



# Polyp Segmentation with Fully Convolutional Deep Dilation Neural Network Evaluation Study

Yunbo Guo<sup>(✉)</sup> and Bogdan J. Matuszewski

Computer Vision and Machine Learning (CVML) Research Group, School of Engineering,  
University of Central Lancashire, Preston, UK  
{ybguo1, bmatuszewski1}@uclan.co.uk

**Abstract.** Analyses of polyp images play an important role in an early detection of colorectal cancer. An automated polyp segmentation is seen as one of the methods that could improve the accuracy of the colonoscopic examination. The paper describes evaluation study of a segmentation method developed for the Endoscopic Vision Gastrointestinal Image ANALysis – (GIANA) polyp segmentation sub-challenges. The proposed polyp segmentation algorithm is based on a fully convolutional network (FCN) model. The paper describes cross-validation results on the training GIANA dataset. Various tests have been evaluated, including network configuration, effects of data augmentation, and performance of the method as a function of polyp characteristics. The proposed method delivers state-of-the-art results. It secured the first place for the image segmentation tasks at the 2017 GIANA challenge and the second place for the SD images at the 2018 GIANA challenge.

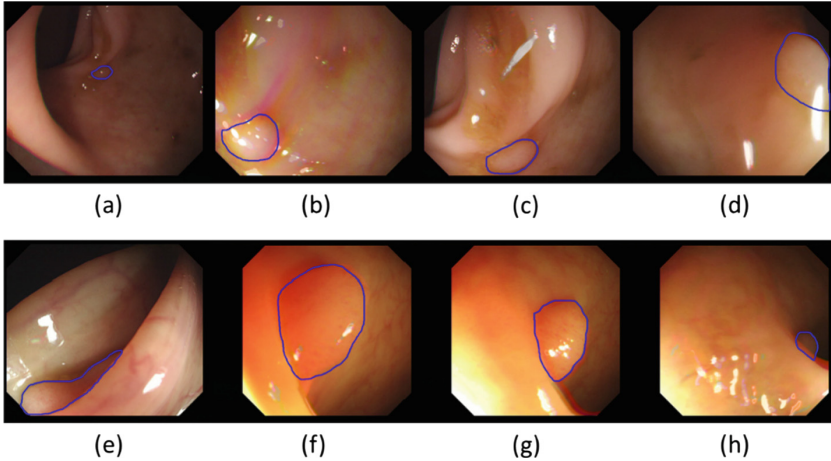
**Keywords:** Fully convolutional dilation neural networks · Polyp segmentation · Data augmentation · Cross-Validation · Ablation tests

## 1 Introduction

Colorectal cancer is one of the leading causes of cancer deaths worldwide. To decrease mortality, an assessment of polyp malignancy is performed during colonoscopy examination, so it can be removed at an early stage. Currently, during colonoscopy, polyps are usually examined visually by a trained clinician. To automate analysis of colonoscopy images, machine learning methods have been utilised and shown to support polyp detectability and segmentation objectivity.

Polyp segmentation is a challenging task due to inherent variability of polyp morphology and colonoscopy image appearance. The size, shape and appearance of a polyp are different at different stages. In an early stage, colorectal polyps are typically small, may not have a distinct appearance, and could be easily confused with other intestinal structures. In the later stages, the polyp morphology changes and the size begins to increase. Illumination in colon screening is also variable, producing local overexposure highlights and specular reflections. Some polyps may look very differently from different

camera positions, do not have a visible transition between the polyp and its surrounding tissue, be affected by intestinal content and luminal regions (Fig. 1), inevitably leading to segmentation errors.



**Fig. 1.** Typical polyps in the GIANA SD training dataset: (a, h) Small size; (b) Blur; (c) Intestinal content; (d) Specular highlights/defocused; (e) Occlusion; (f) Large size; (g) Overexposed areas; (a, e, h) Luminal region.

The research reported here has been motivated by the limitations of previously proposed methods. This paper evaluates a novel fully convolutional neural network designed to accomplish this challenging segmentation task. The developed FCN method outputs polyp occurrence confidence maps. The final polyp delineation is either obtained by a simple thresholding of these maps or the hybrid level-set [1, 2] is used to smooth the polyp contour and eliminate small noisy network responses. The proposed method has been introduced in [3]. This paper aims to provide more in depth analysis of the method characteristics, focusing on the selection of the design parameters, adopted data augmentation scheme as well as overall validation of the proposed method. This analysis has not been published before.

