

Stable Acquisition of Fine-Grained Segments Using Batch Normalization and Focal Loss with L1 Regularization in U-Net Structure

Seoung-Ho Choi¹  and Sung Hoon Jung² 

¹Department of Electronics and Information Engineering, Hansung University, Seoul, Korea

²Division of Mechanical and Electronics Engineering, Hansung University, Seoul, Korea



Abstract

Acquisition of fine-grained segments in semantic segmentation is important in most semantic segmentation applications, especially for clothing images composed of fine-grained textures. However, most existing semantic segmentation methods based on fully convolutional network (FCN) were not enough to acquire fine-grained segments because they are based on a single resolution and can not well distinguish between objects in the images. To stabilize the acquisition of fine-grained segments, we propose a method that is composed of two additional components in the U-Net structure for processing multi-scale fine-grained segments. The first component is to use normalization at all layers. We found from experiments that normalization is a key process in stabilizing the acquisition of fine-grained segments, especially in the U-Net based methods because they operate on a multi-scale fine-grained segment. An additional component is to use model prediction correction using focal loss with L1 regularization. Focal loss can be used to control the model prediction term as regularization in the training process. From experiments, we found that our method was better than the existing methods.

Keywords: Multi-scale segments, Batch normalization, Model prediction correction

1. Introduction

Most images consist of various sizes, shapes, and textures. Therefore, it is necessary for specific image processing to operate appropriately under images with various characteristics. Since the textures of clothing images are very fine-grained and diverse, acquisition of fine-grained segments is particularly important. Finally, obtaining accurate semantic segments of images is an important goal to achieve in semantic segmentation researches [1].

However, there are not many studies that explore the acquisition of fine-grained segments of semantic segmentation in images. Most existing researches on image semantic segmentation has adopted fully convolutional network (FCN) [2, 3]. However, their segmentation results were not good at finding fine-grained segments, semantic segments because the convolution of FCN [2] didn't maintain spatial information. Even though there were related works on fully convolutional network-conditional random field (FCN-CRF) [3], FCN [2] using Atrous convolution, and FCN using the skip diagram in the FCN upsampling process to improve the performances of FCN, most of FCN showed poor results. In the methods, obtaining various sizes of segments are not reflected in model training. Therefore, obtaining various sizes of fine-grained segments is a crucial factor in processing segmentation for acquisition of

Received: Sept. 18, 2019
Revised : Jan. 27, 2020
Accepted: Feb. 10, 2020

Correspondence to: Sung Hoon Jung
(shjung@hansung.ac.kr)
©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

fine-grained segments in images.

To overcome the problem, we adopt two additional components on a U-Net [4] based structure for acquisition of fine-grained segments of semantic segmentation. The first additional component is the use of normalization at all layers in the training process. Normalization alleviates the variation of multi-scale information, especially in the multi-scale processing of U-Net. We took the well-known batch normalization (BN) [5] in all U-Net layers. As second additional component, we take the model prediction correction using focal loss [6] with L1 [7] regularization in training. Existing loss of cross-entropy does not reflect the fine-grained segments because it does not balance the model prediction space. From observation, we consider the model prediction correction for fine-grained segments and propose a novel loss composed of focal loss [6] with L1 regularization [7]. As a result, we introduce a novel structure based on U-Net [4] that trained with BN and a devised novel loss.

To measure performances of our method and previous methods, we experimented with three models such as U-Net [4], Attention U-Net [8], and U-Net BN including the existing FCN [2] models on the ATR dataset [9]. U-Net BN refers to a model that applies BN to all layers of the existing U-Net. We also tested our method with various combinations of loss and provided their results with some measures such as intersection over union (IOU), precision, and recall. From extensive experiments, we find that our method with two additional components have generated correct fine-grained segments, especially in small and complex textures such as hands, feet, and glasses. Also, the overall performances of our method were considerably better than those of the existing methods based on FCN in terms of accuracy using IOU, precision, and recall measures.

The rest of the paper is organized as follows. In Section 2, we introduce related works. The U-Net structure with two additional components of BN and novel combined focal loss with L1 regularization is addressed in Section 3. Section 4 shows the experiment results and discussion. We conclude in Section 5 with further works.

2. Previous Works of Segmentation

The segmentation model is divided into semantic segmentation [2, 4] and instance segmentation [10]. Semantic segmentation [2–4, 8, 11] is to provide a same label to semantically same objects, while instance segmentation is to assign a new label for each instance. Since the aim of the paper is for the acquisition

of fine-grained segments in semantic segmentation, we focus on semantic segmentation. In semantic segmentation, FCN [2] is a conventional method. Although FCN has used in semantic segmentation, there is a limitation that it can only extract features without maintaining the shapes of images because it uses a single resolution. To solve limitations, Hao and Kang [12] have proposed a method for extracting multi-feature maps using pyramid pooling based on FCN-VGG for semantic segmentation. It is confirmed that the improved features enhanced the performance through fusion with the extracted features.

