

Dual Meta-Learning with Longitudinally Generalized Regularization for One-Shot Brain Tissue Segmentation Across the Human Lifespan

Yongheng Sun¹, Fan Wang²(✉), Jun Shu¹, Haifeng Wang¹, Li Wang³(✉), Deyu Meng¹, Chunfeng Lian¹(✉)

¹School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

²The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China

³UNC Chapel Hill, Chapel Hill, NC 27599, United States

chunfeng.lian@xjtu.edu.cn, li.wang@med.unc.edu, fan.wang@xjtu.edu.cn

Abstract

Brain tissue segmentation is essential for neuroscience and clinical studies. However, segmentation on longitudinal data is challenging due to dynamic brain changes across the lifespan. Previous researches mainly focus on self-supervision with regularizations and will lose longitudinal generalization when fine-tuning on a specific age group. In this paper, we propose a dual meta-learning paradigm to learn longitudinally consistent representations and persist when fine-tuning. Specifically, we learn a plug-and-play feature extractor to extract longitudinal-consistent anatomical representations by meta-feature learning and a well-initialized task head for fine-tuning by meta-initialization learning. Besides, two class-aware regularizations are proposed to encourage longitudinal consistency. Experimental results on the iSeg2019 and ADNI datasets demonstrate the effectiveness of our method. Our code is available at <https://github.com/ladderlab-xjtu/DuMeta>.

1. Introduction

Accurate brain tissue segmentation is essential in diverse neuroscience and clinical studies, e.g., population analyses of brain cortical architectures and individualized diagnosis of brain diseases. Due to spatiotemporally dynamic changes in brain structures and functions across the human lifespan, brain magnetic resonance images (MRIs) present longitudinally heterogeneous appearances, leading to varying segmentation difficulties at different periods [37]. Infancy and the elderly are two particularly challenging times for tissue segmentation, as a result of (atypical) developmental and degenerative processes. For example, as shown in Fig. 1 (a), the brains of infants before six months old have inverse contrasts between gray matter (GM) and white matter (WM)

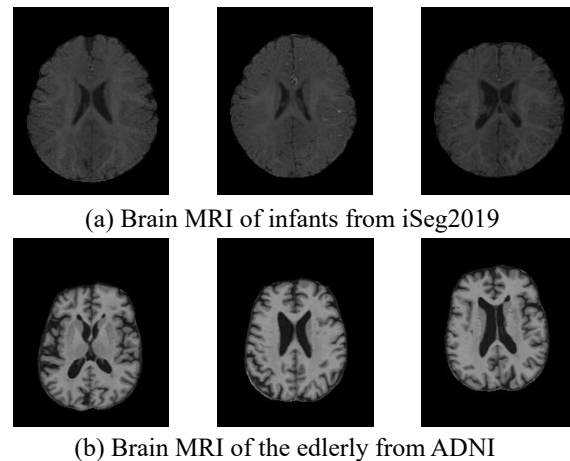


Figure 1. (a) Brain morphology and tissue contrast of 6-month-old infants from the iSeg2019 dataset. (b) Aging and Alzheimer processes of the elderly from the ADNI dataset.

compared with following periods, and the isointense phase around six-month-old exhibits extremely low inter-tissue contrast. In contrast, the brains of the elderly and patients with Alzheimer's disease have enlarged cerebrospinal fluid (CSF) and atrophied GM, as shown in Fig. 1 (b). These challenges significantly hamper the generalization and longitudinal consistency of existing learning-based methods for automatic brain tissue segmentation [25, 39, 18, 9], especially considering that labeled training samples for specific time points could be very limited in practice.

To address the above challenges, some methods in the literature [32, 29] designed contrastive self-supervised learning (SSL) strategies to learn longitudinally consistent representations, which are then fine-tuned on a specific age group for downstream tasks like tissue segmentation. Such SSL methods commonly have three technical limitations from the application perspective. That is, they usually re-

quire longitudinally paired images of the same subjects in the pre-training phase, which is practically hard to satisfy. Besides, the fine-tuned segmentation models are typically restricted to match up with the appearance of a specific age group, sacrificing/losing the generalization capacity to other groups. Moreover, fine-tuning a segmentation model is itself a challenging task, especially when there is not enough labeled training data from the target age group.

In this paper, we revisit and reformulate the idea of longitudinally generalized (or age-agnostic) representation learning. By nature, it is reasonable to make two fundamental assumptions: **1)** independent of changing MRI appearances, the high-order differences between different tissue types (e.g., in semantic space) are relatively stable in terms of the time trajectory; **2)** accurate tissue segmentation requires a reliable mapping function to seamlessly fuse semantic information with age-specific high-resolution image details. In line with such assumptions, we propose a unified meta-learning framework to concurrently learn a universal feature extractor (e.g., the encoder) for age-agnostic (i.e., longitudinally consistent) anatomical representation learning and a well-initialized segmentation head (e.g., the decoder) that can be flexibly adapted by few age-specific samples (e.g., one labeled MRI) to establish accurate segmentation models generalizable across the lifespan.