## 2 Related Work

In literature on colonoscopy image analysis, various terms have been used to describe similar objectives. For example, some of the reported polyp detection and localisation methods provide heat maps and/or different levels of polyp boundary approximations, which could be interpreted as segmentation. On the other hand segmentation tools could be also seen as providing polyp detection and localisation functionality. Most of the reported techniques relevant to polyp segmentation can be divided into two main approaches based on either apparent shape or texture, with the methods using machine learning gradually gaining popularity. Some of the early approaches attempted to fit a

predefined polyp shape models. Hwang et al. [4] used ellipse fitting techniques based on image curvature, edge distance and intensity values for polyp detection. Gross et al. [5] used Canny edge detector to process a prior-filtered images, identifying the relevant edges using a template matching technique for polyp segmentation. Breier et al. [6, 7] investigated applications of active contours for finding polyp outline. Although these methods perform well for typical polyps, they require manual contour initialisation.

The above mentioned techniques rely heavily on a presence of complete polyp contours. To improve the robustness, further research was focused on the development of robust edge detectors. Bernal et al. [8] presented a “depth of valley” concept to detect more general polyp shapes, then segment the polyp through evaluating the relationship between the pixels and detected contour. Further improvements of this technique are described in [9–11]. In the subsequent work, Tajbakhsh et al. [12] put forward a series of polyp segmentation method based on edge classification, utilising the random forest classifier and a voting scheme producing polyp localisation heat maps. In the follow-up work [13, 14] that approach was refined via use of several sub-classifiers.

Another class of polyp segmentation methods is based on texture descriptors, typically operating on a sliding window. Karkanis et al. [15] combined Grey-Level Co-occurrence Matrix and wavelets. Using the same database and classifier, Iakovidis et al. [16] proposed a method, which provided the best results in terms of area under the curve metric.

More recently, with the advances of deep learning, a hand-crafted feature descriptors are gradually being replaced by convolutional neural networks (CNN) [17, 18]. Ribeiro et al. [19] compared CNN with the state-of-art hand-crafted features on polyp classification problem, and found that CNN has superior performance. That method is based on a sliding window approach. The general problem with a sliding window technique is that it is difficult to use image contextual information and approach is very inefficient. This has been addressed by the so called fully convolutional networks (FCN), with two key architectures proposed in [20, 21]. These methods can be trained end-to-end and output complete segmentation results, without a need for any post-processing. Vázquez et al. [22] directly segmented the polyp images using an off-the-shelf FCN architecture. Zhang et al. [23] use the same FCN, but they add a random forest to decrease the false positive. The U-net [21] is one of the most popular architectures for biomedical image segmentation. It has been also used for polyp segmentation. Li et al. [24] designed a U-net architecture for polyp segmentation to encourage smooth contours.

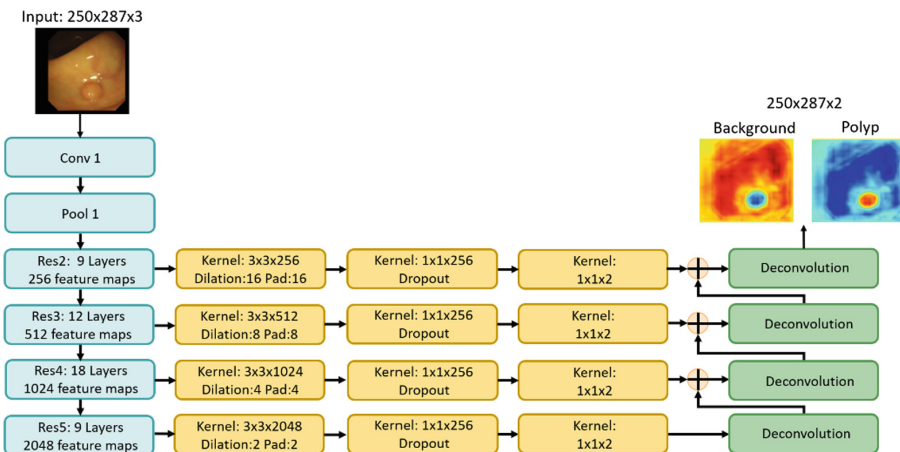
In recent years, it has been noticed that there is a relationship between size of CNN receptive field and the quality of segmentation results. A new layer, called dilation convolution, has been proposed [25] to control the CNN receptive field in a more efficient way. Chen et al. [26] utilised dilation convolution and developed architecture called atrous spatial pyramid pooling (ASPP) to learn the multi-scale features. The ASPP module consists of multiple parallel convolutional layers with different dilations.