However, their method has another limitation since the pyramid pooling is performed after the last layer of FCN-VGG, only abstract features are used for pyramid pooling. As a result, deconvolution using these abstract features in deconvolution layers generated poor segments with much distortion. There are a lot of researches on segmentation without distortion, especially for medical images [4, 13, 14]. Even if they showed good segmentation results with less distortion, they also had other limitations in the viewpoints of finding fine-grained segments. In other words, these methods did not find fine-grained segments, although their distortion was small.

It was well known that multi-scale processing helped segmentation algorithms and a hierarchical structure could implement multi-scale processing. Salembier [15] showed that morphology could be efficiently differentiated using a hierarchical structure. Wang et al. [16] propose spawning and using low-rank subspace features from the hierarchical structure. There are no previous researches on how to affects multiscale processing as acquisition of fine-grained segments of semantic segmentation. We propose a new method to use two additional components to improve performance in U-Net.

3. Proposed Multi-Scale and Fine-Grained Segmentation

We propose a U-Net [4] based semantic segmentation method for acquisition of multi-scale fine-grained segments in semantic segmentation with two additional components. U-Net is a well-known model that can find fine-grained information through efficient reflection of multi-scale information. However, semantic segmentation using only U-Net does not find fine-grained segments, especially for the images composed of various shapes and textures. To overcome the limitation, we adopt two additional components, BN to efficiently learn stabilized multi-scale fine-grained segments and model prediction correction using focal loss with L1 regularization. We found from our extensive

