An Embedded Recurrent Neural Network-based Model for Endoscopic Semantic Segmentation

Mahmood Haithami[a], Amr Ahmed[a], Iman Yi Liao[a] and Hamid Jalab[a]

[a]*Computer Science Department,University of Nottingham Malaysia Campus*
[b]*Computer System and Technology Department, University of Malaya*

Abstract

Detecting cancers at their early stage would decrease mortality rate. For instance, detecting all polyps during colonoscopy would increase the chances of a better prognoses. However, endoscopists are facing difficulties due to the heavy workload of analyzing endoscopic images. Hence, assisting endoscopist while screening would decrease polyp miss rate. In this study, we propose a new deep learning segmentation model to segment polyps found in endoscopic images extracted during Colonoscopy screening. The propose model modifies SegNet architecture to embed Gated recurrent units (GRU) units within the convolution layers to collect contextual information. Therefore, both global and local information are extracted and propagated through the entire layers. This has led to better segmentation performance compared to that of using state of the art SegNet. Four experiments were conducted and the proposed model achieved a better intersection over union "IoU" by 1.36%, 1.71%, and 1.47% on validation sets and 0.24% on a test set, compared to the state of the art SegNet.

Keywords

SegNet, GRU, Embedded RNN, Polyp Segmentation

1. Introduction

According to National Institute of Diabetes and Digestive and Kidney Diseases [1], 60 to 70 million people are affected by a gastrointestinal disease. Malignancies such as esophageal and colorectal cancer are in an increasing rate in western countries [2, 3, 4]. Early detection and removal of such malignant tissues using endoscopy would reduce the risk of developing a cancer. However, endoscopists are facing difficulties due to the heavy workload of analyzing endoscopic images [5], subtle lesions, or lack of experience [6]. Therefore, researchers have been proposing deep learning models to help endoscopists marking malignancies during the screening [7].

Polyp segmentation is considered to be a challenging task due to the non-uniformity of the gastrointestinal tract. Hence, researchers tend to modify on-the-shelf deep learning models that proved to be efficient in a specific domain [8, 9]. Also, the lack of big and representative datasets in the endoscopy domain is a persistent challenge which clamp the performance of

3rd International Workshop and Challenge on Computer Vision in Endoscopy (EndoCV2021) in conjunction with the 18th IEEE International Symposium on Biomedical Imaging ISBI2021, April 13th, 2021, Nice, France
hcxmh1@nottingham.edu.my (M. Haithami); Amr.Ahmed@nottingham.edu.my (A. Ahmed); Iman.Liao@nottingham.edu.my (I. Y. Liao); hamidjalab@um.edu.my (H. Jalab)
0000-0001-6340-8183 (M. Haithami); 0000-0002-7749-7911 (A. Ahmed); 0000-0001-5165-4539 (I. Y. Liao); 0000-0002-4823-6851 (H. Jalab)

