Attention For Damage Assessment

Tashnim Chowdhury [1] **Maryam Rahnemoonfar** [1]

Abstract

Due to climate change the hurricanes are getting stronger and having longer impacts. To reduce the detrimental effects of these hurricanes faster and accurate assessments of damages are essential to the rescue teams. Like other computer vision techniques semantic segmentation can identify the damages and help in proper and prompt damage assessment. Current segmentation methods can be classified into attention and non-attention based methods. Existing non-attention based methods suffers from low accuracy and therefore attention based methods are becoming popular. Self-attention based methods can map the mutual relationship and dependencies among pixels of an image and thus improve semantic segmentation accuracy. In this paper, we present a self-attention semantic segmentation method on UAV imageries to assess the damages inflicted by a natural disaster. The proposed method outperforms four state-of-art segmentation methods both quantitatively and qualitatively with a mean IoU score of 84.03%.

1. Introduction

Recently there have been numerous natural disasters which have brought both personal injuries and economic losses to several countries all over the world. In 2020 alone USA has inflicted with 22 natural disasters which have cost around 96.4 billions dollars (noa). Current tropical storms are persisting far longer and doing more damage than in the past. In 2017, USA' fourth largest city Houston, Texas, was inundated when Hurricane Harvey settled over the city for several days, and was dumped with 127 billion tonnes of water (Van Oldenborgh et al., 2017). Moisture from warm and tropical oceans acts as the fuel of the hurricanes that drives the intense winds. Due to climate change, the air over the oceans can hold more of this moisture and intensify the storms at sea. Consequently, when these storms reach land, the hurricanes should decay very quickly due to lack of fuel from the seas. However, the timescale of decay of the North Atlantic land-falling hurricanes has almost doubled over the past 50 years due to climate change (Li & Chakraborty, 2020). According to (Li & Chakraborty, 2020), the slower decay is fuelled by an increased amount of moisture that is stocked in the hurricane from its passage over ocean prior to landfall.

In this scenario, during and even after a natural disaster, a proper assessment of the damages is imperative which can facilitate the rescue efforts. Semantic segmentation, a fundamental task in computer vision, aims at assigning semantic labels to each pixels of an image. It can assist in rescue efforts by providing quick and accurate damage assessment. DCNNs (Deep Convolutional Neural Networks) has gained popularity in semantic segmentation because of its improved performance compared to traditional computer vision techniques (Chen et al., 2017; 2018; Paszke et al., 2016; Zhao et al., 2017). Besides non-disaster scenarios, the potential of DCNNs are also being utilized by several researchers (Lopez-Fuentes et al., 2017; Doshi et al., 2018; Rahnemoonfar et al., 2018; Rudner et al., 2019; Gupta & Shah, 2020; Gupta et al., 2020; Zhu et al., 2020) in natural disaster damage assessment. Different DCNN semantic segmentation methods have been proposed including encoder-decoder and pyramid based methods (Ronneberger et al., 2015; Zhao et al., 2017; Paszke et al., 2016; Chen et al., 2018). Besides these methods, self-attention based methods are showing excellent performance. Self-attention mechanism captures the spatial dependencies between any two positions of the feature maps and thus contribute to the mutual improvement (Fu et al., 2019).

In this work, we implement a self-attention based semantic segmentation method named ReDNetPlus on a natural disaster dataset FloodNet (Rahnemoonfar et al., 2021). Our results show that our method performs considerably better than four other segmentation methods including both attention and non-attention based methods. This suggests that our approach is feasible and promising.

[1]Computer Vision and Remote Sensing Laboratory (Bina Lab), University of Maryland Baltimore County, Baltimore, MD, USA. Correspondence to: Maryam Rahnemoonfar <maryam@umbc.edu>.

Tackling Climate Change with Machine Learning Workshop at ICML 2021.

