Computers in Biology and Medicine 143 (2022) 105227

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Meta-learning with implicit gradients in a few-shot setting for medical image segmentation

Rabindra Khadka [a,d], Debesh Jha [a,b,*], Steven Hicks [a,d], Vajira Thambawita [a,d], Michael A. Riegler [a,b], Sharib Ali [c,e,**,1], Pål Halvorsen [a,d,1]

[a] SimulaMet, Oslo, Norway
[b] UiT the Arctic University of Norway, Tromsø, Norway
[c] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[d] Oslo Metropolitan University, Oslo, Norway
[e] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:
Meta-learning
Few-shot learning
Colonoscopy
Polyp segmentation
Wireless capsule endoscopy
Skin lesion segmentation
Generalization

ABSTRACT

Widely used traditional supervised deep learning methods require a large number of training samples but often fail to generalize on unseen datasets. Therefore, a more general application of any trained model is quite limited for medical imaging for clinical practice. Using separately trained models for each unique lesion category or a unique patient population will require sufficiently large curated datasets, which is not practical to use in a real-world clinical set-up. Few-shot learning approaches can not only minimize the need for an enormous number of reliable ground truth labels that are labour-intensive and expensive, but can also be used to model on a dataset coming from a new population. To this end, we propose to exploit an optimization-based implicit model agnostic meta-learning (iMAML) algorithm under few-shot settings for medical image segmentation. Our approach can leverage the learned weights from diverse but small training samples to perform analysis on unseen datasets with high accuracy. We show that, unlike classical few-shot learning approaches, our method improves generalization capability. To our knowledge, this is the first work that exploits iMAML for medical image segmentation and explores the strength of the model on scenarios such as meta-training on unique and mixed instances of lesion datasets. Our quantitative results on publicly available skin and polyp datasets show that the proposed method outperforms the naive supervised baseline model and two recent few-shot segmentation approaches by large margins. In addition, our iMAML approach shows an improvement of 2%–4% in dice score compared to its counterpart MAML for most experiments.

1. Introduction

Automatic lesion segmentation can help in accurate quantification of the area covered by anomalies, precise surgical removal, and treatment. Unlike manual processes, which are usually subjective and sub-optimal, automated methods can provide a more objective analysis of the lesions and their risks. Machine Learning (ML) and Deep learning (DL)-based models have already shown successful results in the clinical settings [1–3].

Data shift and availability of labeled data are major bottlenecks in medical image analysis. Other challenges that medical image analysis has to deal with are: 1) difficulty to get domain experts to perform annotations, 2) heterogeneous data, e.g., it could consist of multiple organs (skin, gastrointestinal organs) and varied disease types (melanoma in skin and polyp in the colon), and 3) large variability between expert and novice annotations. The lack of publicly available datasets as well as their quality (e.g., missing and erroneous labels) pose additional challenges. In addition, widely used supervised deep learning approaches require large amounts of training samples with labels and often fail to generalize when tested on different datasets due to data shifts caused by different data distribution. Data shift can arise due to population variation (e.g., different demographics), acquisition difference (e.g., devices

* Corresponding author. SimulaMet, Oslo, Norway.
** Corresponding author. Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK.
E-mail addresses: debesh@simula.no (D. Jha), sharib.ali@eng.ox.ac.uk (S. Ali).
[1] Shared senior authorship.

https://doi.org/10.1016/j.compbiomed.2022.105227
Received 30 August 2021; Received in revised form 5 January 2022; Accepted 5 January 2022
Available online 12 January 2022

or imaging protocols), prevalence shift (e.g., environmental factors affecting organs) and selection bias (e.g., inclusion criteria for study) [4].

The state-of-the-art DL models require a large number of high-quality and diverse datasets with pixel-wise masks for segmentation that is difficult to generate. Additionally, publicly available datasets are still limited and often include only a few samples of each unique class, case or part of a population. Some example datasets include KvasirCapsule-SEG [5] (55 samples), ETIS-Larib [6] (196 samples), PH^2 [7]) (200 samples), EDD2020 [8] (386 samples), Kvasir-instrument [9] (590 samples), and CVCClinicDB [10] (612 samples). With the available datasets, it is still possible to build a ML model by leveraging semi-supervised or few-shot learning methods [11], but these datasets listed above do not cover all lesion categories or include data from multiple sources; for example, rare disease cases, patient variabilities and multi-center data sources. Therefore, it is challenging to design a model that generalizes well on unseen datasets during clinical deployment. The possible solutions to the dataset mismatch can be domain adaptation [12] and domain generalization [13]. Domain adaptation utilizes a labeled source training dataset and unlabeled target data to develop a model that performs well on the target environment. Implementing such adaptation techniques helps to increase the generalization capacity of the model towards unseen target domain configuration. On the other hand, domain generalization capitalizes on using multi-source training datasets to design a classifier that generalizes well on unseen target (test) datasets. However, the problem of data scarcity is not resolved by any of these techniques in a classical supervised setting as they require large training datasets [14–16]. In addition, in domain adaptation methods, the learnt features have a similar embedding for both source and target dataset, and hence, this trade-off leads to compromises in the generalization capacity of the model [17,18].

To mitigate the problem of data scarcity and domain generalization, meta-learning under few-shot settings has emerged as a potential solution [19,20], especially in limited data settings. Meta-learning enables learning model weights by leveraging prior knowledge from various tasks [21] and can be implemented in different task objectives such as few-shot learning or multi-task learning. It is advantageous to use meta-learning in few-shot settings, and it has been primarily used in image classification [22,23]. Few-shot learning is a method that uses few annotated examples (support set) to make predictions on unlabeled examples (query set) and is the most appropriate choice when only limited data samples are available. An episodic training in a meta-learning setting can exploit to generalize to such limited data settings and become a natural choice for other tasks such as segmentation. Few-shot learning for segmentation has mostly been explored for natural images [24,25]. Recently, it is gaining more attention in the medical image segmentation [11,26–30]. Recent work by Ref. [11] used a semi-supervised few-shot learning approach to perform skin lesion segmentation by feeding the learner with unlabeled surrogate tasks [31]. applied a few-shot technique with a squeeze and excite block architecture to perform volumetric segmentation of multiple organs in medical images. In the work proposed by Ref. [32]; few-shot segmentation with a self-supervised method has been used to eliminate the need for having annotated medical images. They used an adaptive local pooling module in conjunction with prototypical networks to perform segmentation.

There are also some studies done to address the data scarcity and data mismatch problems in medical imaging field based on Wasserstein generative adversarial networks (GAN) where it was adopted for image reconstruction [33–35]. Some studies have been carried out to mitigate the generalization problem of the ML model in the medical domain, like the work done by Ref. [36] where they developed multi-scale deep convolutional networks that perform segmentation of overlapping cytoplasm. The work done by Ref. [37] proposes an automatic method to segment overlapping bacteria regions where they also incorporate Markov random field for unsupervised segmentation of small objects. These methods show improved generalization capacity. However, these methods are not tested under a few-shot setting. Furthermore, all of these works based on the few-shot learning approach use the same data source and similar tasks for inference, which means that the data shift problem has not been tackled.

