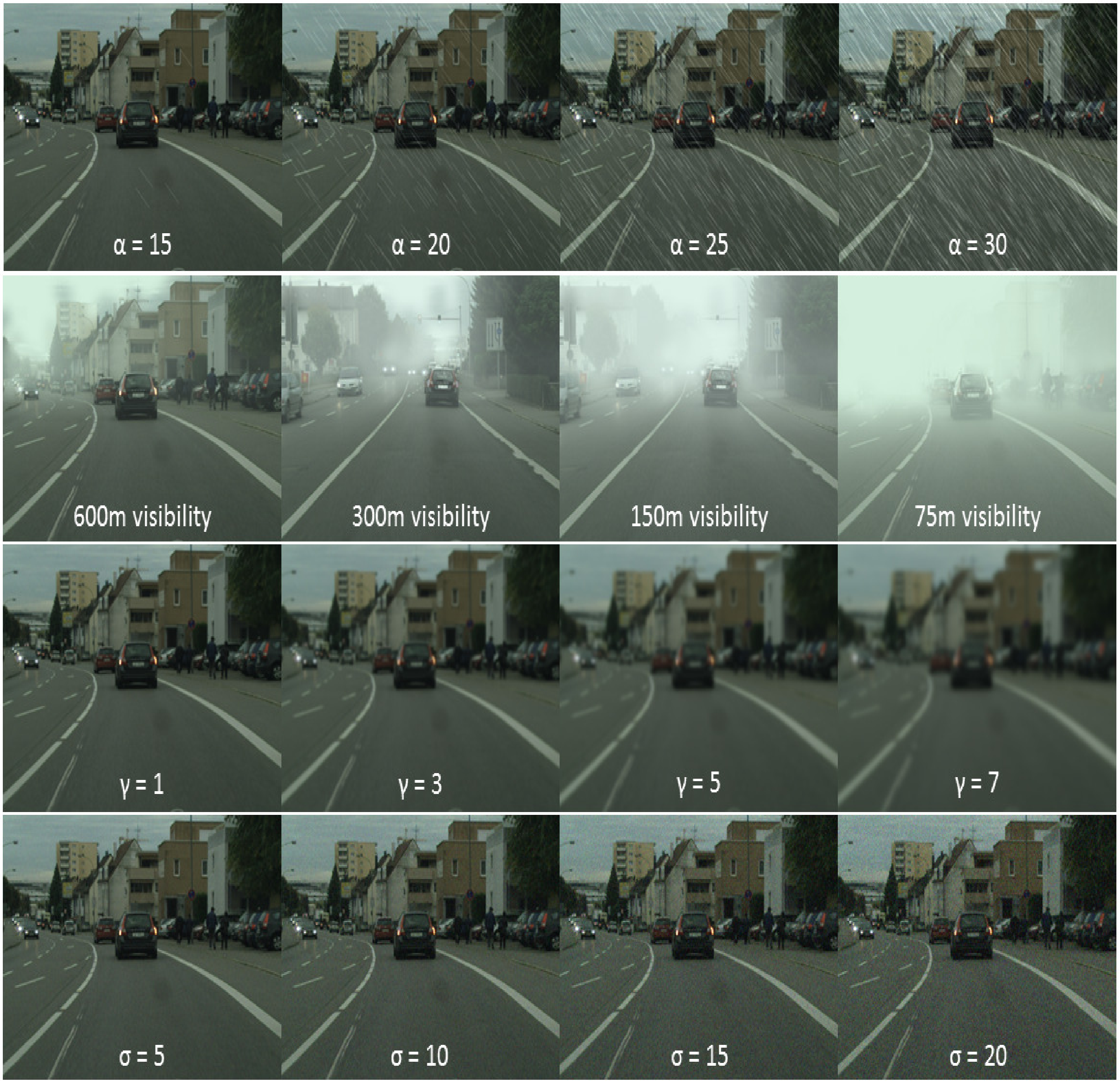


Fig. 4. The First Row shows Example Images from the Rainy Cityscapes with Varying Rain Density $\alpha$. The Second Row shows Example Images from the Foggy Cityscapes with Varying fog Density. The Third Row shows Example Images from the Blurred Cityscapes with Varying Blur Density $\gamma$. The Fourth Row shows Example Images from the Noisy Cityscapes with Varying Noise Density $\sigma$.

the Gaussian kernel represents the density of blurring. By increasing $\gamma$ blurring density increases. We created four blurred datasets with $\gamma$ of 1, 3, 5, and 7.

