Few-Shot Segmentation of Microscopy Images Using Gaussian Process

Surojit Saha[1(✉)], Ouk Choi[2], and Ross Whitaker[1]

[1] Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, USA
{surojit,whitaker}@cs.utah.edu
[2] Department of Electronics Engineering, Incheon National University, Incheon, Republic of Korea
ouk.choi@inu.ac.kr

Abstract. Few-shot segmentation has received recent attention because of its promise to segment images containing novel classes based on a handful of annotated examples. Few-shot-based machine learning methods build generic and adaptable models that can quickly learn new tasks. This approach finds potential application in many scenarios that do not benefit from large repositories of labeled data, which strongly impacts the performance of the existing data-driven deep-learning algorithms. This paper presents a few-shot segmentation method for microscopy images that combines a neural-network architecture with a Gaussian-process (GP) regression. The GP regression is used in the latent space of an autoencoder-based segmentation model to learn the distribution of functions from the encoded image representations to the corresponding representation of the segmentation masks in the support set. This regression analysis serves as the prior for predicting the segmentation mask for the query image. The rich latent representation built by the GP using examples in the support set significantly impacts the performance of the segmentation model, demonstrated by extensive experimental evaluation.

Keywords: Few-shot segmentation · Gaussian process · Microscopy images

1 Introduction

Recent advances in deep learning technology have accelerated the automation of medical image segmentation. Deep learning models deliver good results with a sufficiently large training data set consisting of pairs of medical images and their manual annotations. However, obtaining sufficient training data for a specific domain is challenging. Algorithms that quickly learn the target task from a

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16961-8_10.

Y. Huo et al. (Eds.): MOVI 2022, LNCS 13578, pp. 94–104, 2022.
https://doi.org/10.1007/978-3-031-16961-8_10

limited number of labeled examples without overfitting address this issue. This learning paradigm is known as the few-shot learning.

Microscopy image is an important modality in the field of medical imagining. Segmentation of the microscopy image for nuclei, mitochondria, and cells [1,3,7,18,21,32] enables scientists to quantitatively analyze cell counts, size, and shape over time. Availability of annotated examples in large numbers has always been a challenge in this domain. Supervised methods apply data augmentation [21,32] or rely on synthetic data [32] in the scenarios of limited data. The use of semi-supervised and unsupervised methods [7,13,18,30] are potential workarounds to deal with limited labeled data.

Although a domain-specific data set may not be sufficient to train a deep neural network, we can utilize a variety of data sets of different but related domains to enhance the generalization power of the models. Few-shot learning is a meta-learning method based on this principle. Few-shot learning is a two-stage algorithm, where in the first stage, known as *meta-training*, the model is trained with a large amount of annotated examples from domains related to but different from the target domain for solving tasks different from the target task. For example, if the target task is to learn the segmentation of spleens from MRI images, the model can undergo meta-training for segmentation of different organs (other than the spleen) using both MRI and CT images. Then there is a *fine-tuning* stage where very few examples from the target task are used to update the model parameters for better performance. Fine-tuning is not essential on every occasion, as many algorithms developed for classification and segmentation of natural images [9,20,23,28] can achieve reasonably good performance with the meta-trained models. However, the fine-tuning stage has proved to be beneficial for medical imaging data [3,15,24]. Few-shot learning algorithms can be broadly categorized as model-based (black-box) [20,22], metric-based (non-parametric) [11,25,28] or optimization-based methods [5,17]. Training for all these methods is performed using *support* and *query* examples, except for optimization-based methods [5,17] which do not require *query* examples.

Among the few-shot image segmentation methods, [3,9,23,27,29,31,34,35], the method of [9], DGPNet, which relies on Gaussian processes in the latent space, is particularly promising, because of the high degree of adaptivity (to the number of training examples) and the estimates of uncertainty provided by the GP regression [19,26,33]. Chen et al. [2] introduced a GP in in their formulation of the scalable functional Bayesian neural network for estimation of the uncertainty in the segmentation output, which is critical in the medical imaging community.

In this paper, we present a few-shot segmentation method for microscopy images using a GP in the latent space, named as the GP-UNet. The GP-UNet, illustrated in Fig. 1, uses a GP in the latent space of the UNet-based backbone network. This is motivated by the method introduced by Johnander [9]. Our choice of the U-Net [21] as the backbone network is due to its impressive performance in segmenting medical images. The GP-UNet, similar to other medical imaging applications [3,15,24], also does fine-tuning on examples from the target tasks. Through extensive empirical evaluation, we showcase the strength of the GP-UNet over other competing methods.