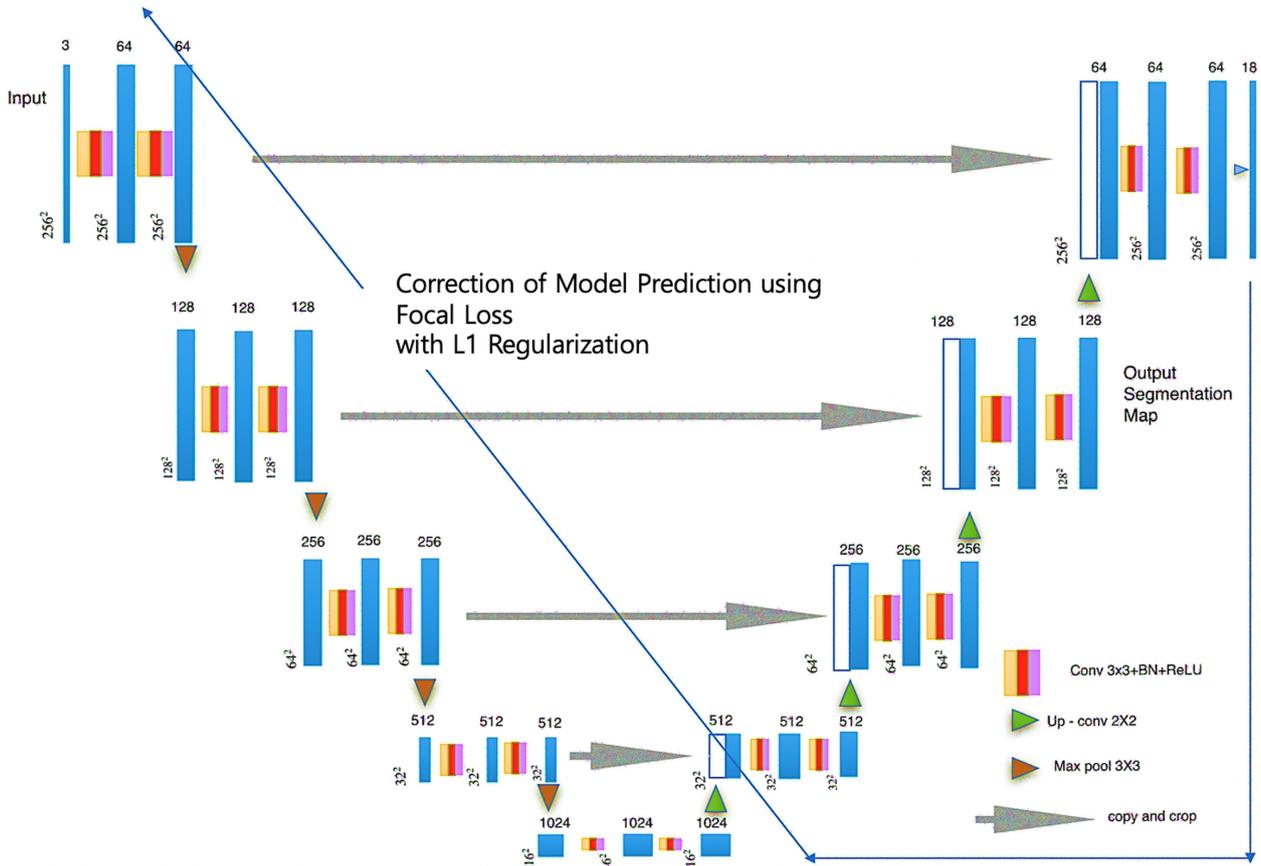


Figure 1. Proposed U-Net structure.

experiments that BN was a crucial process in finding multi-scale fine-grained segments because the normalization of data was more important to precisely calculate fine-grained segments for multi-scale data in all layers. Also, we use a combining method of focal loss and L1 regularization to balance the model prediction correction that enables the segmentation to stabilize acquired fine-grained segments in multi-scale data. We confirmed the results through extensive experimentation in Section 4

Figure 1 shows the structure of our U-Net based semantic segmentation. The overall structure is nearly the same as the original U-Net except for the BN of all layers, depicted as yellow arrows and model prediction correction using focal loss [6] with L1 [7] regularization. s^2 , where $s = 16, 64, 128,$ and 256 represents the width x height of filter maps. The number written above the blue block is the number of filter maps for the block. The gray arrow represents a skip diagram, in which the information of an existing block translated and transformed into a white block in the blue frame. The green, orange, and yellow arrows represent the up-sampling convolution, the 3x3 max pooling,

and the 3x3 convolution, respectively. The proposed focal loss with L1 regularization is applied to the outputs of U-Net [4], and the loss is used to train our structure. As mentioned before, the BN in all layers and the focal loss with L1 regularization enable our structure to acquire multi-scale fine-grained segments. Acquisition of fine-grained segments through BN on all layers can be calculated more precisely because the same size can be calculated in all spaces. To show the performances of our method with combination of regularization coefficient parameters, we extensively experimented and compared the results in Section 4

4. Experiment Results and Discussion

4.1 Experiment Environment

The experimental environments are as follows. We conducted our experiments with a learning rate of 0.001 and a batch size of 4. We chose an ATR dataset [9] of clothing images because they are composed of various sizes, shapes, and textures and are therefore adequate to get fine-grained segments. Background

images excluded during the evaluation. In order to evaluate the performances with various measures, we used recall [17], precision [17], F1 score [17], and IOU.

In all experiments, we used zero-padding to make all the images the same size as the long side of the image. We compared our method with the existing method of FCN and experimented on three U-Net models: U-Net without BN and focal loss with L1 regularization, U-Net BN and focal loss with L1 regularization, and attention U-Net using focal loss with L1 regularization. Attention U-Net has an U-Net structure with an attention gate. To verify focal loss with L1 regularization for optimization using regularization, we compared the performance of existing cross-entropy and proposed method, that is, focal loss with L1 regularization. In experiments, we tested proposed focal loss with L1 regularization, focal loss with L2 regularization, and focal loss with L1 and L2 to compare the regularization. L1 and L2 regularization coefficient are added at a ratio of 0.5.

4.2 Experimental Results and Discussion

We tested the performances of acquisition of fine-grained segment of semantic segmentation using previous cross-entropy loss and proposed focal loss with L1 regularization on three semantic segmentation models such as FCN, Attention U-Net, and U-Net BN on the clothing images. Figure 2 shows the experimental results of the three models on four combination of regularization coefficient parameters. As mentioned before, we choose the the regularization coefficient parameters combining 0 and 0, 0 and 0.5, 0.5 and 0, 0.5 and 0.5. In Figure 2, the blue and red circles are incorrect predictions, and red circles are the best among incorrect predictions. As you can see, the most results using focal loss with L1 regularization are better than those using the cross-entropy loss, especially on fine-grained segments shapes such as hands, foods, shoes, and neck.

The best result of all experiments is shown in panel (VII) of Figure 2, which uses focal loss with L1 regularization. That is, the L2 regularization is not helpful for acquisition of fine-grained segments of semantic segmentation. L1 regularization was more robust than L2 regularization concerning the outlier. Many fine-grained segments in the images with various sizes, shapes, and textures are outliers. Therefore, L1 regularization is better than L2 regularization for acquisition of fine-grained segments of semantic segmentation.