Noise is defined as aberrant pixels. This means that the pixels are not representing the color or the exposure of the scene correctly. Noise in images can make it impossible to determine the objects in the scene. To determine the performance of the semantic segmentation models with noisy images, we chose to add noise to the clear images from Cityscapes.

One kind of noise that occurs in all recorded images to a certain extent is Gaussian noise. This noise can be modeled with an independent, additive model, where the noise has a zero-mean Gaussian distribution and described by its standard deviation $\sigma$. We used the standard deviation $\sigma$ of the Gaussian model to represent the noise density. As $\sigma$ increases noise density increases. We created four noisy datasets with $\sigma$ of 5, 10, 15, and 20.

## IV. Experiments

In this section, we evaluated the performance of DeepLabv3+ and PSPNet methods using foggy, rainy, blurred, and noisy datasets. We used intersection-over-union metric IoU to measure the methods' performance.

$$IoU = \frac{TP}{(TP + FP + FN)} \tag{2}$$

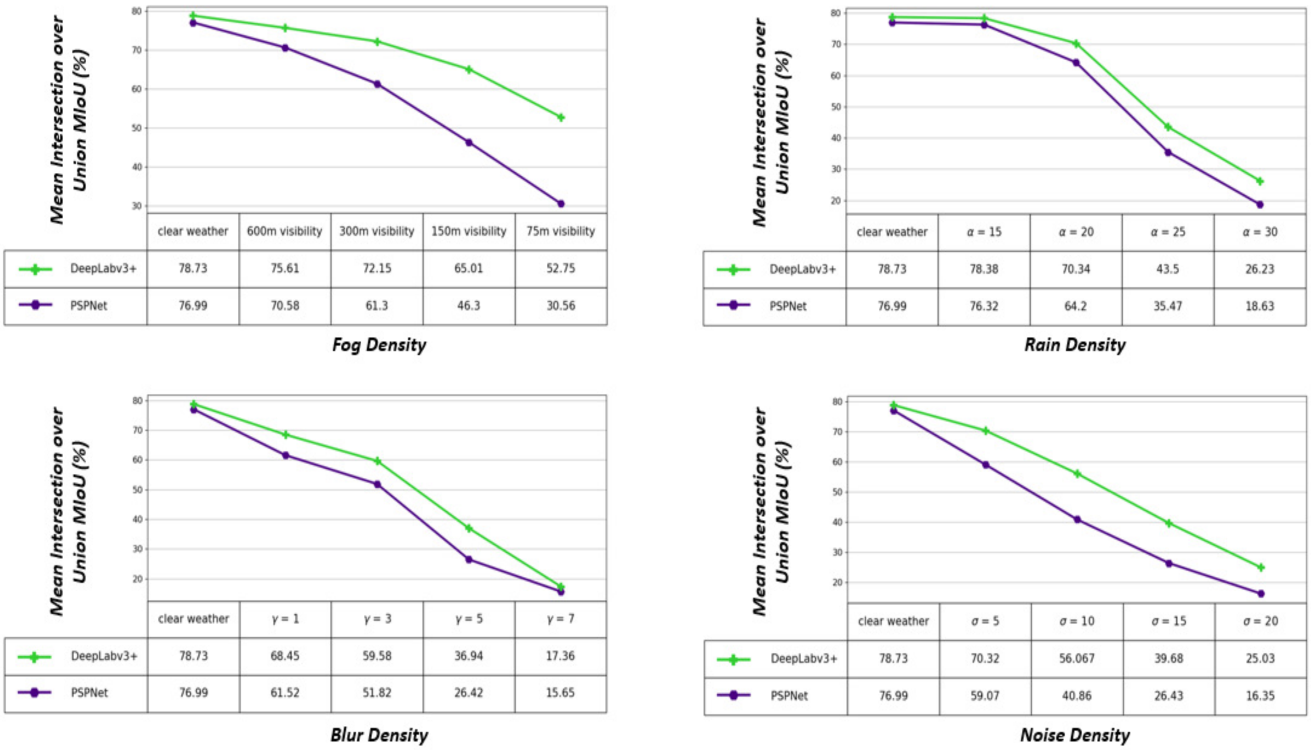
where TP is the true positive labeled pixels, FP is the false positive labeled pixels, and FN is the false negative. mIoU is the mean intersection-overunion of the whole evaluation set.