A recent study [38] suggested that the supervised transfer learning method with fine-tuning can handle the data mismatch better than semi-supervised methods. The few-shot semi-supervised method adopted by Ref. [11] does not show a promising result, the predicted mask stands just at 62.40% of the target mask (ISIC dataset) under the 5-shot setting. Thus, in our work, meta-learning is adopted for domain generalization by further optimizing model weights via a meta-optimizer to overcome the shortcomings of few-shot learning. Recent work by Ref. [14] used the gradient-based meta-learning algorithm known as Model Agnostic Meta Learning (MAML), where the idea was to operate in the semantic feature space and learn semantically invariant features across training domains. They evaluated their method with Magnetic Resonance Imaging (MRI) images of the brain from different datasets that inherited domain shifts. They showed consistent results across all the datasets. However, the approach has not been tested under few-shot settings, i.e., less number of samples given during training to adapt to generalization capability in resource constraint settings during inference. Also, the training and test set included instances from the same anatomy. Being able to generalize well over another lesion type by training on one lesion type can be advantageous in medical imaging to tackle data scarcity problems. Additionally, the used MAML algorithm by Ref. [14] has some caveats related to computation and memory efficiency, which makes it difficult to scale up the accuracy as it requires several optimization steps [39]. The Implicit Model Agnostic Meta Learning (iMAML) algorithm [39,40], on the other hand, can provide faster and improved optimization during the meta-learning since the solution depends only upon the inner optimization and not the path taken by an inner optimization algorithm.

Primarily, this work explores the efficacy of the iMAML algorithm for medical image segmentation with the objective of handling the problem of data scarcity and data shift. We propose to demonstrate the use of iMAML in medical image segmentation and compare the results with other semi-supervised approaches. During this study, the convexity of the dice loss function is improved by applying Lovàsz extension [41]. We also compared the iMAML algorithm with the MAML algorithm under the same setting. The requirement of a few-shot learning paradigm to tackle data limitations is well established. However, the fine-tuning of the weight parameters has been revisited in several studies showing the ability of the model agnostic meta-learning approach. To this end, no studies have used an implicit model agnostic approach for medical image segmentation. Our contribution includes (i) incorporation of attention-UNet [42] mechanism for inner optimization of the weights using segmentation tasks on two different datasets during episodic meta-training, (ii) utilizing an analytical solution (conjugate gradient) for computing meta-gradients to achieve optimized weights, and (iii) a comprehensive analysis of the efficacy of iMAML on publicly available skin and polyp datasets from multiple sources. Our paper is structured into the following sections: Section 2 details the datasets used in this work, in Section 3, we present our iMAML segmentation approach and the compound loss function, Section 4 contains the experimental details and results, in Section 5 we provide comprehensive discussion and finally in Section 6 we conclude the paper.

Table 1
Publicly available medical imaging datasets used in our experiments. Here we provide the number of image samples, size of images and imaging type that were incorporated in these datasets.

Dataset	# of Images	Input size	Imaging type
Kvasir-SEG [43]	1000	Variable	Colonoscopy
KvasirCapsule-SEG [5]	55	Variable	Video capsule endoscopy
CVC-ClinicDB [10]	612	384 × 288	Colonoscopy
SIC-2018 [44,45]	2596	384 × 512	Dermoscopy
PH^2 [7]	200	768 × 560	Dermoscopy

2. Datasets

We use five widely used publicly available datasets, namely Kvasir-SEG [43], KvasirCapsule-SEG [5], CVC-612 [10], ISIC-2018 [44,45], and PH^2 [7]. A combination of these datasets has been used for the meta-training stage and tested on a holdout dataset to evaluate our proposed iMAML segmentation approach. Table 1 presents information of each dataset used in our experimental setup.

Kvasir-SEG [43] is a widely used publicly available colonoscopy dataset. It consists of 1000 polyp images, their corresponding ground truth segmentation masks and bounding boxes information of the area covered by polyp. The example images from the Kvasir-SEG dataset can be found in Fig. 3. The size of each polyp varies from 332 × 487 to 1920 × 1072. The original images from the Kvasir-SEG are captured during a colonoscopy examination using the ScopeGuide colonoscope (Olympus). The dataset can be downloaded from https://datasets.simula.no/kvasir-seg/.

KvasirCapsule-SEG [5] is the wireless video capsule endoscopy dataset. This dataset was developed by annotating the ground truth segmentation maps from the polyp images found in the Kvasir-Capsule dataset [46]. The dataset consists of 55 polyp images and their corresponding ground truth segmentation masks and bounding boxes. The example of KvasirCapsule-SEG can be found in Fig. 4. The dataset can be downloaded from https://www.kaggle.com/debeshjha1/kvasircapsuleseg.

CVC-ClinicDB [10], also known as CVC-612, is another popular polyp segmentation dataset. It consists of 612 polyp images from 31 colonoscopy videos and their corresponding ground truth masks. The sample images from CVC-ClinicDB can be found in stage 1 of Fig. 1. The dataset is available at https://www.dropbox.com/s/khtlmehjgv1b07z/cvc612.zip?dl=0.

The ISIC-2018 dataset [44,45] includes both benign and malignant skin lesion images. It consists of 2596 dermoscopy images and their corresponding ground truth masks. The example samples can be observed in Fig. 3. The image resolution is 384 × 512, and the dataset can be downloaded from https://challenge.isic-archive.com/data.

The PH2 [7] dataset consists of dermoscopic images. It consists of 200 images of melanocytic lesions. The ground truth segmentation mask for each image is provided. The dataset can be downloaded from https://www.dropbox.com/s/k88qukc20ljnbuo/PH2Dataset.rar. More details about the dataset can be found on the webpage.[2]

3. Methodology

This section describes the algorithm and the adopted method used to obtain the empirical results.

3.1. iMAML algorithm

In general, MAML approaches are trained through a meta-learning objective function [20]. However, due to the requirement of back-propagation during model training with high-order meta-gradients, MAML can suffer from vanishing gradients. In order to eliminate this problem [39], suggested to use a bi-level optimization, where an *inner* optimization is focused on computing weights through the Convolutional Neural Network (CNN) model and an analytic solution is used for the *outer* meta-gradient estimation (see Eq. (1)).

$$\theta^*_{ML} := \equiv argmin_\theta \underbrace{\frac{1}{M}\sum_{i=1}^{M} \overbrace{L(Alg_i(\theta, D_i^{tr}), D_i^{val})}^{inner-level}}_{outer-level}, \quad \text{with}$$
$$Alg_i(\theta) := \arg\min_{\varphi\in\theta} \; L_i(\varphi) + \frac{\lambda}{2}\|\varphi - \theta\|^2 \tag{1}$$

In Eq. (1), D_i^{tr} and D_i^{val} represent training (support set) and validation (query set) in the meta-training phase for the ith task. The task-specific parameters in the inner optimization level are represented by φ while the optimized weights after meta-training, i.e., the meta-parameters, are represented by θ. The final optimized meta-parameters are represented as θ^*_{ML}. In order to avoid overfitting and help anchor, the task parameter φ to the meta-parameter θ, an L2-regularization is used for the model training Alg_i.