2 Method

For the discussion of few-shot learning, we denote the source data set containing N different classes as $D_{source} = \{D^1_{source}, D^2_{source}, .., D^N_{source}\}$, where $D^c_{source} = \{(x, y)^i\}_{i=1}^{|D^c_{source}|}$ are the annotated examples for class c. Likewise, the target data set containing M different classes is defined as $D_{target} = \{D^1_{target}, D^2_{target}, .., D^M_{target}\}$, where $D^d_{target} = \{(x, y)^i\}_{i=1}^{|D^d_{target}|}$ are the annotated examples for class d. The source and target data contain examples from completely different data sets or types with no overlap. The *support* set is produced following the N-way K-shot structure, where N and K represent the number of classes and examples (or shots) from each class, respectively. For the segmentation tasks in this paper, we use $N = 1$ as in [23], i.e., for a task, we select a class at random and sample K examples from the selected class as the *support* set and a single example from the same class as the *query* set.

For a given task in the K-shot segmentation algorithm, the *support* set is denoted as $S = \{(x_s, y_s)^i\}_{i=1}^{K}$, where $x^i_s \in R^{H\times W\times 3}$ is a *support* image and $y^i_s \in \{0, 1\}^{H\times W}$ is its corresponding mask. The *query set* is denoted as $Q = \{(x_q, y_q)\}$, where $x_q \in R^{H\times W\times 3}$ is the *query* image and $y_q \in \{0, 1\}^{H\times W}$, its corresponding mask.

The GP-UNet has three trainable blocks, the image encoder (IE), mask encoder (ME), and mask decoder (D), shown in Fig. 1. The image encoder is used to encode both the *support* and *query* images. The mask encoder is used for only the *support* masks because the *query* image mask is treated as unknown. The mask decoder predicts the segmentation mask for the *query* image using the output of the GP regression and skip connections.

The encoded representation of a *support* image and its corresponding mask, denoted as z_s and $z^{'}_s$, respectively, are defined as follows:

$$z_s = IE(x_s) \in R^{H^{'}\times W^{'}\times C} \text{ and } z^{'}_s = ME(y_s) \in R^{H^{'}\times W^{'}\times C^{'}} \tag{1}$$

Similarly, the encoded representation of a *query* image is defined as follows:

$$z_q = IE(x_q) \in R^{H^{'}\times W^{'}\times C} \tag{2}$$

The GP regression used only in the bottleneck layer of the GP-UNet learns a distribution of functions that regress from the encoded image space $\in R^C$ to the encoded mask space $\in R^{C^{'}}$. For an encoded query input, z_q, parameters of the posterior distribution, $\mu_{q|S}$(mean) and $\Sigma_{q|S}$ (co-variance), computed using the encoded representation of all examples in the support set $(z_S, z^{'}_S)$ as the prior, are defined as follows [19]:

$$\mu_{q|S} = K^T_{Sq}(K_{SS} + \sigma^2_z\mathbf{I})^{-1}z^{'}_S \in R^{H^{'}\times W^{'}\times C^{'}} \tag{3}$$

$$\Sigma_{q|S} = K_{qq} - K^T_{Sq}(K_{SS} + \sigma^2_z\mathbf{I})^{-1}K_{Sq} \in R^{H^{'}W^{'}\times H^{'}W^{'}}, \tag{4}$$

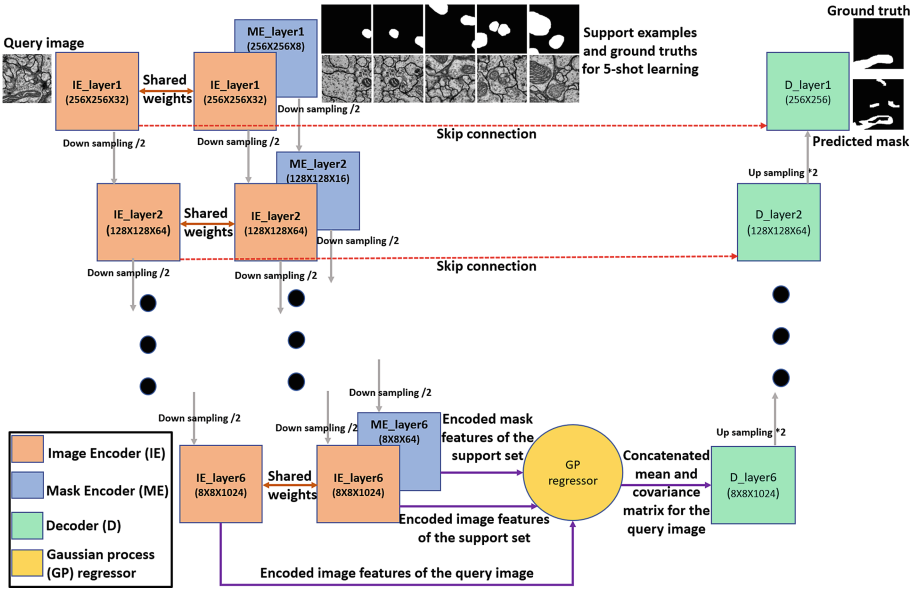

Fig. 1. Architecture of the GP-UNet model for few-shot segmentation of microscopy images. Shared weights indicate the use of the same IE for both source and query inputs. Each block has the details of the output size.