DeepLabv3+ and PSPNet score mIoU of 78.73% and 76.99% respectively on Cityscapes clear evaluation set. Our experiment evaluates the performance of these models throughout different density degrees of fog, rain, blur, and noise.

By comparing the performance of these two methods, we found that DeepLabv3+ performance overcomes PSPNet performance. Even that the two methods have approximately the same performance on clear Cityscapes dataset, DeepLabv3+ has a higher performance than PSPNet on foggy, rainy, blurred, and noisy Cityscapes datasets. The two methods showed a stable performance on light fog and rain, while the performance harshly degraded on excessive amounts of fog and rain. Also, the performance of the two models decreased at a high rate with low densities of blur or noise.

Although DeepLabv3+ shows a higher performance than PSPNet during the evaluation of different semantic segmentation challenges, our experiments show clearly that these two semantic segmentation methods don't show robust performance with foggy, rainy, blurred, and noisy images. We demonstrated that our challenging datasets killed the performance of both methods. Fig. 5 shows the mIoU of the two methods among the different density degrees of fog, rain, blur, and noise.

In order to have safe autonomous vehicles, systems on these vehicles should work efficiently in all the different weather conditions. Also, semantic segmentation methods in autonomous vehicles systems should show robustness against different types of noise in road images. Fig. 6, Fig. 7, Fig. 8, and Fig. 9 show some qualitative results examples of DeepLabv3+ and PSPNet with our challenging datasets.



| | clear weather | 600m visibility | 300m visibility | 150m visibility | 75m visibility |
|---|---|---|---|---|---|
| DeepLabv3+ | 78.73 | 75.61 | 72.15 | 65.01 | 52.75 |
| PSPNet | 76.99 | 70.58 | 61.3 | 46.3 | 30.56 |

Fog Density

| | clear weather | α = 15 | α = 20 | α = 25 | α = 30 |
|---|---|---|---|---|---|
| DeepLabv3+ | 78.73 | 78.38 | 70.34 | 43.5 | 26.23 |
| PSPNet | 76.99 | 76.32 | 64.2 | 35.47 | 18.63 |

Rain Density

| | clear weather | γ = 1 | γ = 3 | γ = 5 | γ = 7 |
|---|---|---|---|---|---|
| DeepLabv3+ | 78.73 | 68.45 | 59.58 | 36.94 | 17.36 |
| PSPNet | 76.99 | 61.52 | 51.82 | 26.42 | 15.65 |

Blur Density

| | clear weather | σ = 5 | σ = 10 | σ = 15 | σ = 20 |
|---|---|---|---|---|---|
| DeepLabv3+ | 78.73 | 70.32 | 56.067 | 39.68 | 25.03 |
| PSPNet | 76.99 | 59.07 | 40.86 | 26.43 | 16.35 |

Noise Density

Fig. 5. Performance of DeepLabv3+ and PSPNet with Foggy, Rainy, Blurred, and Noisy Cityscapes Evaluation Datasets.