The meta-training and meta-testing stages are shown in Fig. 1. During the meta-training stage, tasks are generated. The tasks contain a support set (train) and query set (validation) with few-shot instances. This means that only a few samples are chosen, such as 5 for 5-shot and 10 for 10-shot. We then initialize our attention U-Net segmentation model with random weights θ_o for the ith task. We then computed the loss L between the predicted mask and the ground truth mask in the support set with $L2$ -regularization. Validation loss on the query data completes the task for which the optimized φ_i is fed to the meta-learner where meta-gradients are analytically computed and updated as in Eq. (2). This is then fed to the model weights of the attention U-Net architecture for further backpropagation and optimization. Such a two-level optimization scheme is iterative and done for two different datasets in our case (see Fig. 1, top). The meta-training stage is completed once the set number of tasks M are completed to obtain the final meta-learned parameters θ^*_{ML}.

$$\theta \leftarrow \theta - \eta\frac{1}{M}\sum_{i=1}^{M}\frac{d\, Alg_i(\theta)}{d\theta}\; \nabla_\varphi\; L_i(Alg_i(\theta)) \tag{2}$$

The second stage consists of a simple fine-tuning step on the unseen data where optimized weight θ^*_{ML}, say θ for simplicity, is used to optimize the loss function L in a few-shot setting. The resulting final weights are then used in the final inference for direct segmentation map prediction as shown in Fig. 1 (bottom).

3.2. Loss function

A compound loss was used during training which comprises of both *log-cosh-dice loss* and *binary cross entropy loss*. It attenuates the problem of class imbalance through dice-loss. The final loss function is devised as:

$$L = L_{BCE} + L_{lc-dce} + \lambda\|\theta\|_2^2, \quad \text{with}$$
$$L_{lc-dce} = log(cosh(L_{Dice})) \tag{3}$$

where;

$$L_{BCE} = -(y\log(\widehat{y}) + (1-y)\log(1-\widehat{y})) \tag{4}$$

$$L_{Dice} = 1 - \frac{\left(2\sum_i y_i\widehat{y}_i\right) + 1}{\sum_i y_i + \sum_i \widehat{y}_i + 1} \tag{5}$$

Here, $\widehat{y}_i$ and y_i refer to the pairs of corresponding pixel values of prediction and ground-truth, respectively.

L_{Dice} and L_{BCE} have usual meanings for dice loss and binary cross-

[2] https://www.fc.up.pt/addi/ph2%20database.html.

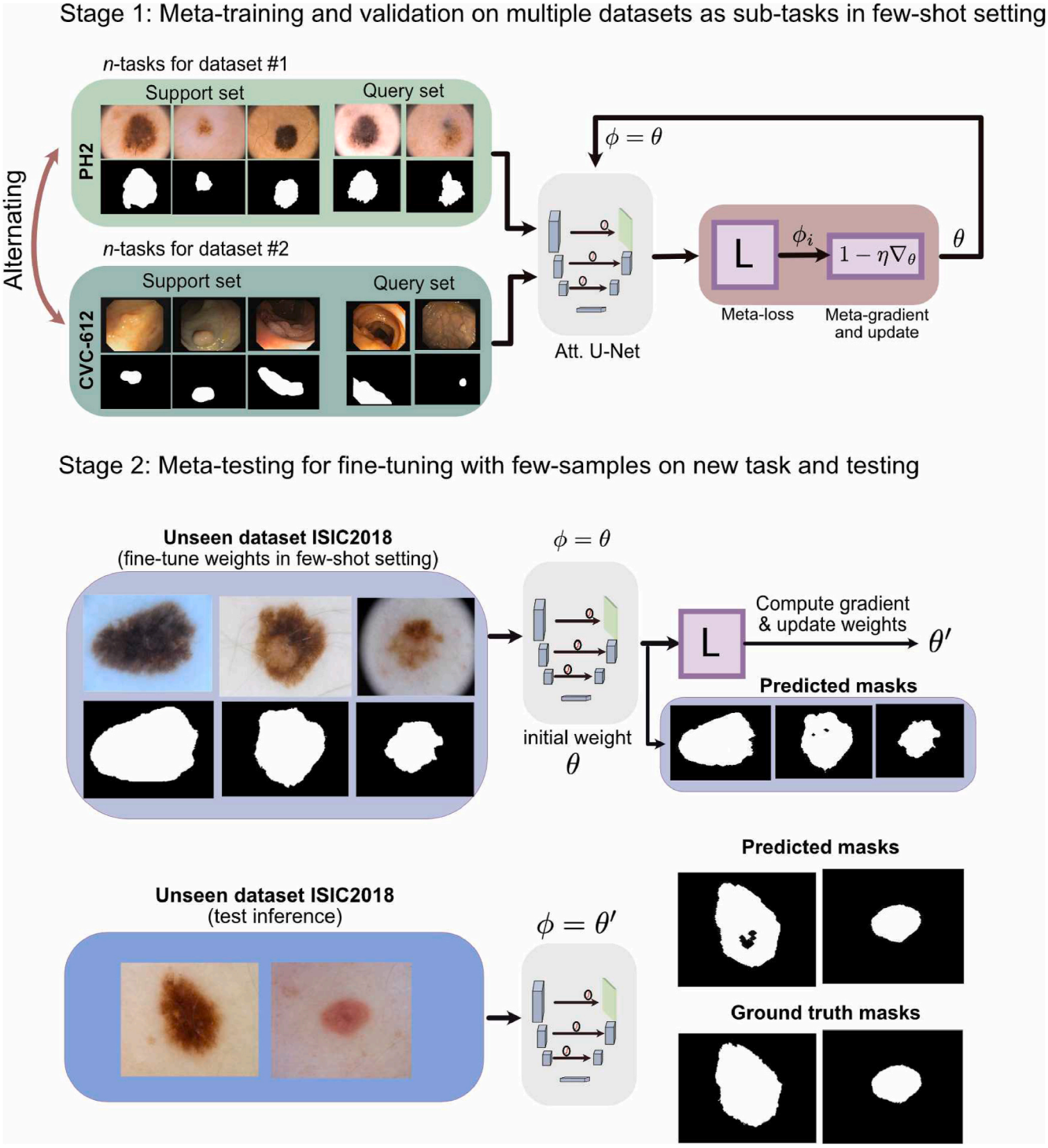

Fig. 1. Meta-learning with an implicit gradient optimization on medical imaging datasets: Meta training is done as episodic tasks on two public datasets (#1 and #2). In the first stage, a few-shot learning framework for each task is used for the support set, and validation is done on the query set. During the meta-testing stage, an unseen task from the third dataset is provided with the optimized weights obtained from the first stage, #1 and the gradient of the computed loss is used to readjust the final weights on only few samples of this dataset. Finally, the fine-tuned weight is used for the inference of the test samples. In all these settings, we use attention U-Net [42] to achieve segmentation maps.

entropy loss classically used in segmentation approaches [47]. Binary Cross-entropy [48] quantifies the difference between two probability distributions for a given random variable (eqn (4)). It is popularly adopted for object classification or pixel-level classification during segmentation. Dice loss (L_{Dice}) [49] is based on dice coefficient, which measures the overlap between predicted and ground-truth masks (eqn (5)). Unlike classical dice loss, L_{lc-dce} is the Lovàsz extension [41] that tackles the non-convex nature of dice loss by smoothing it and making the function tractable and easy to differentiate. Additionally, we have added a weight decay function as an L_2 regularization with λ as regularization hyper-parameter, and θ is the model weight. This allows to encapsulate better generalizability on test samples.