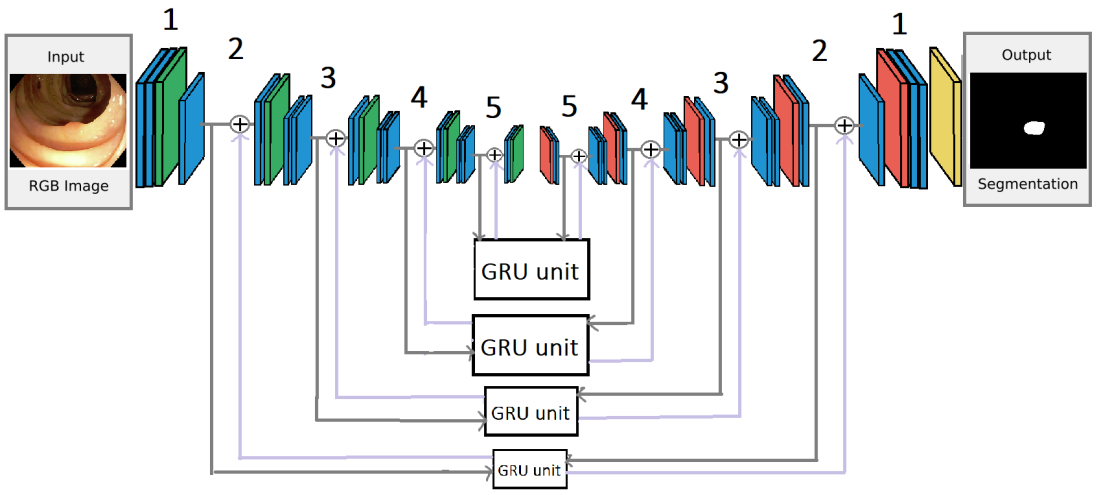

Figure 1: Proposed model that employ SegNet [16] architecture with VGG-16 as a backbone. One GRU unit is used for each stage to sweep over the obtained feature map horizontally and vertically. The corresponding encoder-decoder stages share the same GRU unit.

deep learning models. Therefore, transfer learning is frequently used to enhance a model's capability to segment diseases [10] and artifacts [11] found in endoscopic images. In other cases, augmentation is adopted instead as in [12, 13].

However it is noticed that Recurrent Neural Network (RNN) models are not studied enough in the literature. According to a recent survey [14], only 4 published studies employed RNN models to detect and segment colorectal polyps. Therefore in this research, we attempted to study the feasibility of using RNN. Gated recurrent unit (GRU) [15] - a type of RNN model- is selected to enhance the capabilities of SegNet [16] model to segment polyps. The proposed model contains four GRU units which are embedded into SegNet [16] to add contextual information to the feature maps. We conducted four experiments using EndoCV21 [17] dataset which was posted for "3rd International Endoscopy Computer Vision challenge and workshop" challenge [17]. The proposed model achieved intersection over union "IoU" better than the state-of-the-art SegNet [16] by a 1.51% on average for the validation sets and 0.24% on unseen test set.

2. Methodology

In this section, the used dataset as well as the proposed model would be discussed and elaborated in a detailed fashion.

2.1. Dataset

The used dataset is EndoCV21 which was created using different datacenters [17]. The database can be used for polyp segmentation as well as detection since both the mask and bounding box annotations in VOC format were provided. An interesting fact about this database is that it is

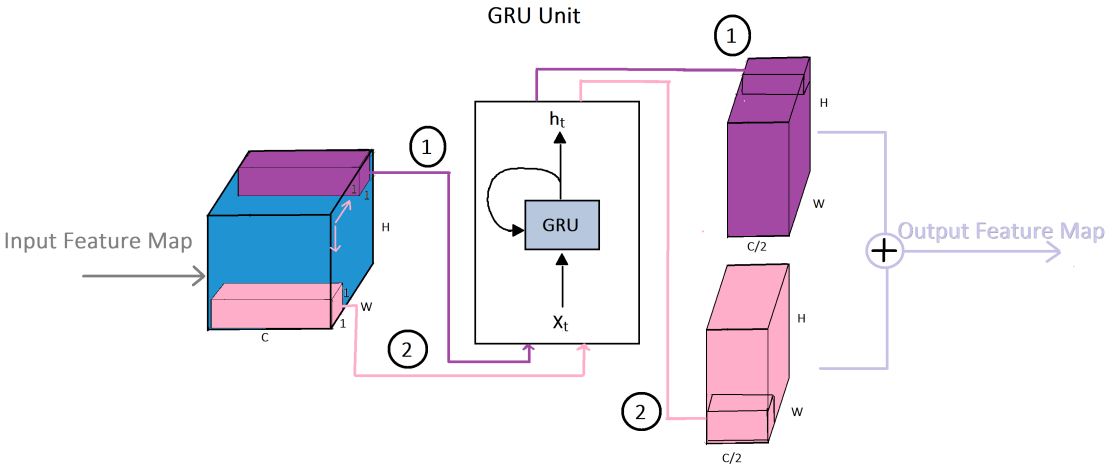

Figure 2: One GRU unit sweeps over the obtained feature map horizontally and vertically for each stage. Same GRU units are used for encoder and decoder.

divided into five sets, each corresponding to a medical center. Center-wise split is provided to encourage researchers to develop generalizable methods [17, 18]. The five annotated single frame datasets are Data_C1, Data_C2, Data_C3, Data_C4, and Data_C5 containing 256, 305, 457, 227, and 207 images, respectively. The images may have more than one polyp or no polyp at all.