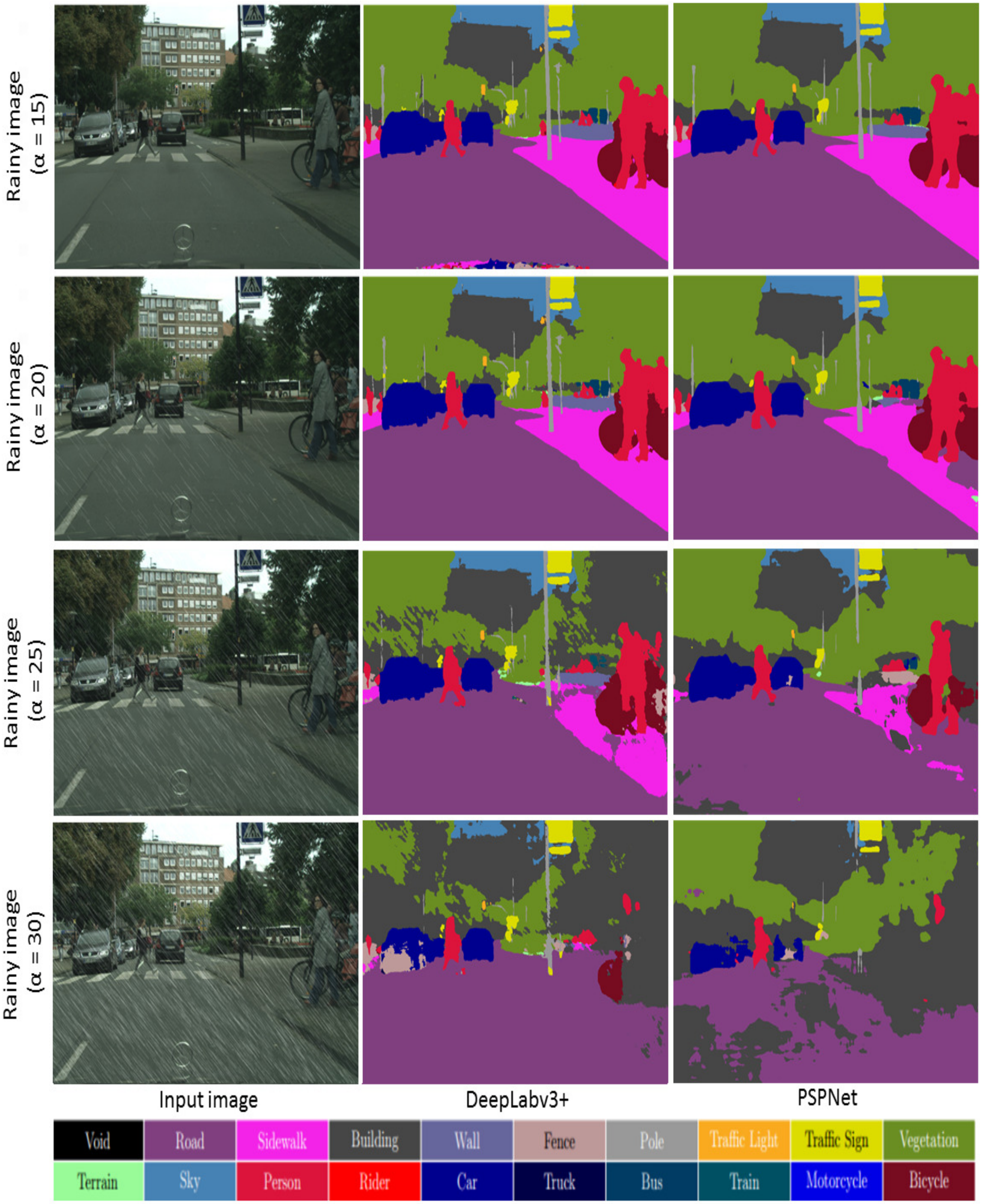


Fig. 6. Example of the Qualitative Results of DeepLabv3+ and PSPNet with Samples from the Rainy Dataset.