The attention U-Net and U-Net BN are better than the FCN. Because the U-Net stably acquires multi-scale fine-grained segments. Of the three methods, the U-Net BN shows the best

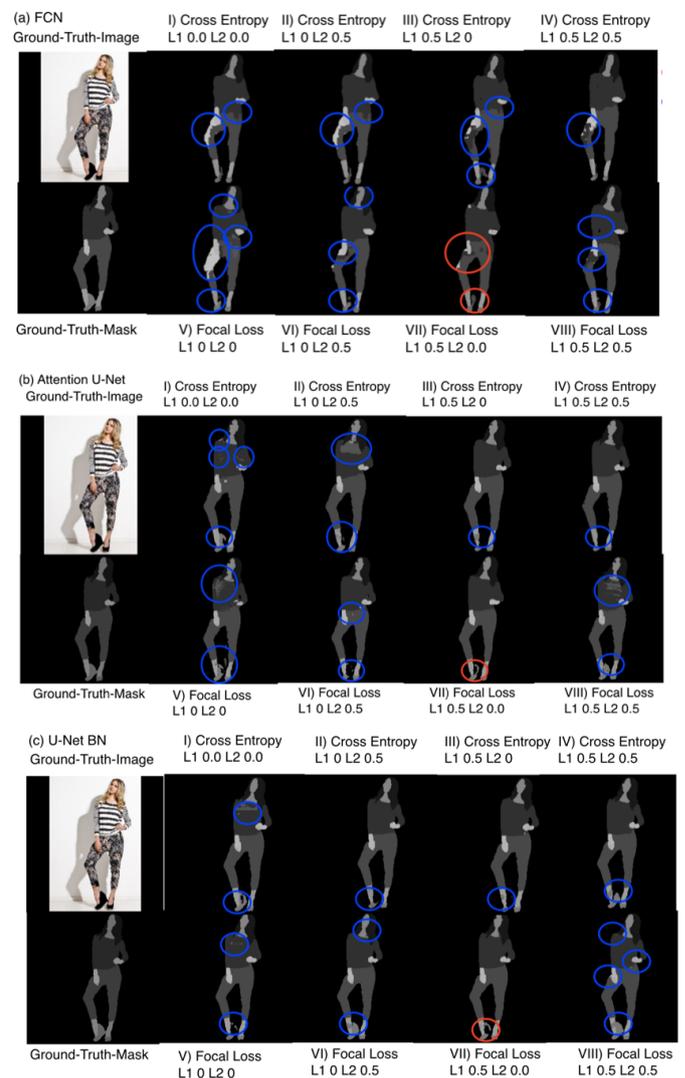


Figure 2. Result of focal loss regularization model: (a) FCN, (b) attention U-Net, and (c) U-Net BN. (I) Cross-entropy with L1 0.0 and L2 0.0 regularization coefficient, (II) cross-entropy with L1 0.0 and L2 0.5 regularization coefficient, (III) cross-entropy with L1 0.5 and L2 0.0 regularization coefficient, (IV) cross-entropy with L1 0.5 and L2 0.5 regularization coefficient, (V) Focal loss with L1 0.0 and L2 0.0 regularization coefficient, (VI) focal loss with L1 0.0 and L2 0.5 regularization coefficient, (VII) focal loss with L1 0.5 and L2 0.0 regularization coefficient, and (VIII) focal loss with L1 0.5 and L2 0.5 regularization coefficient.

results of all experiments because the normalization of all layers helps find fine-grained segments. From these results, we can ascertain that the focal loss with L1 regularization and the normalization are useful for acquisition of fine-grained segmentation. In our experiments, we tested the intrinsic U-Net structure with cross-entropy loss, but the training loss does not decrease, as shown in the result Figures 3 and 4 As a result,