Overall, the technical and practical contributions of this paper are four-fold:

- We propose a dual meta-learning (DuMeta) paradigm for the construction of longitudinally generalized segmentation networks. Our DuMeta unifies the advantages of both meta-feature learning and meta-initialization learning [22, 14] in a compact bi-level optimization framework, which contributes to the joint learning of an age-agnostic plug-and-play feature extractor and a reliably pre-trained segmentation head that can be efficiently adapted to different age groups.
- For the purpose of learning to learn a universal feature extractor generalizable across the lifespan, we design an intra-tissue temporal similarity regularization and an inter-tissue spatial orthogonality regularization, which are combined together to encourage longitudinal consistency in hierarchically multi-scale representation learning. In contrast to previous SSL works, our design explicitly considers class information and has no need of longitudinally paired training data.
- Our DuMeta coupled with the two class-aware regularization terms features a practically attractive meta-learning strategy. It only needs cross-sectional training samples in the meta-learning stage and as less as one labeled image from an unseen age group to establish an accurate brain tissue segmentation model.

- Under the challenging experimental setting of one-shot segmentation, our method significantly outperformed the state-of-the-art longitudinally consistent learning methods on both the infant and elderly datasets.

2. Related works

2.1. Meta-Feature Learning

Meta-learning typically consists of two fundamental stages, i.e., meta-training and meta-test [14]. As a representative type of meta-learning methods, meta-feature learning (MFL) [22] aims to learn a common feature extractor in the meta-training stage to share across various (correlated) tasks, based on which a task-specific head is further learned in the meta-test stage for downstream (unseen) tasks. The common feature extractor can be formulated as an explicit meta-learner, usually meta-trained on various tasks/datasets in a bi-level optimization fashion. In terms of the strategies to compute the gradients (*w.r.t.* meta-learner parameters), MFL is typically split into two categories: one updates the meta-learner with gradients in the implicit function relation by performing several steps of gradient descent on the loss function [11, 4, 35, 2, 24], the other derives the hyper-gradients through the implicit function theory [8, 13, 28, 12, 1]. The former is widely used to solve the bi-level optimization problem by automatic differentiation. For example, Franceschi *et al.* [13] proposed to treat the final layer and remaining layers of a classification network as the base-learner and meta-learner, respectively. The learnable weights of the meta-learner are iteratively updated by gradient descent in the back-propagation fashion. In this paper, we separate a segmentation network as a plug-and-play feature extractor (i.e., the meta-learner) and a task-specific segmentation head (i.e., the base-learner). We meta-train the plug-and-play meta-learner by gradient descent in the implicit function relation for longitudinally consistent learning of anatomical representations, under the constraints by two dedicated class-aware regularizations.

2.2. Meta-Initialization Learning

As another representative type of meta-learning, meta-initialization learning (MIL) aims to learn a good initialization of network parameters shared across (correlated) tasks [22]. In contrast to most MFL approaches, MIL typically does not have an explicit meta-learner, while the parameters initialized by MIL can be regarded as an implicit meta-learner. In addition, the implicit meta-learner of MIL is only explicitly related to the base-learner in the initial state. A mainstream of MIL methods is to update the initialization parameters with the implicit gradients in the outer-loop of the bi-level optimization framework, with the implicit function relationships coming from the inner-loop. As a pioneering MIL research, MAML [10] uses the implicit gradi-

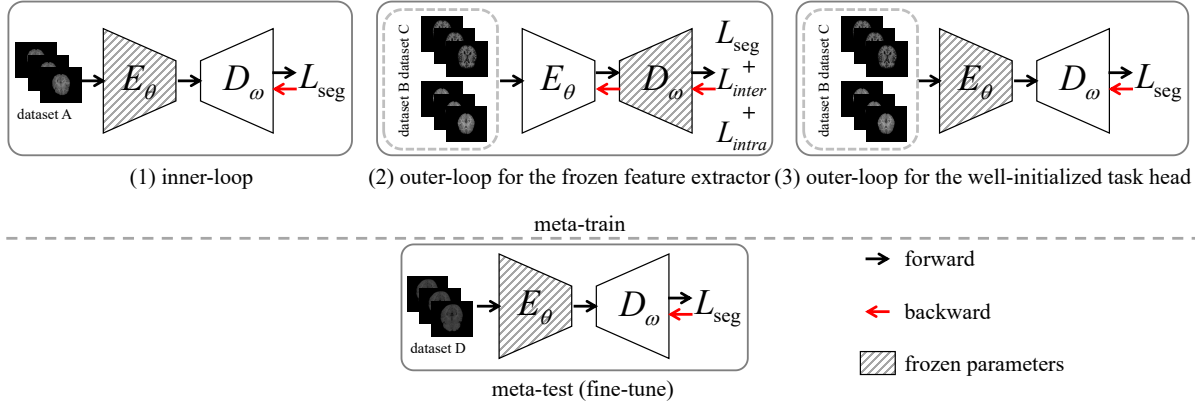


Figure 2. Overview of DuMeta. The meta-train of our DuMeta is divided into three steps (one shared inner-loop and two different outer-loops) with a feature extractor and a task head. Steps (1) and (2) constitute MFL for the frozen feature extractor; Steps (1) and (3) constitute MIL for the well-initialized task head. We finally only fine-tune the task head on an unseen dataset.