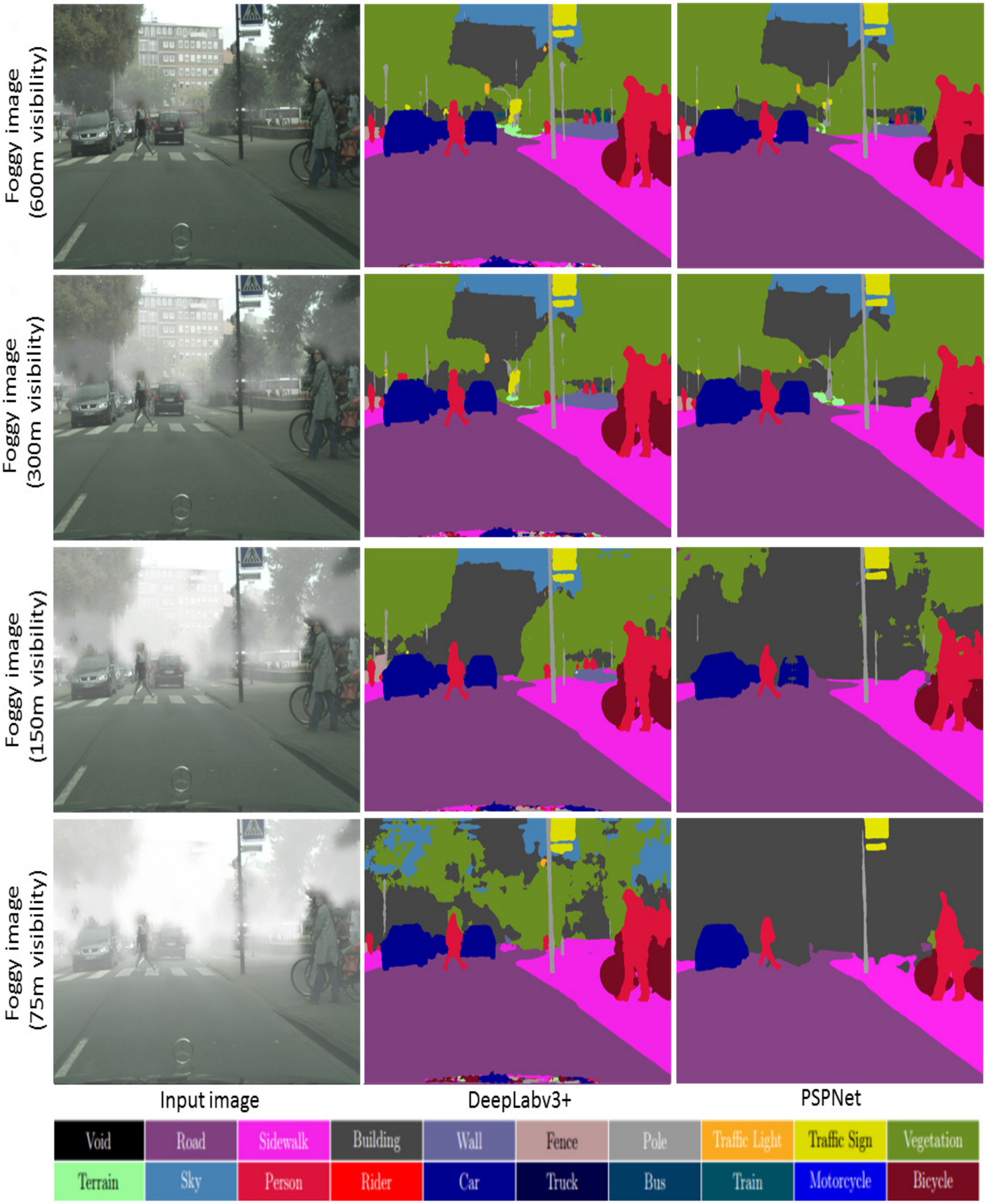


Fig. 7. Example of the Qualitative Results of DeepLabv3+ and PSPNet with Samples from the Foggy Dataset.