3.3. Network architecture

Our proposed model architecture is shown in Fig. 1. Our figure is divided into stage 1 and stage 2. In stage 1, meta-training is done on the support set, and the validation is done on the query set. Similarly, in stage 2, meta-testing is done on the test set. From the figure, we can observe that meta-training is performed as episodic tasks on two public datasets (PH^2 [7], and CVC-ClinicDB [10]). During the meta-testing stage (stage 2), an unseen task from the third dataset is provided with the optimized weights obtained from the first stage, #1 and the gradient of the computed loss is used to readjust the final weights on only a few samples of this dataset (please refer to stage 2 part of Fig. 1). The network consists of a sampler for creating support and query set for the few-shot setting of our experiment and for specific tasks. In all these settings, we use attention U-Net [42] as the meta learner to achieve segmentation maps. The attention U-Net is used for each task's inner-level parameter optimization φ_i. We have a meta-gradient optimizer for computing the optimized weights fed to the attention U-Net model. Finally, the fine-tuned weight is used for the inference of the test samples, and the ground truth masks are predicted.

4. Experiments and results

This section will describe the experimental setup, implementation details, and our results on each dataset.

4.1. Setup

4.1.1. Experimental design

All experiments in this work use few-shot supervised settings for which *N*-way, *K*-shot tasks are randomly generated from two publicly available datasets. In this context, *N* refers to the number of classes and *K* refers to samples from each class. The number of classes *N* corresponds to the number of different data pools, making our experiments a 2-way *K*-shot task. Finally, the learned parameters were fine-tuned over an entirely new task drawn from the hold-out data pool for the meta-testing. We present three sets of experiments: (i) tasks that comprised of samples exclusively from the *Kvasir-SEG* (polyp) dataset or from the PH^2 (skin) dataset, (ii) tasks that are comprised of mixed samples, and (iii) tasks trained on the same class datasets and tested on an entirely different class, such as meta-training on skin datasets and meta-testing on polyp dataset.

4.1.2. Implementation details

The meta-parameters were initialized with pre-trained weights from U-Net trained on brain MRI scans [50]. The meta-gradient is computed by applying conjugate gradient (CG), and the meta-parameters are updated using the Adam optimizer [51] with a learning rate of 10^{-5} and a weight decay of 0.000 5. Our convergence criteria is reached when the loss function does not change more than 0.001 over ten epochs. Fig. 2 shows the training convergence at the 50th epoch for a model trained in 2-way 5-shot and 2-way 10-shot settings. For the regularization of the computed learned weights, we fixed $\lambda = 100$. The images and their corresponding ground truth were normalized in the range of $[-1, 1]$ and resized to 256×256. All implementations were done using the PyTorch framework, and experiments were conducted on NVIDIA Tesla V100-SXM3.

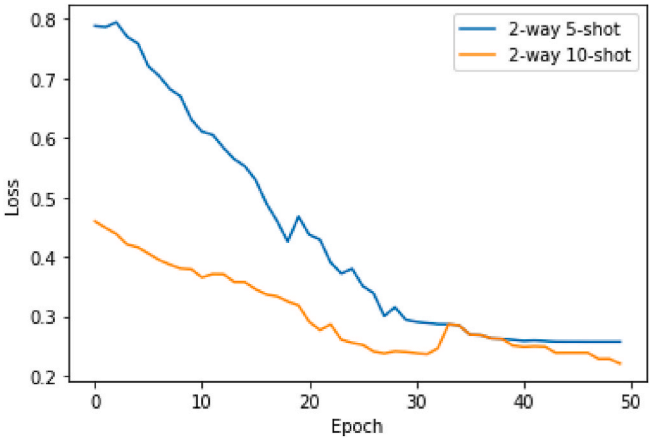

Fig. 2. Comparison of learning curves between 2-way 5-shot and 2-way 10-shot meta-learning settings that correspond to Table 2 for our proposed iMAML approach. Loss up to 50 epochs are provided to illustrate the convergence.

4.2. Results

We present results for three different experimental setups to illustrate the model efficacy compared to naive supervised attention U-Net and two recent SOTA few-shot methods used for medical image segmentation.

4.2.1. Meta-training with samples drawn exclusively from two unique datasets and unique categories

Table 2 presents the episodic training of our meta-learning approach on the PH^2 and Kvasir-SEG datasets consisting of skin and polyp categories, respectively. It can be observed that on the unseen ISIC dataset for test, our proposed iMAML-based segmentation outperformed the naive baseline U-Net by a very large margin of 25% and by nearly 23% and 16% on the dice coefficient compared to the baseline semi-supervised method [11] and the recent mask guided few-shot segmentation approach (PMG baseline) [30], respectively. The qualitative results (Fig. 3, left) also provide insight that our method provided optimal segmentation masks for different skin lesion types. The proposed meta-learning-based segmentation obtained the highest dice coefficient of 77.39%, 79.17% and 83.26% for different *K*-shots, i.e., 5, 10, and 20 shots, respectively.

4.2.2. Tasks comprising mixed samples of two unique datasets

Table 3 presents quantitative results for a different setting where the samples are mixed from two datasets (PH^2 and Kvasir-SEG). Clearly, there is evidence of a performance drop in our meta-learning method. Nevertheless, the proposed algorithm consistently outperformed baseline methods. The best dice score of 72.48% is obtained on the ISIC (skin) dataset under 2-way 20-shot setting, which is nearly 11.69% and 5.48% on the dice coefficient compared to the baseline semi-supervised method and PMG baseline model, respectively. Similarly, under this experimental setup, the segmentation results on the KvasirCapsule-SEG dataset using the iMAML and MAML algorithms are also captured in Table 5. It can be observed that both the meta-learning algorithms iMAML and MAML, outperforms the baseline models by 45% and 42%, respectively, that was naively trained under classical supervised setting with limited 44 images that were available for the KvasirCapsule-SEG dataset.

4.2.3. Tasks comprising samples from two unique datasets of the same class

Tables 4 and 6 represent meta-training on two unique datasets, but with the same categories and tested on a different class dataset. The categories here refers to a particular disease type (polyp or skin lesion). It can be observed that for episodic training conducted on the polyp datasets (CVC-ClinicDB and Kvasir-SEG) and tested on the skin dataset (see Table 4), our method is still able to generalize better than the naive baseline approach trained on 800 samples and the recent semi-supervised approach. The best dice score of 66.71% is obtained on the ISIC (skin) dataset under a 2-way 20-shot setting which is better by nearly 5.92% compared to the baseline semi-supervised method and competitive to the PMG baseline. Similar observations can be found when the method is trained on the skin datasets such as ISIC-2018 and

