KD-ResUNet++: Automatic Polyp Segmentation via Self-Knowledge Distillation

Jaeyong Kang[1], Jeonghwan Gwak[1,2,*]
[1] Department of Software, Korea National University of Transportation, Chungju 27469, South Korea
[2] Department of IT · Energy Convergence (BK21 FOUR), Korea National University of Transportation, Chungju 27469, South Korea

ABSTRACT

In this paper, we present our method for Medico automatic polyp segmentation challenge at MediaEval 2020. In our method, we utilized the knowledge distillation technique to improve ResUNet++ which performs well on automatic polyp segmentation. In our experiment, our proposed model called KD-ResUNet++ outperforms ResUNet++ in terms of Jaccard index, Dice similarity coefficient, and recall. Our best models achieved Jaccard index, Dice similarity coefficient, and FPS of 0.6196, 0.7089, and 107.8797 respectively on the official test dataset in the challenge.

1 INTRODUCTION

Automatic polyp segmentation is a challenging task due to variations in the shape and size of polyps. In this paper, we propose KD-ResUNet++, which is based on the ResUNet++ architecture [9] and knowledge distillation for Medico automatic polyp segmentation challenge at MediaEval 2020 [7]. Knowledge distillation is a method to transfer knowledge from one architecture (e.g., teacher) to another (e.g., student). In particular, we use self-knowledge distillation where teacher and student architectures are the same.

2 RELATE WORKS

2.1 ResUNet++

U-Net is a very popular deep learning architecture for biomedical image segmentation. U-Net won the 2015 ISBI cell tracking challenge. ResUNet [14] is an improved U-Net architecture which takes advantage of strengths from both the U-Net architecture and deep residual learning. ResUNet++ [9] is an improved ResUNet architecture that further takes advantage of attention blocks, Atrous Spatial Pyramidal Pooling (ASPP), and squeeze and excitation blocks. As being reported in [9], ResUNet++ shows the state-of-the-art performance on automatic polyp segmentation.

2.2 Knowledge distillation

Knowledge distillation aims at transferring dark knowledge from a teacher model such as wide [11] or deep [2, 10, 12], or an ensemble of models [5] to a student model which is typically thin and small. Trained in this way, the student model can mimic the behavior of the teacher model such as class probability distribution to achieve better performance than the model trained with hard labels independently. Self-knowledge distillation refers to the special case where the teacher and student architectures are the same. It has been consistently reported [1, 4, 6, 13] that student models trained with self-knowledge distillation show better performance than their teacher models by significant margins in several language modeling and computer vision tasks.

MediaEval'20, December 14-15 2020, Online

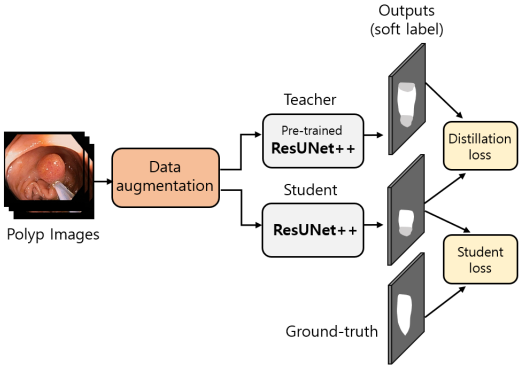

Figure 1: Our proposed KD-ResUNet++ architecture

3 METHODS

In this section, the architecture of our proposed method for automatic polyp segmentation is first presented. After that, we describe the details of the key components in the following subsections. The overall architecture of our proposed model is shown in Figure 1. First, input images are augmented by the data augmentation module. Second, augmented images are used as the input of both the student model and the teacher model. Third, distillation loss between the output of the student model and the output of the teacher model and the student loss between the output of the student model and ground-truth label is calculated to train the student model.

3.1 Data augmentation

Deep learning models require a large amount of training data to work effectively. However, the size of the provided colonoscopy dataset is not very large. To solve this problem, data augmentation can be used to make the relatively smaller dataset a large one. It is reported that the performance of the deep learning model can be improved by augmenting the existing data rather than collecting new data. In our data augmentation step, we used 2 augmentation strategies (rotation and horizontal flipping) to generate new training sets. The rotation operation used for data augmentation is done by randomly rotating the input by 90 degrees zero or more times. The rotation operation fills the area of rotated images where there was no image pixel with black. In addition, we applied horizontal flipping to each of the rotated images.

Table 1: Results on validation set

Model	Jaccard	DSC	Recall	Precision
ResUNet++	0.7342	0.8120	0.8260	0.8892
KD-ResUNet++	0.7530	0.8310	0.8495	0.8701

Table 2: Official results on polyp segmentation task

Model	Jaccard	DSC	Recall	Precision
KD-ResUNet++	0.6196	0.7089	0.7287	0.7914

Table 3: Official results on algorithm efficiency task

Model	FPS	Mean time taken
KD-ResUNet++	107.8797	0.0093

3.2 Training with Self-knowledge distillation

In our proposed approach, we use self-knowledge distillation where teacher network and student network are the same. We use ResUNet++ for both teacher and student networks. In knowledge distillation, the teacher network is first trained to transfer knowledge to the student network. Also, the loss function consists of 1) the distillation loss and 2) the student loss. The distillation loss L_{dist} can be calculated using dice loss between the output of the student model y_s and the output of pre-trained teacher model y_t, and the student loss L_s can be calculated using dice loss between the output of the student model y_s and the ground-truth label y_{true} as follows:

$$L_{dist} = 1 - Dice(y_s, y_t) \quad (1)$$

$$L_s = 1 - Dice(y_s, y_{true}) \quad (2)$$

The total loss L_{total} is then calculated as the joint of the distillation and student losses as follows:

$$L_{total} = 0.1 * L_{dist} + L_s \quad (3)$$

4 EXPERIMENTS AND RESULTS

4.1 Dataset

We trained our proposed model using the Kvasir-SEG dataset [8], the benchmark dataset for the 2020 Medico automatic polyp segmentation challenge. It consists of 1,000 polyp images and their corresponding ground truth masks annotated by expert endoscopists from Oslo University Hospital, Norway.

