



OPEN

Bidirectional meta-Kronecker factored optimizer and Hausdorff distance loss for few-shot medical image segmentation

Yeongjoon Kim¹, Donggoo Kang², Yeongheon Mok¹, Sunkyu Kwon¹ & Joonki Paik^{1,2}✉

To increase the accuracy of medical image analysis using supervised learning-based AI technology, a large amount of accurately labeled training data is required. However, the supervised learning approach may not be applicable to real-world medical imaging due to the lack of labeled data, the privacy of patients, and the cost of specialized knowledge. To handle these issues, we utilized Kronecker-factored decomposition, which enhances both computational efficiency and stability of the learning process. We combined this approach with a model-agnostic meta-learning framework for the parameter optimization. Based on this method, we present a bidirectional meta-Kronecker factored optimizer (BM-KFO) framework to quickly optimize semantic segmentation tasks using just a few magnetic resonance imaging (MRI) images as input. This model-agnostic approach can be implemented without altering network components and is capable of learning the learning process and meta-initial points while training on previously unseen data. We also incorporated a combination of average Hausdorff distance loss (AHD-loss) and cross-entropy loss into our objective function to specifically target the morphology of organs or lesions in medical images. Through evaluation of the proposed method on the abdominal MRI dataset, we obtained an average performance of 78.07% in setting 1 and 79.85% in setting 2. Our experiments demonstrate that BM-KFO with AHD-loss is suitable for general medical image segmentation applications and achieves superior performance compared to the baseline method in few-shot learning tasks. In order to replicate the proposed method, we have shared our code on GitHub. The corresponding URL can be found: <https://github.com/YeongjoonKim/BMKFO.git>.

Medical image segmentation plays a key role in numerous clinical applications, including diagnosis and treatment assistance, surgical planning, and disease progression monitoring. Precise delineation of organs and lesions from medical modalities, such as MRI and CT scans, is essential for these applications. For example, accurate tumor segmentation is crucial for gauging disease extent, devising radiation therapy strategies, and evaluating treatment efficacy in oncology. In addition, the segmentation of organs such as the brain, liver, and heart can aid in surgical planning and monitoring disease progression. Therefore, medical image segmentation is a critical component of modern healthcare, necessitating the development of precise and efficient algorithms to enhance patient outcomes.

Traditional image segmentation methods, such as region growing, graph cut¹, active contour model², and active shape model³, have been widely used in medical image analysis and have produced reliable segmentation results in many clinical applications. Software using these methodologies includes 3D Slicer, ITK-SNAP, and BrainSuite. However, these methods have limitations such as sensitivity to noise, computational cost, and the need for manual intervention.

Owing to recent advancements in deep learning technology, the field of semantic segmentation has seen substantial progress in various computer vision applications, including medical image segmentation. The U-Net⁴ model-based semantic segmentation method has emerged as a key tool for various clinical procedures and medical imaging research using MRI or CT. Therefore, it is necessary to develop a deep-learning-based application to complement the limitations of the traditional methodologies.

¹Department of Artificial Intelligence, Chung-Ang University, 84 Heukseok-ro, Seoul 06974, Korea. ²Department of Image, Chung-Ang University, 84 Heukseok-ro, Seoul 06974, Korea. ✉email: paikj@cau.ac.kr

Recent advances in deep learning have become an inflection point for AI to achieve human-level performance. Before the deep learning era, the expert system⁵ is a representative computer-based system that complements human decision-making. However, with the arrival of the era of deep learning, research beyond hand-craft-based methods has been conducted in various fields. In particular, FCN⁶, U-Net⁴, DeepLab v3 plus⁷, and SegNet⁸ models have improved the performance of semantic segmentation. These works are widely applied to medical image segmentation required for image reading and diagnosis in the medical field^{9–13}. The reason is that medical data requires medical experts, which incurs high costs for collection and annotation. Like other fields of research, it also raised a common problem that good results are obtained when learning with a lot of data. Therefore, massive data in medical image segmentation is a crucial part in research and often becomes an important factor that makes it difficult to achieve the purpose.

In the semantic segmentation method of medical images, problems with such a large amount of data are summarized as follows.

- In the field of medicine, a fundamental challenge arises from the scarcity of data required for research, as data from healthy subjects are readily available, while acquiring patient data proves to be difficult.
- The disclosure of patient's medical data for research and the privacy policy of patients conflict with the ethical aspects of medical researchers, which causes another problem that makes research more difficult.
- Medical imaging data requires expertise in the classification and judgment of diseases, and it induces a lot of cost and difficulty in the annotation.

To solve these problems, few-shot learning^{14–19} methods that learn only use a limited dataset have been proposed. Few-shot learning predicts unlabeled query sets by quickly optimizing the model with only a small number of labeled support sets. Therefore, applying few-shot learning to medical images has the advantage of achieving efficient image segmentation with only a small amount of data in order to semantically segment invisible lesions. As shown in Fig. 1, superpixels perform self-supervised learning by creating pseudolabels and solve the problem of data shortage with meta-learning.

Recently, meta-learning methods are attracting attention to utilize these technologies^{20–22}. Meta-learning is well known as “learning how to learn”. Because meta-learning teaches the learning process to solve a problem, it is important in that it aims to implement universal AI. Various meta-learning approaches are divided into model-based approaches²⁰ widely used in reinforcement learning, metric-based approaches used for data clustering in embedding spaces²¹, and optimization-based approaches²² that optimize initial parameters. In this paper, we

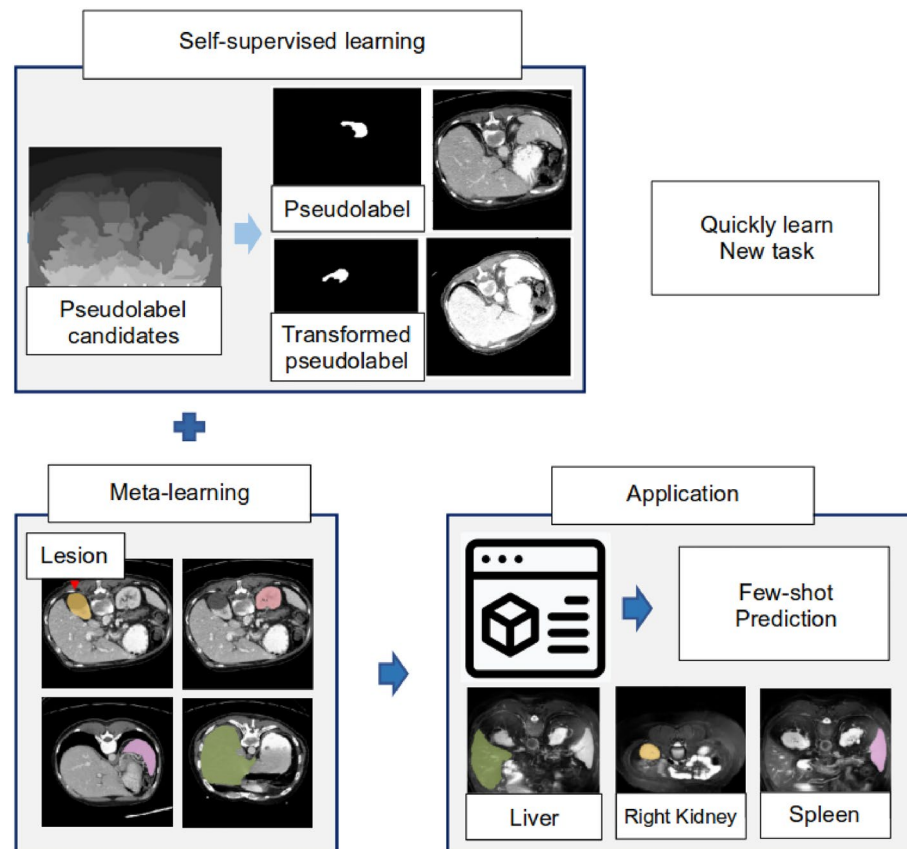


Figure 1. Overall architecture of the proposed model.

used the optimization-based approach of model-agnostic meta-learning (MAML) to find the generalized optimal initial point by considering model parameters and meta-optimizer simultaneously²². In addition, we used the Kronecker-factorized decomposition algorithm to efficiently decompose network parameters²³. Therefore, the proposed method can optimize the (tens of thousands) parameters of the network stably and show better performance than MAML algorithm only²³.

MAML only updates optimal initial values of model-parameters and meta-optimizer parameters. However, this approach does not make the best use of limited data. To solve this problem, we propose a novel meta-optimizer algorithm that considers validation and test set as well as train set in meta-learning. The proposed meta-optimizer consists of two optimizers which are updated separately: meta-train optimizer and meta-test optimizer. Through few-shot learning, the parameters of a few-shot learner are only updated with a k -shot support set. It leads to a highly unstable learning curve and consequently hard to optimize. To handle this difficulty, we use a meta learning scheme. The original MAML finds optimal initial values on the inner loop, which updates parameters of meta optimizer independently. Inspired by this scheme, we organize the inner loop to train-validation-test separately, then parameters of the meta optimizer are updated like few-shot learning procedure. We called the train phase of inner loop as a meta-train optimizer and the validation-test phase as meta-test optimizer. By this new inner loop design, we can train robust few-shot learner.

In the semantic segmentation of medical images, the conformation between Ground Truth (GT) and prediction is important. To handle this issue, we introduce the Hausdorff distance loss as our objective function. It measures the distance of the segment's boundary and minimizes them. Therefore, this objective function focuses on the shape of the segment rather than simply the area of the intersection.