where K_{SS}, K_{qq}, and K_{Sq} are the co-variance matrices computed using $KH^{'}W^{'}$ examples from z_S and $H^{'}W^{'}$ examples from z_q in the encoded space. The noise in the labeled data is represented by σ_z. Considering, the influence of the spatial neighbors$(O \times O)$, $\Sigma_{q|S}$ is reshaped to $R^{H^{'} \times W^{'} \times (O^2)}$ as in [9]. Each entry in the co-variance matrices $\in R$ is a measure of similarity between the pairs of encoded points $\in R^C$ which is computed using the widely used squared exponential kernel, defined as follows:

$$Kernel_{SE}(z_1, z_2) = \sigma^2 \exp \frac{-\|z_2 - z_1\|_2^2}{2l^2} \tag{5}$$

In Eq. 5, l and σ represent the kernel bandwidth and scaling factor, respectively. The decoder predicts the final mask $\in R^{H \times W}$ using the mean $(\mu_{q|S})$ and co-variance $(\Sigma_{q|S})$ of the posterior distribution (concatenated along the channel axis $\in R^{H^{'} \times W^{'} \times (C^{'} + O^2)}$) and encoding of the query image at different levels obtained through the skip connections. The estimated mask for the query image and its ground truth is used for computing the weighted cross entropy loss to update the trainable model parameters using stochastic gradient descent.

3 Experimental Setup

In this section, we discuss the data set and its use in training and evaluating the competing methods for a $K-$shot experimental setup. In addition, we discuss

the deep neural architecture used in the GP-UNet along with the related hyper-parameters for training the models.

3.1 Training and Evaluation

To evaluate the few-shot segmentation methods, we use the data set of microscopy images curated by Dawoud et al. [3]. This publicly available data set contains examples from five different types of microscopy images, B5 [12], B39 [12], TNBC [16], EM [14], and ssTEM [6], with different image resolutions. Table 1 of the supplementary has the details. Microscopy images are usually quite large in size. Therefore, using patches extracted from original images for training deep neural networks is a standard practice [1,10]. Moreover, the following reasons advocate the use of patches for training: computational benefits and managing the batch-size/image-size trade off, relatively stationary statistics (less variation between patches, unlike many natural images), and different sizes/resolutions of the input images. The strategy for generating patches (of size 256×256) used in [3] is also followed in our implementation for *meta-training* and *fine-tuning* stage.

For our experiments, out of five classes in the data set, four classes define the source set, D_{source}, and the remaining one class is considered as the target set, D_{target}. In the meta-training stage of a $K-$shot experiment, the model is trained on patches extracted from all the images in D_{source}, following the strategy outlined in Sect. 2. In the fine-tuning stage, the model is trained on patches extracted only from K images in D_{target} (containing N images) for adapting it to the target task, following the same training procedure. Average number of patches extracted from K images in D_{target} (for the fine-tuning operation) are reported in Table 4 for all the target classes in the data set. Typical values of K used in our experiments are $K \in \{1, 5, 10\}$. However, for the evaluation of a model, full-sized images are used instead of patches. The use of patches in the training stages (meta-training and fine-tuning) and full-sized images for evaluation [10] is a typical setup used in the study of microscopy images, unlike natural images. The architecture of the GP-UNet is capable of handling different settings used in the training and evaluation phase.

For evaluation of a model for the K-shot setup, K full-sized images in D_{target} (containing N images) form the *support* set (which are also used in the fine-tuning stage) and prediction of the segmentation mask is done for the remaining $N - K$ examples (*query* set) in D_{target}. We use mean intersection over union (mIoU) for evaluating the performance (higher is better). For statistical significance of the mIoU evaluated for a target class, 10 different sets of K examples are constructed.

3.2 Implementation Details

The U-Net [21] architecture adapted for the image encoder and decoder in the GP-UNet has 6 layers in the encoder-decoder structure (increased from 5 layers in the U-Net [21]) with 32 and 1024 channels in the first convolution and bottleneck layer, respectively. The channel depth of the mask encoder for the first double

Table 1. Performance of the GPUNet with and without fine-tuning in terms of the mIoU.