ent of the initialization process to update the meta-learner. Considering that calculating the hessian matrix in implicit gradient descent is computationally heavy, the following works of MAML have actively investigated the simplification of the MIL steps, e.g., finding ways to directly omit the expensive second-order derivatives [20, 27, 36]. For example, ANIL [31] proposed to remove the inner loop for all but the (task-specific) head of the underlying neural network. Reptile [27] proposed a variety of first-order algorithms to avoid the consumption of the second-order derivatives. As for the applications of MIL, some methods proposed to use MIL in domain generalization tasks [21, 23, 17]. For example, SAML [21] explicitly simulates domain shifts with virtual meta-train and meta-test during the training procedure to improve the generalization of learned networks. Notably, most of these existing MIL strategies were formulated to learn the entire network initialization, which by nature is a practically challenging task. In contrast, our DuMeta simplifies the task: Based on the learning of a frozen feature extractor for longitudinally consistent representation learning, we only need to further learn a lightweight segmentation head with reliable initializations. To this end, we simulate domain shift during fine-tuning by MIL.

2.3. Longitudinally Consistent Learning

Self-supervised learning (SSL) is the most actively studied strategy to learn longitudinally consistent or generic representations from data acquired at different time points [5, 3, 40, 30]. Such SSL methods typically design contrastive regularization terms to learn on longitudinal data the initialized parameters of a deep network, which will then be fine-tuned for downstream tasks. For example, as an extension of SimSiam [6], Ren *et al.* [32] proposed a patch-wise spatiotemporal similarity loss to encourage longitudinal consistency, and leveraged orthogonality, variance, and covariance regularizations to avoid SSL mode col-

lapse. Ouyang *et al.* proposed an LNE method [29] that adopts longitudinal neighbourhood embedding to promote the low-dimensional representation to be instructive while keeping a smooth progression trajectory field in the latent space. For lifespan segmentation, Chen *et al.* [7] proposed a joint feature regularization to encourage longitudinal consistency. Compared with these SSL methods, we take the goal of longitudinally consistent representation learning further. Specifically, our DuMeta jointly learns a plug-and-play feature extractor with promising longitudinal consistency and a segmentation head with reliable initializations. Only the lightweight segmentation head needs to be fine-tuned for specific age groups. Also, the meta-training of the shared feature extractor does not require longitudinally paired training data. Moreover, instead of using existing patch- and image-based [29, 32] regularizations, we propose two class-aware regularizations dedicated for brain tissue segmentation.

3. Method

3.1. DuMeta Paradigm

We propose a dual meta-learning (DuMeta) paradigm for accurate brain tissue segmentation across the lifespan, with the schematic diagram shown in Fig. 2. In the meta-learning (or pre-training) stage, our DuMeta attempts to learn concurrently a universal plug-and-play feature extractor (i.e., $E_\theta(\cdot)$) for longitudinally consistent learning of anatomical representations and a well-initialized segmentation head (i.e., $D_\omega(\cdot)$). They are formulated as two different outer-loops under a unified inner-loop in the bi-level meta-learning optimization framework. In the meta-test (or fine-tuning) stage, only the segmentation head needs to be fine-tuned with few labeled images to efficiently adapt to specific unseen age groups.

3.1.1 Meta-Learning of Universal Feature Extractor

To learn a longitudinally consistent (i.e., age-agnostic) $E_\theta(\cdot)$ that can be plug-and-play for unseen age groups, we regard it as an explicit meta-learner and treat $D_\omega(\cdot)$ as the base-learner. Such a universal $E_\theta(\cdot)$ is therefore learned in the MFL fashion by leveraging the idea of implicit gradient function. Specifically, as shown in Fig. 2 (1), in an episode of meta-train, we first freeze $E_\theta(\cdot)$ to update $D_\omega(\cdot)$ in the inner loop. To this end, we randomly sample a dataset from the meta-training dataset pool, on which the gradient descent to optimize $D_\omega(\cdot)$ is defined as:

$$\omega_t^*(\theta_{t-1}) = \omega_{t-1} - \alpha * \frac{\partial L_{\text{inner}}(\omega_{t-1}, \theta_{t-1})}{\partial \omega_{t-1}}, \quad (1)$$

where L_{inner} is the loss function (Dice + CE) of the inner-loop; θ_{t-1} and ω_{t-1} denote the parameters of the meta-learner and base-learner at the $(t-1)$ th iteration, respectively; and α is the learning rate. It is worth noting that L_{inner} is actually a function of θ_{t-1} . Thus, the parameter $\omega_t^*(\theta_{t-1})$, updated with the gradient generated by L_{inner} , is still a function of θ_{t-1} . Therefore, Eq. (1) can form an implicit function relationship to assist the updating of $E_\theta(\cdot)$ in the subsequent outer-loop.