The contributions of this paper are summarized as follows:

- We proposed an optimization-based meta-learning algorithm to few-shot medical image segmentation and novel algorithms that optimize few-shot like meta-learner consisting of meta-train optimizer and meta-test optimizer.
- We used a Kronecker-factored decomposition algorithm to compute optimal points of important learning parameters in optimizer-based meta-learning. This outperforms the baseline in few-shot learning with efficient parameter computation with MAML.
- We used Hausdorff-distance loss that focus on the shape of segment rather than area. It reduces noise of prediction and shows favorable results, especially in qualitative aspects that are important to assist a doctor's judgment.

The paper is organized as follows: After the “[Introduction](#)” section, we briefly summarize “[Related work](#)” section. The “[Proposed method](#)” section include in-depth description of model-agnostic meta-learning with Kronecker-factored decomposition and Hausdorff loss. After presenting “[Experimental results](#)” section with both quantitative and qualitative evaluation on kidney, liver, and spleen images, we concluded the paper.

Related work

Meta-learning. The goal of meta-learning is to train a model on different tasks so that it can solve unseen tasks using only a few training samples. Each task consists of dataset $D \in \mathbb{R}^2$, and optimal model parameters are computed by:

$$\theta = \operatorname{argmin}_{\theta} \mathbb{E}_{D \sim p(D)} \left[L_{D \sim p(D)}^{\text{validation}} \left(\theta - \alpha \nabla L_{D \sim p(D)}^{\text{train}}(\theta) \right) \right], \quad (1)$$

where θ represents the set of model parameters, $\mathbb{E}_{D \sim p(D)}$ the expectation over a distribution of dataset D , and $L(\theta)$ the loss function for a given dataset D and model parameters θ . $\theta - \alpha \nabla L_{D \sim p(D)}^{\text{train}}(\theta)$ denotes the gradient update for training data, where α represents the learning rate, and $\nabla L_{D \sim p(D)}^{\text{train}}(\theta)$ is the gradient of the loss function L with respect to θ on the training data. $L_{D \sim p(D)}^{\text{validation}}(\theta - \alpha \nabla L_{D \sim p(D)}^{\text{train}}(\theta))$ represents the loss function on the validation data, where $\theta - \alpha \nabla L_{D \sim p(D)}^{\text{train}}(\theta)$ is the updated parameter using the training data. We compute the loss function on the validation data using the updated parameter. Thus, Optimizing the minimized $\operatorname{argmin} \theta$ becomes the objective function. Although this formula looks similar to the general learning method, we need to find a model parameter that is optimized for all n tasks since each task is different here. This *learning-to-learn* meta-learning typically includes metric-based²¹, model-based^{24–31} and optimization-based²² approaches.

The goal of metric-based meta-learning is similar to nearest neighbors (k -NN) and kernel density estimation algorithm³². First, k support sets and query set are projected onto the embedding space through the feature extractor, and then learned in the direction to minimize the metric or distance between the support set and the query set.

On the other hand, the core of model-based meta-learning is to design a model architecture specific to fast learning. This approach aims to update its parameters rapidly with a few training steps. Shaban et al. proposed a fully convolutional network (FCN)-based few-shot segmentation method²⁴, which produces a set of parameters using a pre-trained feature extractor learned by a support set and then multiply to parameters from FCN that passed query image. Kate et al. proposed a conditional network for few-shot segmentation²⁵, which is conditioned on an annotated support set of images via feature fusion to inference on an unseen query image. Recently, attention-based approach^{26,27} and graph convolutional network-based approach^{28,29}, have become popular as research subjects because of their efficient use of feature relationships. In addition, the method using the global correlation network with discriminative embedding (GCN-DE)³³ and the method using the location-sensitive local prototype network³⁴ show impressive results in the few-shot task. Sun et al. proposed a medical image segmentation method that incorporated discriminative embedding into a global correlation network. This method

employs a deep convolutional network trained episodically, incorporating an efficient global correlation module to capture the correlation between support and query images. Furthermore, it enhances the discriminability of deep embedding for efficient feature domain clustering. Yu et al. proposed a location-sensitive local prototype network, a prototype-based few-shot segmentation method that leverages spatial priors in medical imaging, addressing the need for large amounts of expert-level annotated data. Their methodology involves two main steps: (1) location-sensitive local prototype extraction and (2) grid-based few-shot segmentation. Their approach divides the challenging problem of global image segmentation into smaller, more manageable sub-problems of local region segmentation. The proposed method is different in that it focuses on the meta-learning algorithm and the objective function. Wang et al. proposed a network that uses both metric-based and model-based meta-learning methods jointly called PANet³⁰, which learns class-specific prototype representations from support sets into an embedding space and then performs segmentation over the query set by matching each pixel to the learned prototypes. Ouyang et al. proposed a method using self-supervised learning exploits to superpixels called SSL-ALPNet³¹, which is especially effective in medical imaging because most CT or MRI images are suitable to apply a superpixel algorithm.

Optimization-based meta-learning methods assume that there are some internal representations that are more transferable than others. Under this assumption, it might learn internal features that are widely acceptable to all tasks rather than a single task. The pioneering work in this aspect is MAML²², which aims to find model parameters called meta-gradient that are sensitive to changes in the task, such that small changes in the parameters will produce large impacts on the loss function of any task. The meta-gradient update involves a gradient through a gradient. More detailed explanation will be given later.

Kronecker decomposition. There are many different matrix decomposition or factorization schemes related to solving systems of linear equations, such as Cholesky factorization, QR decomposition, and LU decomposition. The purpose of decomposition is to implement efficient matrix algorithms. For example, when solving linear equations $Ax = b$, the matrix A can be decomposed using LU decomposition, which factorizes a matrix into a lower triangular matrix L and an upper triangular matrix U . The systems $L(Ux) = b$ and $Ux = L^{-1}b$ require fewer additions and multiplications to solve than original equations.

Denton et al. proposed applying truncated singular value decomposition to the convolutional neural network³⁵, which is one of the earliest works to apply decomposition scheme to a neural network. Kronecker decomposition is another factorization scheme. It can replace a large matrix with two smaller Kronecker factor matrices that best approximate the original matrix. This decomposition scheme effectively reduces model parameters and consequently dimensionality. These properties of the matrix decomposition scheme are similar to the goal of meta-learning. Decomposition is interpreted as noise reduction and extract principle components. Therefore, matrix decomposition schemes assist to find the optimal meta initial point in a quick and stable manner.

Hausdorff distance. Loss functions play an important role in determining the model performance. There are various loss functions based on the distance to measure the quality of model prediction. Although many deep learning models adopt cross-entropy loss function, the cross-entropy loss is not suitable for segmentation task in that it only considers intersection area, not the whole object shape. Therefore, in some cases, quantitative evaluation is high, but qualitative evaluation often shows poor results. To solve this problem, Sudre et al. proposed dice loss³⁶, which is inspired from dice coefficient, a metric to evaluate segmentation results. It considers the region and scale of the predicted segment. To consider the segment's boundary, Hayder et al. devise shape-aware loss³⁷, which is a variant of cross-entropy loss by adding a shape based coefficient used in case of segment boundary. The average Hausdorff distance is widely used for a measure of two points sets³⁸. Inspired by this concept, Hausdorff distance loss was designed to compare ground truth segmentation boundary and predicted segmentation boundary³⁸. The directed Hausdorff distance from boundary point set X to Y is given by the sum of all minimum distances from all points from X to Y . Average Hausdorff distance can be computed by the bidirectional form of directed Hausdorff distance $D(\cdot, \cdot)$, such that mean of $D(X, Y)$ and $D(Y, X)$.

Proposed method

The proposed model-agnostic method is compatible with any other models without network component manipulation. We employed a state-of-the-art super-pixel-based self-supervised learning method, SSL-ALPNet³¹, as our baseline network for few-shot abdominal organ segmentation in both CT and MRI. In order to highlight the significance of meta-learning in medical imaging segmentation, we provide a concise overview of the distinctive features of medical imaging. CT and MRI images are acquired using highly sensitive medical equipment. Although each acquisition process is slightly different, it usually takes 10–15 min to obtain the result. Since each process has a unique characteristics, such as breathing patterns, movement, and slight body tremors, the correspondingly acquired image is also unique. Unfortunately, the uniqueness makes image segmentation difficult, and degrades the optimization performance of few-shot learning. To solve this problem, we proposed a bidirectional meta-Kronecker factored optimizer (BM-KFO) framework.