In summary, colonoscopy image analysis (including polyp segmentation) is becoming more and more automated and integrated. Deep feature learning and end-to-end architectures are gradually replacing the hand-crafted and deep features operating on a sliding window. Polyp segmentation can be seen as a semantic instance segmentation problem and therefore, a large number of techniques developed in computer vision for generic semantic segmentation could be possibly adopted, providing effective and more accurate methods for polyp segmentation.

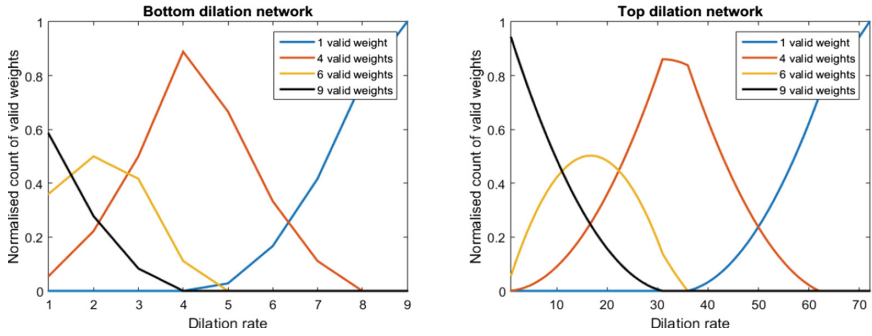
### 3 Method

The full processing pipeline of the proposed methodology is described in [3]. This section provides only the key information necessary for understating of the method evaluation described in the subsequent sections.

The proposed Dilated ResFCN polyp segmentation network is shown in Fig. 2. This architecture is inspired by [20, 26], and the Global Convolutional Network [27]. The proposed FCN consists of three sub-networks performing specific tasks: feature extraction, multi-resolution classification, and fusion (deconvolution). The feature extraction sub-network is based on the ResNet-50 model [28]. The ResNet-50 has been selected, as for the polyp segmentation problem it has showed to provide a reasonable balance between network capacity and required resources. The multi-resolution classification sub-network consists of four parallel paths connected to the outputs from Res2 - Res5. Each such parallel path includes a dilation convolutional layer, which is used to increase the receptive field without increasing computational complexity. The larger receptive fields are needed to access contextual information about polyp neighborhood areas. The dilation rate is determined by the statistics of polyp size in the database used for training. For the lowest resolution path (the bottom path in Fig. 2) the  $3 \times 3$  kernel can only represent a part of most polyps and the  $7 \times 7$  kernel is too large. Therefore,  $5 \times 5$  kernel, corresponds to dilation rate of 2, has been experimentally selected, as it can adequately represent 91% of all polyps in the training dataset. The regions of dilation convolutions should be overlapping and therefore the dilation rates increase with resolution. The dilation rates for sub-nets connected to Res5-Res2 are 2, 4, 8, 16 and the corresponding kernel sizes are 5, 9, 17, and 33. The fusion sub-network, corresponds to the deconvolution layers of FCN model. The segmentation results from each classification sub-network are up-sampled and fused by a bilinear interpolation.