More specifically, considering that the inner-loop can be regarded as a simulated step to fine-tune the base-learner (i.e., the segmentation head) in our task, we hope that the meta-learner (i.e., the feature extractor) can be good enough to produce promising segmentation results when combined with such a fine-tuned base-learner. For this purpose, as shown in Fig. 2 (2), we further sample another two datasets from the meta-training pool to update $E_\theta(\cdot)$. Thus, the gradient generated by Eq. (1) in updating $D_\omega(\cdot)$ is merged to update the universal meta-learner, such as:

$$\theta_t = \theta_{t-1} - \alpha * \frac{\partial L_{\text{outer}}}{\partial \theta_{t-1}}; \quad (2)$$

$$\begin{aligned} \frac{\partial L_{\text{outer}}}{\partial \theta_{t-1}} &= \frac{\partial L_{\text{outer}}(\theta_{t-1}, \omega_t^*(\theta_{t-1}))}{\partial \theta_{t-1}} \\ &+ \frac{\partial L_{\text{outer}}(\theta_{t-1}, \omega_t^*(\theta_{t-1}))}{\partial \omega_t^*} \frac{\partial \omega_t^*}{\partial \theta_{t-1}}, \end{aligned} \quad (3)$$

where L_{outer} is the combination of the segmentation loss (i.e., Dice + CE) and class-aware regularizations (will be introduced in Sec. 3.2) for training the universal meta-learner in such an outer-loop. According to Eq. (3), we can see that the outer-loop gradient to determine the updated θ_t is actually a combination of two parts, i.e., the direct gradient and the indirect gradient, which jointly make sure the optimization of $E_\theta(\cdot)$ working in the right direction to deduce a universal meta-learner generalizable across datasets from different age groups. That is, the direct gradient part is determined by minimizing L_{outer} , which allows the updated $E_\theta(\cdot)$ to perform well in the current outer-loop. On the other hand, the indirect gradient part comes from the

implicit function formulated by Eq. (1), which allows the updated $E_\theta(\cdot)$ to match up with the updated $D_\omega(\cdot)$ to perform well for the simulated fine-tuning step and therefore unseen age groups. In more detail, the indirect gradient part of Eq. (3) can be unrolled as explicit second-order derivatives:

$$\begin{aligned} \frac{\partial \omega_t^*}{\partial \theta_{t-1}} &= \frac{\partial(\omega_{t-1} - \alpha * \frac{\partial L_{\text{inner}}(\omega_{t-1}, \theta_{t-1})}{\partial \omega_{t-1}})}{\partial \theta_{t-1}} \\ &= \alpha * \frac{\partial^2 L_{\text{inner}}(\omega_{t-1}, \theta_{t-1})}{\partial \omega_{t-1} \partial \theta_{t-1}}. \end{aligned} \quad (4)$$

Finally, by substituting Eqs. (3) and (4) into Eq. (2), we get the detailed gradient-descent strategy to optimize $E_\theta(\cdot)$:

$$\begin{aligned} \theta_t &= \theta_{t-1} - \alpha * \frac{\partial L_{\text{outer}}(\theta_{t-1}, \omega_t^*(\theta_{t-1}))}{\partial \theta_{t-1}} \\ &- \alpha^2 * \frac{\partial L_{\text{outer}}(\theta_{t-1}, \omega_t^*(\theta_{t-1}))}{\partial \omega_t^*} \frac{\partial^2 L_{\text{inner}}(\omega_{t-1}, \theta_{t-1})}{\partial \omega_{t-1} \partial \theta_{t-1}}, \end{aligned} \quad (5)$$

which consists of a direct first-order gradient for the current outer-loop and an indirect second-order gradient for the simulated fine-tuning step.

3.1.2 Meta-Learning of Initialized Segmentation Head

Given the plug-and-play feature extractor $E_\theta(\cdot)$ determined by Sec. 3.1.1, we further learn a well-initialized segmentation head $D_\omega(\cdot)$ in the MIL fashion. In this step, $E_\theta(\cdot)$ is fixed, while the initial weights of $D_\omega(\cdot)$, say ϕ_{t-1} , can be regarded as an implicit meta-learner required to be updated.