Figure 2 the overall framework of the proposed model, which uses ResNet as a feature extractor, and takes an MRI image as input. Although a general few-shot learning model uses only support images to train the network, we additionally use self-supervised learning scheme with super-pixel pseudo labels to use a limited number of data. Each pseudo label is created from the super-pixel version of support images. The model is then trained using both original support and pseudo labels. Each unlabeled image is also created as a super-pixel pseudo-label image. Then, the original image and pseudo-label are used as support and query sets³¹. Proposed BM-KFO framework falls into the category of optimization-based meta-learning for faster optimization of a few-shot task. The proposed framework is divided into meta-train and meta-test optimizers, and the parameters are efficiently and

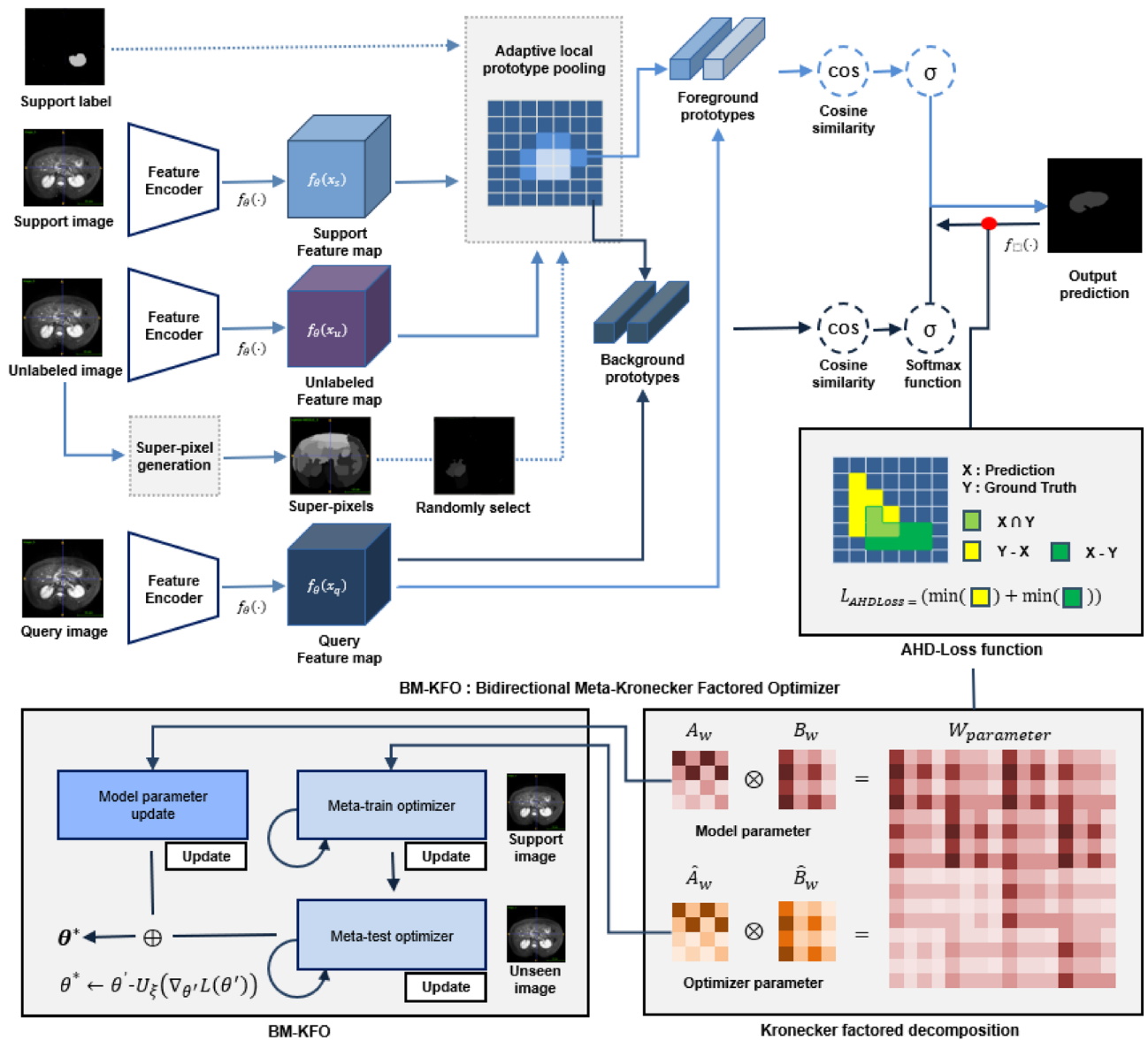


Figure 2. Overall architecture of the proposed model.

stably learned using kronecker-factored decomposition. We use objective function which combines cross-entropy and Hausdorff distance (HD) losses to minimize the difference between GT and prediction in the segmentation process. In the following section, we will describes the BM-KFO and HD loss, respectively.

BM-KFO: bidirectional meta-Kronecker factored optimizer. *MAML: model-agnostic meta-learning.* The MAML algorithm is the basis of our method. We aim to find an objective function that optimizes the meta initial parameter θ which can be quickly applied to an unknown task. The training data, denoted as D , used for meta-learning include a support and query sets. The loss function with the meta-initial parameter θ is expressed as $L(\theta)$, and task-specific parameters are then computed as follows:

$$\theta' = \mathbb{E}_{D \sim p(D)}[(\theta - \alpha \nabla L_D(f_{\theta}))], \tag{2}$$

where α represents the learning rate, and $p(D)$ the distribution of each task’s training data. Meta-initial parameter is updated using loss which calculated by task-specific adapted parameters with gradient descent as:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{D \sim p(D)} L_{D_i}(f_{\theta_i}), \tag{3}$$

where β represents the meta step size. $p(D)$ is the distribution of the data and the sampled new task D_i is used to update the parameter θ of f_{θ} . Thus, it can be explained in three steps: (1) Calculate the gradient of the loss function $L_{D_i}(f_{\theta_i})$ concerning the model parameters θ' , evaluated on the task data D_i , (2) compute the sum of the

loss over all tasks sampled from $p(D)$, and (3) determine the learning rate β and update θ . The basic algorithm of the same method as in the general case is summarized in Algorithm 1.

Algorithm 1 MAML: Model-Agnostic Meta-Learning

Input: $p(D)$: distribution over tasks, α, β : step size hyper-parameters, θ : randomly initialized

while not done **do**

Sample batch of tasks $D_i \sim p(D)$

for step $i = 1, \dots, D$ **do**

Evaluate $\nabla_{\theta} L_{D_i}(f_{\theta})$ with respect to K examples

Compute adapted parameters with gradient descent: $\theta' = \theta - \alpha \nabla_{\theta} L_{D_i}(f_{\theta})$

end for

Update $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{D_i \sim p(D)} L_{D_i}(f_{\theta'})$

end while

The inner loop statement optimizes the task-specific parameter θ' through gradient descent during meta-training. The outer loop parameter θ' is computed by the inner loop code. The meta-initial parameter is updated by gradient of loss function computed over each task, which is referred as *meta-testing*. By this inner-outer loop design, network can start optimization from generalized initial point. It is an algorithm for meta-testing with the optimal point θ' as input. D_i is tested with unseen data, and better generalization is possible.

Meta-Kronecker factored optimizer. Inspired by a MAML algorithm called META-KFO²³, we devised a meta-optimizer that is independent of the model parameters. The function U_{ξ} is a meta-optimizer that defines for updating model parameters²³. The meta-optimizer is to train the model parameters and optimizer parameters jointly. With optimizer parameters ξ , we use this optimizer to gradient updates:

$$L(\theta) = \mathbb{E}_{D \sim p(D)} [loss_D(\theta - \alpha U_{\xi}(\nabla_{\theta} loss_D(\theta)))], \quad (4)$$

$$U_{\xi}(\nabla_{\theta} loss(\theta)) = f(\nabla_{\theta} loss(\theta); \phi), \quad (5)$$

where $L(\theta)$ represents the objective function, which is defined as the expectation over a distribution of datasets D drawn from a probability distribution $p(D)$. The function $loss_D(\theta)$ represents the loss function for a particular dataset D and model parameters θ . The term $\alpha U_{\xi}(\nabla_{\theta} loss_D(\theta))$ denotes the update rule for the meta-parameters θ , where α is the learning rate and $U_{\xi}(\nabla_{\theta} loss_D(\theta))$ represents the update for θ obtained from the meta-optimizer with optimizer parameter ξ . In Eq. (5), $f(\cdot)$ represents a small parameterized neural network that does not increase the entire model parameters that are independent of the model parameter θ . Where ξ denotes the hyper-parameters of the function U_{ξ} and ϕ are values used to optimize other hyper-parameters. Based on these two methods, we differentiated the parameters of the model and optimizer and constructed a meta-train optimizer and a meta-test optimizer. Unlike the MAML method, which only updates model parameters during learning, model parameters, and optimizer parameters are separated, and opti-parameters are simultaneously updated starting with the initial model parameters. These updates are applied to the meta-train optimizer and the meta-test optimizer, respectively, and finally, the updated parameter creates θ^* . The proposed method can be found in the Bidirectional meta-Kronecker factored optimizer (BM-KFO) section.

META-KFO provides better meta-learning capabilities with a shallow network. However, META-KFO is an external meta-optimizer that works inside the network. Therefore, the meta-optimizer U_{ξ} have tens of thousands of parameters, which directly impacts to the network complexity. To reduce the computational complexity and memory usage, we used the Kronecker product to decompose the matrix as a convenient representations. Specifically, given any $m \times n$ matrix A and $p \times q$ matrix B , the Kronecker product between A and B is defined as:

$$A = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mn} \end{bmatrix} \in Mat(m, n; \mathbb{R}), \quad B = \begin{bmatrix} B_{11} & \dots & B_{1q} \\ \vdots & \ddots & \vdots \\ B_{p1} & \dots & B_{pq} \end{bmatrix} \in Mat(p, q; \mathbb{R}), \quad (6)$$

$$A \otimes B \in Mat(mp, nq; \mathbb{R}), \quad A \otimes B = \begin{bmatrix} A_{11}B & \dots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \dots & A_{mn}B \end{bmatrix} \in \mathbb{R}^{mp \times nq}, \quad (7)$$

where A is an $m \times n$ matrix, and B is a $p \times q$ matrix. $A \otimes B$ denotes the Kronecker product of A and B , which is a matrix of size $mp \times nq$ with real entries. The entries of $A \otimes B$ are obtained by taking each entry A_{ij} of A and multiplying it by the entire matrix B . More specifically, the (i, k) -th entry of $A \otimes B$ is given by $(A \otimes B)_{ik} = A_{ij}B_{kl}$, where i and j are such that $1 \leq i \leq m$ and $1 \leq j \leq n$, and k and l are such that $1 \leq k \leq p$ and $1 \leq l \leq q$.