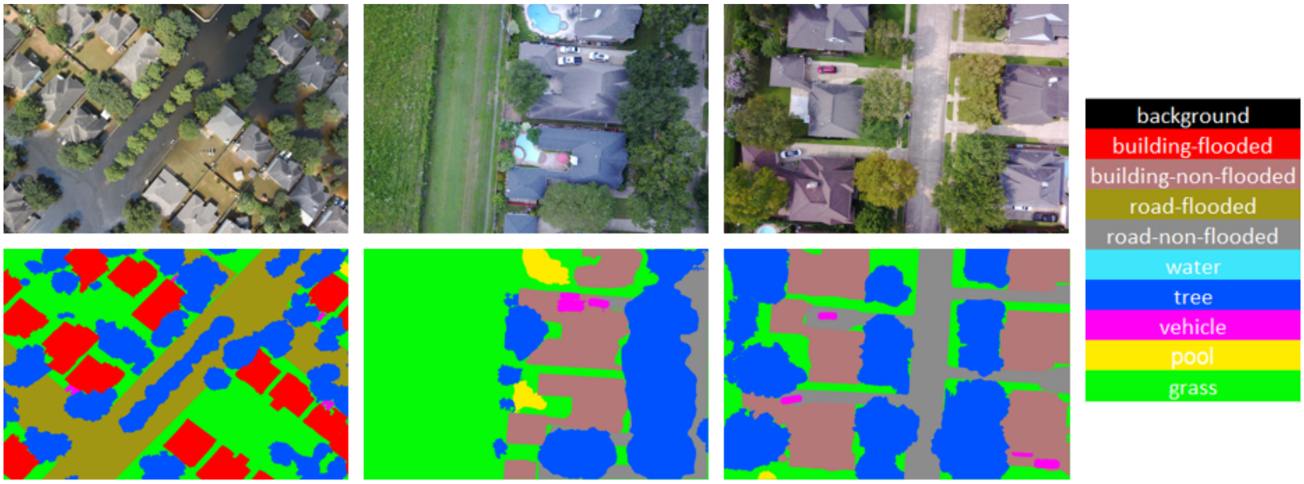

Figure 1. Illustration of after disaster scenes of FloodNet dataset. First row shows original image and the second row shows the corresponding annotations.

2. Related Works

2.1. Self Attention Method

State-of-art semantic segmentation methods can be classified into encoder-decoder based method (Ronneberger et al., 2015), pyramid pooling based method (Zhao et al., 2017; Chen et al., 2018), and attention based method (Huang et al., 2019; Fu et al., 2019). Encoder-decoder based methods like U-Net adopt local context from middle and lower level features through encoder-decoder architecture. These methods generate sharp object boundaries or small details based on local context and make high resolution prediction. On the other hand, pyramid based modules (Zhao et al., 2017; Chen et al., 2018) create global context using pooling based operations like global average pooling, pyramid pooling (Zhao et al., 2017), and atrous convolution (Chen et al., 2018). These methods generate rich global contextual information at different resolutions or scales. In computer vision, self-attention module calculates the context at one position as weighted sum of all positions in a sentence or an image. Several self-attention based works (Huang et al., 2019; Fu et al., 2019; Yuan & Wang, 2018) have been proposed in computer vision field. These approaches use different self-attention mechanisms to aggregate contextual information in order to augment feature representation. Authors in (Fu et al., 2019) propose a position attention module along with a channel attention module to capture interdependencies of features among spatial and channels dimension respectively. (Huang et al., 2019) proposes a criss-cross attention module to gather contextual information in spatial domain. The work in (Yuan & Wang, 2018) calculates object based contexts using self-attention based mechanism.

2.2. Natural Disaster Damage Assessment

Different research works have attempted to assess the damages inflicted by different natural disasters by implementing and proposing different deep learning methods. Authors in (Doshi et al., 2018) perform segmentation on satellite images to detect maximal damaged areas. Rahnemoonfar *et al.* in (Rahnemoonfar et al., 2018) propose a densely connected recurrent neural network in order to segment UAV images for flood area detection. Multi3Net is proposed in (Rudner et al., 2019) for rapid segmentation of flooded buildings by fusing multiresolution, multisensor, and multitemporal satellite imageries. RescueNet is proposed by Gupta *et al.* in (Gupta & Shah, 2020) for joint building segmentation. To assess the damages of the buildings a multilevel instance segmentation method named MSNet is proposed by Zhu *et al.* in (Zhu et al., 2020).

In this work, ReDNetPlus is applied on FloodNet (Rahnemoonfar et al., 2021) dataset which is a high resolution UAV dataset collected after hurricane Harvey. Unlike existing segmentation methods proposed for natural disaster damage assessment, we perform segmentation not only on buildings and roads but also on pools, vehicles, trees, and grass.