**Fig. 2.** Architecture of the proposed Dilated ResFCN network: feature extraction sub-network shown in blue, multiresolution feature classification sub-network shown in yellow, and fusion sub-network shown in green. (Color figure online)



**Fig. 3.** The number of valid weights in bottom and top dilation networks in Fig. 2. (Color figure online)

Following the methodology described in [29], the number of active kernel weights at the top and bottom paths of the classification subnetwork are shown in Fig. 3. It can be seen, that with the dilation rate too high, the  $3 \times 3$  kernel is effectively being reduced to a  $1 \times 1$  kernel. On the other hand too small dilation rate leads to a small receptive field negatively affecting performance of the network. The selected dilation rates of 2 and 16 respectively for the “bottom” and “top” networks provide compromise with a sufficient number of kernels having 4–9 valid weights.

## 4 Implementation

### 4.1 Dataset

The proposed polyp segmentation method has been developed and evaluated on the data from the 2017 Endoscopic Vision GIANA Polyp Segmentation Challenge [30]. That data consist of Standard Definition (SD) and High Definition (HD) colonoscopy databases. The SD database has two datasets: training dataset, consisting of 300 low resolution, 500-by-574 pixel RGB images with the corresponding ground truth binary images. The images in that training dataset were obtained from 15 video sequences showing different polyps. The SD test dataset consists of 612 images with 288-by-384 pixels resolution. The HD database is composed of independent high-resolution RGB images of 1080-by-1920 pixels. The HD database includes 56 training images (with the corresponding ground truth) and 108 images used for testing. The results reported in this paper are based on a cross-validation approach using the training datasets only. Selected results obtained on the SD test dataset were reported in [3].

### 4.2 Data Augmentation

For the purpose of the method validation the SD and HD training datasets have been combined giving in total 355 training images. The performance of the CNN-based methods relies heavily on the size of training data used. Clearly, a set of 355 training images is very limited, at least from the perspective of a typical training set used in a context of

deep learning. Moreover, some polyp types are not represented in the database, and for some others there are just a few exemplar images available. Therefore, it is necessary to enlarge the training set via data augmentation. Data augmentation is designed to provide more polyp images for CNN training. Although this method cannot generate new polyp types, it can provide additional data samples based on modelling different image acquisition conditions, e.g. illumination, camera position, and colon deformations.

All HD and SD images are rescaled to a common image size (250-by-287 pixels) in such a way that image aspect ratio is preserved. This operation includes random cropping equivalent to image translation augmentation. Subsequently, all images are augmented using four transformations. Specifically, each image is: (i) rotated with the rotation angle randomly selected from  $[0^\circ-360^\circ]$  range, (ii) scaled with the scale factor randomly selected between 0.8 and 1.2, (iii) deformed using a thin plate spline (TPS) model with a fixed  $10 \times 10$  grid and a random displacement of the each grid point with the maximum displacement of 4 pixels, (iv) colour adjusted, using colour jitter, with the Hue, Saturation, Value randomly changed, with the new values drawn from the distributions derived from the original training images [31]. In total after augmentation the training dataset consists of 19,170 images (Fig. 4).

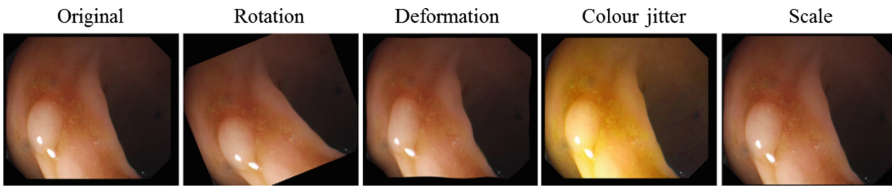


Fig. 4. A sample of augmented images using rotation, local deformation, colour jitter, and scale.

### 4.3 Evaluation Metrics

For a single segmented polyp, Dice coefficient (also known as F1 score), Precision, Recall, and Hausdorff distance are used to compare the similarity between the binary segmentation results and the ground truth. Precision and Recall are standard measures used in a context of binary classification:

$$Precision = \frac{TP}{TP+FP} \quad Recall = \frac{TP}{TP+FN} \tag{1}$$

where:  $TP$ ,  $FN$ , and  $FP$  denotes respectively true positive, false negative and false positive. Precision and Recall could be used as indicators of over- and under- segmentation. Dice coefficient is often used in a context of image segmentation and is defined as:

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{2}$$

Hausdorff Distance is the measure used to determine similarities between the boundaries  $G$  and  $S$  of two objectives. It is defined as:

$$H(G, S) = \max \{ \sup_{x \in G} \inf_{y \in S} d(x, y), \sup_{x \in S} \inf_{y \in G} d(x, y) \} \tag{3}$$

where:  $d(x, y)$  denotes the distance between points  $x$  and  $y$ . The best result of this measure is 0, which means that the shapes of two objectives are completely overlapping.