2.2. Proposed model

Semantic segmentation requires integration and balancing of local information as well as global information at various spatial scales [19]. Conventional CNN models have some degree of spatial invariance due to pooling, though, they neglect global context information [19]. Therefore, different techniques are proposed in the literature to make CNN aware of contextual information such as applying post-processing refinement, dilated convolution, multi-scale convolution\aggregation, and applying sequence modeling "RNN". Inspired by [20], we propose a segmentation model that employ Gated Recurrent Unit (GRU) [15] within the layers of SegNet [16].

As opposed to [20], our proposed model embeds a GRU unit in each convolution stage within SegNet as shown in Figure2. A GRU unit is employed before the last convolution layer within each convolution stage. The total number of stages is five for the encoder as well as the decoder as depicted in Figure2. The GRU units are used to produce relevant global information [20]. The output feature map from the GRU units is then concatenated with the input feature map and passed to the next stage. By doing so the local features at each pixel position with respect to the whole input feature map would be implicitly encoded [20]. At each stage, only one GRU unit sweeps over the obtained feature map horizontally then vertically as depicted in Figure 2. The size of GRU hidden state is half of the channel size of the input feature map. Hence concatenating the GRU's output feature map after sweeping it in two directions will produce an output feature map with a size equal to the input feature map, as depicted in Figure 2.

The encoder as well as the decoder share the same GRU units for each corresponding stage as illustrated in Figure 2.

3. Experimental Results

The used Hyperparameters as will as the used metrics will be explained. Then the experimental results will be presented.

3.1. Hyperparameters

Both the proposed model and SegNet [16] had the same hyperparameters. Pytorch and Torchvision were used to conduct all experiments. The learning rate lr=0.005 and the batch size was 4 images. Adam optimization method was employed, and weighted Cross Entropy loss was used. The loss was weighted to mitigate the effect of imbalanced classes (i.e., class 0 "healthy" and class 1 "polyps"). Learning rate decay was employed with Multiplicative factor "gamma=0.8". Training and validation percentages are 80% and 20%, respectively.

3.2. Metrics

Three metrics were used in the experiments, namely, pixel accuracy and Intersection over Union "IoU". The pixel accuracy is defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP and TN are the true positive "polyp" and true negative "background", respectively. While FP and FN are false positive and false negative, respectively. As opposed to pixel accuracy metrics, only TP "polyp" is considered in calculating the IoU:

$$IoU = \frac{TP}{TP + FP + FN}$$

Finally, the Dice similarity is calculated based on IoU as follows:

$$Dice = \frac{2IoU}{IoU + 1}$$

EndoCV2021 leaderboard also included the generalisation metrics similar to detection generalisation defined in [18] for assessing performance gaps.

3.3. Results

It is important to mention that only polyp class is considered for calculating the IoU. Adding a background class "healthy" in the IoU's calculation would give us exaggerated results since the images' pixels are biased toward the background. The images as well as the masks are resized for computational efficiency reasons. One experiment was conducted with a resize factor of 3 and the rest were conducted with a resize factor of 6 (i.e., H=1080/6, W=1380/6).