The parameters of the optimizer are decomposed to the following Kronecker factorization.

$$W = R^T \otimes L \in \mathbb{R}^{m \times n}, \text{ for } R \in \mathbb{R}^{m \times n} \text{ and } L \in \mathbb{R}^{n \times n}. \quad (8)$$

Among several different decomposition methods, we used Kronecker-factored decomposition since: (1) the computational and memory cost of the Kronecker-product are acceptable, (2) $R^T \otimes L$ is full-rank whenever L, R are full-rank, and (3) the identity matrix lies in the span of Kronecker-factored matrices. In particular, this last motivation allows meta-optimizers to recover the gradient descent update by letting R, L be the identity²³. More specifically, in Kronecker-factored decomposition, the rank is closely associated with the dimensionality reduction of the decomposed tensor. When the learning parameters are decomposed and the rank of the decomposed tensor is lower, dimensionality reduction and information compression takes place. In other words, it is possible to reduce unnecessary information and obtain core information. On the other hand, if the rank of the decomposed tensor is full-rank, the dimension is equal to the size of the original tensor, and the original tensor can be completely recovered. Therefore, rank plays a major role in terms of information loss and efficient decomposition in Kronecker-factored decomposition. Lower ranks do not perfectly reproduce the original dimensionality, so a balance must be struck between dimensionality reduction and loss of information.

The effect of the number of convolutional layers on the adaptation performance can be compared with the KFO and the MAML method. First, the performance of both methods improves as the model size increases. In addition to better meta-learning, the improvement may come from an increase in the capacity of the model to learn the target task. Second, the net gain of meta-kfo tends to decrease with increasing layer size. In other words, the advantage of directly transforming the gradients using an external meta-optimizer is reduced because the upper layers of the larger model have more capacity to meta-learn to control their lower layers. Third, it is more performance to use Kronecker decomposition of parameters compared to MAML²³.

Bidirectional meta-Kronecker factored optimizer (BM-KFO). In the inner loop of MAML, Meta-KFO and other meta-learning frameworks only computes meta-optimizer parameters by the training loop. After obtaining the computed meta-optimizer parameters, a test loop in the inner loop assesses their effectiveness. To achieve this, we need to divide the k support sets into training and testing sets. While this method is appropriate for numerous k -shot situations, in low-shot learning contexts, such as the one-shot learning setting of our target dataset, we have only one training and testing set available for training the meta-optimizer. Consequently, in a low-shot setting, it is necessary to employ a test set within the inner loop. To solve this problem, we present a meta-learning framework consisting of meta-train optimizer and meta-test optimizer. It is inspired by a general learning process that train-validation-test and actor-critic algorithm in reinforcement learning³⁹.

Algorithm 2 BM-KFO: Bidirectional Meta-Kronecker Optimizer

Input: Fast learning rate α , Initial parameters $\theta^{(init)}$, Optimizer Opt
while $\theta^{(init)}, \xi^{(init)}$ not converged **do**
 Sample batch of tasks $D \sim p(D)$
 $\theta_1 = \theta^{(init)}, \xi_1 = \xi^{(init)}$
 for train step $i = 1, \dots, D$ **do**
 Compute loss $l_D(\theta_i)$
 Compute adapted parameters with gradient descent: $\nabla_{\theta_i} l_D(\theta_i)$ and $\nabla_{\xi_i} l_D(\theta_i)$
 Update the meta-optimizer parameters $\xi_{i+1} = \xi_i - \alpha \nabla_{\xi_i} l_D(\theta_i)$
 Compute the model update $U_{\xi_{i+1}}(\nabla_{\theta_i} l_D(\theta_i))$
 Update the model parameters $\theta_{i+1} = \theta_i - U_{\xi_{i+1}}(\nabla_{\theta_i} l_D(\theta_i))$
 end for
 Update model and meta-optimizer initializations
 $\theta^{(train)} \leftarrow \theta^{(init)} - Opt(\nabla_{\theta^{init}} l_D(\theta_D))$
 $\xi^{(train)} \leftarrow \xi^{(init)} - Opt(\nabla_{\xi^{init}} l_D(\theta_D))$
 Sample batch of tasks $D \sim p(D)$
 $\theta'_1 = \theta^{(train)}, \xi'_1 = \xi^{(train)}$
 for test step $k = 1, \dots, D$ **do**
 Compute loss $l_D(\theta'_k)$
 Compute gradient descent: $\nabla_{\theta'_k} l_D(\theta'_k)$ and $\nabla_{\xi'_k} l_D(\theta'_k)$
 Update the meta-optimizer parameters $\xi'_{k+1} = \xi'_k - \alpha \nabla_{\xi'_k} l_D(\theta'_k)$
 Compute the model update $U_{\xi'_{k+1}}(\nabla_{\theta'_k} l_D(\theta'_k))$
 Update the model parameters $\theta'_{k+1} = \theta'_k - U_{\xi'_{k+1}}(\nabla_{\theta'_k} l_D(\theta'_k))$
 end for
 Update model and meta-optimizer initializations
 $\theta^{*(test)} \leftarrow \theta^{(train)} - Opt(\nabla_{\theta^{train}} l_D(\theta'_D))$
 $\xi^{*(test)} \leftarrow \xi^{(train)} - Opt(\nabla_{\xi^{train}} l_D(\theta'_D))$
end while

Algorithm 2 takes the parameters of the model and optimizer as input. Meta-train-optimizer and meta-test-optimizer are processed during batch. It is divided into two blocks θ' and ξ' , output parameters from meta-train feed to the next block as input. For-loop statements proceed in the form of updating meta-optimizer and model parameters, respectively. The output of two blocks, θ^* and ξ^* , respectively, are final meta-initial point. Figure 3 is visualization of Algorithm 2.

The proposed BM-KFO takes the form:

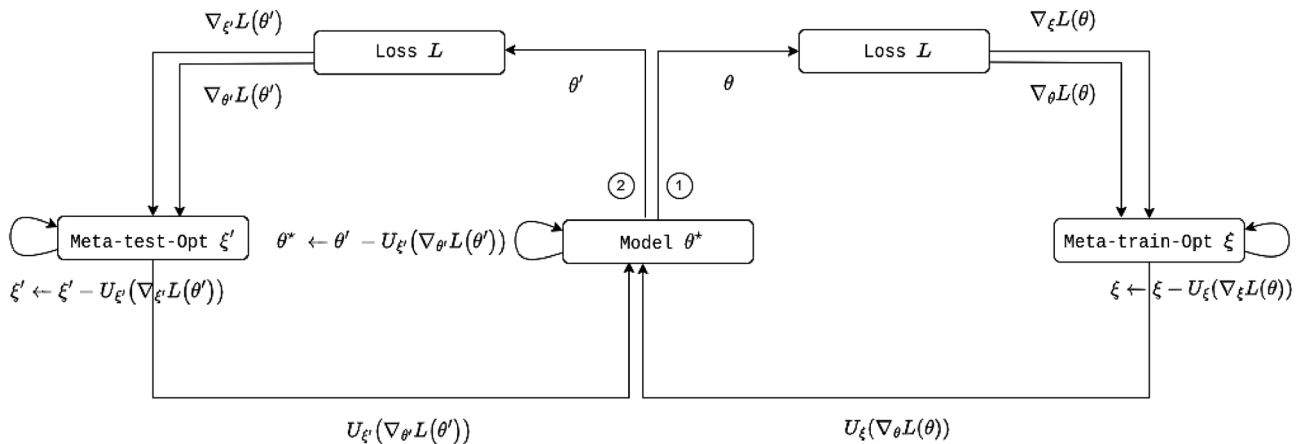


Figure 3. Bidirectional meta-Kronecker factored optimizer.

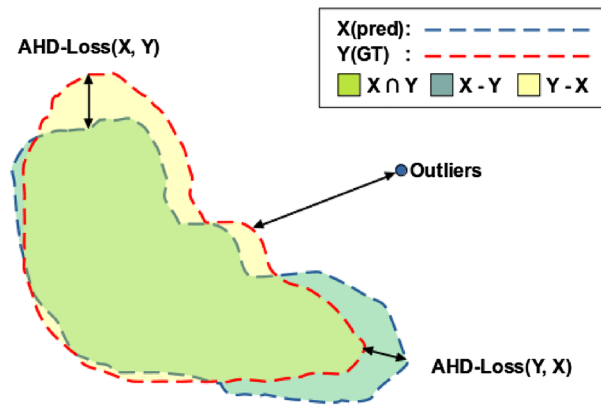


Figure 4. Visualization of AHD-Loss. It computes the distance between GT and prediction from both perspectives. Thus, it is robust to outliers.

$$\theta' \leftarrow \theta - \beta \frac{\partial L(\theta)}{\partial \theta}, \xi' \leftarrow \xi - \beta \frac{\partial L(\theta)}{\partial \xi}, \text{ and (BM - KFO) : } \theta^* \leftarrow \theta' - U_{\xi'}(\nabla_{\theta'} \text{Loss}(\theta')), \quad (9)$$

where θ' and ξ' take θ and ξ of each parameter as inputs, and produce $\nabla_{\theta} \text{Loss}(\theta)$. This process is a meta-train-optimizer, and the process of entering θ' and ξ' once again and finding θ^* becomes the meta-test-optimizer. θ^* is the final model parameter passed through meta-train optimizer and meta-test optimizer. The detailed BM-KFO framework pseudo-code is described in Algorithm 2, and the overall meta-optimizer flow is shown in Fig. 3.