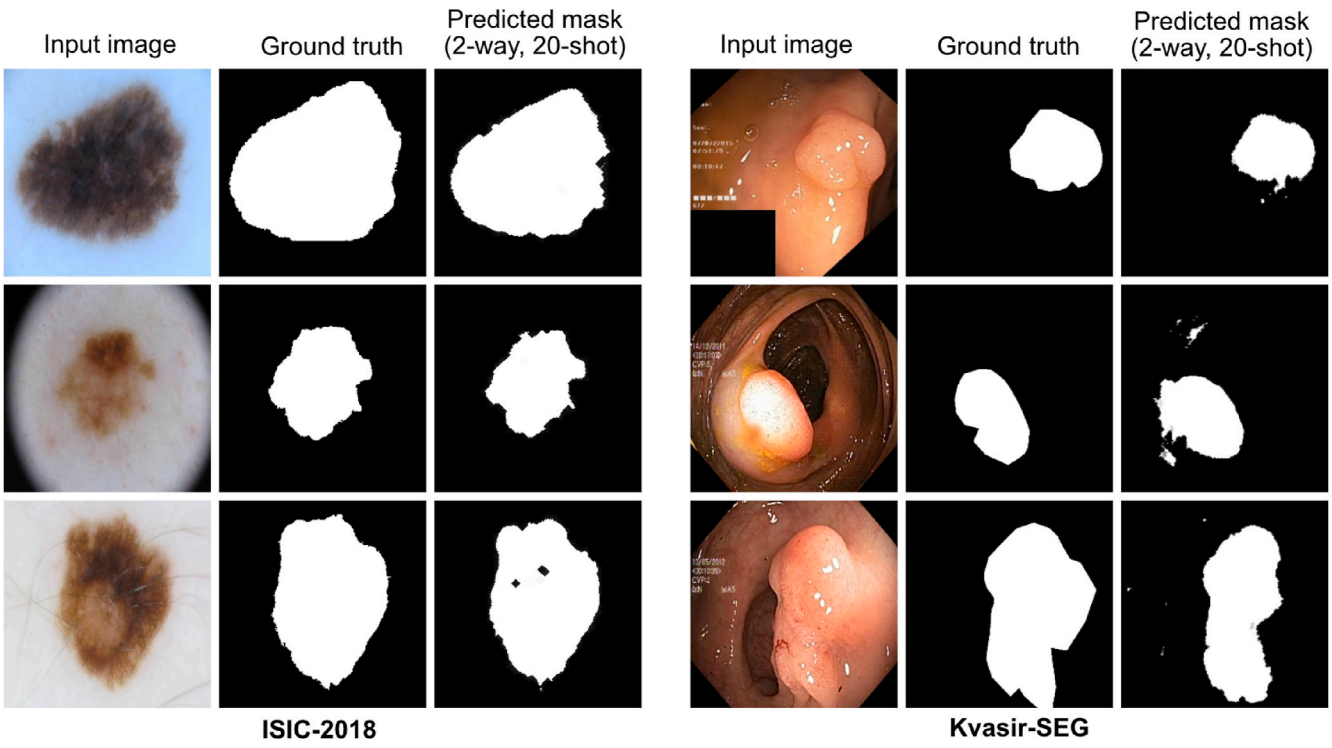

Fig. 3. Qualitative results of the proposed method on ISIC-2018 [44,45] (left) presented in Table 2 and results on Kvasir-SEG [43] (right) corresponding to Table 6. (Left) represents our first experimental configuration, i.e., training with samples drawn uniquely from two datasets from two different class categories (in this case PH^2 and Kvasir-SEG) while (right) corresponds to our third setup where samples comprise of two different datasets but of unique class (only skin datasets in this case).

PH^2 datasets and tested on the Kvasir-SEG and KvasirCapsule-SEG polyp segmentation datasets (please refer to Fig. 3 and Table 6).

Additionally, we further investigated the effect of Lovász extension and standard dice loss function in a meta-learning setting using our first episodic training setup. Based on the experimental results (see Table 7), Lovàsz extension was chosen, which improved the segmentation result by nearly 3.00%.

5. Discussion

Owing to the challenges such as data scarcity and data mismatch in the medical field while applying deep learning techniques, the generalization capacity of the trained model is reduced during deployment. Furthermore, various biases are introduced during the data collection process that can induce data shift at test time and derail the trained model's performance during a clinical deployment.

Acknowledging these challenges, the ML community has carried out some studies, which includes the work by Ref. [11] which is based on a semi-supervised few-shot learning method. Similarly, work by Ref. [14] uses the meta-learning method MAML to tackle the challenge of data shifts due to various data sources. However, these previous works have not been tested for completely different anatomies and under a few shot settings. We propose a meta-learning method with an implicit gradient (iMAML) to overcome these challenges under a few shot settings. The adopted meta-learning method is model agnostic and can take any other segmentation network as the meta-learner to learn the segmentation mask. For our experiment, we select the most popular segmentation network, U-Net, with an attention module as the meta-learner. The two SOTA methods used in comparison use few-shot learning approaches and hence can be directly compared. Adding any other supervised models would direct us to similar accuracy gains when used in a meta-learning framework. To test the efficacy of the iMAML algorithm, we arranged three different experiment setups (see Section 4.1). For carrying out the experiments, two datasets for skin, two datasets of normal colonoscopy and a dataset from video capsule endoscopy, which is a different modality, were used. The idea was to perform meta-training with tasks that are comprised of instances either from the same medical categories or different medical categories; to observe the generalization capacity of the algorithm. So, we picked two datasets that provided enough variability for episodic training. The results in Table 2 are from the first experimental setup where tasks are homogeneously comprised either only from the PH^2 (skin) dataset or from the Kvasir-SEG (polyp) dataset and then tested on the ISIC dataset. The segmentation results from the iMAML algorithm outperformed all the baseline models with the largest improvement of over 25% compared to the naive baseline model. Furthermore, iMAML has an improvement of nearly 2%–4% over the standard MAML approach. Similarly, the results from the second experiment are tabulated in Table 3 where the meta-learning algorithm is trained on tasks comprised of both the PH^2 (skin) and the Kvasir-SEG (polyp) datasets together. The segmentation performance on the test task from the ISIC (skin) dataset shows that the iMAML algorithm outperforms the naive baseline model by nearly 15% and shows distinct performance gains over all other methods in our comparison. The overall degradation in performance of both meta-learning algorithms compared to the previous setup (see Table 2) can be due to the increased variability in samples presented during the episodic meta-training that can make the network difficult to converge optimally to two different dataset attributes.

The third experiment setup aims to test the generalization capacity of the meta-learning model on an entirely never seen task. The task is comprised of polyp datasets, namely from CVC-ClinicDB and Kvasir-SEG. Table 4 depicts the result of the third experiment setup where the performance of the meta-learning algorithm is further degraded. Again, this could be because of training on a completely different dataset acquired from a different device and a different class category. Thus, the proposed iMAML generalizes well and provides an improved result compared to naive baseline by 8.61% and still eclipses the semi-supervised baseline method [11] by 5.92%. We further compared the

