Few-Shot Semantic Segmentation of Wireless Capsule Endoscopy Images

Abhijeet Dhupia, Kevin Raj. P, and Chandra Sekhar Seelamantula

Department of Electrical Engineering, Indian Institute of Science, India
{abhijeetd,kevinraj,css}@iisc.ac.in

Abstract. The main challenge in few-shot learning is the translation of latent representations learned from the support image using a feature extractor to the query image. We propose an end-to-end Bayesian learning framework for few-shot learning, which utilizes the information from both the support and query images. The features extracted from the support (prior) and query image (posterior) are modeled as a multivariate Gaussian Mixture Model (GMM) using an autoencoder coupled with a shallow convolutional neural network. The support and query prototypes are sampled from the learned GMM distribution and fused with the extracted query features to estimate the final segmentation map. Joint optimization of the feature maps and the GMM parameters results in rich feature extraction and robust distribution estimation of the input samples. It also alleviates the network from finding the local optima, strengthening the overall stability of the network. The proposed technique is extensively validated on two publicly available wireless capsule endoscopy datasets, KID-1 consisting of 77 images of 9 different abnormalities & KID-2 consisting of 593 images of 4 different abnormalities proving the efficacy of our technique. The code will be released on GitHub.

Keywords: Few-shot · Segmentation · Gaussian Mixture Models · Wireless Capsule Endoscopy.

1 Introduction

The conventional diagnostic methods for the assessment of the gastrointestinal (GI) tract are exhausting and uncomfortable for the patients; these issues are resolved by the introduction of wireless capsule endoscopy (WCE) by Iddan et al. WCE also increased the reach to the small bowel region, where the possibility of various diseases is higher. A patient must ingest a WCE capsule for 7-8 hours, which transmits the frames back to the receiver resulting in 50,000 to 60,000 frames [1]. It is a daunting task for gastroenterologists to analyze such a vast amount of data per patient. To address the aforementioned issue, there has been a lot of advancement occurring to automate the process.

1.1 Our Contribution

Our proposed work closely aligns with two recent works: Zhang et al. [8] and Yang et al. [7]. The key difference compared to Zhang et al. [8] is that instead of class-wise distribution estimation, we used image-wise distribution as medical images can vary within the same classes and exploited the query image distribution for better translation of latent representations from support to query image. Compared to Yang et al. [7], an end-to-end learning framework is proposed resulting in better feature extraction, further leading to better distribution estimation and performance. Our contribution consists of three folds,

1. We propose a novel Bayesian network for a few-shot segmentation task by utilizing the semantics of both the support and query image.
2. End to end few-shot pipeline is proposed for better feature extraction and estimation of GMM distribution.
3. Compared to previous works in WCE segmentation, we provide an extensive validation on two different public WCE datasets, consisting a total of 11 abnormalities.

2 Proposed Method

2.1 Architecture

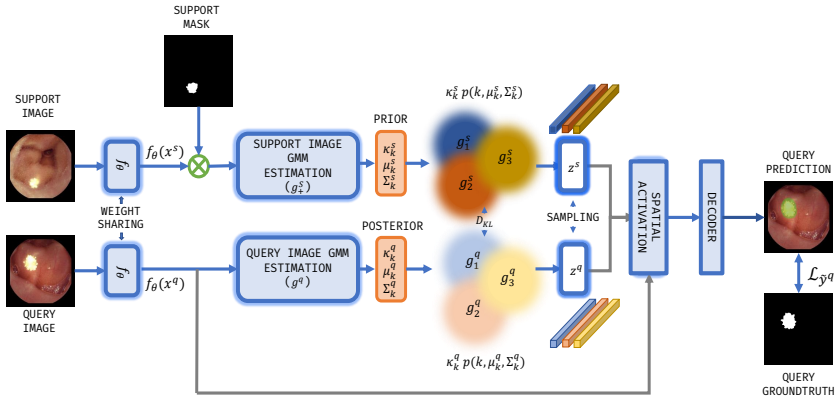

Fig. 1. [Color online] Proposed network consists of two branch weight sharing feature extractor f_θ one for support & other for query image(s). Foreground support features $f_\theta(x_+^s)$ are spatially partitioned using the support mask y^s. Further spatially activated query regions using the sampled latent representations from the estimated GMM are passed through the decoder block resulting in the final segmentation map $\tilde{y}^q$.