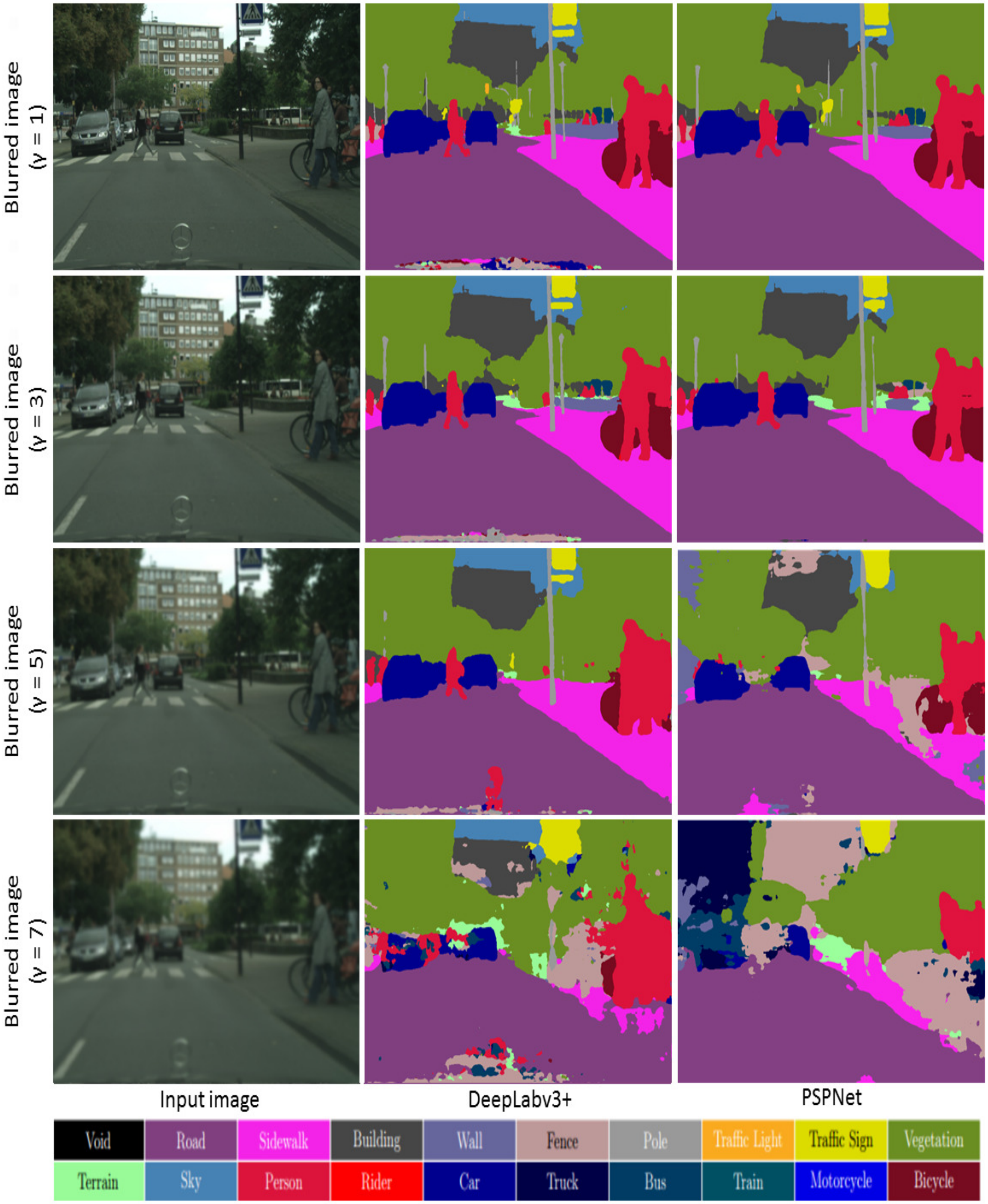


Fig. 8. Example of the Qualitative Results of DeepLabv3+ and PSPNet with Samples from the Blurred Dataset.