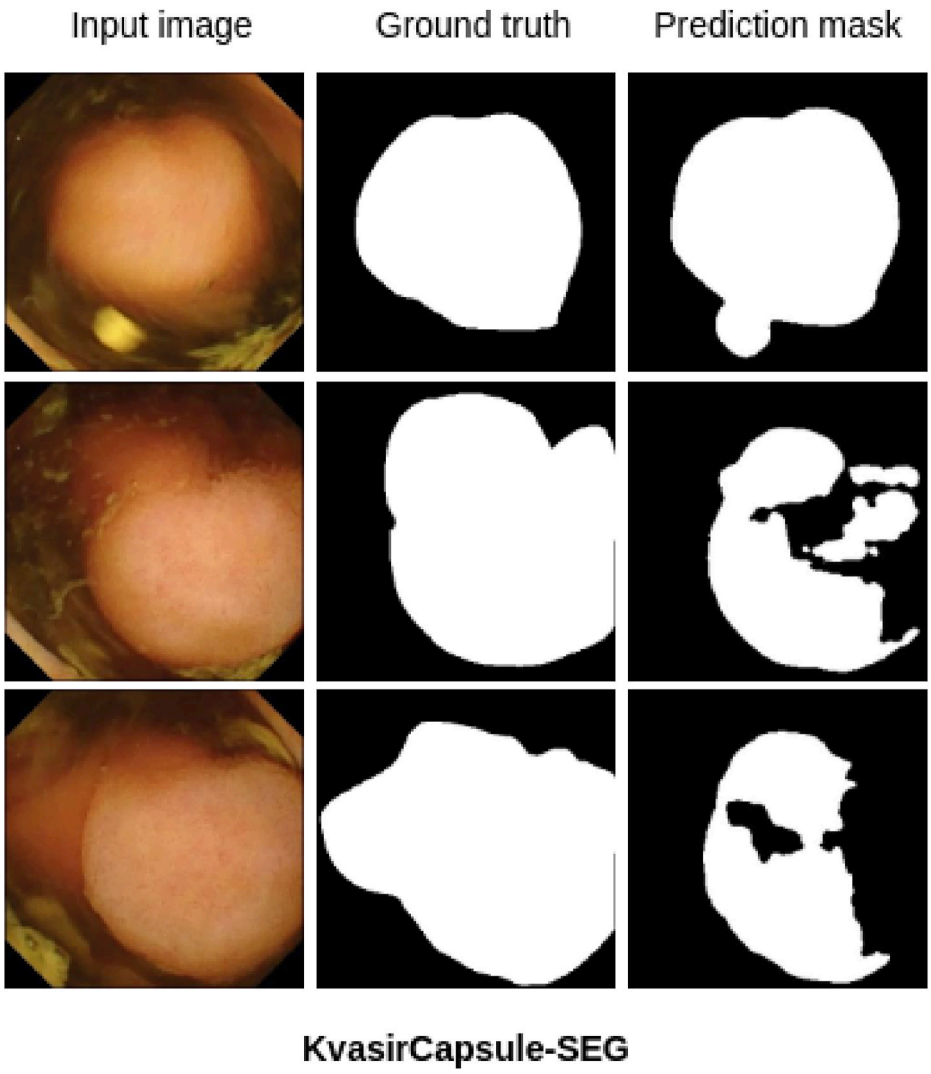

Fig. 4. Qualitative results on the KvasirCapsule-SEG dataset corresponding to Table 6. The illustrated prediction maps (left) refer to the results from our third setup where samples comprised of two different datasets but with unique class (i.e., skin datasets, PH^2 and ISIC-2018, in this case).

Table 2
Quantitative results as DSC metric for our first experimental setup. Here, episodic training for meta-learning is done independently with 50 tasks, first on PH^2 (skin) and then on Kvasir-SEG (polyp). Here, naive baseline (i.e., attention UNet) is trained on 800 image samples while 5 shot (referring to a few-shot training using 5 samples) results for PMG Baseline is reported [30]. Similarly, for meta-learning approaches we provide results on 5, 10 and 20 shot episodic training. Test samples consist of only unseen ISIC data samples.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Naive Baseline	800	–	ISIC	58.10
Semi-supervised. baseline (Feyjie et al., 2020)	5	–	ISIC	61.38
	10	–	ISIC	61.40
	20	–	ISIC	60.79
PMG Baseline (Xiao et al., 2021)	5	–	ISIC	67.00
Meta-learned (MAML)	5	50	ISIC	75.62
	10	50	ISIC	77.31
	20	50	ISIC	**79.60**
Meta-learned (iMAML)	5	50	ISIC	**77.39**
	10	50	ISIC	**79.17**
	20	50	ISIC	**83.26**

Table 3
Episodic training on tasks comprised of both PH^2 (skin) and Kvasir-SEG (polyp) instances. This refers to our second experimental setup. Similar to Table 2, here we present DSC metric scores for 5, 10, and 20 shots for meta-learning approaches again tested on unseen ISIC datasets.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Naive Baseline	800	–	ISIC	58.10
Semi-supv. Baseline [11]	5	–	ISIC	61.38
	10	–	ISIC	61.40
	20	–	ISIC	60.79
PMG Baseline [30]	5	–	ISIC	67.00
Meta-learned (MAML)	5	50	ISIC	66.19
	10	50	ISIC	68.54
	20	50	ISIC	70.61
Meta-learned (iMAML)	5	50	ISIC	**70.15**
	10	50	ISIC	**71.69**
	20	50	ISIC	**72.48**

test results between Kvasir-SEG and KvasirCapsule-Seg while training on tasks comprised of skin datasets only. The results captured in Table 5 demonstrate that even with datasets with few examples like KvasirCapsule-SEG, the meta-learning algorithms can perform better than the naive baseline models.

From the empirical observations, we can note that the DSC score is

Table 4
Episodic training on CVC-612 (polyp) and Kvasir-SEG (polyp) dataset. Here we provide quantitative results from our third experimental setup (i.e., tasks comprising samples from two unique datasets of the same class). Similar to Table 2, here we present DSC metric scores for 5, 10, and 20 shots for meta-learning approaches again tested on the unseen ISIC dataset.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Naive Baseline	800	–	ISIC	58.10
Semi-supv. Baseline [11]	5	–	ISIC	61.38
	10	–	ISIC	61.40
	20	–	ISIC	60.79
PMG Baseline [30]	5	–	ISIC	67.00
Meta-learned (MAML)	5	50	ISIC	59.70
	10	50	ISIC	62.43
	20	50	ISIC	64.70
Meta-learned (iMAML)	5	50	ISIC	**63.56**
	10	50	ISIC	**65.09**
	20	50	ISIC	**66.71**

Table 5
Episodic meta-training on Kvasir-SEG (polyp) and PH^2 (skin) dataset from the second experimental setup (i.e., tasks comprising mixed samples of two unique datasets). Meta-testing is done on instances from unseen KvasirCapsule-SEG (wireless capsule endoscopy polyp) dataset. Here, naive baseline (attention U-Net) is trained on KvasirCapsule-SEG using 44 samples (80%) and tested on remaining samples (20%) as done for other meta-learning approaches.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Naive Baseline	44	–	KvasirCapsule-SEG	16.23
Meta-learned (MAML)	5	50	KvasirCapsule-SEG	53.33
	10	50	KvasirCapsule-SEG	56.10
	20	50	KvasirCapsule-SEG	58.47
Meta-learned (iMAML)	5	50	KvasirCapsule-SEG	**56.39**
	10	50	KvasirCapsule-SEG	**59.34**
	20	50	KvasirCapsule-SEG	**61.28**