Target class	1−shot		5−shot	
	Without fine-tuning	With fine-tuning	Without fine-tuning	With fine-tuning
B5	0.460 ± 0.128	$\mathbf{0.842 \pm 0.139}$	0.232 ± 0.000	$\mathbf{0.884 \pm 0.022}$
B39	0.635 ± 0.000	$\mathbf{0.870 \pm 0.052}$	0.571 ± 0.001	$\mathbf{0.907 \pm 0.006}$
TNBC	0.140 ± 0.002	$\mathbf{0.370 \pm 0.060}$	0.133 ± 0.004	$\mathbf{0.455 \pm 0.017}$
EM	0.246 ± 0.000	$\mathbf{0.736 \pm 0.041}$	0.374 ± 0.001	$\mathbf{0.857 \pm 0.007}$
ssTEM	0.265 ± 0.002	$\mathbf{0.741 \pm 0.033}$	0.333 ± 0.007	$\mathbf{0.838 \pm 0.007}$

Table 2. Comparison of the GP-UNet with DGPNet+ initialized randomly or with pre-trained model parameters in terms of the mIoU.

Target class	1−shot			5−shot		
	DGPNet+ (random)	DGPNet+ (pre-trained)	GP-UNet	DGPNet+ (random)	DGPNet+ (pre-trained)	GP-UNet
B5	0.392 ± 0.102	0.496 ± 0.167	$\mathbf{0.842 \pm 0.139}$	0.320 ± 0.025	0.452 ± 0.044	$\mathbf{0.884 \pm 0.028}$
B39	0.255 ± 0.094	0.379 ± 0.144	$\mathbf{0.870 \pm 0.052}$	0.313 ± 0.054	0.751 ± 0.022	$\mathbf{0.907 \pm 0.006}$
TNBC	0.150 ± 0.063	0.200 ± 0.112	$\mathbf{0.370 \pm 0.060}$	0.190 ± 0.037	0.313 ± 0.042	$\mathbf{0.455 \pm 0.017}$
EM	0.015 ± 0.013	0.647 ± 0.025	$\mathbf{0.736 \pm 0.041}$	0.006 ± 0.004	0.756 ± 0.012	$\mathbf{0.857 \pm 0.007}$
ssTEM	0.028 ± 0.010	0.641 ± 0.069	$\mathbf{0.741 \pm 0.033}$	0.025 ± 0.011	0.772 ± 0.026	$\mathbf{0.838 \pm 0.007}$

convolution block is 8, which is doubled in successive layers with the maximum channel depth of 64 in the bottleneck layer. We replaced the max pool layers in the image and mask encoder, used for down sampling, with strided convolution layers. In the *meta-training* stage, we trained the models for 20 epochs using the Adam optimizer with the learning rate of $1e-03$ (weight decay $= 5e-04$) and a batch size of 8. In the *fine-tuning* stage, we set the learning rate to $1e-04$ (weight decay $= 5e-04$), and train the model for 20 epochs with a batch size of 8. Since there is a chance of over-fitting in the fine-tuning stage, we use the model with the best validation loss for evaluation. The hyper-parameters of the GP used in all the experiments are as follows: $\sigma_z = 0.1, \sigma = 1, l = \sqrt{C}$, and $O = 5$. Table 2 of the supplementary has additional implementation details.

4 Results

In this section, we first study the effect of fine-tuning on the performance of GP-UNet. Results reported in Table 1 for 1−, and 5−shot segmentation tasks clearly demonstrates the effectiveness of the fine-tuning operation. This observation is consistent with other few-shot applications in the medical imaging community [3,15,24]. We observe similar behavior in the performance of the DGPNet [9] (refer to results reported in Table 3 of the supplementary), which in its original formulation did not adopt the refinement policy. Hence, we consider the DGPNet [9] with *fine-tuning* stage as a potential method for segmentation of microscopy images, and named it as the DGPNet+.

Table 3. Comparison of the GP-UNet with GBMLCS [3] in terms of the mIoU.

Target	1−shot		5−shot		10−shot	
class	GBMLCS*	GP-UNet	GBMLCS*	GP-UNet	GBMLCS*	GP-UNet
B5	**0.85**	0.842 ± 0.139	**0.90**	0.884 ± 0.028	**0.95**	0.917 ± 0.016
B39	0.85	$\mathbf{0.870 \pm 0.052}$	0.88	$\mathbf{0.907 \pm 0.006}$	0.90	$\mathbf{0.919 \pm 0.002}$
TNBC	**0.41**	0.370 ± 0.060	**0.47**	0.455 ± 0.017	**0.52**	0.512 ± 0.017
EM	0.56	$\mathbf{0.736 \pm 0.041}$	0.65	$\mathbf{0.857 \pm 0.007}$	0.70	$\mathbf{0.869 \pm 0.006}$
ssTEM	0.47	$\mathbf{0.741 \pm 0.033}$	0.65	$\mathbf{0.838 \pm 0.007}$	0.70	$\mathbf{0.847 \pm 0.013}$

We use the ResNet [8], pretrained on ImageNet data set [4], as the backbone for the DGPNet+. To evaluate its effect, we initialized the ResNet parameters randomly, similar to the other blocks of the DGPNet+. From the results reported in Table 2 for 1− and 5−shot setup, we can infer that random initialization has adverse effects on the performance of the DGPNet+.