To this end, we first assign the initial weight ϕ_{t-1} to the segmentation head parameter ω :

$$\omega_{t-1} = \phi_{t-1}. \quad (6)$$

Then, we randomly sample a dataset from the meta-training pool to update ω_{t-1} . In line with the settings of MAML, here we only update ω_{t-1} by one inner-loop iteration, i.e.,

$$\omega_t^* = \omega_{t-1} - \alpha * \frac{\partial L_{\text{inner}}}{\partial \omega_{t-1}}, \quad (7)$$

It is worth noting that, by the nature of network pre-training, the parameters of the base-learner can be seen as a function (identity function) of the meta-learner parameters. Hence, after the inner-loop optimization and given the fixed $E_\theta(\cdot)$, we update the initialization parameters of the segmentation head (i.e., the implicit meta-learner) in the subsequent outer-loop. Specifically, we use the updated ω_t^* to calculate the loss function on another two datasets from the meta-learning pool and update the initialization parameter ϕ_{t-1} with gradient descent, such as

$$\phi_t = \phi_{t-1} - \alpha * \frac{\partial L_{\text{outer}2}}{\partial \phi_{t-1}}; \quad (8)$$

Algorithm 1 Dual Meta-Learning (DuMeta)

Require: Meta-learning pool (of three datasets: A, B, and C); feature extractor $E_\theta(\cdot)$ parameterized by θ ; segmentation head $D_\omega(\cdot)$ parameterized by ω .

- 1: Randomly initialize θ and ω ;
 - 2: **while** Not converged **do**
 - 3: **The inner loop:**
 - 4: Sample a mini-batch from a random dataset;
 - 5: Update ω based on Eq. (1);
 - 6: **The outer-loop for the frozen feature extractor:**
 - 7: Sample a mini-batch from the other two datasets;
 - 8: Update θ based on Eq. (5);
 - 9: **The outer-loop for the well-initialized task head:**
 - 10: Sample a mini-batch from the other two datasets;
 - 11: Update ω based on Eq. (11).
 - 12: **end while**
-

$$\frac{\partial L_{\text{outer2}}}{\partial \phi_{t-1}} = \frac{\partial L_{\text{outer2}}}{\partial \omega_t^*} \frac{\partial \omega_t^*}{\partial \phi_{t-1}}; \quad (9)$$

$$\begin{aligned} \frac{\partial \omega_t^*}{\partial \phi_{t-1}} &= \frac{\partial(\omega_{t-1} - \alpha * \frac{\partial L_{\text{inner}}}{\partial \omega_{t-1}})}{\partial \phi_{t-1}} \\ &= 1 - \alpha * \frac{\partial^2 L_{\text{inner}}}{\partial \phi_{t-1}^2}, \end{aligned} \quad (10)$$

where L_{outer2} is the same (Dice + CE) segmentation loss as L_{inner} . By substituting Eqs. (9) and (10) into Eq. (8), the final gradient descent in the current outer-loop iteration is:

$$\phi_t = \phi_{t-1} - \alpha * \frac{\partial L_{\text{outer2}}}{\partial \omega_t^*} (1 - \alpha * \frac{\partial^2 L_{\text{inner}}}{\partial \phi_{t-1}^2}). \quad (11)$$

Considering that $\alpha * \frac{\partial^2 L_{\text{inner}}}{\partial \phi_{t-1}^2}$ significantly increases computational complexity and may cause gradient explosion, such a high-level differentiation can be omitted in practice. Besides, Reptile [27] points out that the first-order term can still enhance generalization by increasing the inner product between gradients of different minibatches.

By summarizing the inner-loop and outer-loop iterations in both the MFL step (i.e., Sec. 3.1.1) and the MIL step (i.e., Sec. 3.1.2), we can see that these two meta-learning steps actually have the same format of the inner-loop operation (i.e., Eqs. (1) and (7)). Therefore, the MFL of a universal feature extractor and the MIL of a well-initialized segmentation head are highly correlated and can be merged to share one single inner-loop for simplicity, resulting in a compact dual meta-learning (DuMeta) paradigm in Algorithm 1.

3.2. Class-Aware Regularizations

In Sec. 3.1.1, our DuMeta adopts L_{outer} that combines general segmentation losses with two class-aware regularizations to meta-learn the plug-and-play feature extractor

$E_\theta(\cdot)$ for age-agnostic (i.e., longitudinally consistent) learning of anatomical representations. These two class-aware regularizations are intra-tissue temporal similarity and inter-tissue spatial orthogonality, which are designed under a fundamental assumption that, independent of changing time points and MRI appearances, the high-order differences between different tissue types are relatively stable across the lifespan. Different from existing SSL approaches [32, 29] that typically implement their contrastive losses to align random MRI patches or images regardless of class information, our class-aware regularizations can be seen as a supervised contrastive learning design to align directly the anatomical presentations of each tissue type, which could be more straightforward for the downstream segmentation task. Besides, we implement such regularizations at multiple learning scales of $E_\theta(\cdot)$ to encourage hierarchically multi-scale longitudinal consistency across age groups.