4.2 Experimental Setting

The dataset is split into 88 % for learning the weights and 12 % for validating the model during the training step. Before the training step, we augment input images using our data augmentation module described in Section 3.1 and converted the images to the size of 256×256 pixels. The validation set is only normalized. The learning rate is set to 0.001. We use Adam as our optimizer.

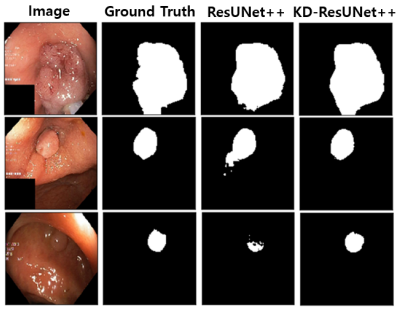

Figure 2: Examples of three different segmentations produced by ResUNet++ and KD-ResUNet++

4.3 Results

Our results on the validation set are presented in Table 1. Also, the official results on polyp segmentation and algorithm efficiency tasks on the same test dataset are shown in Table 2 and Table 3, respectively. Table 1 shows that ResUNet++ achieved slightly better precision than KD-ResUNet++. However, KD-ResUNet++ outperforms ResUNet++ in terms of Jaccard index and Dice similarity coefficient which is an important metric for semantic segmentation task. Table 1 shows that our proposed model outperforms ResUNet++ in terms of Jaccard index, Dice similarity coefficient, and recall. Table 2 and 3 show that our proposed model achieved Jaccard index, Dice similarity coefficient, and FPS of 0.6196, 0.7089, and 107.8797 respectively on the official test dataset. Besides, examples of three different segmentations produced by ResUNet++ and KD-ResUNet++ are depicted in Figure 2. Figure 2 shows that the result of KD-ResUNet++ are more similar with ground truth than the result of ResUNet++.

5 CONCLUSION

In this paper, we presented KD-ResUNet++ for automatic polyp segmentation. In our proposed framework, the data augmentation technique is applied to input images. Also, we use self-knowledge distillation where teacher and student networks are the same. We use the ResUNet++ model for our student and teacher networks. Our proposed model is evaluated on the validation set as well as the official test set. Our experimental results show that our proposed model outperforms ResUNet++ in terms of Jaccard index, Dice similarity coefficient, and recall. These results indicate that our proposed method can capture polyp segmentation boundary well and could be potentially used in clinical settings. In the future, we plan to use different knowledge types in our knowledge distillation. Also, we plan to modify the ResUNet++ architecture to incorporate the model pre-trained on a large image dataset (e.g., ImageNet [3]) to reduces the long training time which is normally required to train deep learning model from scratch, and also to remove the requirement of having a large training dataset.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2020R1I1A3074141).

REFERENCES

[1] Sungsoo Ahn, Shell Xu Hu, Andreas Damianou, Neil D Lawrence, and Zhenwen Dai. 2019. Variational information distillation for knowledge transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 9163–9171.

[2] Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan Chandraker. 2017. Learning efficient object detection models with knowledge distillation. In *Advances in Neural Information Processing Systems*. 742–751.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[4] Tommaso Furlanello, Zachary C Lipton, Michael Tschannen, Laurent Itti, and Anima Anandkumar. 2018. Born again neural networks. *arXiv preprint arXiv:1805.04770* (2018).

[5] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).

[6] Thi Kieu Khanh Ho and Jeonghwan Gwak. 2020. Utilizing Knowledge Distillation in Deep Learning for Classification of Chest X-Ray Abnormalities. *IEEE Access* 8 (2020), 160749–160761.

[7] Debesh Jha, Steven A. Hicks, Krister Emanuelsen, Håvard D. Johansen, Dag Johansen, Thomas de Lange, Michael A. Riegler, and Pål Halvorsen. 2020. Medico Multimedia Task at MediaEval 2020:Automatic Polyp Segmentation. In *Proc. of MediaEval 2020 CEUR Workshop*.

[8] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. 2020. Kvasir-SEG: A Segmented Polyp Dataset. In *Proc. of International Conference on Multimedia Modeling (MMM)*. 451–462.

[9] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. 2019. ResUNet++: An Advanced Architecture for Medical Image Segmentation. In *Proc. of International Symposium on Multimedia*. 225–230.

[10] Jaeyong Kang and Jeonghwan Gwak. 2020. Ensemble Learning of Lightweight Deep Learning Models Using Knowledge Distillation for Image Classification. *Mathematics* 8, 10 (2020), 1652.

[11] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2014. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550* (2014).

[12] Gregor Urban, Krzysztof J Geras, Samira Ebrahimi Kahou, Ozlem Aslan, Shengjie Wang, Rich Caruana, Abdelrahman Mohamed, Matthai Philipose, and Matt Richardson. 2016. Do deep convolutional nets really need to be deep and convolutional? *arXiv preprint arXiv:1603.05691* (2016).

[13] Chenglin Yang, Lingxi Xie, Siyuan Qiao, and Alan L Yuille. 2019. Training deep neural networks in generations: A more tolerant teacher educates better students. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5628–5635.

[14] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. 2018. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters* 15, 5 (May 2018), 749–753. https://doi.org/10.1109/lgrs.2018.2802944