The method of Semantic Segmentation improvement using RescueNet based PSPNet

Habibullaev Sobirjon[a] and [*]**Young Sil Lee**[a]
[a]Department of Computer Engineering, International College, Dongseo University, 47 Jurye-ro, Sasang-gu, Busan, South Korea (47011), sobirjonh777@gmail.com,lys0113@dongseo.ac.kr

Abstract

Due to climate change increasing rate of natural disasters worldwide can be observed. These catastrophes are causing severe damage to humanity and nature. As a result of the disasters, economic losses are steadily increasing, and the importance of fast and accurate auxiliary tools for rescue teams is growing in saving people's lives and eliminating financial losses. In this regard, by using deep learning algorithms, it is possible to assess the consequences of any disaster that can accurately understand the affected areas. Until now, several successful works have been done by several researchers. The following paper focuses on the experiment to enhance PSPNet model results on RescueNet datasets, which were collected by UAV, by modernizing configurations and datasets.

Keywords: Semantic Segmentation, Post-disaster damage assessment, PSPNet, RescueNet

1 Introduction

In recent years, due to drastic changes in the climate and some factors, the number of natural disasters faced by humanity and nature has increased immensely. Many such natural disasters result in severe loss of life and economic losses to governments. Modern deep learning techniques and computer vision models are at the center of attention in eliminating losses and financial losses. Semantic segmentation is essential for computer vision techniques to assist in accurate damage assessment and it aims to classify each pixel of an image.

Datasets for disaster damage assessment can be collected from 3 different sources: satellite[1-2], social media[3-5], or UAV(Unmanned Aerial Vehicle)[6-9]. But UAV datasets [8-9] have some advantages for semantic segmentation in post-disaster damage assessment compared to satellite datasets: 1) Higher Spatial Resolution: UAVs can capture imagery at much higher spatial resolutions than satellites. This enables a more complete and precise investigation of the affected areas, making identifying and classifying specific objects or regions within the scene easier. 2) Higher Quality Data: UAV imagery tends to have higher quality due to its proximity to the target area. The images are less affected by atmospheric conditions, such as clouds or haze, which can degrade satellite imagery. This leads to sharper and clearer data, improving semantic segmentation accuracy.

Datasets such as RescueNet[9] and FloodNet[8] are recent years' highest resolutions and well-annotated datasets. Several experiments with different architectures PSPNet[10], DeepLabv3+[12], and ENet[13] were conducted on both datasets, and research papers with successful results were published. However, interestingly, despite using the same PSPNet[10] architectures on the two datasets, two very different results were generated. Effects such as 79.43% (See Table 1.) and 79.69% achieved by the PSPNet model on the HRUD[11] dataset, which are similar to the RescueNet dataset, and on the FloodNet[8] dataset, the possibility to improve the PSPNet modeling results for the RescueNet dataset motivated us to conduct these experiments. During this research, we have focused on the investigation to enhance PSPNet model results[9] on RescueNet[9] datasets through modernizing configurations and datasets. the possibility to improve the PSPNet modeling results for the RescueNet dataset motivated us to conduct these experiments. During this research, we have focused on the experiment to enhance PSPNet model results[9] on RescueNet[9] datasets through modernizing of configurations and datasets.

Method	Debris	Water	Building Non Total Destruction	Building Total Destruction	Vehicle	Road	Tree	Pool	Sand	mIoU
ENet[13]	45.97	75.84	66.16	39.52	36.74	61.19	71.64	28.47	61.77	54.15
DeepLabv3+[12]	65.8	85.8	84.5	57.3	51.3	73.3	75.9	55.7	77.4	69.67
PSPNet	**88.76**	**67.98**	**85.75**	**80.51**	**65.83**	**82.81**	**94.53**	**72.61**	**76.04**	**79.43**

Table 1. Per-class results on HRUD[11] testing set

Method	Building Non Flooded	Road Flooded	Road Non Flooded	Water	Tree	Vehicle	Pool	Grass	mIoU
ENet[13]	47.35	12.49	48.43	48.95	68.36	32.26	42.49	76.23	42.61
DeepLabv3+[12]	72.8	52.00	70.2	75.2	77.00	42.5	47.1	84.3	61.53
PSPNet	**89.75**	**82.16**	**91.18**	**92.00**	**89.55**	**46.15**	**64.19**	**93.29**	**79.69**

Table 2. Per-class results on FloodNet[8] testing set

2 Related Works

Several studies have previously tested the PSPNet architecture on several datasets related to post-disaster damage assessment. First, several experiments on the effects of PSPNet learning rate changes and classes on the final result were conducted on the HRUD dataset[11] by M. Rahnemoonfar *et al.* in [11]. In these experiments, the model was successful in learning rates such as 1e-3 and 1e-4 (See Table 1). Secondly, in the experiments conducted on the FloodNet dataset[8] with several different ENet[13], DeepLabv3+[12], and PSPNet network architectures, PSPNet showed the most successful result among all the network architectures in the experiment[8] (See Table 2). However, in experiments conducted by M. Rahnemoonfar *et al.* on the RescueNet dataset[9], the PSPNet model failed slightly more than the other models in the experiment. Through this study, we tried to improve the results of the PSPNet network architecture on this particular RescueNet dataset[9].