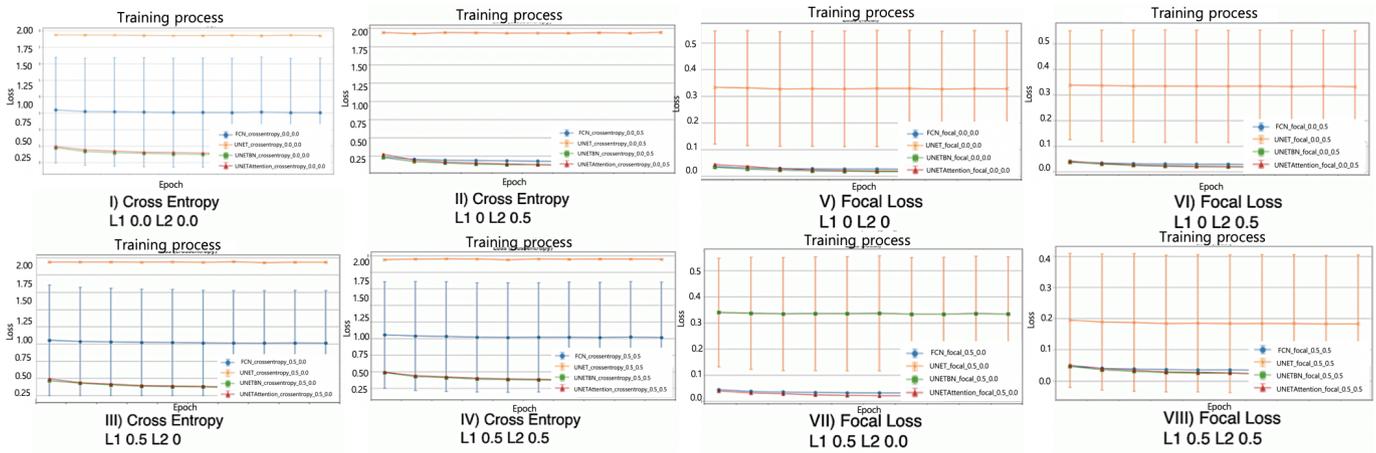


Figure 3. Comparison with or without BN about two loss function types and various regularization coefficients in loss function within training time. (I) Cross-entropy with L1 0.0 and L2 0.0 regularization coefficient, (II) cross-entropy with L1 0.0 and L2 0.5 regularization coefficient, (III) cross-entropy with L1 0.5 and L2 0.0 regularization coefficient, (IV) cross-entropy with L1 0.5 and L2 0.5 regularization coefficient, (V) focal loss with L1 0.0 and L2 0.0 regularization coefficient, (VI) focal loss with L1 0.0 and L2 0.5 regularization coefficient, (VII) focal loss with L1 0.5 and L2 0.0 regularization coefficient, and (VIII) Focal loss with L1 0.5 and L2 0.5 regularization coefficient.

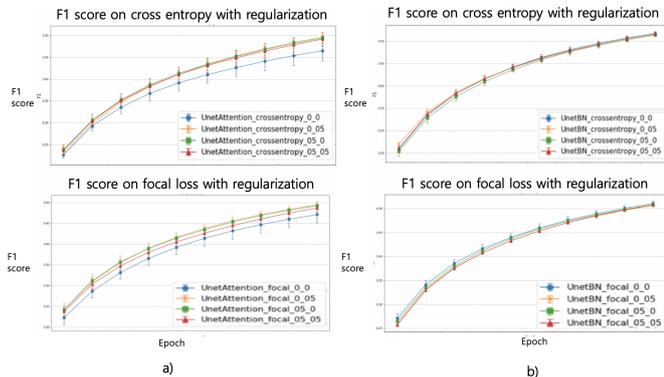


Figure 4. Comparison of F1 score according to addition of attention gate on segmentation model within training time: (a) in the attention gate included and (b) in the attention gate non-included.

the U-Net BN using focal loss with L1 regularization shows excellent results.

To ensure whether the U-Net BN shows the best result for various clothes images, we tested four methods on four clothing images, as shown in Figure 5. In these experiments, we added the results of U-Net using focal loss with L1 0.5 regularization. Figure 5 shows the experimental results of four methods on four images. As shown in Figure 5, U-Net BN using focal loss with L1 0.5 and L2 0.0 regularization shows the best results for the leg, skirt, and neck. The results of Figure 5(a) are the worst because the colors in the background are similar to the colors wrong by the women. Even the U-Net BN using focal

loss with L1 0.5 and L2 0.0 generates the best quality because the localization effect of normalization makes it possible for the method to better distinguish the colors in the background from those of the women. In Figure 5(c), the U-Net BN using focal loss with L1 0.5 and L2 0.0 regularization also shows the best results, especially on the small parts of the shoes, when compared to the other methods. The U-Net BN using focal loss with L1 0.5 and L2 0.0 regularization in Figure 5(d) generates more precise fine-grained segments in the area of the sunglasses than the other methods and even better than the ground-truth mask.

To more specifically analyze the effects of BN with two loss and two loss with some regularization coefficients on the ATR dataset, we showed the loss for the four models in Figure 3. Orange line indicates loss of U-Net structure. As you can see, the loss of the U-Net structure showed inferior results and nearly did not decrease. The intrinsic U-Net with cross-entropy and without normalization does not work well for fine-grained segments of semantic segmentation. The variation of loss of intrinsic U-Net is not very large in the cases of cross-entropy loss but is quite large in the focal loss because the model prediction correction of focal loss is sometimes successful in training the intrinsic U-Net.