Table 6
Episodic meta-training on ISIC (skin) and PH^2 (skin) datasets from the third experimental setup (i.e., tasks comprising samples from two unique datasets of the same class). The meta-testing is done on instances from Kvasir-SEG and KvasirCapsule-SEG dataset. Here, both naive baseline attention-UNet, MAML and iMAML meta-learning approaches are compared.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Naive Baseline	800	–	Kvasir-SEG	60.53
	44	–	KvasirCapsule-SEG	16.23
Meta-learned (MAML)	5	50	Kvasir-SEG	59.30
	10	50	Kvasir-SEG	61.72
	20	50	Kvasir-SEG	64.09
Meta-learned (iMAML)	5	50	Kvasir-SEG	**62.00**
	10	50	Kvasir-SEG	**65.10**
	20	50	Kvasir-SEG	**66.58**
Meta-learned (MAML)	5	50	KvasirCapsule-SEG	52.26
	10	50	KvasirCapsule-SEG	54.09
	20	50	KvasirCapsule-SEG	57.47
Meta-learned (iMAML)	5	50	KvasirCapsule-SEG	**53.80**
	10	50	KvasirCapsule-SEG	**55.35**
	20	50	KvasirCapsule-SEG	**58.19**

Table 7
Quantitative results on the study of the effect of Lovász extension compared to the standard dice loss function in a meta-learning setting with 5 shot 2 way. The meta-training was done on two datasets (i.e., 2- way) namely CVC-612 and PH^2, and tested on unseen ISIC dataset.

Algorithm	K-shots	# Tasks	Target Dataset	DSC
Dice Loss	5	20	ISIC	73.90
Log(cosh(Dice Loss))	5	20	ISIC	**76.85**

higher with ISIC as the target dataset in comparison with KvasirCapsule-SEG. This is because Kvasir-SEG and Kvasir Capsule datasets come from colorectal inspection (inside body), where the obtained images are often specular and have variable contrast based on their location. In contrast, the ISIC dataset is obtained from dermatoscopy, which is usually taken from the exposed skin region with polarised or non-polarised light sources and are concentrated closer to the area of interest and often have diffused reflection.

Empirically, we showed that the iMAML algorithm could efficiently handle tasks with higher variations of instances during deployment. The method is model agnostic and should be replicable with other imaging modalities. However, we have chosen skin and polyp datasets as the domain shift in these data are very observant due to 1) patient or population variability, 2) imaging type (e.g., colonoscopy vs capsule endoscopy) and 3) class and color variability in skin images (e.g., PH2 and ISIC). It also illustrates that iMAML can be applied effectively in a complex problem like segmentation. During each of the experiment setups, the performance of the meta-learning algorithm is further improved by increasing the number of training tasks or the number of instances in each task which was a trade-off between training time and accuracy. This provides a robust method to handle data scarcity problems while training a deep neural network.

The findings of the empirical studies suggest that optimization-based meta-learning can alleviate the problem of data generalization and data scarcity which is prominent in the medical domain. We showed that the idea of meta-learning is a plausible concept that can benefit medical image segmentation under few-shot settings. In the future, we want to investigate how prior information about feature embedding from each task could be used to reduce the training time.

6. Conclusion

We proposed a novel model-agnostic meta-learning segmentation method in a few-shot setting that uses an implicit gradient-based optimization technique for improved model parameter estimation and generalization over unseen datasets with unique and seen categories. The proposed method improved performance and generalization capabilities compared to naive supervised techniques and the most recent few-shot segmentation approaches. We also demonstrated that the iMAML algorithm performs better than a popular meta-learning approach, MAML. Our method allowed the exploitation of available medical imaging datasets for training such that the trained model can be applied on an unseen dataset without requiring ample ground truth labels. Thus, the proposed method eliminates the need for abundant data for each specialized medical imaging category. However, the adopted meta-learning algorithm (iMAML) showed only marginal performance gain when trained with tasks comprised of instances from various medical categories. The generalization capacity of the iMAML algorithm is reduced when trained with skewed tasks, for example, tasks comprising of instances from skin and polyp datasets. To address such an issue, we will aim to shuffle channels or mix embedded features between instances of datasets while performing meta-training in future work. Nevertheless, this meta-learning approach could potentially contribute to developing clinically deployable systems for real-world application in the future.

Author contribution

RK, DJ and SA wrote most of the manuscript with input from all the authors. RK conducted all the experiments reported in the paper. RK and SA revised the manuscript with valuable feedback for corrections from DJ, SH, VT, MAR, and PH. All authors agreed for submission.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgment

D. Jha is funded by the PRIVATON project (#263 248) which is funded by Research Council of Norway (RCN). S. Ali is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). Our experiments were performed on the Experimental Infrastructure for Exploration of Exascale Computing (eX3) system, which is financially supported by RCN under contract 270 053. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