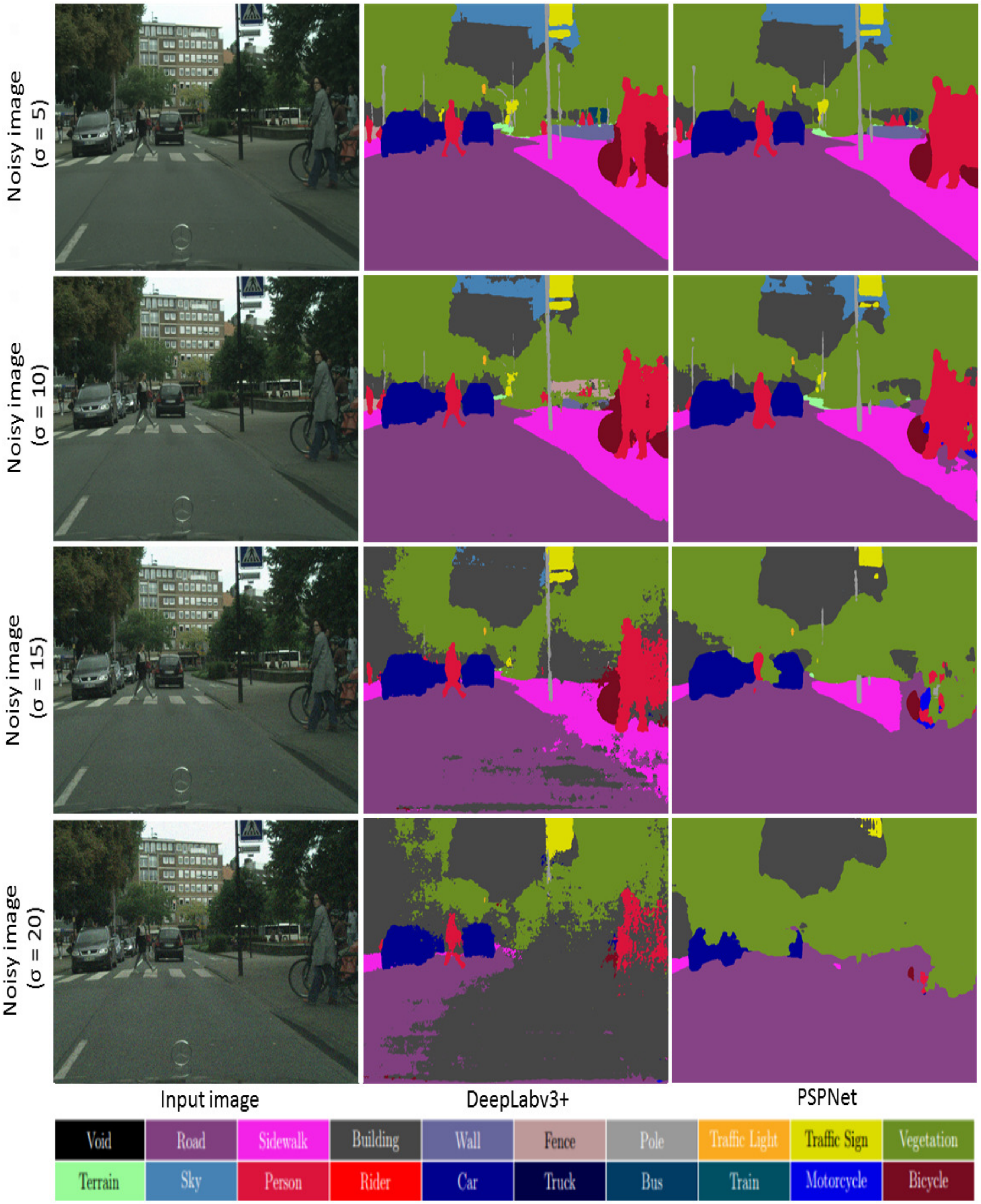


Fig. 9. Example of the Qualitative Results of DeepLabv3+ and PSPNet with Samples from the Noisy Dataset.