We can see in Table 2 that the GP-UNet outperforms DGPNet+ in all the scenarios. Therefore, though DGPNet+ and GP-UNet both use the GP regression in the latent space, the backbone architecture of the GP-UNet is a better fit for the segmentation of microscopy images.

Comparison of the GP-UNet with GBMLCS is reported in Table 3. For the GBMLCS [3], approximate mIoU scores (indicated by $*$) for different experimental setups are obtained from the bar-graphs of the UNet-based model. For a typical setup, say 5−shot learning for the target class, EM, we have considered the mIoU score of the best performing model out of all the regularized few-shot learning models proposed in [3]. For all but a couple of the target classes, the GP-UNet outperforms GBMLCS in all scenarios, with significant improvement observed for the EM and ssTEM classes. Bar graphs in Fig. 2(a) that report the average mIoU over all the shots for the target classes in the data set corroborate our observation. Figure 2(b), summarizing the results for different shots, shows that the performance of the GP-UNet is better than GBMLCS in all the scenarios, with an average overall gain of $+\mathbf{7.4}$ mIoU. Poor performance for the target class, TNBC [16], could be attributed to the high variability in the data. Moreover, we did not do any post-processing as in [16]. Figure 3 shows the segmentation masks predicted by the GP-UNet for the 5−shot experiment. The supplementary material has the prediction of the segmentation masks for 1−shot and 10−shot learning.

In an ablation study, we removed the GP from the latent space of the GP-UNet. This represents a model similar to the encoder-decoder architecture used in the vanilla U-Net [21] with skip connections where we perform transfer learning, referred to herein as the TL-UNet. For a K-shot setup, we train the network on patches extracted from all the images in D_{source} and then *fine-tune* the model on the patches extracted from K images in D_{target}. From the results reported in

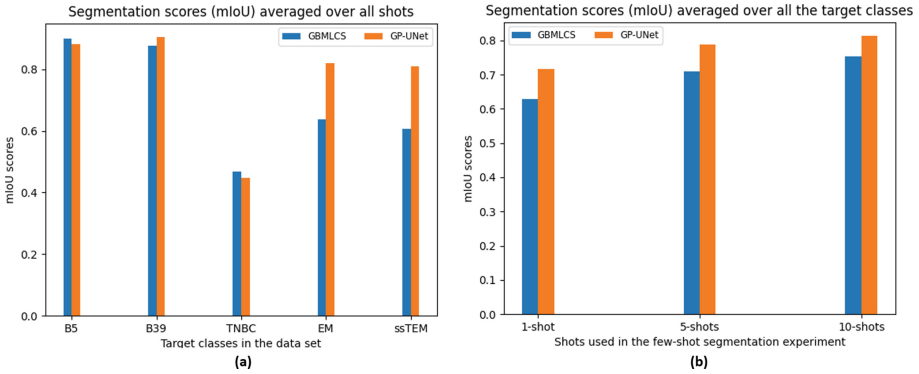

Fig. 2. Comparison of the GP-UNet with GBMLCS (a) segmentation scores (mIoU) averaged over 1−, 5−, and 10−shot experiments (b) segmentation scores (mIoU) averaged over all the target classes.

Table 4. Comparison of the GP-UNet with TL-UNet in terms of the mIoU. The average number of patches (over 10 different sets) used in the *fine-tuning* stage is reported under the column Patches.

Target	1−shot			5−shot		
class	Patches	TL-UNet	GP-UNet	Patches	TL-UNet	GP-UNet
B5	135	$\mathbf{0.843 \pm 0.109}$	0.842 ± 0.139	660	0.879 ± 0.032	$\mathbf{0.884 \pm 0.028}$
B39	102	0.850 ± 0.090	$\mathbf{0.870 \pm 0.052}$	495	$\mathbf{0.908 \pm 0.006}$	0.907 ± 0.006
TNBC	100	0.367 ± 0.060	$\mathbf{0.370 \pm 0.060}$	490	$\mathbf{0.469 \pm 0.015}$	0.455 ± 0.017
EM	390	$\mathbf{0.747 \pm 0.034}$	0.736 ± 0.041	1910	$\mathbf{0.857 \pm 0.007}$	$\mathbf{0.857 \pm 0.007}$
ssTEM	525	0.726 ± 0.023	$\mathbf{0.741 \pm 0.033}$	2680	0.836 ± 0.006	$\mathbf{0.838 \pm 0.007}$