HD-loss: Hausdorff distance loss. Hausdorff distance is a representation of the difference in the boundary between GT and prediction as a distance. It measures the distance between the set of coordinates of the difference between GT and prediction, and is widely used in medical image segmentation. Average Hausdorff distance (AHD-Loss) is applied to measure the performance of various applications, including cerebrovascular segmentation⁴⁰, extra-cranial facial nerve segmentation⁴¹. We used the average Hausdorff distance as our objective function to alleviate the unstable learning problem in few-shot learning. HD-Loss defines the mismatch of two sets as a nonlinear function that measures the distance. The max–min method using a general distance function is as follows:

$$D_{\text{distance}}(x, y) = \|x - y\|. \quad (10)$$

Given two sets $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, the Hausdorff distance, denoted as $D(x, Y)$, from a point in X to Y is defined as:

$$D_{\text{distance}}(x, Y) = \min_{y \in Y} d(x, y). \quad (11)$$

The hausdorff distance from set X to Y is:

$$HD(X, Y) = \max_{x \in X} D_{\text{distance}}(x, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|. \quad (12)$$

Conversely, the distance from set Y to X is:

$$HD(Y, X) = \max_{y \in Y} D_{\text{distance}}(y, X) = \max_{y \in Y} \min_{x \in X} \|y - x\|. \quad (13)$$

When $Prediction = X, GT = Y$, the final expression is as follows:

$$L_{HD\text{Loss}}(Pred, GT) = \max(HD(Pred, GT), HD(GT, Pred)). \quad (14)$$

AHD-loss: average Hausdorff distance loss. The average Hausdorff distance between two finite sets of points X and Y is average of (14). HD-loss has a problem that it is sensitive to outliers. On the other hand, the average Hausdorff distance in the direction from point set X to Y is specified as the sum of all minimum distances from all points from point set X to Y divided by the number of points in X . The AHD-Loss can be calculated as both directions. Figure 4 is a visualization of AHD-Loss. It computes the distance between GT and predictions. It is form-aligned and robust to outliers than general Hausdorff distance. It is defined as follows using the AHD from X to Y and the AHD from Y to X :

$$L_{AHD\text{Loss}}(X, Y) = \left(\frac{1}{X} \sum_{x \in X} \min_{y \in Y} D_{\text{dist}}(x, y) + \frac{1}{Y} \sum_{y \in Y} \min_{x \in X} D_{\text{dist}}(x, y) \right) / 2. \quad (15)$$

$$D_{\text{manhattan}}(X, Y) = \sum_i^n (|x_i, y_i|). \quad (16)$$

$$D_{\text{euclidean}}(X, Y) = \left(\sum_i^n (|x_i, y_i|^2) \right)^{\frac{1}{2}}. \quad (17)$$

$$D_{\text{chebyshev}}(X, Y) = \max_i (|x_i, y_i|). \quad (18)$$

As shown in (15), AHD-Loss calculates the difference between GT and Prediction from both perspectives, and can make the model focused more on the whole shape of the organ than intersections. It is also robust to outlier points, and the distance measurement methods include Euclidean, Chebyshev, and Manhattan distances. Equation (16) is the Manhattan distance. It is also called L1 distance and it is equal to the sum of the differences of each coordinate. Equation (17) is the Euclidean distance. It is also called L2 distance, and it is the shortest distance between coordinates. Compared to the Eqs. (16,17) distances, we use Chebyshev distance because it can be measured the distance in all directions.

Our final objective function is sum of AHD- and cross-entropy losses.

$$L_{CE} = - \sum_{i=1}^n y_i \log(\hat{y}_i), \quad (19)$$

where y_i is GT and \hat{y} is prediction.

$$L_{\text{Total}} = L_{CE} + L_{\text{AHDLoss}}. \quad (20)$$

Experimental results

Implementation details. *Dataset.* To evaluate the proposed BM-KFO and Hausdorff distance loss, we used the abdominal MRI (Abd-MRI) segmentation dataset for training. It is part of Combined Healthy Abdominal Organ Segmentation Challenges¹² from ISBI 2019¹². CHAOS dataset can be downloaded via the following link. Go to the next URL: <https://chaos.grand-challenge.org/Download/>. It contains 40 different patients' images of resolution 224×224 . This dataset does not include any tumors or lesions at the borders of the annotated organs of interest (i.e. liver, kidney, and spleen). The number of slices per person is between 26 and 50 (average 36). It provides 1594 slices for training and 1537 slices for testing.

Training. The proposed method was implemented with PyTorch based on SSL-ALPNet³¹. As a backbone for feature extraction, we used fully-convolutional ResNet101, which was pre-trained using part of MS-COCO dataset. Although ResNet101 is not the latest model, we did not have any modifications on it for a fair comparison with existing methods.

We used an NVIDIA RTX 3090 GPU for training network. The learning rate was 0.001 with a stepping decay rate of 0.98 per 1000 iterations with Stochastic Gradient Descent (SGD) and Adam optimizers. After 100,000 iterations, the proposed objective function converged, and it took about hours.

Evaluation metric. To evaluate the proposed method elaborately, we used *precision*, *recall*, and *dice coefficients* as metrics. Precision and recall are trade-offs. If the network predicts a very large prediction about 2–3 times than the ground truth, it is sufficient to cover all ground truth. In this case, precision becomes close to 1, meaning a near perfect match. On the other hand, recall is very low because its denominator is a number of predictions.

Precision, also known as positive predictive value and specificity, describes the purity of positive predictions relative to the ground truth. It measures how many predictions actually have matching ground truth segments.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{TP}{\# \text{ground truth}}. \quad (21)$$

Recall, also known as sensitivity, effectively describes the completeness of our predictions relative to the ground truth. It measures all predictions given and the number that captures true positives.

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{TP}{\# \text{prediction}}. \quad (22)$$

Dice similarity, also known as F1 score or Sørensen–Dice index, is used to gauge the similarity of two segments. It is defined as the harmonic mean between precision and recall.

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{(TP + FP) + (TP + FN)}, \quad (23)$$

where $|X|$ and $|Y|$ represent the cardinality of segments.

Method	Lower		Upper		Mean
	LK	RK	Spleen	Liver	
SE-Net ⁴⁴	62.11	61.32	51.80	27.43	50.66
Vanilla PANet ³⁰	53.45	38.64	50.90	42.26	46.33
ALPNet-init	19.28	14.93	23.76	37.73	23.93
ALPNet	53.21	58.99	52.18	37.32	50.43
SSL-PANet	47.71	47.95	58.73	64.99	54.85
SSL-ALPNet	73.63	78.39	67.02	73.05	73.02
CRAPNet ⁴⁵	74.66	82.77	70.82	73.82	75.52
Proposed method	79.89	86.86	72.13	73.43	78.07

Table 1. Experiment results(in Dice score) on abdominal MRI images under setting 2. Significant values are in [bold].

Quantitative evaluation. To evaluate 2D segmentation on 3D images, we follow the evaluation method established by⁴³, where support and query sets are data of different patients. To evaluate the generalization ability in few-shot learning, the medical image segmentation method follows the evaluation method established by Roy et al.'s setting 1⁴³.

However, setting 1 has a problem in that the test class appears in the background of the training data during training. In other words, few-shot model learns each class respectively using inner-loop of BM-KFO. By updating meta-initial points computed by inner-loop, we can optimize quickly to other class. Each class is independently used for training, but input images are the same. Thus, the same input images cause unwanted spatial bias to the meta-optimizer. To handle this issue, we performed additional experiments with our method using setting 2, which removes test class relevant data. In setting 2, the model trained with the images in the test class is completely removed from the training data. In setting 1, label partitioning is unnecessary to train. On the other hand, in setting 2, data are divided into the upper and lower abdomen with liver, spleen, and left and right kidneys. It can make a separate group for testing. In each experiment, all the training data belonging to the test data group are removed. To simulate the sparsity of labeled data in applications, all experiments are conducted in one-way one-shot settings³¹.

Tables 1 and 2 show experimental results with setting 1 and setting 2, respectively. In both settings, the proposed method outperforms the other methods by a large margin. Especially, setting 1 provides relatively higher improvements than setting 2. Here, setting 1 and setting 2 refer to the setup to evaluate the generalization ability for unseen testing classes during training. Setting 1 corresponds to a scenario where the testing class can appear as the background of training data in a few-shot segmentation experiment for medical images. On the other hand, setting 2 refers to a situation in which the test class is completely invisible by excluding images that contain the testing class from the training dataset³¹. Table 2 presents two methods for few-shot medical image segmentation: SSL-RPNet⁴⁶ and CRAPNet⁴⁵. SSL-RPNet is a supervised approach that uses two core modules to capture local relational features and iteratively refines a segmentation mask. This method employs a prototype network with a context relational encoder, leading to improved performance in few-shot medical image segmentation. On the other hand, CRAPNet addresses the issue of neglecting the connection between the query set and support set, which is present in existing medical image segmentation studies using Prototype Networks. It represents the latest few-shot methodology that preserves pixel-related relationships between medical images.

The proposed method is compared with various state-of-the-art methodologies, including CRAPNet and SSL-RPNet. In setting 1, CRAPNet achieves a 4.3% higher performance than the proposed method for spleen segmentation, but the overall average is slightly lower. SSL-RPNet has lower overall performance than CRAPNet. In setting 2, CRAPNet shows an approximately 0.4% improved performance over the proposed method for liver segmentation.