FCN showed poor performances in the case of L1 0.5 regularization. In the process of obtaining the segments in the FCN, the calculated difference between the predicted value and the correct value is not reflected in the process of reflecting the dif-



Figure 5. Four methods on four clothing images: (a) case 1, (b) case 2, (c) case 3, and (d) case 4. The proposed method is focal loss L1 0.5 regularization coefficient. (I) FCN, (II) U-Net, (III) attention U-Net, and (IV) U-Net BN.

ference value. The variation of loss of U-Net BN becomes too large when L1 0.5 and L2 regularization. The L1 regularization reflects the differences that have diverse values according to experiments. Therefore, it is crucial to reflect the difference value efficiently. Overall, the methods with focal loss showed better results than those with cross-entropy loss. As shown in panel (VII) of Figure 3, the methods using both L1 and L2 regularization showed the worst performances. Because too much regularization provides a reverse effect.

We analyzed the influence of the attention gate in terms of an F1 score. Figure 4 showed the results of experiments: Figure 4(a) is with attention gate and Figure 4(b) is without attention gate. The result of U-Net without the attention gate was higher than U-Net with the attention gate as shown in Figure 4. Because the attention gate in the initial steps of training tries to find the strongest characteristics in the images, but it find erroneous characteristics.

Table 1 shows the impact of four regularization methods along with IOU, precision, and recall measurements for three models: U-Net, attention U-Net, and U-Net BN. The values shown in Table 1 are the mean and standard deviation of the results obtained three times on the same model. As shown in Table 1, the overall performance of U-Net is worse than those of the other models. However, the U-Net with L1 0.5 and L2 0.0 regularization Table 1, which showed about 23% improvement over U-Net with other regularization methods. We think that the regularization with L1 0.5 and L2 0.0 coefficient is effective in the attention of U-Net. From the results, the attention is not helpful, and the performance of U-Net BN is improved by about 22% over those of U-Net. Therefore we can be deduced that attention gate in the process of learning induces learning in the wrong direction. From these results, it is confirmed that a focal loss with L1 0.5 and L2 0.0 regularization proves useful for the acquisition of fine-grained segments in semantic segmentation.

Table 1. Comparison of both focal loss about U-Net models

		IOU	Precision	Recall
U-Net	(i)	0.496 ± 0.000	0.001 ± 0.000	0.000 ± 0.000
	(ii)	0.496 ± 0.000	0.001 ± 0.001	0.000 ± 0.000
	(iii)	0.632 ± 0.011	0.344 ± 0.029	0.333 ± 0.004
	(iv)	0.496 ± 0.000	0.003 ± 0.002	0.000 ± 0.000
Attention U-Net	(i)	0.715 ± 0.008	0.536 ± 0.002	0.420 ± 0.002
	(ii)	0.724 ± 0.005	0.555 ± 0.005	0.444 ± 0.013
	(iii)	0.723 ± 0.002	0.554 ± 0.006	0.441 ± 0.004
	(iv)	0.723 ± 0.005	0.548 ± 0.012	0.437 ± 0.013
U-Net BN	(i)	0.731 ± 0.002	0.565 ± 0.007	0.464 ± 0.005
	(ii)	0.728 ± 0.004	0.564 ± 0.004	0.458 ± 0.004
	(iii)	0.729 ± 0.003	0.567 ± 0.003	0.459 ± 0.007
	(iv)	0.730 ± 0.001	0.564 ± 0.006	0.461 ± 0.001

(I) L1 0.0 and L2 0.0 regularization coefficient, (ii) L1 0.0 and L2 0.5 regularization coefficient, (iii) L1 0.5 and L2 0.0 regularization coefficient, and (iv) L1 0.5 and L2 0.5 regularization coefficient.

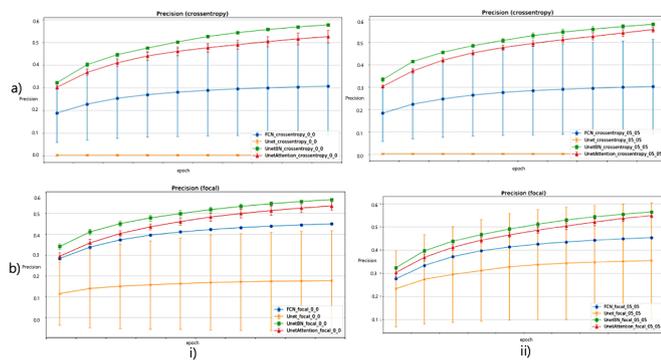


Figure 6. Comparison of regularization effect using non-regularization and regularization with L1 and L2 regularization: (a) cross-entropy loss and (b) focal loss. (i) L1 0.0 and L2 0.0 regularization coefficient, (ii) L1 0.5 and L2 0.5 regularization coefficient.