Table 4 for 1−shot and 5−shot setup, we can say that the GP-UNet mostly outperforms this conventional approach when the amount of data in the target set (D_{target}) is limited (i.e., fewer shots or fewer patches for fine-tuning). However, the difference in performance decreases with increasing training data. This observation is consistent with [3]. Comparable performance by the TL-UNet could be attributed to the sufficient number of patches used in the fine-tuning stage. The average number of patches extracted from K full-sized images in D_{target} (used in the *fine-tuning* stage) are reported in Table 4.

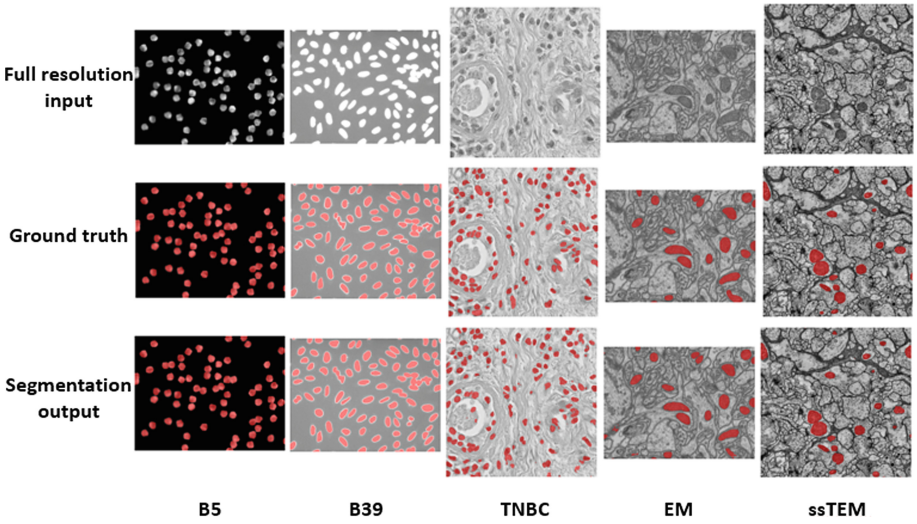

Fig. 3. Segmentation output of the 5−shot GP-UNet for different target classes.

5 Conclusion

In this paper, we present a few-shot segmentation method for microscopy images using a GP in the latent space, GP-UNet. We demonstrate that the *fine-tuning* stage of the few-shot learning framework is beneficial for microscopy images. The GP-UNet outperforms the DGPNet+ in all scenarios. The overall performance of the GP-UNet is better than the GBMLCS. The ablation study manifests the strength of the GP-UNet to produce compelling results in scenarios having a limited amount of data. The GP-UNet sets a new state-of-the-art for the few-shot segmentation of the complex target classes, EM and ssTEM, in the microscopy image data set [3], considered in this work. We plan to investigate more microscopy image data sets and extend this method to other challenging modalities in medical imaging.

References

1. Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W., Rusu, M.: A deep learning-based algorithm for 2-d cell segmentation in microscopy images. BMC Bioinform. **19**(365), 1050–1065 (2018)
2. Chen, X., Zhao, Y., Liu, C.: Medical image segmentation using scalable functional variational Bayesian neural networks with gaussian processes. Neurocomputing **500**, 58–72 (2022)
3. Dawoud, Y., Hornauer, J., Carneiro, G., Belagiannis, V.: Few-shot microscopy image cell segmentation. In: Dong, Y., Ifrim, G., Mladenić, D., Saunders, C., Van Hoecke, S. (eds.) ECML PKDD 2020. LNCS (LNAI), vol. 12461, pp. 139–154. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67670-4_9