Method	Kidneys		Spleen	Liver	Mean
	LK	RK			
SE-Net ⁴⁴	45.78	47.96	47.30	29.02	42.51
Vanilla PANet ³⁰	30.99	32.19	40.58	50.40	38.53
ALPNet	44.73	48.42	49.61	62.35	51.28
SE-FSS ⁴³	62.56	65.81	48.93	40.32	54.38
SSL-PANet	58.83	60.81	61.32	71.73	63.17
SSL-ALPNet	81.92	85.18	72.18	76.10	78.84
SSL-RPNet ⁴⁶	71.46	81.96	73.55	75.99	75.74
CRAPNet ⁴⁵	81.95	86.42	74.32	76.46	79.79
Proposed method	84.14	87.07	70.79	77.42	79.85

Table 2. Experiment results(in Dice score) on abdominal MRI images under setting 1. Significant values are in [bold].

Method	Training time (s)	Method—baseline	Training time (ms/iter)
SSL-ALPNet + CE	809	–	40.01
SSL-ALPNet + CE + Boundary loss	1114	305	55.71
SSL-ALPNet + CE + Boundary loss + HDLoss	1416	607	70.81
SSL-ALPNet + MAML	899	90	44.95
Proposed method	1522	713	76.10

Table 3. Evaluations on the various meta-learning methods and loss functions to the baseline and proposed method. The required memory usage during training (min–max), the required training time, and the time required for 1 iteration are compared. It is based on 20 epochs. Significant values are in [bold].

Method	Dice score	Precision	Recall	Mean
SSL-ALPNet + CE	71.8	70.1	74.3	72.06
SSL-ALPNet + CE + Boundary loss ⁴⁷	66.3	90.3	52.9	69.83
SSL-ALPNet + CE + Boundary loss + HDLoss	69.4	92.4	56.2	72.66
Proposed method (+ CE + AHD-Loss)	74.0	70.6	79.2	74.6

Table 4. Impact of each loss function on the baseline model and the proposed method, excluding the meta-learning approach. Significant values are in [bold].

Method	Dice score	Precision	Recall	Mean
SSL-ALPNet	73.42	78.43	69.87	73.90
SSL-ALPNet + MAML	69.50	80.79	62.32	70.87
SSL-ALPNet + Meta-SGD ⁴⁸	73.00	82.23	70.20	75.14
SSL-ALPNet + Meta-curvature ⁴⁹	80.00	96.30	71.80	82.70
Proposed method(+ BM-KFO)	83.40	87.10	81.01	83.83

Table 5. Result of experimenting with various meta-learning methods without loss function. Significant values are in [bold].

Ablation study. We carried out an ablation study that focused on three different perspectives, such as training time, loss function, and meta-learning method, to conduct a more comprehensive analysis of the proposed method.

Furthermore, we conducted additional experiments on training time per iteration, and training time to achieve convergence as shown in Table 3. We devised methods that integrate different combinations of components to ensure a fair comparison of each element of the proposed approach. Compared to the baseline method, the proposed method exhibits increased resource consumption in terms of memory and training time. This can be attributed to the extra inner loop that trains the meta-test optimizer and the computation time necessary for AHD loss. In contrast, MAML solely trains the meta-train optimizer in the inner loop.

Table 4 shows the result of experiments conducted to evaluate the impact of the loss function on both the baseline model and the proposed method. We utilized Recall, Precision, and Dice Score metrics to evaluate the performance of the proposed method. In each experiment, the baseline loss, cross-entropy loss (CE-Loss), boundary loss, and HD loss functions are compared. Boundary loss led to low Dice and recall scores but high precision scores due to its focus on the ground truth boundary. However, this led to the network struggling to fit irregular shapes, resulting in low recall scores. In the medical image segmentation task, identifying abnormal organ shapes is more crucial than normal ones. Combining Boundary loss with HD-Loss resulted in a slight increase in Dice and recall scores. Although the proposed method exhibits lower precision than other methods, it demonstrates high performance in terms of Dice, recall, and mean. This indicates that the proposed method produces more balanced outputs compared to other methods.

Table 5 presents a comparison of meta-learning methods including MAML, Meta-SGD⁴⁸, and Meta-Curvature⁴⁹ with the baseline. When using MAML alone, precision is higher than baseline and Dice and Recall score is slightly lower. The method incorporating Meta-Curvature to the baseline has the highest precision evaluation, but the average of all evaluations showed that it is lower than the proposed method using BM-KFO. This suggests that the proposed BM-KFO approach significantly outperforms other methods.

Qualitative evaluation. Figures 5, 6, 7, and 8 show visual comparison with SSL-ALPNet³¹, proposed method, and corresponding ground truth segmentation on the setting 2. For all experiments, the baseline did

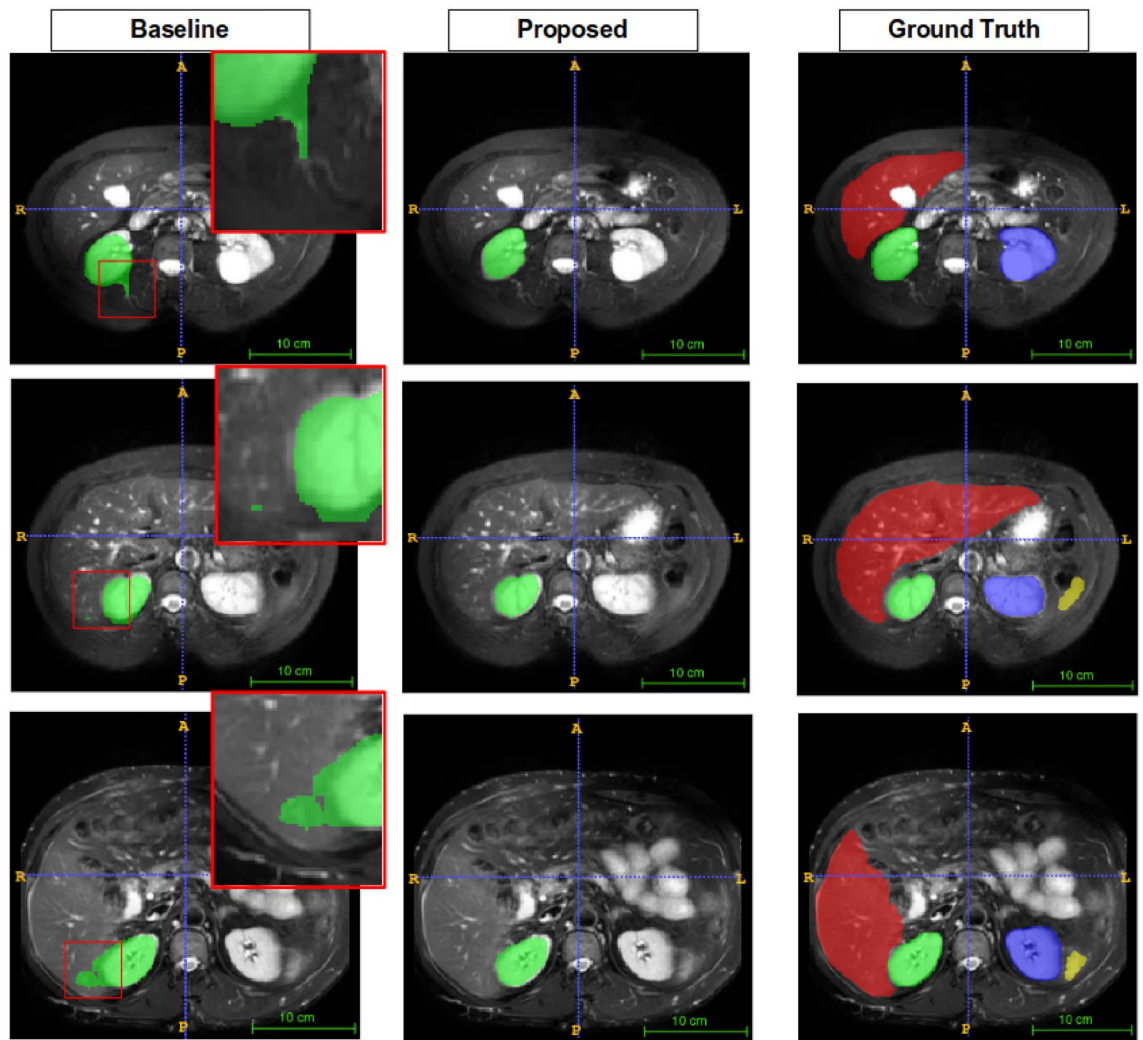


Figure 5. Visual comparison to right kidney.

not consider the shape of organs. On the other hand, the proposed BM-KFO prediction preserved the shape of the object through fast optimization.

As shown in Fig. 5, the proposed method accurately segmented the right kidney, but the baseline produced an irregular shape. Compared to the baseline methods, the proposed method better focused on the difference in boundaries and the shape of organs, and also produced better results in various shapes and sizes.

As shown in Fig. 6, the shape of the left and right kidneys is closer to GT.

Figure 7 compares the segmentation performance of the liver, which has various shapes per person. For this reason, it is hard to optimize quickly using few-shot learning scheme. Predictions of baseline did not cover corresponding ground truth liver. This class shape variance is still a challenging issue in few-shot learning research. But, we observe that the proposed method segments the liver relatively well.

Figure 8 is a segmented image of the spleen. It has a problem that other organs have similar brightness values. This feature confuses the network and eventually leads to noisy predictions. Baseline suffered from this problem, but the proposed method shows shape-aware outputs.