#### 4.4 Cross-Validation Data

For the purpose of validation original training images are divided into four V1-V4 cross-validation subsets with 56, 96, 97 and 106 images respectively. Following augmentation corresponding sets have 4784, 4832, 4821 and 4733 images for training. Following standard 4-fold validation scheme any three of these subsets are used for training (after image augmentation) and the remaining subset (without augmentation) for validation. Frames extracted from the same video are always in the same validation sub-set, i.e. they are not used for training and validation at same time.

## 5 Results

### 5.1 Comparison with Benchmark Methods

Two reference network architectures FCN8s [20] and ResFCN have been selected as benchmarks for evaluation of the proposed method. Whereas FCN8s is a well known fully convolutional network, the ResFCN is a simplified version of the network from Fig. 2 with the dilation kernels removed from the parallel classification paths. Table 1 lists the results (mean and standard deviation) for all three tested methods and all four-evaluation metrics. As it can be seen from that table the Dilated ResFCN achieves the best mean results for all the four metrics (the highest value for dice, precision, recall and the smallest value for the Hausdorff distance), as well as the smallest standard deviations for all the metrics, demonstrating the stability of the proposed method.

**Table 1.** Mean values and standard deviation obtained for different metrics on 4-fold validation data using FCN8s, ResFCN and Dilated ResFCN.

	Dice		Precision		Recall		Hausdorff	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
FCN8s	0.63	0.11	0.68	0.1	0.65	0.12	193	76
ResFCN	0.71	0.08	0.75	0.07	0.74	0.09	201	110
Dilated ResFCN	<b>0.79</b>	0.08	<b>0.81</b>	0.07	<b>0.81</b>	0.09	<b>54</b>	21

Figure 5 demonstrates results' statistics for all the methods and all the metrics using box-plot, with median represented by the central red line, the 25th and 75th percentiles represented by the bottom and top of each box and the outliers shown as red points. It can be concluded that the proposed method achieves better results than the benchmark methods. For all the metrics the true medians for the proposed method are better, with the 95% confidence, than for the other methods.

Significantly smaller Hausdorff distance measure obtained for the Dilated ResFCN results indicates a better stability of the proposed method with boundaries of segmented polyps better fitting to the ground truth data.

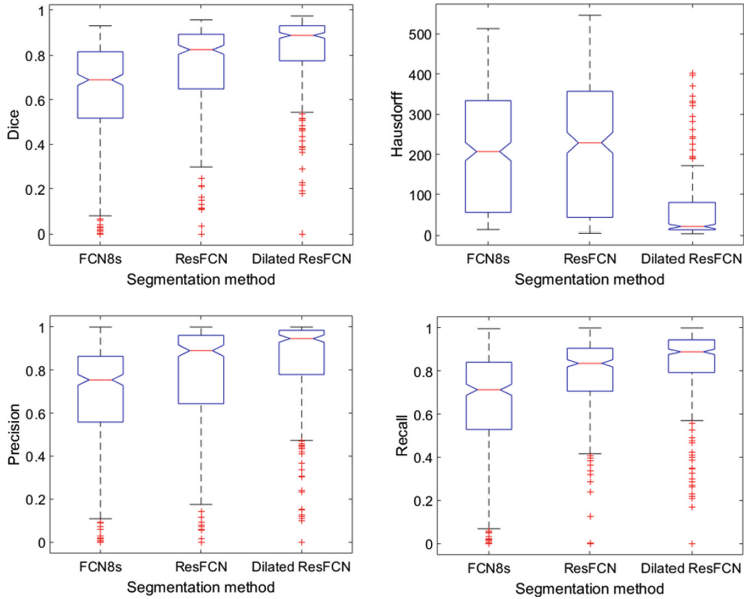


Fig. 5. The box-plot for different evaluation metrics. (Color figure online)