4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning, vol. 70, pp. 1126–1135 (2017)
6. Gerhard, S., Funke, J., Martel, J., Cardona, A., Fetter, R.: Segmented anisotropic ssTEM dataset of neural tissue. In: figshare (2013)
7. Han, L., Yin, Z.: Unsupervised network learning for cell segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 282–292. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_27
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Johnander, J., Edstedt, J., Felsberg, M., Khan, F.S., Danelljan, M.: Dense Gaussian processes for few-shot segmentation (2021)
10. Kassim, Y.M., Glinskii, O.V., Glinsky, V.V., Huxley, V.H., Palaniappan, K.: Patch-based semantic segmentation for detecting arterioles and venules in epifluorescence imagery. In: IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–5 (2018)
11. Koch, G.: Siamese neural networks for one-shot image recognition. Master's thesis, University of Toronto (2015)
12. Lehmussola, A., Ruusuvuori, P., Selinummi, J., Huttunen, H., Yli-Harja, O.: Computational framework for simulating fluorescence microscope images with cell populations. IEEE Trans. Med. Imaging **26**(7), 1010–1016 (2007)
13. Liu, D., et al.: Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4243–4252 (2020)
14. Lucchi, A., Li, Y., Fua, P.: Learning for structured prediction using approximate subgradient descent with working sets. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1987–1994. IEEE (2013)
15. Mahajan, K., Sharma, M., Vig, L.: Meta-dermdiagnosis: few-shot skin disease identification using meta-learning. In: Computer Vision and Pattern Recognition Workshops. IEEE (2020)
16. Naylor, P., Laé, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Trans. Med. Imaging **38**(2), 448–459 (2019)
17. Nichol, A., Achiam, J., Schulman, J.: On first-order meta-learning algorithms (2018)
18. Nishimura, K., Ker, D.F.E., Bise, R.: Weakly supervised cell instance segmentation by propagating from detection response. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 649–657. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_72
19. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press, Cambridge (2006)
20. Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: International Conference on Learning Representations (2017)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

22. Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., Lillicrap, T.: Meta-learning with memory-augmented neural networks. In: International Conference on Machine Learning (2016)
23. Shaban, A., Shray, Liu, B.Z., Essa, I., Boots, B.: One-shot learning for semantic segmentation. In: British Machine Vision Conference. BMVA Press (2017)
24. Singh, R., Bharti, V., Purohit, V., Kumar, A., Singh, A.K., Singh, S.K.: MetaMed: few-shot medical image classification using gradient-based meta-learning. Pattern Recogn. **44**(2), 1050–1065 (2021)
25. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
26. Snell, J., Zemel, R.S.: Bayesian few-shot classification with one-vs-each pólya-gamma augmented gaussian processes. In: International Conference on Learning Representations (2021)
27. Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R., Jia, J.: Prior guided feature enrichment network for few-shot segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **44**(2), 1050–1065 (2022)
28. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems, vol. 29 (2016)
29. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: PANet: Few-shot image semantic segmentation with prototype alignment. In: IEEE/CVF International Conference on Computer Vision, pp. 9196–9205 (2019)
30. Wu, H., Wang, Z., Song, Y., Yang, L., Qin, J.: Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11666–11675 (2022)
31. Xie, G.S., Liu, J., Xiong, H., Shao, L.: Scale-aware graph neural network for few-shot semantic segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5471–5480 (2021)
32. Xie, W., Noble, J.A., Zisserman, A.: Microscopy cell counting and detection with fully convolutional regression networks. Comput. Methods Biomech. Biomed. Eng. Imaging Visual. **6**(3), 283–292 (2018)
33. Ze, W., Zichen, M., Xiantong, Z., Qiang, Q.: Learning to learn dense gaussian processes for few-shot learning. In: Advances in Neural Information Processing Systems (2021)
34. Zhang, C., Lin, G., Liu, F., Guo, J., Wu, Q., Yao, R.: Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation. In: IEEE/CVF International Conference on Computer Vision, pp. 9586–9594 (2019)
35. Zhang, C., Lin, G., Liu, F., Yao, R., Shen, C.: CANet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5212–5221 (2019)

Supplementary: Few-shot segmentation of microscopy images using Gaussian process

Surojit Saha[1], Ouk Choi[2], and Ross Whitaker[1]

[1] Scientific Computing and Imaging Institute, University of Utah, Salt Lake City USA
{surojit,whitaker}@cs.utah.edu
[2] Dept. of Electronics Engineering, Incheon National University, Republic of Korea
ouk.choi@inu.ac.kr

Table 1: Description of the data set used for the study of the few-shot segmentation of microscopy images.

Target class	B5	B39	TNBC	EM	ssTEM
Cell type	Synthetic strain	Nuclei	Nuclei	Mitochondria	Mitochondria
Resolution	696X520	696X520	512X512	768X1024	1024X1024
Count	1200	200	50	165	20