Conclusion

In this paper, we presented the BM-KFO framework and AHD-Loss for meta-learning-based medical image segmentation. The proposed method addresses the problem of small data in image segmentation by learning a bidirectional few-shot task that optimizes both the model parameters and the optimizer parameters during training. We applied Kronecker factored and Hausdorff distance loss for efficient segmentation of medical images as well as parameter optimization. By combining these methods with meta-learning, we improved the efficiency of parameter computation and the accuracy of semantic segmentation in few-shot tasks for medical image segmentation. Moreover, this approach may be useful for identifying unseen lesions in medical applications. Compared to existing state-of-the-art methods, the proposed meta-learning framework for medical image segmentation demonstrated an average performance improvement of 1.2%, BM-KFO method by 9.07%, and AHD-Loss function by 2.54%, while producing faster inference and better segmentation results. Although the proposed method increases training time costs due to bidirectional optimization of the model and optimizer for

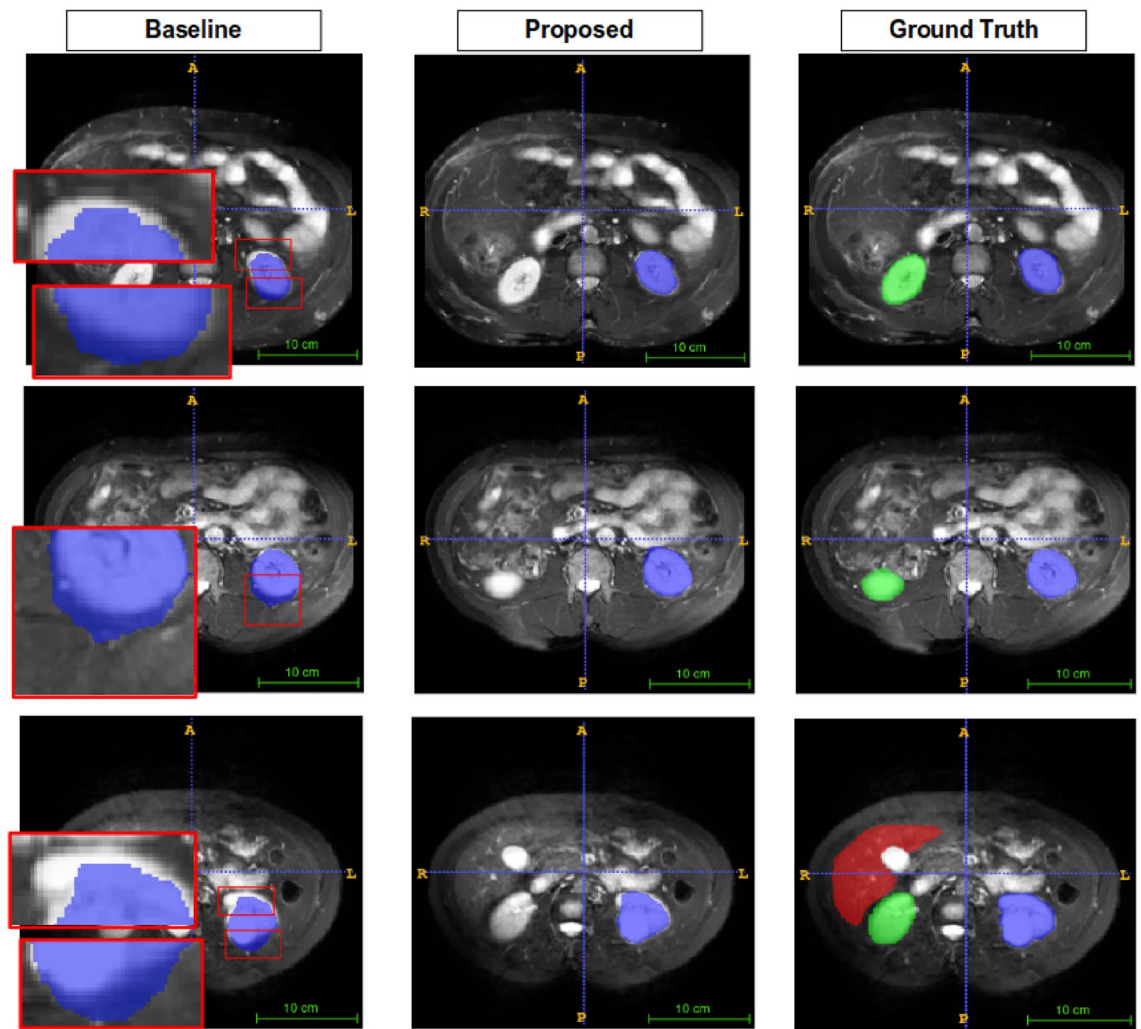


Figure 6. Visual comparison to left kidney. The baseline does not predict segments well, but the proposed method predicts segments well despite of only few data available.

few-shot parameters during training, this can be addressed by changing the structure of the network architecture and improving hardware. Additionally, as the learning time is acceptable in real-world few-shot tasks, the meta-learning approach can assist in the evaluation of medical images across various domains. Thus, the development of meta-medical applications becomes an important field for future research.

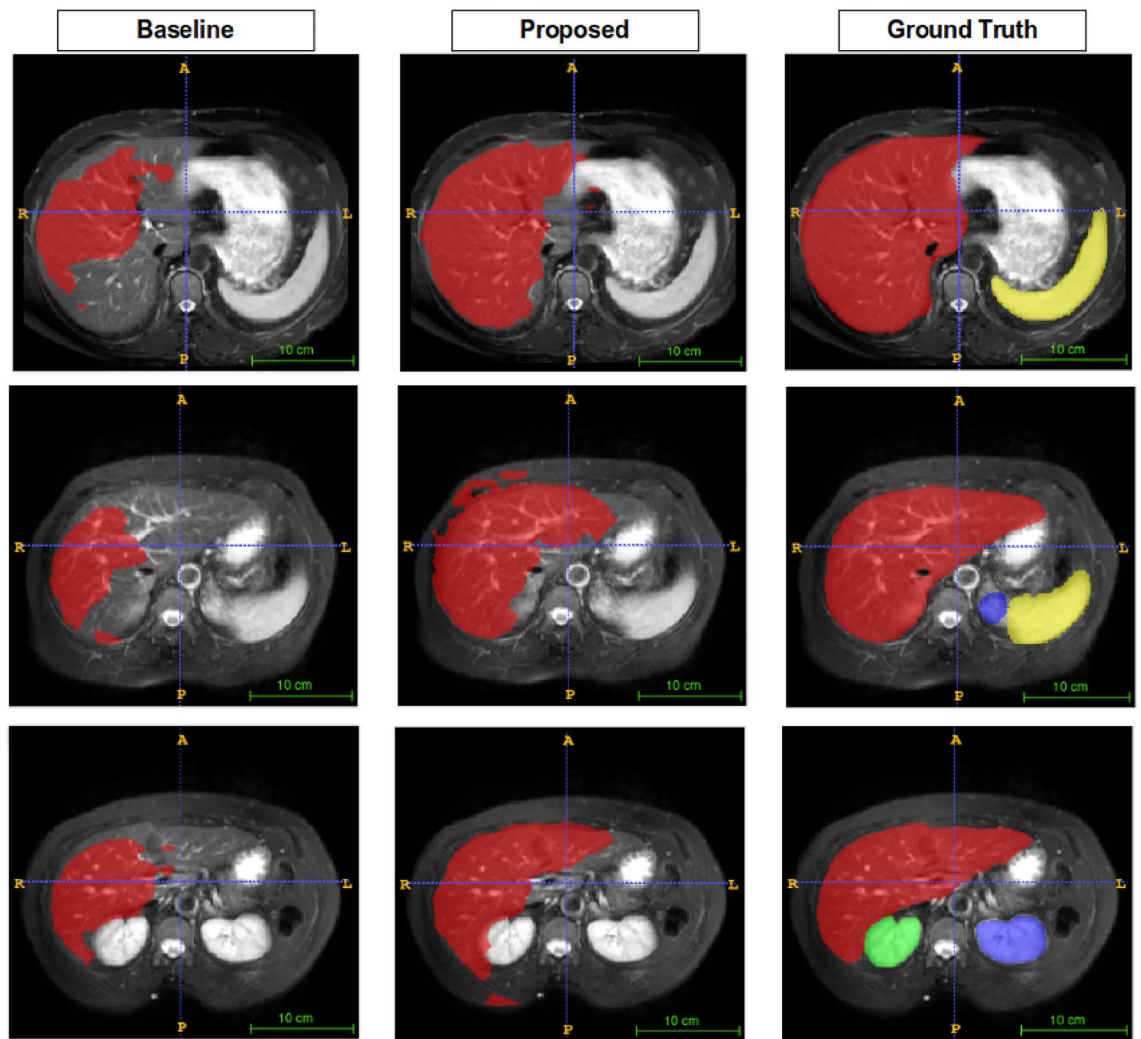


Figure 7. Visual comparison to the liver. It has various shapes and low brightness. The baseline does not predict accurately, but the proposed method predicts shape-aware direction.