3. Dataset

FloodNet (Rahnemoonfar et al., 2021) is collected with small UAV platform, DJI Mavic Pro quadcopters, after *Hurricane Harvey*. Hurricane Harvey has made disasterous impact near Texas and Louisiana on August, 2017, as a Category 4 hurricane. The FloodNet consists of imagery taken from several flights conducted between August 30 - September 04, 2017, at Ford Bend County in Texas and

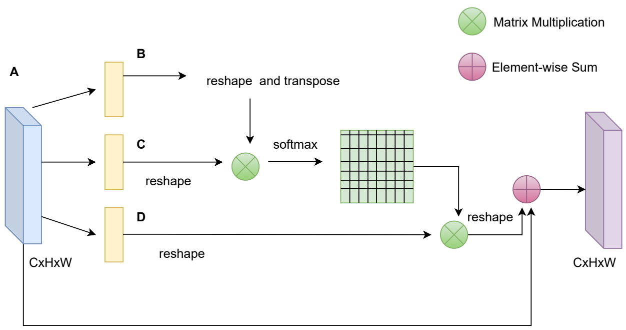

Figure 2. The details of the position attention module.

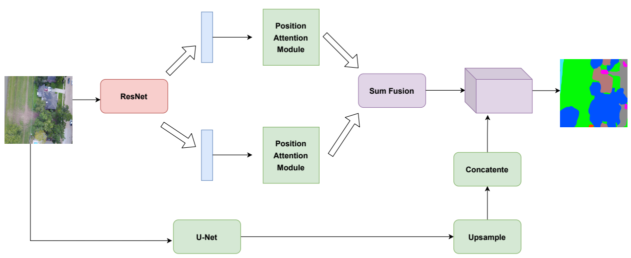

Figure 3. Overview of ReDNetPlus. The network consists of two position attention modules (PAMs) and an U-Net.

other directly impacted areas. All flights were flown at 200 feet AGL, as compared to manned assets which normally fly at 500 feet AGL or higher. FloodNet has in total 2343 images have been annotated with 9 classes which include building-flooded, building-non-flooded, road-flooded, road-non-flooded, water, tree, vehicle, pool, and grass. A buildings is classified as flooded when at least one side of a building is touching the flood water. Although we have classes created for flooded buildings and roads, to distinguish between natural water and flood water, "water" class has been created which represents any natural water body like river and lake. For the classification task, each image is classified either "flooded" or "non-flooded". If more than 30% area of an image is occupied by flood water then that area is classified as flooded, otherwise non-flooded.

4. Method

4.1. Position Attention Module (PAM) and ReDNet

Local receptive field is calculated by convolutional operations. However, lack of global context results in intra-class inconsistency, and eventually hurts the recognition accuracy of network models (Fu et al., 2019). We use ResNet-101 (He et al., 2016) as base recognition model. Different stages of the ResNet-101 has different recognition capabilities. Smaller receptive field of lower stages can encode fine spatial information but semantic consistency is poor. On other hand, larger receptive field of higher stages is able to maintain good semantic consistency but results in poor encoding of spatial information (Yu et al., 2018). Keeping this in mind, a self-attention network which combines feature extraction from both lower and higher stages of ResNet-101 is implemented.

The self-attention block is a position attention module (PAM) which creates global context over local features. Two position attention modules use two different feature maps of ResNet-101 (layer 2 and layer 3), and generate two self-attention maps which are rich in global contextual information. Position attention module is shown in Figure 2. Given a local feature map A with dimension $C \times H \times W$, three new feature maps are generated using a convolution layer with similar shape. Matrix multiplication is performed on transpose of B and C. Then using a softmax layer spatial attention map S is calculated using the following formula.

$$s_{ji} = \frac{exp(B_i \cdot C_j)}{\sum_{i=1}^{N} exp(B_i \cdot C_j)} \quad (1)$$

Then matrix multiplication is performed between transpose of S and D. The output from the matrix multiplication is multiplied with α. Finally element-wise sum is performed with feature A to generate final output with shape $C \times H \times W$. This architecture is coined as ReDNet (a ResNet (He et al., 2016) based Dual attention Network).

4.2. ReDNetPlus

The two position attention modules are generated using the methodology discussed in section 4.1. These two modules are placed in parallel and their attention maps are added together using element-wise sum operation which constitutes ReDNet architecture.