## V. Conclusion

In this paper, we studied the performance of state-of-the-art semantic segmentation methods with different severe imaging conditions and challenges. We used Cityscapes dataset which consists of clear images only to create new challenging datasets. We created foggy, rainy, blurred, and noisy Cityscapes datasets. We evaluated the performance of DeepLabv3+ and PSPNet methods using our new challenging datasets. We showed that although DeepLabv3+ and PSPNet have good performance with clear images, these two methods don't show a reliable performance with different challenging datasets.

Our created dataset can be used to boost the performance of semantic segmentation models. This could be done by fine-tuning these models during training using images from our datasets.

In this work, we prove that semantic segmentation methods must be evaluated with different kinds of severe imaging conditions to ensure the robustness of the methods and so the safety of autonomous vehicles.

## References

[1] F. Yu and V. Koltun, ”Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.

[2] G. Lin, A. Milan, C. Shen, and I. D. Reid, ”Refinenet: Multi-path refinement networks for high-resolution semantic segmentation,” *CoRR*, abs/1611.06612, 2016.

[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ”Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ”Pyramid scene parsing network,” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[5] 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)*, pages 801–818, 2018.

[6] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, ”Gated-scnn: Gated shape cnns for semantic segmentation,” In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5229–5238, 2019.

[7] A. Tao, K. Sapra, and B. Catanzaro, ”Hierarchical multi-scale attention for semantic segmentation,” *arXiv preprint arXiv:2005.10821*, 2020.

[8] A. Geiger, P. Lenz, and R. Urtasun, ”Are we ready for autonomous driving? the kitti vision benchmark suite,” In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al, ”Imagenet large scale visual recognition challenge,” *International journal of computer vision*, 115(3):211–252, 2015.

[10] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, ”The pascal visual object classes challenge: A retrospective,” *International journal of computer vision*, 111(1):98–136, 2015.

[11] 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,” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[12] C. Sakaridis, D. Dai, and L. Van Gool, ”Semantic foggy scene understanding with synthetic data,” *International Journal of Computer Vision*, 126(9):973–992, 2018.

[13] K. Jeong and B. Song, ”Image synthesis algorithm for road object detection in rainy weather,” *IEIE Transactions on Smart Processing and Computing*, 7:342–349, 2018.

[14] S. Hasirlioglu and A. Riener, ”Challenges in object detection under rainy weather conditions,” In *First International Conference on Intelligent Transport Systems*, pages 53–65, 2018.

[15] H. Chiang, Y. Ge, and C. Wu, ”Multiple object recognition with focusing and blurring,” Technical report, technical report, Stanford Univ., http://cs231n. stanford. edu/reports/2016/pdfs/259_Report. pdf, 2016.

[16] P. Halkarnikar, H. Khandagle, S. Talbar, and P. Vasambekar, ”Object detection under noisy condition,” In *AIP Conference Proceedings*, pages 288–290, 2010.

[17] S. Milyaev and I. Laptev, ”Towards reliable object detection in noisy images,” *Pattern Recognition and Image Analysis*, 27(4):713–722, 2017.

[18] E. Medvedeva, ”Moving object detection in noisy images,” In *2019 8th Mediterranean Conference on Embedded Computing (MECO)*, pages 1–4, 2019.

[19] S. Sharma, C. Goodin, M. Doude, C. Hudson, D. Carruth, B. Tang, and J. Ball, ”Understanding how rain affects semantic segmentation algorithm performance,” Technical report, technical report, SAE Technical Paper, 2020.

[20] H. Imam, B. A. Abdullah, and H. E. A. El Munim, ”Semantic segmentation under severe imaging conditions,” In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7. IEEE, 2019.