Specifically, assume that $\{\mathbf{F}_k\}_{k=1}^K$ is a set of the feature maps at K different scales of $E_\theta(\cdot)$. We have $\mathbf{F}_k \in \mathbb{R}^{BS \times NC_k \times H_k \times W_k \times D_k}$, where BS , NC_k , and $H_k \times W_k \times D_k$ stand for mini-batch size, channel size, and spatial size, respectively. For each \mathbf{F}_k , we first resize the (pseudo) tissue maps (say \mathbf{M}) to its spatial resolution, denoted as \mathbf{M}_k . Then, according to the label information in \mathbf{M}_k , we get the spatial indices of the voxels belonging to the same classes, based on which we average the corresponding voxel-wise representations from \mathbf{F}_k to obtain the anatomical representations of different tissue types (i.e., GM, WM, and CSF), denoted as \mathbf{f}_k^{GM} , \mathbf{f}_k^{WM} , and $\mathbf{f}_k^{\text{CSF}}$ ($\in \mathbb{R}^{BS \times NC_k}$), respectively.

In terms of \mathbf{f}_k^{GM} , \mathbf{f}_k^{WM} , and $\mathbf{f}_k^{\text{CSF}}$, we define the inter-tissue spatial orthogonality loss as

$$\begin{aligned} L_{\text{inter}} &= \frac{1}{3K} \sum_{k=1}^K (\text{cossim}(\mathbf{f}_k^{\text{CSF}}, \mathbf{f}_k^{\text{GM}}) + \text{cossim}(\mathbf{f}_k^{\text{CSF}}, \mathbf{f}_k^{\text{WM}}) \\ &\quad + \text{cossim}(\mathbf{f}_k^{\text{GM}}, \mathbf{f}_k^{\text{WM}})), \end{aligned} \quad (12)$$

where $\text{cossim}(\cdot)$ denotes the cosine similarity that we use to encourage the anatomical representations of different tissue types to be separated for each spatial scale.

On the other hand, to promote the anatomical representations of the same tissue at different time points to be consistent, we define the intra-tissue temporal similarity loss as

$$\begin{aligned} L_{\text{intra}} &= -\frac{1}{3K} \sum_{k=1}^K (\text{cossim}(\mathbf{f}_{B_k}^{\text{CSF}}, \mathbf{f}_{NC_k}^{\text{CSF}}) \\ &\quad + \text{cossim}(\mathbf{f}_{B_k}^{\text{GM}}, \mathbf{f}_{NC_k}^{\text{GM}}) + \text{cossim}(\mathbf{f}_{B_k}^{\text{WM}}, \mathbf{f}_{NC_k}^{\text{WM}})), \end{aligned} \quad (13)$$

where $\mathbf{f}_{B_k}^{\text{CSF}}$ and $\mathbf{f}_{NC_k}^{\text{CSF}}$ denote the k th-scale tissue anatomical representations quantified on two datasets (from different age groups), respectively.

Finally, the overall loss function in the outer-loop of our MFL stage to update the universal feature extractor $E_\theta(\cdot)$

can be specified as:

$$L_{\text{outer}} = L_{\text{seg}} + \beta L_{\text{inter}} + \gamma L_{\text{intra}}, \quad (14)$$

where β and γ are two hyper-parameters, which were set as 0.1 and 0.001, respectively, in our experiments.

4. Experiments

4.1. Datasets

To evaluate the efficacy of our method, we applied it to brain tissue segmentation on two difficult age groups, i.e., 6-month-old infancy and the elderly, under the challenging setting of one-shot segmentation. Specifically, we meta-trained the segmentation network (consisting of a feature extractor and a segmentation head) on three public datasets, including OASIS3, IBIS12M, and IBIS24M, and meta-tested it on another two public datasets, i.e., ADNI and iSeg-2019. The network inputs were T1w MRIs undergone a series of pre-processing, including skull stripping, bias field correction, and intensity normalization. To obtain high-quality meta-training labels, we segmented the pre-processed T1w images automatically by using an advanced pipeline, i.e., iBEAT [38]. These tissue maps produced by iBEAT can be regarded as pseudo labels from the aspect of semi-supervision. To obtain accurate meta-test labels, iBEAT was first applied, followed by manual corrections by experienced neuroradiologists to produce the ground-truth tissue maps. Each of the three meta-training datasets was split according to an 80-20 training-validation ratio, where the validation set was used for model and hyper-parameter selection. After meta-training, one training subject from each of the two meta-test datasets was randomly sampled to fine-tune the segmentation head. We reported the metrics on the test set from the meta-test datasets.

OASIS3 [16]: A publicly accessible dataset consisting of 1,639 T1w MRIs from 992 longitudinally scanned participants (each with 1–5 temporal acquisitions across a 5-year observation window). The participants include Alzheimer’s disease (AD) patients, and healthy controls or minimally impaired abilities, with the ages ranging from 42 to 95.

ADNI [26]: A multisite dataset consisting of 2,389 longitudinal T1w MRIs (at least two visits per subject). Participants with AD, those who may acquire AD, and healthy controls without any evidence of cognitive impairment were included, with the ages ranging from 20 to 90.