Figure 6 shows regularization effects of cross-entropy and focal loss using non-regularization and regularization with L1 and L2. This confirms that regularization does not significantly affect on cross-entropy results of U-Net. However, the performance of acquisition of fine-grained segments of semantic segmentation is improved when the regularization is combined to focal loss. The focal loss has a term that reflects the reverse of the probability from the model prediction. That means that model correction prediction is more affected by focal loss unlike the existing cross-entropy.

5. Conclusion

We proposed a method based on the U-Net structure with two additional components to acquisition of fine-grained segments. For acquisition of fine-grained segments, we added normalization to all layer of the U-Net structure and proposed combined component composed of focal loss with L1 regularization. We experimented with proposed methods on an ATR dataset and analyzed their results. Experimental results showed that proposed methods were better than the previous FCN and intrinsic U-Net. These results allowed us to know that the U-Net was a structure for semantic segmentation, adopted normalization about all layer on the U-Net was beneficial for semantic segmentation, and the model prediction correction using focal loss with L1 regularization was good at acquiring the fine-grained segments in semantic segmentation. In the future, we will proceed with the semantic segmentation structure using the generative model to obtain a more robust acquisition of fine-grained segment of semantic segmentation.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This research was financially supported by Hansung University.

References

- [1] L. Huang, J. Peng, R. Zhang, G. Li, and L. Lin, "Learning deep representations for semantic image parsing: a comprehensive overview," *Frontiers of Computer Science*, vol. 12, no. 5, pp. 840-857, 2018. <https://doi.org/10.1007/s11704-018-7195-8>
- [2] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 2017. <http://doi.org/10.1109/TPAMI.2016.2572683>
- [3] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2015, Available <https://arxiv.org/pdf/1412.7062.pdf>
- [4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [5] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, Available <https://arxiv.org/abs/1502.03167>
- [6] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," 2017, Available <https://arxiv.org/abs/1708.02002>
- [7] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada, 2004. <https://doi.org/10.1145/1015330.1015435>
- [8] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., "Attention U-net: learning where to look for the pancreas," 2018, Available <https://arxiv.org/abs/1804.03999>
- [9] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin, and S. Yan, "Deep human parsing with active template regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2402-2414, 2015. <https://doi.org/10.1109/TPAMI.2015.2408360>
- [10] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2961-2969. <https://doi.org/10.1109/ICCV.2017.322>
- [11] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-net for lesion segmentation," in *Proceedings of 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, Venice, Italy, 2019, pp. 683-687. <https://doi.org/10.1109/ISBI.2019.8759329>
- [12] B. Hao and D. S. Kang, "Research on image semantic segmentation based on FCN-VGG and pyramid pooling module," *Journal of Korean Institute of Information Technology*, vol. 16, no. 7, pp. 1-8, 2018. <https://doi.org/10.14801/jkiit.2018.16.7.1>
- [13] N. Sharma, A. K. Ray, S. Sharma, K. K. Shukla, S. Pradhan, and L. M. Aggarwal, "Segmentation and classification of medical images using texture-primitive features: application of BAM-type artificial neural network," *Journal of Medical Physics*, vol. 33, no. 3, pp. 119-126, 2008. <https://dx.doi.org/10.4103%2F0971-6203.42763>
- [14] R. Sharma and A. Sungheetha, "Segmentation and classification techniques of medical images using innovated hybridized techniques: a study," in *Proceedings of 2017 11th International Conference on Intelligent Systems and Control (ISCO)*, Coimbatore, India, 2017, pp. 192-196. <https://doi.org/10.1109/ISCO.2017.7855979>
- [15] P. Salembier, "Morphological multiscale segmentation for image coding," *Signal Processing*, vol. 38, no. 3, pp. 359-386, 1994. [https://doi.org/10.1016/0165-1684\(94\)90155-4](https://doi.org/10.1016/0165-1684(94)90155-4)
- [16] Y. Wang, C. He, X. Liu, and M. Liao, "A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for PolSAR imagery classification," *Remote Sensing*, vol. 10, no. 2, article no. 342, 2018. <https://doi.org/10.3390/rs10020342>
- [17] G. Csurka, D. Larlus, and F. Perronnin, "What is a good evaluation measure for semantic segmentation?" in *Proceedings of the British Machine Vision Conference*, Bristol, UK, 2013.
- [18] S. Zhu, S. Fidler, R. Urtasun, D. Lin, and C. C. Loy, "Be your own prada: fashion synthesis with structural coher-