### 5.2 Data Augmentation Ablation Tests

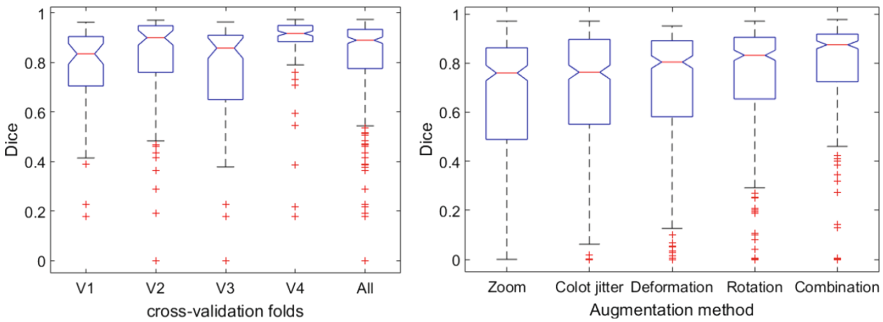
As mentioned above, due to a very small training dataset, the data augmentation is an important step required for a suitable network training. In this section various data augmentations are investigated with the proposed Dilated ResFCN architecture. The result obtained after combining all the augmentations is also presented. Table 2 shows the mean Dice index obtained on each cross-validation subset along the overall mean dice index averaged across the four subsets. It is clear that the rotation seems to be the most informative augmentation method, followed by local deformations and colour jitter. It is also evident that the combination of different augmentation methods improves overall performance. It should be noted that for the “combined” augmentation, the same number of augmented images are used as for any other augmentation method tested.

Table 2. Mean Dice index obtained on 4-fold validation data using Dilated ResFCN network

Network	V1	V2	V3	V4	Mean
Combination	0.7583	<b>0.7420</b>	0.6086	<b>0.8518</b>	<b>0.7402</b>
Rotation	<b>0.7602</b>	0.7146	<b>0.6145</b>	0.8361	0.7314
Deformation	0.6772	0.7058	0.5917	0.7483	0.6807
Color jitter	0.6241	0.6957	0.5696	0.8019	0.6728
Scale	0.6536	0.6368	0.4742	0.7817	0.6366



The box-plots of the augmentation ablation tests are shown in Fig. 6. This confirms the conclusions drawn from the Table 2. Furthermore, it also demonstrates that the combined augmentation significantly improves the segmentation results when compared to any other standalone augmentation, with the real combined method median, being better than any other individual augmentation median with the 95% confidence level. Figure 6 shows also the distribution of the results as a function of the cross-validation folds. It can be seen that the results obtained on the fourth and third folds are respectively the best and worst. A closer examination of these folds reveals that images in the fourth fold are mostly showing larger polyps, whereas images in third fold are mostly depicting small polyps.

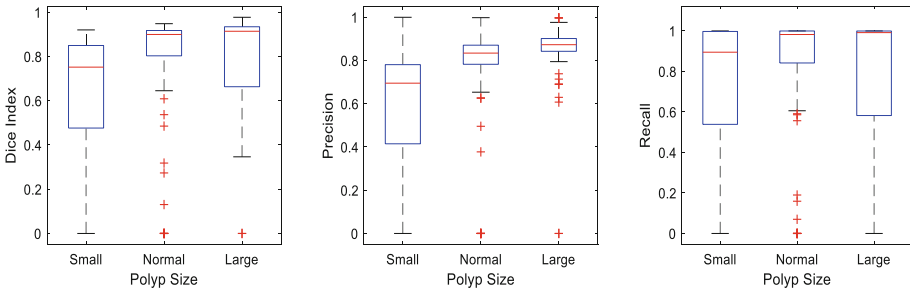


**Fig. 6.** Dice coefficient of Dilated ResFCN for cross-validation folds (left); and data augmentation ablation tests (right).