IBIS: An infant brain imaging dataset that collects 1,272 structural T1w/T2w MRIs from 552 babies throughout 3 to 36 months old, including both controls and infants at

high risk for Autism Spectrum Disorder (ASD). These subjects can be assigned into different groups according to their ages. For example, IBIS12M and IBIS24M denote the 12-month-old and 24-month-old groups, respectively.

iSeg-2019 [37]: The benchmark of the MICCAI 2019 grand challenge on 6-month infant brain MRI segmentation from multiple sites [37]. The iSeg-2019 training and test sets include 10 and 13 subjects, respectively. Each subject consists of T1w and T2w MRIs. Notably, 6-month-old infant brain segmentation is challenging, considering that the GM and WM intensity ranges are substantially overlapped (particularly around the cortical areas) at this iso-intense stage, which results in the ambiguities that present the greatest obstacle for tissue segmentation.

4.2. Implementation Details

We used 3D U-Net [34] as our baseline model, which performs five down-samplings and five up-samplings in the encoding and decoding path, respectively. We used InstanceNorm3d for normalization and ReLU for nonlinear activation. The network was trained by the SGD optimizer with Nesterov momentum ($\mu = 0.99$), and poly learning rate schedule (initial 0.01). The weight decay was set as 3e-5 to avoid overfitting. The size of input image patches was $128 \times 128 \times 128$, and the mini-batch size was 2. We conducted deep supervision on five scales, with the weights of 0.0625, 0.125, 0.25, 0.5, and 1 (from coarse to fine), respectively. The training data was extensively augmented using the pipeline of nnUNet [15]. Two commonly used metrics were selected to quantify segmentation performance: Dice ratio and Average Symmetric Surface Distance (ASD). A larger dice value indicates better overlapping between the prediction and ground truth. In contrast, smaller ASD values suggest better performance in terms of the boundary consistency between the prediction and ground truth.

4.3. Comparison Results

We compared our method with other methods addressing longitudinally consistent/generalized representation learning. We also compared our method with other meta-learning methods addressing few-shot learning. In order to verify the performance of our method on longitudinal data, we performed challenging one-shot brain tissue segmentation on the ADNI and iSeg-2019 datasets in the meta-test stage. Both data sets have longitudinal domain shifts due to dynamic neuro-degeneration/development processes.

One-shot segmentation on iSeg-2019: The iSeg-2019 dataset consists of 6-month-old infant MRIs from multiple sites, which present a flip GM/WM contrast with a huge domain shift compared with other time points. Our DuMeta meta-trained (pre-trained) the 3D U-Net on IBIS12M, and

Table 1. One-shot segmentation on the official validation set of the iSeg-2019 dataset.

Exp	Dice \uparrow			ASD \downarrow		
	CSF	GM	WM	CSF	GM	WM
RandInitUnet.2D	0.8767 \pm 0.0113	0.8402 \pm 0.0220	0.7965 \pm 0.0035	0.3317 \pm 0.0215	0.5578 \pm 0.0825	0.8479 \pm 0.1930
RandInitUnet.3D	0.9029 \pm 0.0066	0.8616 \pm 0.0176	0.8200 \pm 0.0071	0.2652 \pm 0.0162	0.5497 \pm 0.0658	0.6487 \pm 0.1324
Context Restore [5]	0.9070 \pm 0.0065	0.8615 \pm 0.0118	0.8245 \pm 0.0155	0.2641 \pm 0.0239	0.5420 \pm 0.0560	0.6298 \pm 0.1068
LNE [29]	0.9315 \pm 0.0115	0.8937 \pm 0.0147	0.8654 \pm 0.0141	0.1753 \pm 0.0163	0.4482 \pm 0.0373	0.5274 \pm 0.0710
GLCL [3]	0.9141 \pm 0.0103	0.8718 \pm 0.0185	0.8337 \pm 0.0088	0.2337 \pm 0.0147	0.5017 \pm 0.0546	0.6596 \pm 0.1024
PCL [40]	0.9139 \pm 0.0089	0.8688 \pm 0.0148	0.8313 \pm 0.0133	0.2326 \pm 0.0130	0.5161 \pm 0.0484	0.6692 \pm 0.0943
PatchNCE [30]	0.9444 \pm 0.0096	0.9043 \pm 0.0081	0.8782 \pm 0.0223	0.1360 \pm 0.0153	0.3820 \pm 0.0376	0.4520 \pm 0.0760
MAML [10]	0.9433 \pm 0.0107	0.9025 \pm 0.0100	0.8768 \pm 0.0185	0.0174 \pm 0.0139	0.3931 \pm 0.0493	0.5374 \pm 0.0859
Reptile [27]	0.9450 \pm 0.0129	0.9047 \pm 0.0121	0.8777 \pm 0.0143	0.0117 \pm 0.0123	0.3896 \pm 0.0451	0.5324 \pm 0.0825
Ours	0.9611\pm0.0059	0.9313\pm0.0083	0.9145\pm0.0126	0.1082\pm0.0158	0.2916\pm0.0423	0.3318\pm0.0483

Table 2. One-shot segmentation on the ADNI dataset.