3 Semantic Segmentation improvement method

A) Improvement on the size of train set

RescueNet[9] dataset provides pixel-level annotation of 11 distinct categories, including debris, water, building, vehicle, road, tree, pool, and sand. On both RescueNet[9] and FloodNet[8] datasets, the number of annotations of small objects like "vehicle" and "pool" is very lack. In Fig 4, as you can see, PSPNet, a pyramid pooling-based method, achieved only 9.83% Intersection over Union (IoU) in the "pool" class. This indicates that encoder-decoder-based methods outperformed PSPNet in small classes such as "vehicle" and "pool" within the experiments. In Table 2, you can see the same issue with PSPNet on the FloodNet[8]. To overcome the problem of identifying small shapes in the RescueNet dataset, we have also added images from the FloodNet datasets that match the annotations of the RescueNet datasets to the train set.

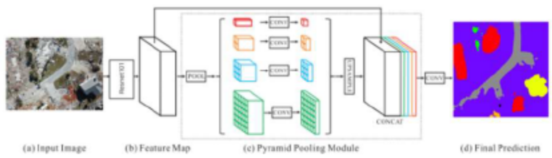

Figure 1: The network architecture of PSPNet by using resnet101 as the backbone.

B) *Improvement on configuration*

We have selected PSPNet (Pyramid Scene Parsing Network), which seeks to classify each pixel in an input image. A resnet101 backbone is used in the architecture to extract features from the input picture, and a pyramid pooling module is applied to collect multi-scale contextual information. You can see Figure 1. In the PSPNet architecture, the resnet101 backbone serves as the feature extractor. ResNet (Residual Network) is a well-known deep neural network architecture noted for its skip connections, which enable improved gradient flow during training. resnet101 has 101 layers, indicating a deep architecture capable of collecting complex visual data. The input image was resized to 713x713, suitable for the network.

Resizing ensures that the image dimensions match the input requirements of the network. The resized input image is passed through the resnet101 backbone, which consists of 5 convolutional layers structured in ascending order. These layers gradually reduce the image's spatial resolution while increasing the number of learned features. The backbone produces a collection of feature maps with varying spatial resolutions.

The PSPNet architecture relies heavily on the pyramid pooling module. It solves the problem of gathering multi-scale contextual data. The module accepts the resnet101 backbone's output feature maps and performs pooling operations at multiple scales. This enables the network to aggregate both global and local information. Transposed convolution was used for upsampling the last convolutional layer's output to the original input image size. This upsampling technique results in a dense prediction map, with each pixel representing the predicted label for the associated pixel in the input image. Finally, the dense prediction map is subjected to a softmax activation with a CrossEntropy loss function to turn the pixel-wise predictions into probabilities. This allows the result to be interpreted as a probability distribution across 11 classes, assigning a semantic label to each pixel.

Method	Debris	Water	Building no Damage	Building minor Damage	Building major Damage	Building total Destruction	Vehicle	Road	Tree	Pool	Sand	mIoU %
PSPNet-old	65.94	61.76	51.18	50.03	61.79	58.27	23.26	67.12	74.10	9.83	81.07	54.94
PSPNet-new	**76.04**	**81.80**	**75.34**	**68.6**	**72.70**	**70.24**	**58.15**	**80.64**	**73.13**	**36.83**	**82.98**	**70.58**