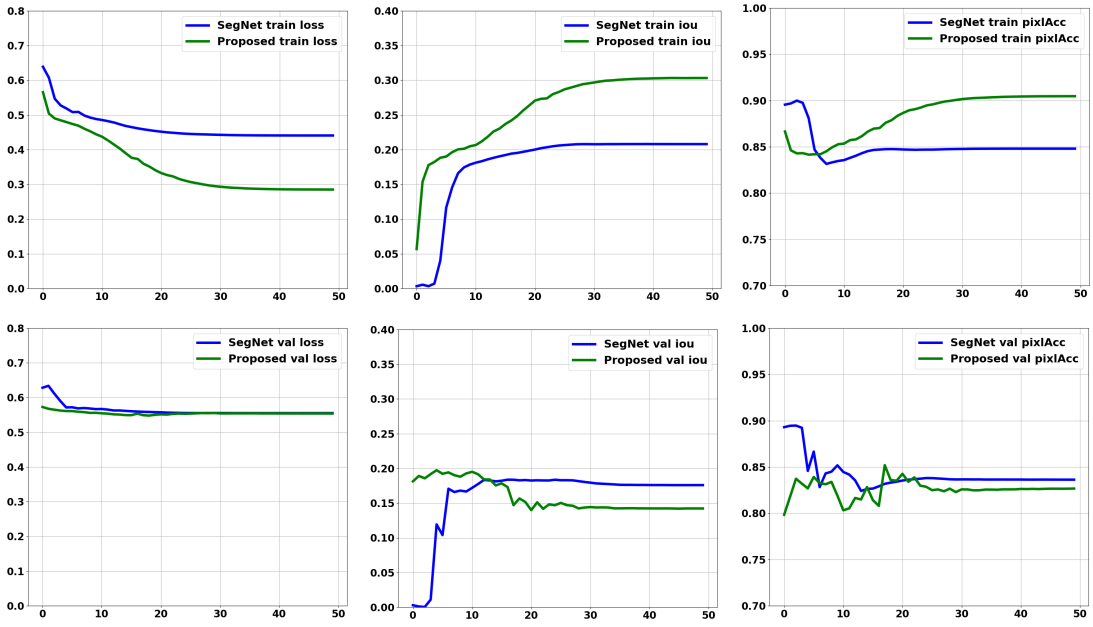

Figure 3: First experiment. `data_C1`, `data_C2`, and `data_C3` datasets were use. The images\masks are re-sized to a factor of 3. The first row is reserved for the training metrics while the second row is for validation metrics.

Four experiments in total were conducted to assess the performance of the proposed model against the state-of-the-art SegNet [16]. For the first experiment and the second experiment, Data_C1, Data_C2, and Data_C3 datasets were used with an image resize factor of 3 and 6, respectively. The aim is to see whether resizing the images would affect the performance of the proposed model in terms IoU metric. The total number of single frame images in this dataset is 1452. Figure 3 depicts metric curves (i.e., loss, IoU, and pixels accuracy) for the training and validation across 50 epochs. The third experiment employed different dataset in which Data_C1, Data_C4, and Data_C5 are used for training and validation. The images are resized by a factor of 6 and the total number of images is 690. The obtained highest IoU for the training and the validation set are summarized in Table 1 and Table 2, respectively. The mean and the standard deviation are calculated across the three experiments for the validation set. Dice similarity are calculated for each experiment based on the obtained IoU.

The fourth experiment was conducted to test the generalizability of the proposed model to correctly segment the polyp with contrast to SegNet as depicted in Table3. The best models' weights are selected based on the IoU obtained from the second experiment. The test set is composed of Data_C4 and Data_C5 with a total of 434 images. Note that in the second experiment we didn't use Data_C4 and Data_C5.

Table 1
The highest IoU results obtained from the training set. The best results are highlighted. The mean and the standard deviation are calculated across the three experiments

Model	Experiment1	Experiment2	Experiment3	Mean	Std
SegNet	0.2082	0.2927	0.2905	0.2638	0.0393
Proposed	**0.3033**	**0.3189**	**0.3180**	**0.3134**	**0.0072**

Table 2
The highest IoU results obtained from the validation set for each model. The Dice similarity was calculated directly from the obtained IoU. The best results are highlighted. The mean and the standard deviation are calculated across the three experiments

	Experiment1		Experiment2		Experiment3		Mean		Std	
Model\Metric	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
SegNet	0.1838	0.3105	0.1709	0.2919	0.1690	0.2891	0.1746	0.2972	0.0066	0.0095
Proposed	**0.1974**	**0.3297**	**0.1880**	**0.3165**	**0.1837**	**0.3104**	**0.1897**	**0.3189**	**0.0057**	**0.0081**