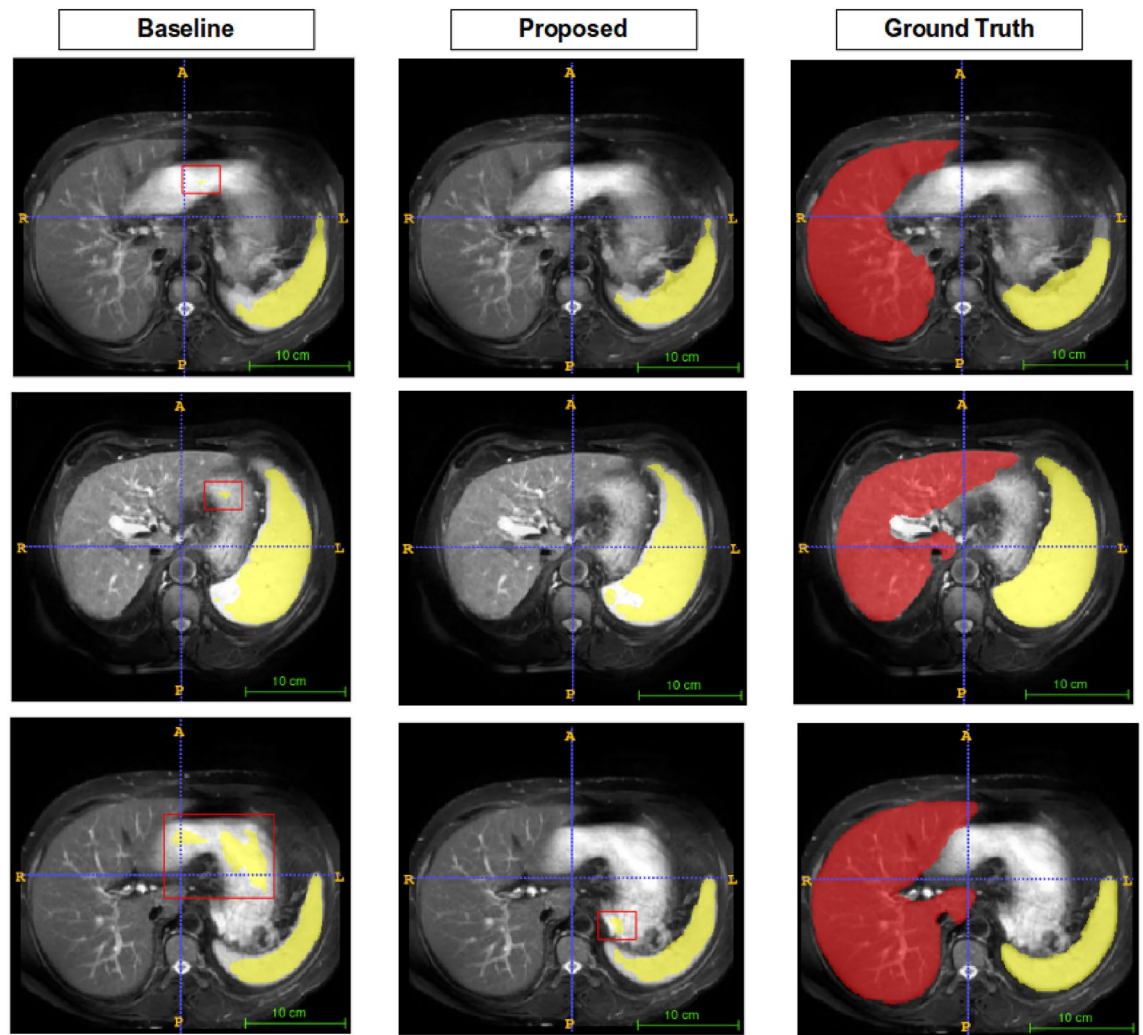


Figure 8. Visual comparison to the spleen. Despite of confused boundary problem, the proposed method predicts shape-aware direction.

Data availability

All data generated or analyzed during this study are included in this published article.

Received: 29 November 2022; Accepted: 16 May 2023

Published online: 19 May 2023

References

1. Boykov, Y. Y. & Jolly, M.-P. Interactive graph cuts for optimal boundary and region segmentation of objects in nd images. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, 105–112 (IEEE, 2001).
2. Chan, T. F. & Vese, L. A. Active contours without edges. *IEEE Trans. Image Process.* **10**, 266–277 (2001).
3. Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. Active shape models—their training and application. *Comput. Vis. Image Underst.* **61**, 38–59 (1995).
4. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (Springer, 2015).
5. Jackson, P. *Introduction to Expert Systems* (Addison-Wesley Longman Publishing Co, 1998).
6. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440 (2015).
7. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder–decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818 (2018).
8. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495 (2017).
9. Isensee, F. *et al.* nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint [arXiv:1809.10486](https://arxiv.org/abs/1809.10486) (2018).
10. Cai, L., Gao, J. & Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **8**, 713 (2020).
11. Zhou, Y. *et al.* Prior-aware neural network for partially-supervised multi-organ segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10672–10681 (2019).

12. Kavur, A. E. *et al.* Chaos challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med. Image Anal.* **69**, 101950 (2021).
13. Ji, X., Henriques, J. F. & Vedaldi, A. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9865–9874 (2019).
14. Snell, J., Swersky, K. & Zemel, R. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, Vol. 30 (2017).
15. Sung, F. *et al.* Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1199–1208 (2018).
16. Garcia, V. & Bruna, J. Few-shot learning with graph neural networks. arXiv preprint [arXiv:1711.04043](https://arxiv.org/abs/1711.04043) (2017).
17. Vinyals, O. *et al.* Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, Vol. 29 (2016).
18. Fei-Fei, L., Fergus, R. & Perona, P. One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 594–611 (2006).
19. Lake, B., Salakhutdinov, R., Gross, J. & Tenenbaum, J. One shot learning of simple visual concepts. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33 (2011).
20. Nagabandi, A., Finn, C. & Levine, S. Deep online learning via meta-learning: Continual adaptation for model-based rl. arXiv preprint [arXiv:1812.07671](https://arxiv.org/abs/1812.07671) (2018).
21. Chen, J., Zhan, L.-M., Wu, X.-M. & Chung, F.-L. Variational metric scaling for metric-based meta-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, 3478–3485 (2020).
22. Finn, C., Abbeel, P. & Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 1126–1135 (PMLR, 2017).
23. Arnold, S., Iqbal, S. & Sha, F. When maml can adapt fast and how to assist when it cannot. In *International Conference on Artificial Intelligence and Statistics*, 244–252 (PMLR, 2021).
24. Shaban, A., Bansal, S., Liu, Z., Essa, I. & Boots, B. One-shot learning for semantic segmentation. arXiv preprint [arXiv:1709.03410](https://arxiv.org/abs/1709.03410) (2017).
25. Rakelly, K., Shelhamer, E., Darrell, T., Efros, A. & Levine, S. Conditional networks for few-shot semantic segmentation. In *ICLR 2018 Workshop* (2018).
26. Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, Vol. 30 (2017).
27. Hu, T. *et al.* Attention-based multi-context guiding for few-shot semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 8441–8448 (2019).
28. Schlichtkrull, M. *et al.* Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, 593–607 (Springer, 2018).
29. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907) (2016).
30. Wang, K., Liew, J. H., Zou, Y., Zhou, D. & Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9197–9206 (2019).
31. Ouyang, C. *et al.* Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In *European Conference on Computer Vision*, 762–780 (Springer, 2020).
32. Terrell, G. R. & Scott, D. W. Variable kernel density estimation. *Ann. Stat.* **20**, 1236–1265 (1992).
33. Sun, L. *et al.* Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Comput. Biol. Med.* **140**, 105067 (2022).
34. Yu, Q., Dang, K., Tajbakhsh, N., Terzopoulos, D. & Ding, X. A location-sensitive local prototype network for few-shot medical image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 262–266 (IEEE, 2021).
35. Denton, E. L., Zaremba, W., Bruna, J., LeCun, Y. & Fergus, R. Exploiting linear structure within convolutional networks for efficient evaluation. In *Advances in Neural Information Processing Systems*, Vol. 27 (2014).
36. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 240–248 (Springer, 2017).
37. Hayder, Z., He, X. & Salzmann, M. Boundary-aware instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5696–5704 (2017).
38. Karimi, D. & Salcudean, S. E. Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans. Med. Imaging* **39**, 499–513 (2019).
39. Araslanov, N., Rothkopf, C. A. & Roth, S. Actor-critic instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8237–8246 (2019).
40. Livne, M. *et al.* A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Front. Neurosci.* **13**, 97 (2019).
41. Guenette, J. P. *et al.* MR imaging of the extracranial facial nerve with the CISS sequence. *Am. J. Neuroradiol.* **40**, 1954–1959 (2019).
42. IEEE international symposium on biomedical imaging. <https://biomedicalimaging.org/2019/>. Online; Accessed 23 November 2023 (2019).
43. Roy, A. G., Siddiqui, S., Pölsterl, S., Navab, N. & Wachinger, C. ‘squeeze and excite’ guided few-shot segmentation of volumetric images. *Med. Image Anal.* **59**, 101587 (2020).
44. Zhang, C., Lin, G., Liu, F., Yao, R. & Shen, C. 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*, 5217–5226 (2019).
45. Ding, H., Sun, C., Tang, H., Cai, D. & Yan, Y. Few-shot medical image segmentation with cycle-resemblance attention. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2488–2497 (2023).
46. Tang, H., Liu, X., Sun, S., Yan, X. & Xie, X. Recurrent mask refinement for few-shot medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3918–3928 (2021).
47. Chen, X., Qi, D. & Shen, J. Boundary-aware network for fast and high-accuracy portrait segmentation. arXiv preprint [arXiv:1901.03814](https://arxiv.org/abs/1901.03814) (2019).
48. Andrychowicz, M. *et al.* Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems*, Vol. 29 (2016).
49. Park, E. & Oliva, J. B. Meta-curvature. *Advances in Neural Information Processing Systems*, vol. 32 (2019).

Acknowledgements

This work was supported partly by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by Korea government (MSIT) [2021-0-01341, Artificial Intelligent Graduate School Program(Chung-Ang University)], and partly by Field-oriented Technology Development Project for Customs Administration through National Research Foundation of Korea(NRF) funded by the Ministry of Science and ICT and Korea Customs Service (2021M3I1A1097911).

Author contributions

Y.K. designed and developed the algorithm and performed experiment. D.K. prepared training data. Y.M. performed experiment and data analysis. S.K. performed data analysis. J.P. guided the project and wrote original draft. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023