The ReDNetPlus segmentation network implements a smaller U-Net (Ronneberger et al., 2015) along with ReDNet. The output from the U-Net is added to the output from the sum of two position attention maps and finally passed through a fully connected layer. The output of fully connected layer generates the segmentation map of the input image.

The visual interpretation of the ReDNetPlus network is shown in Figure 3. The final self-attention map generated from two PAMs contains global contextual information. On the other hand, the U-Net generates local context map and later is added to the global context map generated from the self-attention modules.

5. Experiments

PyTorch has been used for implementation of segmentation network. As hardware we use NVIDIA GeForce RTX 2080

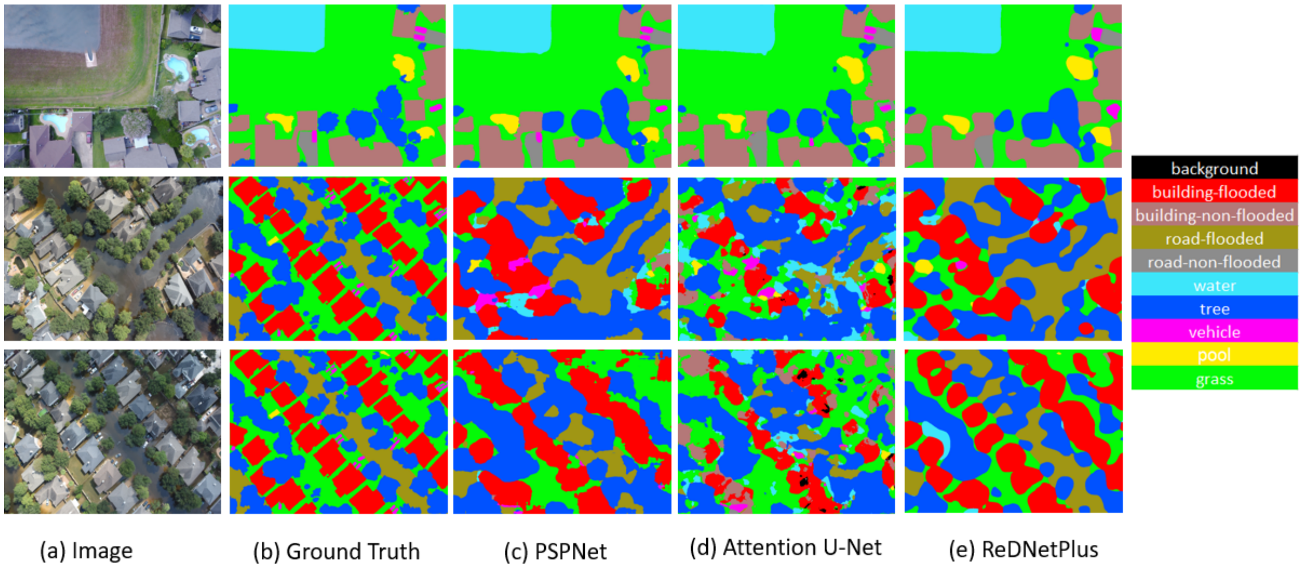

Figure 4. Visual comparison of PSPNet, Attention U-Net, and ReDNetPlus on FloodNet test set.

Table 1. Per-class intersection over union (in %) and their mean value (mIoU) on FloodNet testing set.

Method	Building Flooded	Building Non Flooded	Road Flooded	Road Non Flooded	Water	Tree	Vehicle	Pool	Grass	mIoU
ENet(Paszke et al., 2016)	21.82	41.41	14.76	52.53	47.14	62.56	26.21	16.57	75.57	39.84
DeepLabv3+(Chen et al., 2018)	28.10	78.10	32.00	81.10	73.00	74.50	33.60	40.00	87.10	58.61
PSPNet(Zhao et al., 2017)	65.61	90.92	78.69	90.90	91.25	89.17	**54.83**	66.37	95.45	80.35
Attention U-Net (Oktay et al., 2018)	64.82	86.14	28.20	**92.35**	77.74	90.95	54.20	**71.82**	95.29	73.50
ReDNetPlus	**80.99**	**91.76**	**88.90**	91.90	**95.56**	**91.20**	48.68	70.90	**96.39**	**84.03**

Ti GPU and Intel Core i9 CPU. We use "poly" learning rate with base learning rate 0.001. Momentum, weight decay, and power are set to 0.9, 0.0001, and 0.9 respectively. For augmentation we use random shuffling, scaling, flipping, and random rotation which help models to avoid overfitting. During training, we resize the images to 713×713 since large crop size is useful for the high resolution images. For semantic segmentation we use mean IoU (Intersection over Union) as evaluation metric.