Table 3
SegNet and the proposed model tested on a test set "Data_C4 and Data_C5". The best results are highlighted and Dice similarity was calculated directly from the obtained IoU

Model\Metric	IoU	Dice
SegNet	0.1039	0.1882
Proposed	**0.1063**	**0.1922**

4. Discussion and Conclusion

4.1. Discussion

Experiment1: From Figure 3, it is clear that the proposed model is able to learn from the training data better than SegNet. Specifically speaking, the training IoU graph of the proposed model is higher than of the one obtained by SegNet. Furthermore, the highest IoU on the validation set is 0.1837 and 0.1974 for SegNet and the proposed model, respectively. As it can be seen from Table 1, the recorded highest IoU during the training were 0.2082 and 0.3033 for SegNet and the proposed model, respectively. This is an indication that the embedded GRU units enhanced the model's segmentation capabilities.

Experiment2 and Experiment3: As opposed to Experiment1, the gap between the achieved highest IoU during the training of the two models is less, as it can be observed from Table 1. Note that in Experiment1 the images are resized to a factor of 3 while in Experiment2 and Experiment3 the image are resize by a factor of 6. This observation is an indication that SegNet learns contextual information better with smaller images resolution given the fact that SegNet backbone is based on VGG16 which has only 16 convolutional layers. Reducing the spatial dimensions of input images will enhance the performance of the convolution layers of VGG11 which in turn will enhance the overall segmentation's performance. Furthermore,

using max-pooling layers will create denser feature maps if small images were used instead of larger resolution images. Nonetheless, still the proposed model achieved the highest IoU on the validation set as well with a difference of 1.72%, 1.47% in Experiment2 and Experiment 3, respectively. The IoU results for the training and validation sets are summarized in Table 1 and Table 2.

Experiment4 (test set): The proposed model achieved better IoU on the test set than SegNet with IoU=0.1063 for the former and IoU=0.1039 for the later, as seen in Table 3. However, compared to the other experiments, the difference in IoU between SegNet and the proposed model is relatively small (i.e., the difference is 0.24%). Although the embedded GRU units enhanced the model's ability to segment polyps found in images, they did not considerably help the model, relative to the state of the art SegNet, to generalize the learned segmentation capabilities to new unseen images. Nevertheless, it is clear from the experimental results that the proposed embedding GRU units enhanced the model's ability to learn contextual features, which in turn enhanced the overall segmentation performance. In general, the proposed model achieved IoU better than the state-of-the-art SegNet in all experiments.

4.2. Conclusion

The conclusions can be summarized as follows:

- We proposed a new segmentation model based on SegNet [16] and GRU [] RNN network for colorectal polyp segmentation. Four GRU units are embedded within the SegNet convolution layers and shared between the encoder and decoder.
- The IoU metric demonstrated that the segmentation capabilities are better than the state-of-the-art SegNet. The reason of this enhancement is the ability of GRU units to extract global information within the feature maps.
- The proposed model achieved better IoU than the state-of-the-art SegNet. The mean IoU on the validation sets for the proposed mode and SegNet are 0.1897 and 0.1746, respectively. While in the test set the proposed model and SegNet achieved 0.1063 and 0.1039, respectively.
- The difference in IoU on the test set between the proposed model and SegNet is relatively small even though the proposed model showed better IoU on the validation sets. Therefore, the generalizability property of the embedded GRU units needs to be studied further with different embedding techniques.

References

[1] Digestive diseases statistics for the united states | niddk, 2020. URL: https://www.niddk.nih.gov/health-information/health-statistics/digestive-diseases#all, [Online; accessed 2020-02-12].

[2] A. Nogueira-Rodríguez, R. Dominguez-Carbajales, H. López-Fernández, A. Iglesias, J. Cubiella, F. Fdez-Riverola, M. Reboiro-Jato, D. Glez-Pena, Deep neural networks approaches for detecting and classifying colorectal polyps, Neurocomputing 423 (2021) 721–734.