[1] S. Horng, D.A. Sontag, Y. Halpern, Y. Jernite, N.I. Shapiro, L.A. Nathanson, Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning, PLoS One 12 (2017), e0174708.
[2] J.M. Brown, J.P. Campbell, A. Beers, K. Chang, S. Ostmo, R.P. Chan, J. Dy, D. Erdogmus, S. Ioannidis, J. Kalpathy-Cramer, et al., Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks, JAMA Ophthalmol. 136 (2018) 803–810.
[3] G. Hinton, Deep learning—a technology with the potential to transform health care, JAMA 320 (2018) 1101–1102.
[4] D.C. Castro, I. Walker, B. Glocker, Causality matters in medical imaging, Nat. Commun. 11 (2020) 1–10.
[5] D. Jha, N.K. Tomar, S. Ali, M.A. Riegler, H.D. Johansen, D. Johansen, T. de Lange, P. Halvorsen, Nanonet: real-time polyp segmentation in video capsule endoscopy and colonoscopy, in: Proceedings of IEEE International Symposium on Computer-Based Medical Systems, (CBMS), 2021, pp. 37–43.
[6] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer, Int. J. Comput. Assisted Radiol. Surg. 9 (2014) 283–293.
[7] T. Mendonc̈ya, P.M. Ferreira, J.S. Marques, A.R. Marc̈yal, J. Rozeira, PH 2-A dermoscopic image database for research and benchmarking, in: In 2013 35th Annual International Conference of the IEEE Engineering In Medicine And Biology Society (EMBC), IEEE, 2013 July, pp. 5437–5440.
[8] S. Ali, M. Dmitrieva, N. Ghatwary, S. Bano, G. Polat, A. Temizel, A. Krenzer, A. Hekalo, Y.B. Guo, B. Matuszewski, et al., Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy, Med. Image Anal. 70 (2021) 102002.
[9] D. Jha, S. Ali, K. Emanuelsen, S.A. Hicks, V. Thambawita, E. Garcia-Ceja, M. A. Riegler, T. de Lange, P.T. Schmidt, H.D. Johansen, et al., Kvasir-instrument: diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy, in: Proceedings of IEEE International Conference on Multimedia Modeling, (ISM), 2021, pp. 218–229.
[10] J. Bernal, et al., Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Comput. Med. Imag. Graph. 43 (2015) 99–111.
[11] A.R. Feyjie, R. Azad, M. Pedersoli, C. Kauffman, I.B. Ayed, J. Dolz, Semi-supervised Few-Shot Learning for Medical Image Segmentation, 2020 arXiv preprint arXiv: 2003.08462.
[12] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (2016), 2096–2030.
[13] D. Li, Y. Yang, Y.Z. Song, T.M. Hospedales, Deeper, broader and artier domain generalization, in: Proceedings of the IEEE International Conference on Computer Vision, (ICCV), 2017, pp. 5542–5550.
[14] Q. Dou, D.C. Castro, K. Kamnitsas, B. Glocker, Domain generalization via model-agnostic learning of semantic features, in: Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2019.
[15] M. Ghifary, W.B. Kleijn, M. Zhang, D. Balduzzi, Domain generalization for object recognition with multi-task autoencoders, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
[16] S. Motiian, M. Piccirilli, D.A. Adjeroh, G. Doretto, Unified deep supervised domain adaptation and generalization, in: Proceedings of the IEEE International Conference on Computer Vision, (ICCV), 2017, pp. 5715–5725.
[17] Y.H. Tsai, W.C. Hung, S. Schulter, K. Sohn, M.H. Yang, M. Chandraker, Learning to adapt structured output space for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (CVPR, 2018, pp. 7472–7481.
[18] N. Celik, S. Ali, S. Gupta, B. Braden, J. Rittscher, Endouda: a modality independent segmentation approach for endoscopy imaging, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2021.
[19] S. Ravi, H. Larochelle, Optimization as a Model for Few-Shot Learning, 2016.
[20] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proceedings of the International Conference on Machine Learning, (ICML), 2017, pp. 1126–1135.
[21] S. Thrun, L. Pratt, Learning to Learn, 2012.
[22] S. Ali, B. Bhattarai, T.K. Kim, J. Rittscher, Additive angular margin for few shot learning to classify clinical endoscopy images, in: Proceedings of the International Workshop on Machine Learning in Medical Imaging, (MLMI), 2020, pp. 494–503.
[23] K. Mahajan, M. Sharma, L. Vig, Meta-dermdiagnosis: few-shot skin disease identification using meta-learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020, pp. 730–731.
[24] X. Zhang, Y. Wei, Y. Yang, T.S. Huang, Sg-one: similarity guidance network for one-shot semantic segmentation, IEEE Trans. Cybern. 50 (2020) 3855–3865.
[25] C. Zhang, G. Lin, F. Liu, R. Yao, C. Shen, Canet: class-agnostic segmentation networks with iterative refinement and attentive few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5217–5226.
[26] P. Khandelwal, P. Yushkevich, Domain generalizer: a few-shot meta learning framework for domain generalization in medical imaging, in: Proceedings of the Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, 2020, pp. 73–84.
[27] E.M. Rutter, J.H. Lagergren, K.B. Flores, A convolutional neural network method for boundary optimization enables few-shot learning for biomedical image segmentation, in: Proceedings of the Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, (DART), 2019, pp. 190–198.
[28] Q. Liu, Q. Dou, P.A. Heng, Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, (MICCAI), 2020, pp. 475–485.
[29] P. Zhang, J. Li, Y. Wang, J. Pan, Domain adaptation for medical image segmentation: a meta-learning method, J. Imag. 7 (2021) 31.
[30] J. Xiao, H. Xu, W. Zhao, C. Cheng, H. Gao, A prior-mask-guided few-shot learning for skin lesion segmentation, Computing (2021) 1–23.
[31] A.G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, C. Wachinger, 'squeeze & excite'guided few-shot segmentation of volumetric images, Med. Image Anal. 59 (2020), 101587.
[32] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, D. Rueckert, Self-supervision with super pixels: training few-shot medical image segmentation without annotation, in: Proceedings of European Conference on Computer Vision, (ECCV), 2020, pp. 762–780.
[33] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, IEEE Trans. Med. Imag. 37 (2018) 1348–1357.
[34] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M.W. Vannier, P.K. Saha, E.A. Hoffman, G. Wang, Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle), IEEE Trans. Med. Imag. 39 (2020) 188–203, https://doi.org/10.1109/TMI.2019.2922960.
[35] S. Tian, M. Wang, F. Yuan, N. Dai, Y. Sun, W. Xie, J. Qin, Efficient computer-aided design of dental inlay restoration: a deep adversarial framework, IEEE Trans. Med. Imag. 40 (2021) 2415–2427, https://doi.org/10.1109/TMI.2021.3077334.
[36] Y. Song, E.L. Tan, X. Jiang, J.Z. Cheng, D. Ni, S. Chen, B. Lei, T. Wang, Accurate cervical cell segmentation from overlapping clumps in pap smear images, IEEE Trans. Med. Imag. 36 (2017) 288–300, https://doi.org/10.1109/TMI.2016.2606380.
[37] Y. Song, L. He, F. Zhou, S. Chen, D. Ni, B. Lei, T. Wang, Segmentation, splitting, and classification of overlapping bacteria in microscope images for automatic bacterial vaginosis diagnosis, IEEE J. Biomed. Health Inform. 21 (2017) 1095–1104, https://doi.org/10.1109/JBHI.2016.2594239.
[38] A. Oliver, A. Odena, C.A. Raffel, E.D. Cubuk, I. Goodfellow, Realistic evaluation of deep semi-supervised learning algorithms, in: Proceedings of Advances in Neural Information Processing Systems, 2018.
[39] A. Rajeswaran, C. Finn, S. Kakade, S. Levine, Meta-learning with implicit gradients, in: Proceeding of the Conference on Neural Information Processing Systems (NeurIPS), 2019.
[40] R. Khadka, Meta-learning for Medical Image Segmentation, Master's thesis, 2021.
[41] M. Berman, A.R. Triki, M.B. Blaschko, The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4413–4421.
[42] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: learning where to look for the pancreas, in: 1st Conference on Medical Imaging with Deep Learning, 2018.
[43] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H. D. Johansen, Kvasir-seg: a segmented polyp dataset, in: Proceedings of the International Conference on Multimedia Modeling, (MMM), 2020, pp. 451–462.
[44] N. Codella, V. Rotemberg, P. Tschandl, M.E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al., Skin Lesion Analysis toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (Isic), 2019 arXiv preprint arXiv:1902.03368.
[45] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018) 1–9.
[46] P.H. Smedsrud, V. Thambawita, S.A. Hicks, H. Gjestang, O.O. Nedrejord, E. Næss, H. Borgli, D. Jha, T.J.D. Berstad, S.L. Eskeland, et al., Kvasir-capsule, a video capsule endoscopy dataset, Sci. Data 8 (2021) 1–10.
[47] S. Jadon, A survey of loss functions for semantic segmentation, in: Proceedings of IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology, (CIBCB), 2020, pp. 1–7.
[48] M. Yi-de, L. Qing, Q. Zhi-bai, Automated image segmentation using improved pcnn model based on cross-entropy, in: Proceedings of 2004 International Symposium

on Intelligent Multimedia, Video and Speech Processing, 2004, 2004, pp. 743–746, https://doi.org/10.1109/ISIMP.2004.1434171.

[49] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, 2017. CoRR abs/1707.03237, http://arxiv.org/abs/1707.03237. arXiv: 1707.03237.

[50] N. Pedano, A.E. Flanders, L. Scarpace, T. Mikkelsen, J.M. Eschbacher, B. Hermes, Q. Ostrom, Radiology data from the cancer genome atlas low grade glioma [tcga-lgg] collection, Canc. Imag. Arch. 2 (2016).

[51] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Proceedings of the 3rd International Conference for Learning Representations (ICLR), 2015.