6. Results And Discussion

Table 1 shows the performance evaluation of the ReDNetPlus compared to other state-of-art methods on FloodNet dataset (Rahnemoonfar et al., 2021). The result includes comparison with both non-attention based methods such as ENet (Paszke et al., 2016), DeepLabv3+ (Chen et al., 2018), and PSPNet (Zhao et al., 2017), and attention based method such as Attention U-Net (Oktay et al., 2018). Non attention based methods perform worse than attention based methods except in class "vehicle". In class "vehicle" PSPNet performs best among all methods. Attention U-Net shows improved performance in all classes specially in segmenting small objects. Although Attention U-Net shows better performance than PSPNet in most of the cases, due to its lower performance in class "Road Flooded", its overall performance is lower than PSPNet by 6.85%.

ReDNetPlus combines self-attention based global feature map with local feature map produced by U-Net. Although this method does not present superior performance in smaller object classes, it provides excellent performance in other classes which includes bigger and flooded objects. ReDNetPlus shows superior performance with Mean IoU 84.03% outperforming Attention U-Net by 10.53% and PSPNet by 3.68%.

Qualitative results are shown in Figure 4. In this figure samples of evaluated segmentation from top three performing methods (PSPNet, Attention U-Net, ReDNetPlus) are presented. It can be seen that segmentation results from these three methods are very close to the original ground truths.

7. Conclusion

Climate change is highly associated with stronger and more dangerous hurricanes. Current research works are implementing deep learning based computer vision techniques to assess the damages after any natural disaster. Fast and accurate assessment is instrumental in this scenario to reduce the detrimental impacts. In this work we implement a self-attention based semantic segmentation method named ReDNetPlus on a natural disaster dataset called FloodNet. We compare the method with state-of-art non-attention and attention based methods. The result indicates that ReDNetPlus performs superior than other methods. From the experiments it is evident that combining attention map produced using lower level feature maps, and local context map produced using U-Net can significantly improve the segmentation of flooded objects. The achieved higher accuracy in damage assessment indicates promising application of deep learning techniques in reducing harmful impacts of climate change.

References

Noaa national centers for environmental information (ncei). u.s. billion-dollar weather and climate disasters. `https://www.ncdc.noaa.gov/billions/events`. Accessed: 2021-05-25.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. 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.

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

Doshi, J., Basu, S., and Pang, G. From satellite imagery to disaster insights. *arXiv preprint arXiv:1812.07033*, 2018.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H. Dual attention network for scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3146–3154, 2019.

Gupta, A., Watson, S., and Yin, H. Deep learning-based aerial image segmentation with open data for disaster impact assessment. *arXiv preprint arXiv:2006.05575*, 2020.

Gupta, R. and Shah, M. Rescuenet: Joint building segmentation and damage assessment from satellite imagery. *arXiv preprint arXiv:2004.07312*, 2020.

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

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 603–612, 2019.

Li, L. and Chakraborty, P. Slower decay of landfalling hurricanes in a warming world. *Nature*, 587(7833):230–234, 2020.

Lopez-Fuentes, L., Rossi, C., and Skinnemoen, H. River segmentation for flood monitoring. In *2017 IEEE International Conference on Big Data (Big Data)*, pp. 3746–3749. IEEE, 2017.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

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

Rahnemoonfar, M., Murphy, R., Miquel, M. V., Dobbs, D., and Adams, A. Flooded area detection from uav images based on densely connected recurrent neural networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1788–1791. IEEE, 2018.

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., and Murphy, R. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access*, 2021.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Rudner, T. G., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V., and Biliński, P. Multi3net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 702–709, 2019.

Van Oldenborgh, G. J., Van Der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G., and Cullen, H. Attribution of extreme rainfall from hurricane harvey, august 2017. *Environmental Research Letters*, 12(12):124009, 2017.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., and Sang, N. Learning a discriminative feature network for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1857–1866, 2018.

Yuan, Y. and Wang, J. Ocnet: Object context network for scene parsing. *arXiv preprint arXiv:1809.00916*, 2018.

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

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