- ence,” in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 1680-1688. <https://doi.org/10.1109/ICCV.2017.186>
- [19] W. C. Kang, C. Fang, Z. Wang, and J. McAuley, “Visually-aware fashion recommendation and design with generative image models,” in *Proceedings of 2017 IEEE International Conference on Data Mining (ICDM)*, New Orleans, LA, 2017, pp. 207-216. <https://doi.org/10.1109/ICDM.2017.30>
- [20] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301-320, 2005. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [21] S. Kubo, Y. Iwasawa, and Y. Matsuo, “Generative adversarial network-based virtual try-on with clothing region,” 2018, Available <https://openreview.net/forum?id=B1WLiDJvM>
- [22] X. Zhu, H. Zhou, C. Yang, J. Shi, and D. Lin, “Penalizing top performers: conservative loss for semantic segmentation adaptation,” in *Computer Vision – ECCV 2018*. Cham: Springer, 2018, pp. 587-603. https://doi.org/10.1007/978-3-030-01234-2_35
- [23] N. Sarafianos, X. Xu, and I. Kakadiaris, “Deep imbalanced attribute classification using visual attention aggregation,” in *Computer Vision – ECCV 2018*. Cham: Springer, 2018, p. 708-725. https://doi.org/10.1007/978-3-030-01252-6_42
- [24] Z. H. Zheng, H. T. Zhang, F. L. Zhang, and T. J. Mu, “Image-based clothes changing system,” *Computational Visual Media*, vol. 3, no. 4, pp. 337-347, 2017. <https://doi.org/10.1007/s41095-017-0084-6>
- [25] M. D. Worsley, “Automatic segmentation of clothing for the identification of fashion trends using k-means clustering,” 2009, Available <http://cs229.stanford.edu/proj2009/McDanielsWorsley.pdf>
- [26] E. Simo-Serra, S. Fidler, F. Moreno-Noguer, and R. Urtasun, “A high performance crf model for clothes parsing,” in *Computer Vision – ACCV 2014*. Cham: Springer, 2014, pp. 64-81. https://doi.org/10.1007/978-3-319-16811-1_5
- [27] H. Zhang, Y. Sun, L. Liu, X. Wang, L. Li, and W. Liu, “ClothingOut: a category-supervised GAN model for clothing segmentation and retrieval,” *Neural Computing and Applications*, 2018. <https://doi.org/10.1007/s00521-018-3691-y>
- [28] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On large-batch training for deep learning: Generalization gap and sharp minima,” 2016, Available <https://arxiv.org/abs/1609.04836>
- [29] D. Masters and C. Luschi, “Revisiting small batch training for deep neural networks,” 2018, Available <https://arxiv.org/abs/1804.07612>
- [30] A. Laddha, “Road detection and semantic segmentation without strong human supervision,” M.S. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2016.
- [31] 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*, vol. 40, no. 4, pp. 834-848, 2018. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [32] N. Hurley and S. Rickard, “Comparing measures of sparsity,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723-4741, 2009. <https://doi.org/10.1109/TIT.2009.2027527>
- [33] T. Nguyen, T. Ozaslan, I. D. Miller, J. Keller, G. Loianno, C. J. Taylor, D. D. Lee, V. Kumar, J. H. Harwood, and J. Wozencraft, “U-net for MAV-based penstock inspection: an investigation of focal loss in multi-class segmentation for corrosion identification,” 2018, Available <https://arxiv.org/abs/1809.06576>
- [34] S. Eppel, “Hierarchical semantic segmentation using modular convolutional neural networks,” 2017, Available <https://arxiv.org/abs/1710.05126>
- [35] X. Wang, Z. Wu, P. Luo, and H. Wu, “An image segmentation method combining L1-sparse reconstruction with spectral clustering,” *Procedia Engineering*, vol. 29, pp. 1387-1391, 2012. <https://doi.org/10.1016/j.proeng.2012.01.145>

- [36] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" *Advances in Neural Information Processing Systems*, vol. 31, pp. 2483-2493, 2018.



Seoung-Ho Choi is an M.S. student in Department of Electronics and Information Engineering at Hansung University. Before that, he received the B.S. degree in Department of Electronics and Information Engineering at Hansung University.



Sung Hoon Jung received his B.S.E.E. degree from Hanyang University, Korea, in 1988 and M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), in 1991 and 1995, respectively. He joined the Department of Information and Communication Engineering at the Hansung University in 1996, where he is a professor. His research interests are in the fields of intelligent systems and systems biology. He is a member of the Korean Institute of Intelligent Systems (KIIS) and Institute of Electronics Engineers of Korea (IEEK).