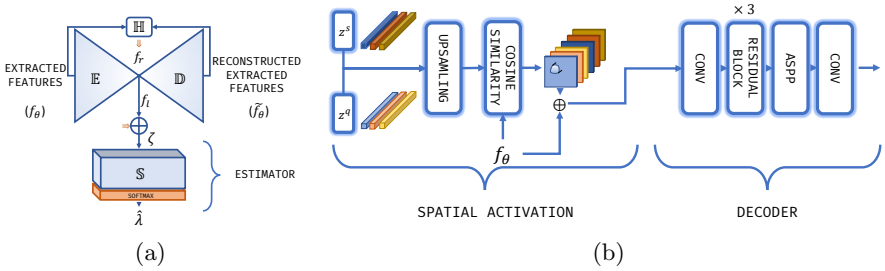

Fig. 2. [Color online] (a) Gaussian Mixture Model comprising of the encoder $\mathbb{E}(\phi_e)$, decoder $\mathbb{D}(\phi_d)$ and a shallow convolutional network $\mathbb{S}(\phi_s)$, (b) Spatial activation and decoder block.

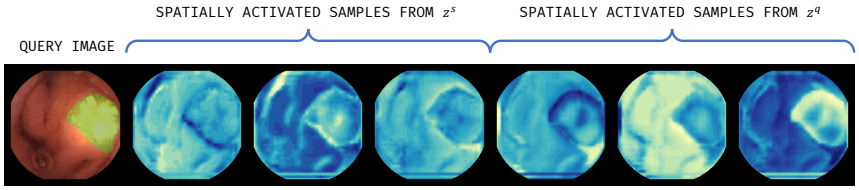

Fig. 3. [Color online] Spatially activated regions of the query image by the sampled latent representation z^s and z^q. Green overlay on the query image indicates the final prediction.

3 Experiments

3.1 Datasets

The proposed technique is validated on two publicly available datasets named KID-1 [4] and KID-2 [5]. KID-1 consists of a total of 77 images containing 9 abnormalities: angioectasias (27), aphthae (5), lymphangiectasia (9), polypoid (6), bleeding (5), chylous (8), stenosis (6), ulcer (9) and, villous oedemas (2) [not considered due to less number of images]. KID-2 consists of 593 images containing four abnormalities: vascular (303), inflammatory (227), polypoid (44) and ampulla-of-vater (19). Polypoid of KID-1 & KID-2 are combined and considered as KID-1. The images from both the datasets are of spatial dimension 360×360, and pixel-wise annotations are used as ground-truth for the few-shot segmentation task. The eight abnormalities in KID-1 are divided into four groups, each group containing two abnormalities. Among the four groups, one group is used for validation, and the rest three groups are used for training. Abnormalities of KID-2 are entirely used for validation.

3.2 Implementation Details

The support and query features are obtained using the DeepLab Resnet-50 [3] architecture, pretrained on Imagenet [6]. The proposed model is trained with four

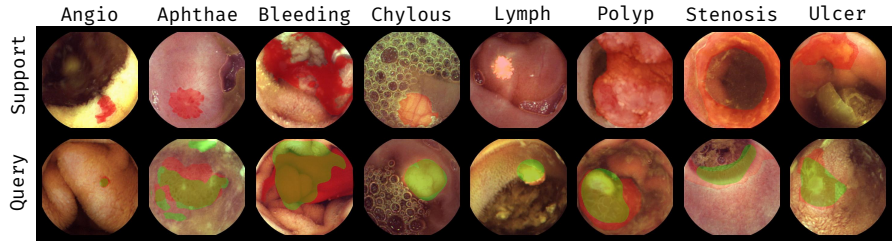

Fig. 4. [Color online] Overlaid segmented results. Green indicates prediction and Red indicates ground-truth.

Table 1. Performance metrics of 6-way 1-shot setting for KID-1 and KID-2 of the proposed technique.

Datasets	Group	Abnormalities	IOU	Dice	Sensitivity	Specificity	Accuracy
KID-1	Group 0	Angioectasia	0.2791	0.3868	0.7867	0.9096	0.907
		Apathe	0.1652	0.2358	0.3446	0.9411	0.8922
	Group 1	Lymphangiectasia	0.5495	0.7061	0.9001	0.9798	0.9743
		Polypoid	0.1627	0.2309	0.2151	0.9635	0.8223
	Group 2	Bleeding	0.3269	0.4767	0.6210	0.7858	0.7478
		Chylous	0.6602	0.7899	0.8130	0.9697	0.942
	Group 3	Stenosis	0.3198	0.4583	0.4133	0.8792	0.7597
		Ulcer	0.4235	0.5322	0.5753	0.9673	0.9435
		Mean	**0.3608**	**0.4770**	**0.5836**	**0.9245**	**0.8736**
KID-2	Group 0	Ampulla-of-vater	0.4593	0.5917	0.8016	0.8997	0.8941
	Group 1	Inflammatory	0.2260	0.3222	0.4673	0.9456	0.8969
	Group 2	Vascular	0.1065	0.1653	0.4290	0.9317	0.9088
		Mean	**0.2639**	**0.3597**	**0.5659**	**0.9256**	**0.8999**

pairs of support and query images per batch by optimizing the final objective function L, using Adam optimizer with a learning rate of $3e-4$. During the training phase, the learning rate is reduced by using cosine decay. The model is trained for 200 epochs with an early stopping, based on the validation dice score with a tolerance of 50 epochs. Due to a fewer number of training images, data augmentation is performed using a library called albumentation [2]. The abnormality classes are randomly chosen and the support-query images from the chosen class are also randomly sampled during the training step.