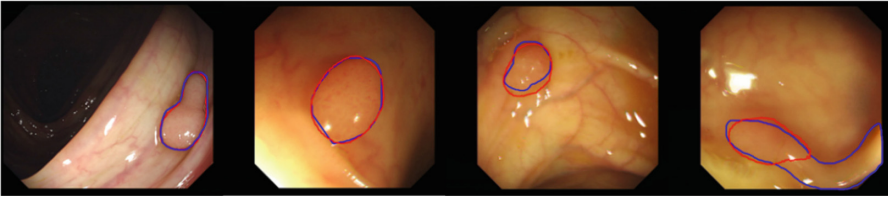
To further investigate the performance of the proposed method as a function of the polyp size, Fig. 7 shows the box-plot showing Dice index as a function of the polyp size. The “Small” and “Large” polyps are defined as having size smaller than the 25<sup>th</sup> and larger than 75<sup>th</sup> percentile of the polyp sizes in the training dataset. The remaining polyps as denoted as “Normal”. The results demonstrate that the small polyps are hardest to segment. However it should be said that the metrics used are biased towards a larger polyps as a relatively small (absolute) over and under segmentation for a small polyps would led to more significant deterioration of the metrics. To combat this effect the authors proposed a secondary network, so called SE-Unet, designed specifically to segment small polyps [3]. The description of that method is though beyond the scope of this paper.

A typical segmentation results obtained using the Dilated ResNet network are shown in Fig. 8 with the blue and red contour representing respectively ground truth and the segmentation results<sup>1</sup>.

<sup>1</sup> It is envisaged that supplementary results of the ongoing research on polyp segmentation will be gradually made available at: <https://github.com/ybguo1/Polyp-Segmentation>.



**Fig. 7.** Validation results obtained for the Dilated ResFCN network grouped as function of polyp size.



**Fig. 8.** Typical results, with Dice index (from left to right) of 0.97, 0.96, 0.71, and 0.69. (Color figure online)

## 6 Conclusion

The paper describes a validation framework for evaluation of the newly proposed Dilated ResFCN network architecture, specifically designed for segmentation of polyps in colonoscopy images. The method has been compared against two benchmark methods: FCN8s and ResFCN. It has been shown that suitably selected dilation kernels can improve performance of polyp segmentation on multiple evaluation metrics. In particular it has been shown that the proposed method matches well the shape of the polyp with the smallest and most consistent value of the Hausdorff distance. Due to a small number of training images, the data augmentation is the key for improving segmentation results. It has been shown that in that case the rotation is the strongest augmentation technique followed by local image deformation and colour jitter. Overall combination of different augmentation techniques has a significant effect on the results. The performance of the method as a function of the polyp size has been also analysed. Although some improvement on segmentation of small polyps have been achieved using architecture not reported in this paper a further improvement is still required, possibly through further optimisation of the dilation spatial pooling. The proposed method has been tested against state-of-the-art at the MICCAI's Endoscopic Vision GIANA Challenges, securing the first place for the SD and HD image segmentation tasks at the 2017 challenge and the second place for the SD images at the 2018 challenge.

**Acknowledgements.** The authors would like to acknowledge the organizers of the Gastrointestinal Image ANALysis – (GIANA) challenges for providing video colonoscopy polyp images.