[3] N. Ghatwary, A. Ahmed, E. Grisan, H. Jalab, L. Bidaut, X. Ye, In-vivo barrett's esophagus digital pathology stage classification through feature enhancement of confocal laser endomicroscopy, Journal of Medical Imaging 6 (2019) 014502.
[4] N. Ghatwary, A. Ahmed, X. Ye, Automated detection of barrett's esophagus using endoscopic images: a survey, in: Annual conference on medical image understanding and analysis, Springer, 2017, pp. 897–908.
[5] Y. S. He, J. R. Su, Z. Li, X. L. Zuo, Y. Q. Li, Application of artificial intelligence in gastrointestinal endoscopy, Journal of digestive diseases 20 (2019) 623–630.
[6] P. Wang, T. M. Berzin, J. R. G. Brown, S. Bharadwaj, A. Becq, X. Xiao, P. Liu, L. Li, Y. Song, D. Zhang, et al., Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study, Gut 68 (2019) 1813–1819.
[7] K. Namikawa, T. Hirasawa, T. Yoshio, J. Fujisaki, T. Ozawa, S. Ishihara, T. Aoki, A. Yamada, K. Koike, H. Suzuki, et al., Utilizing artificial intelligence in endoscopy: a clinician's guide, Expert Review of Gastroenterology & Hepatology 14 (2020) 689–706.
[8] P. Brandao, O. Zisimopoulos, E. Mazomenos, G. Ciuti, J. Bernal, M. Visentini-Scarzanella, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, et al., Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks, Journal of Medical Robotics Research 3 (2018) 1840002.
[9] Y. H. Choi, Y. C. Lee, S. Hong, J. Kim, H.-H. Won, T. Kim, Centernet-based detection model and u-net-based multi-class segmentation model for gastrointestinal diseases., in: EndoCV@ ISBI, 2020, pp. 73–75.
[10] S. Rezvy, T. Zebin, B. Braden, W. Pang, S. Taylor, X. Gao, Transfer learning for endoscopy disease detection and segmentation with mask-rcnn benchmark architecture, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, 17, 2020.
[11] L. D. Huynh, N. Boutry, A u-net++ with pre-trained efficientnet backbone for segmentation of diseases and artifacts in endoscopy images and videos, volume 2595, CEUR-WS, 2020, pp. 13–17.
[12] A. Subramanian, K. Srivatsan, Exploring deep learning based approaches for endoscopic artefact detection and segmentation, volume 2595, CEUR-WS, 2020, pp. 51–56.
[13] Y. B. Guo, Q. Zheng, B. J. Matuszewski, Deep encoder-decoder networks for artefacts segmentation in endoscopy images, volume 2595, CEUR-WS, 2020, pp. 18–21.
[14] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, J. B. Pagador, Deep learning to find colorectal polyps in colonoscopy: A systematic literature review, Artificial Intelligence In Medicine (2020) 101923.
[15] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder-decoder approaches, arXiv preprint arXiv:1409.1259 (2014).
[16] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence 39 (2017) 2481–2495.
[17] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, M. A. Riegler, P. Halvorsen, C. Daul, J. Rittscher, O. E. Salem, D. Lamarque, T. de Lange, J. E. East, Polypgen: A multi-center polyp detection and segmentation dataset for generalisability assessment, arXiv (2021).
[18] S. Ali, F. Zhou, B. Braden, A. Bailey, S. Yang, G. Cheng, P. Zhang, X. Li, M. Kayser, R. D.

Soberanis-Mukul, S. Albarqouni, X. Wang, C. Wang, S. Watanabe, I. Oksuz, Q. Ning, S. Yang, M. A. Khan, X. W. Gao, S. Realdon, M. Loshchenov, J. A. Schnabel, J. E. East, G. Wagnieres, V. B. Loschenov, E. Grisan, C. Daul, W. Blondel, J. Rittscher, An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy, Scientific Reports 10 (2020) 2748. doi:10.1038/s41598-020-59413-5.

[19] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, J. Garcia-Rodriguez, A survey on deep learning techniques for image and video semantic segmentation, Applied Soft Computing Journal 70 (2018) 41–65.

[20] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, A. Courville, Reseg: A recurrent neural network-based model for semantic segmentation, 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2015) 426–433.