3.3 Results

Performance of the proposed technique is evaluated by calculating the standard metrics such as IOU, dice, sensitivity, specificity, and accuracy as given in Table 1. The few-shot paradigm is comparatively new and to our best of our knowledge there is no few-shot segmentation of Wireless Capsule Images. For comparison of the proposed technique we implement a recent few-shot technique FPMMs and it's variant FRPMMs [7] and compare our proposed method as given in

Table 2. Comparison of 6-way 1-shot average dice score for KID-1 and KID-2 dataset abnormality wise with other few-shot techniques.

Datasets	Abnormalities \ Methods	FPMM	FRPMM	VGMM (ours)
KID-1	Angioectasia	0.3310	0.3596	**0.3868**
	Apathe	**0.3893**	0.1285	0.2358
	Lymphangiectasia	0.2600	0.6360	**0.7061**
	Polypoid	0.2177	**0.4239**	0.2309
	Bleeding	0.5393	**0.5945**	0.4767
	Chylous	0.6163	0.5112	**0.7899**
	Stenosis	0.3981	0.2626	**0.4583**
	Ulcer	0.4842	0.3651	**0.5322**
	Mean	0.4044	0.4101	**0.4770**
KID-2	Ampulla-of-vater	0.4358	0.3771	**0.5917**
	Inflammatory	0.2552	0.2968	**0.3222**
	Vascular	**0.1738**	0.1709	0.1653
	Mean	0.2882	0.2816	**0.3597**

Table 3. Comparison of average dice score of 6-way 1-shot and 6-way 5-shot settings.

Datasets	k-shot	FPMMs	FRPMMs	VGMM(ours)
KID-1	1-shot	0.4044	0.4101	**0.4770**
	5-shot	0.4069	0.4255	**0.4669**
KID-2	1-shot	0.2882	0.2816	**0.3597**
	5-shot	0.2941	0.3073	**0.3631**

Table 2. Our technique achieves an increase of 7.25% for 1-shot and 4.86% for 5-shot in average dice score compared to [7]. It also achieves superior performance for seven abnormalities proving the efficacy of the proposed technique. Moreover, the performance of our proposed 1-shot technique is performing better than the 5-shot of [7]. Comparison of k-shot setting is given in Table 3.2. In contrary, performance gain of the proposed method between 1-shot and 5-shot is quite negligible, which can considered as a task for future works. The final segmentation results of our proposed technique is given in Fig 4.

4 Conclusion

As a consequence of the lack of data samples, traditional semantic segmentation of WCE can easily lead to over-fitting, and the generalization capability of the network is significantly compromised. To address the issues mentioned above, we have proposed a few-shot based network. There are no previous works related to the few-shot semantic segmentation of WCE images to the best of our knowledge. Our work is extensively validated on two public datasets, KID-1 and KID-2, containing 11 abnormalities in total, achieving a 7.25% for 1-shot and 4.86% for

5-shot, increase in average dice and also performs better for 7 abnormalities, thus proving the generalization capability and efficacy of the proposed technique.

References

1. Adler, D.G., Gostout, C.J.: Wireless capsule endoscopy. Hospital Physician **39**(5), 14–22 (2003)
2. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information **11**(2), 125 (2020)
3. 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 transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
4. Iakovidis, D.K., Koulaouzidis, A.: Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. Gastrointestinal endoscopy **80**(5), 877–883 (2014)
5. Koulaouzidis, A., Iakovidis, D.K., Yung, D.E., Rondonotti, E., Kopylov, U., Plevris, J.N., Toth, E., Eliakim, A., Johansson, G.W., Marlicz, W., et al.: Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. Endoscopy international open **5**(6), E477 (2017)
6. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)
7. Yang, B., Liu, C., Li, B., Jiao, J., Ye, Q.: Prototype mixture models for few-shot semantic segmentation. In: European Conference on Computer Vision. pp. 763–778. Springer (2020)
8. Zhang, J., Zhao, C., Ni, B., Xu, M., Yang, X.: Variational few-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1685–1694 (2019)