Table 3. Comparison of per-class results on RescueNet test set

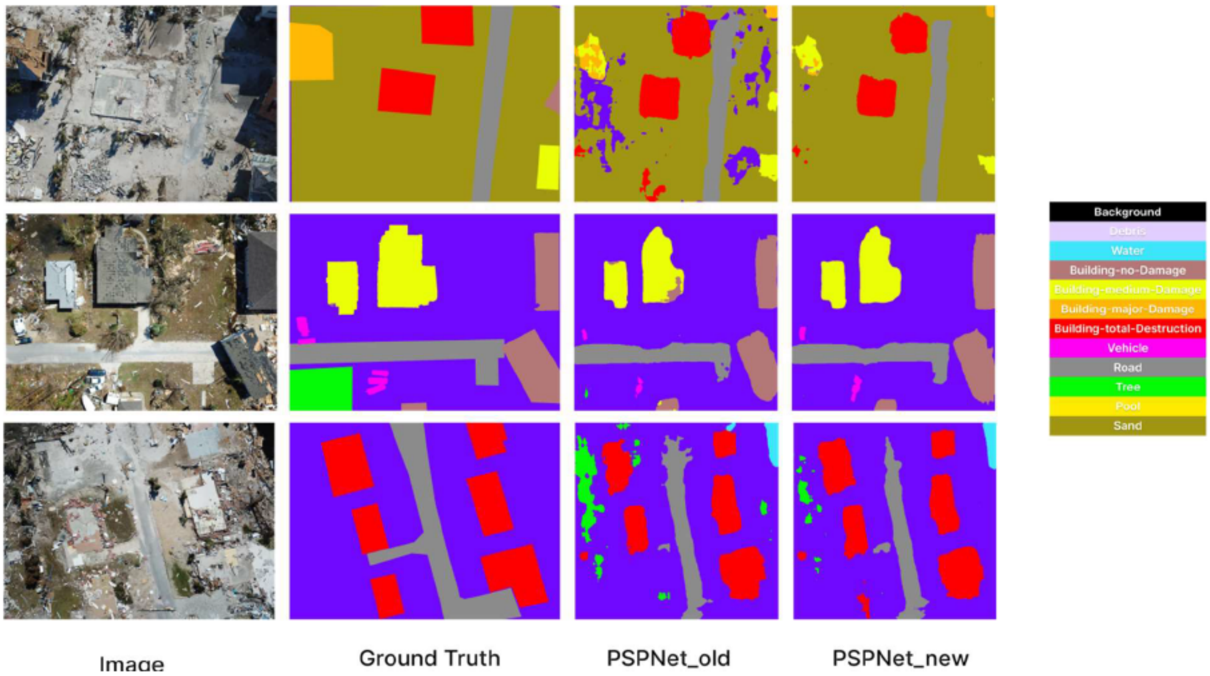

Figure 2. Visual comparison of PSPNet old version model and new version of PSPNet model on RescueNet test set .

4 Experimental Results

Pytorch has been used for the implementation of segmentation networks. For the experiment, the batch size is set to 2 and the crop size is set to 713. Also, momentum, weight decay, power, and weight of auxiliary loss were set to 0.9, 0.001, 0.9, and 0.4, respectively. For augmentation, we use random shuffling, scaling, flipping, and random rotation, which help models avoid overfitting. We also conducted several tests on the changes in learning rate, momentum, and weight decay values. However, the most important changes in the success of the model might be explained by an increase in the train set size for the "vehicle" and "pool" classes of the datasets, as well as an increase in the weight decay from 0.0001 to 0.001 and using a learning rate as 0.001. The training accuracy result (mIoU) on the test set was 70.58%. We can also see a comparison of per-class results on the RescueNet test set in Table 3. All models were trained and tested on the Google Colab Pro version with A100 GPUs.

5 Conclusion

Overall, this study demonstrates the experiment to improve PSPNet results on the RescueNet dataset. The experiment also involved conducting tests to explore the impact of changes in learning rate, momentum, and weight decay values. So appropriate parameter tuning and sufficient training data led to the model's success of 70.58%. The findings provide valuable insights for further advancements in segmentation network research and applications for post-disaster damage assessment activities.

Acknowledgement

This research was a part of the project titled 'Marine digital AtoN information management and service system development(3/5) (20210650)', funded by the Ministry of Oceans and Fisheries, Korea.

References

[1] Chen, A. Escay, C. Haberland, T. Schneider, V. Staneva, and Y. Choe, "Benchmark dataset for automatic damaged building detection from posthurricane remotely sensed imagery," arXiv preprint arXiv:1812.05581,2018.

[2] Gupta, B. Goodman, N. Patel, R. Hosfelt, S. Sajeev, E. Heim, J. Doshi,K. Lucas, H. Choset, and M. Gaston, "Creating xbd: A dataset for assessing building damage from satellite imagery," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2019, pp. 10–17.

[3] D. T. Nguyen, F. Ofli, M. Imran, and P. Mitra, "Damage assessment from social media imagery data during disasters," in Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 2017, pp.569–576.

[4] K. R. Nia and G. Mori, "Building damage assessment using deep learning and ground-level image data," in 2017 14th Conference on Computer and robot vision (CRV). IEEE, 2017, pp. 95–102.

[5] Weber, N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba, "Detecting natural disasters, damage, and incidents in the wild," in European Conference on Computer Vision. Springer, 2020, pp. 331–350.

[6] C. Kyrkou and T. Theocharides, "Deep-learning-based aerial image classification for emergency response applications using unmanned aerial vehicles." in CVPR Workshops, 2019, pp. 517–525.

[7] X. Zhu, J. Liang, and A. Hauptmann, "Msnet: A multilevel instance segmentation network for natural disaster damage assessment in aerial videos," arXiv preprint arXiv:2006.16479, 2020.

[8] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "Floodnet: A high-resolution aerial imagery dataset for post-flood scene understanding," IEEE Access, vol. 9, pp. 89 644–89 654, 2021.

[9] M. Rahnemoonfar, T. Chowdhury, and R. R. Murphy, "RescueNet: A High Resolution UAV Semantic Segmentation Benchmark Dataset for Natural Disaster Damage Assessment" arXiv:2202.12361v1, 2022

[10] 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, 2017, pp. 2881–2890.

[11] T. Chowdhury, M. Rahnemoonfar, R. Murphy and O. Fernandes, "Comprehensive Semantic Segmentation on High-Resolution UAV Imagery for Natural Disaster Damage Assessment," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 3904-3913, doi: 10.1109/BigData50022.2020.9377916.

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

[13] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," arXiv preprint arXiv:1606.02147, 2016.