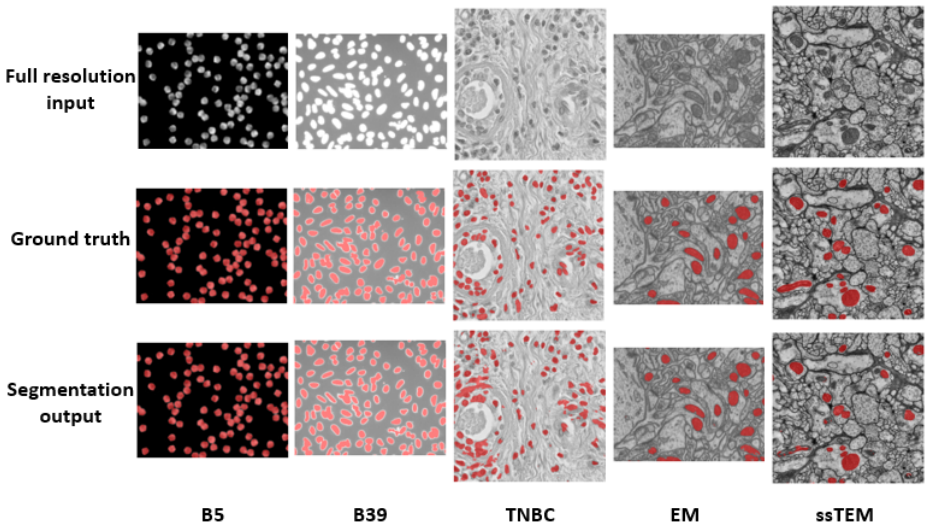

Fig. 1: Segmentation output of the 1−shot GP-UNet for different target classes.

Table 2: Implementation details of the GP-UNet.

Sl. no.	Items	Descriptions
1	Data augmentation	Horizontal flip with probability 0.5
2	Threshold value for the predicted output	0.5
3	**Meta-training:**	
	Time required for 10-shots (maximum)	16 hrs
	GPU memory required for 10-shots (maximum)	9.5 GB
4	**Fine-tuning:**	
	Time required for 10-shots (maximum)	3.5 hrs
	GPU memory required for 10-shots (maximum)	9.5 GB
5	**Inference:**	
	Time required for 10-shots (maximum)	0.95 secs
	GPU memory required for 10-shots (maximum)	7.2 GB
6	GPU used for the experimental results	TITAN V (12 GB)
7	Deep learning framework	PyTorch

Table 3: Comparison of the DGPNet with DGPNet+ in terms of the mIoU.

Target	1−shot		5−shot	
class	DGPNet	DGPNet+	DGPNet	DGPNet+
B5	0.086 ± 0.039	$\mathbf{0.496 \pm 0.167}$	0.050 ± 0.026	$\mathbf{0.452 \pm 0.044}$
B39	0.001 ± 0.000	$\mathbf{0.379 \pm 0.144}$	0.001 ± 0.000	$\mathbf{0.751 \pm 0.022}$
TNBC	0.018 ± 0.003	$\mathbf{0.200 \pm 0.112}$	0.049 ± 0.007	$\mathbf{0.313 \pm 0.042}$
EM	0.123 ± 0.013	$\mathbf{0.647 \pm 0.025}$	0.300 ± 0.003	$\mathbf{0.756 \pm 0.012}$
ssTEM	0.158 ± 0.007	$\mathbf{0.641 \pm 0.069}$	0.444 ± 0.005	$\mathbf{0.772 \pm 0.026}$

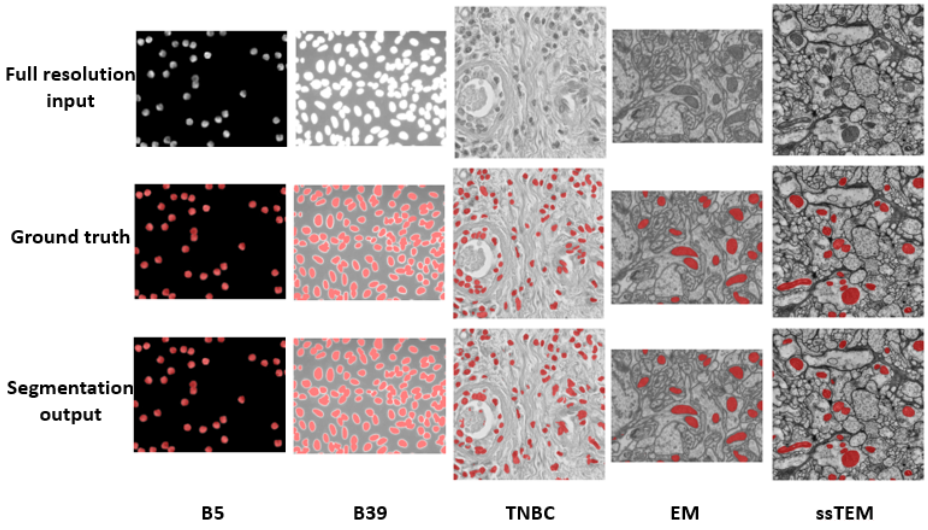

Fig. 2: Segmentation output of the 10−shot GP-UNet for different target classes.