## References

1. Zhang, Y., Matuszewski, B.J., Shark, L.K., Moore, C.J.: Medical image segmentation using new hybrid level-set method. In: 15th International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics, pp. 71–76. IEEE (2008)
2. Zhang, Y., Matuszewski, B.J.: Multiphase active contour segmentation constrained by evolving medial axes. In: 16th IEEE International Conference on Image Processing, pp. 2993–2996. IEEE (2009)
3. Guo, Y., Matuszewski, B.J.: GIANA polyp segmentation with fully convolutional dilation neural networks. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: GIANA, ISBN 978-989-758-354-4, pp. 632–641 (2019). <https://doi.org/10.5220/0007698806320641>
4. Hwang, S., Oh, J., Tavanapong, W., Wong, J., De Groen, P.C.: Polyp detection in colonoscopy video using elliptical shape feature. In: IEEE International Conference on Image Processing, vol. 2, pp. II–465. IEEE (2007)
5. Gross, S., et al.: Polyp segmentation in NBI colonoscopy. In: Meinzer, H.P., Deserno, T.M., Handels, H., Tolxdorff, T. (eds.) *Bildverarbeitung für die Medizin 2009*, pp. 252–256. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-540-93860-6\\_51](https://doi.org/10.1007/978-3-540-93860-6_51)
6. Breier, M., Gross, S., Behrens, A., Stehle, T., Aach, T.: Active contours for localizing polyps in colonoscopic NBI image data. In: *Medical Imaging 2011: Computer-Aided Diagnosis*, vol. 7963, p. 79632 M. International Society for Optics and Photonics (2011)
7. Breier, M., Gross, S., Behrens, A.: Chan-Vese segmentation of polyps in colonoscopic image data. In: Proceedings of the 15th International Student Conference on Electrical Engineering POSTER, vol. 2011 (2011)
8. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. *Pattern Recogn.* **45**(9), 3166–3182 (2012)
9. Bernal, J., Sánchez, J., Vilarino, F.: Impact of image preprocessing methods on polyp localization in colonoscopy frames. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 7350–7354. IEEE (2013)
10. Bernal, J., Sánchez, F.J., Fernández-esparrach, G., Gil, D., Rodríguez, C., Vilarino, F.: WMDOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* **43**, 99–111 (2015)
11. Bernal, J., et al.: Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. *IEEE Trans. Med. Imaging* **36**(6), 1231–1249 (2017)
12. Tajbakhsh, N., Gurudu, S.R., Liang, J.: A Classification-Enhanced Vote Accumulation Scheme for Detecting Colonic Polyps. In: Yoshida, H., Warfield, S., Vannier, M.W. (eds.) *ABD-MICCAI 2013*. LNCS, vol. 8198, pp. 53–62. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-41083-3\\_7](https://doi.org/10.1007/978-3-642-41083-3_7)
13. Tajbakhsh, N., Chi, C., Gurudu, S.R., Liang, J.: Automatic polyp detection from learned boundaries. In: IEEE 11th International Symposium on Biomedical Imaging, pp. 97–100. IEEE (2014)
14. Tajbakhsh, N., Gurudu, Suryakanth R., Liang, J.: Automatic Polyp Detection Using Global Geometric Constraints and Local Intensity Variation Patterns. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014*. LNCS, vol. 8674, pp. 179–187. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-10470-6\\_23](https://doi.org/10.1007/978-3-319-10470-6_23)

15. Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A., Tzivras, M.: Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. Inf Technol. Biomed.* **7**(3), 141–152 (2003)
16. Iakovidis, D.K., Maroulis, D.E., Karkanis, S.A., Brokos, A.: A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In: 18<sup>th</sup> IEEE Symposium on Computer-Based Medical Systems (CBMS) (2005). <https://doi.org/10.1109/cbms.2005.6>
17. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
19. Ribeiro, E., Uhl, A., Hafner, M.: Colonic polyp classification with convolutional neural networks. In: *IEEE 29th International Symposium on Computer-Based Medical Systems*, pp. 253–258. IEEE (2016)
20. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, Alejandro F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
22. Vázquez, D., et al.: A benchmark for endoluminal scene segmentation of colonoscopy images. *J. Healthc. Eng.* **2017** (2017)
23. Zhang, L., Dolwani, S., Ye, X.: Automated Polyp Segmentation in Colonoscopy Frames Using Fully Convolutional Neural Network and Textons. In: Valdés Hernández, M., González-Castro, V. (eds.) *MIUA 2017. CCIS*, vol. 723, pp. 707–717. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-60964-5\\_62](https://doi.org/10.1007/978-3-319-60964-5_62)
24. Li, Q., et al.: Colorectal polyp segmentation using a fully convolutional neural network. In: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, pp. 1–5. IEEE (2017)
25. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint [arXiv:1511.07122](https://arxiv.org/abs/1511.07122) (2015)
26. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018)
27. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters—improve semantic segmentation by global convolutional network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4353–4361 (2017)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
29. Chen, L.C., Papandreou, G., Schroff, F. and Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017)
30. GIANA Challenge page, <https://endovissub2017-giana.grand-challenge.org/polypsegmentation/>. Accessed 02 Mar 2019
31. Must Know Tips/Tricks in Deep Neural Networks, <http://lamda.nju.edu.cn/weixs/project/CNNTricks/CNNTricks.html>. Accessed 02 Mar 2019