Exp	Dice \uparrow			ASD \downarrow		
	CSF	GM	WM	CSF	GM	WM
RandInitUnet.2D	0.9223 \pm 0.0084	0.9051 \pm 0.0098	0.9361 \pm 0.0096	0.1571 \pm 0.0422	0.1573 \pm 0.0338	0.2059 \pm 0.0909
RandInitUnet.3D	0.9459 \pm 0.0062	0.9236 \pm 0.0073	0.9495 \pm 0.0058	0.0933 \pm 0.0205	0.1160 \pm 0.0297	0.1226 \pm 0.0291
Context Restore [5]	0.9500 \pm 0.0058	0.9289 \pm 0.0069	0.9527 \pm 0.0054	0.0821 \pm 0.0179	0.1056 \pm 0.0267	0.1101 \pm 0.0241
LNE [29]	0.9665 \pm 0.0043	0.9456 \pm 0.0065	0.9637 \pm 0.0047	0.0487 \pm 0.0107	0.0711 \pm 0.0168	0.0760 \pm 0.0180
GLCL [3]	0.9531 \pm 0.0055	0.9332 \pm 0.0069	0.9559 \pm 0.0052	0.0736 \pm 0.0151	0.0964 \pm 0.0238	0.1018 \pm 0.0236
PCL [40]	0.9526 \pm 0.0057	0.9314 \pm 0.0068	0.9547 \pm 0.0053	0.0750 \pm 0.0137	0.0995 \pm 0.0255	0.1059 \pm 0.0273
PatchNCE [30]	0.9716 \pm 0.0040	0.9520 \pm 0.0060	0.9687 \pm 0.0043	0.0402 \pm 0.0091	0.0591 \pm 0.0118	0.0634 \pm 0.0157
MAML [10]	0.9680 \pm 0.0062	0.9472 \pm 0.0070	0.9647 \pm 0.0061	0.0460 \pm 0.0197	0.0680 \pm 0.0239	0.0736 \pm 0.0281
Reptile [27]	0.9687 \pm 0.0056	0.9476 \pm 0.0071	0.9656 \pm 0.0049	0.0452 \pm 0.0167	0.0666 \pm 0.0163	0.0702 \pm 0.0176
Ours	0.9809\pm0.0021	0.9678\pm0.0034	0.9796\pm0.0024	0.0222\pm0.0042	0.0315\pm0.0052	0.0322\pm0.0054

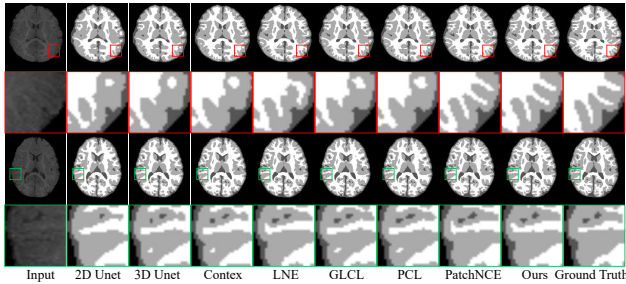


Figure 3. The 2D slice views of representative one-shot segmentation results on the held-out test set of iSeg-2019.

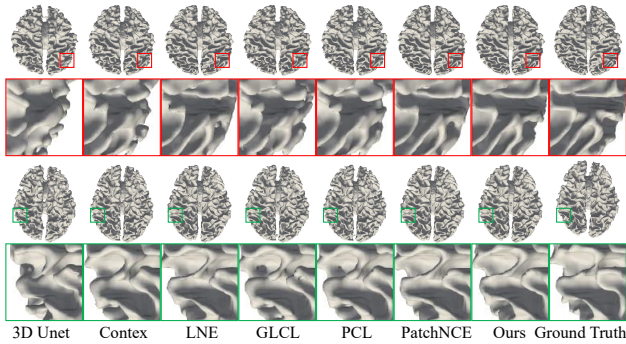


Figure 4. The 3D WM surface views of representative one-shot segmentation results on the held-out test set of iSeg-2019.

IBIS24M (i.e., 12/24-month-old infants), and OASIS3 (i.e., the elderly), and then meta-tested (fine-tuned) it on iSeg-2019. Notably, our model did not see 6-month-old data dur-

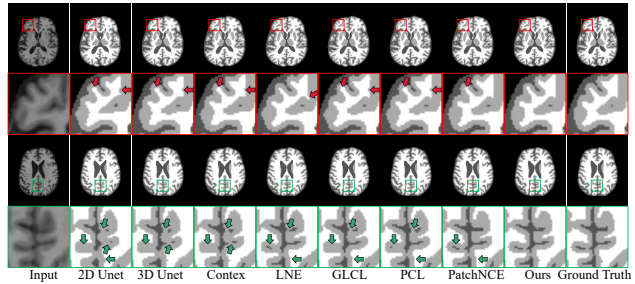


Figure 5. The 2D slice views of representative one-shot segmentation results on the ADNI dataset.

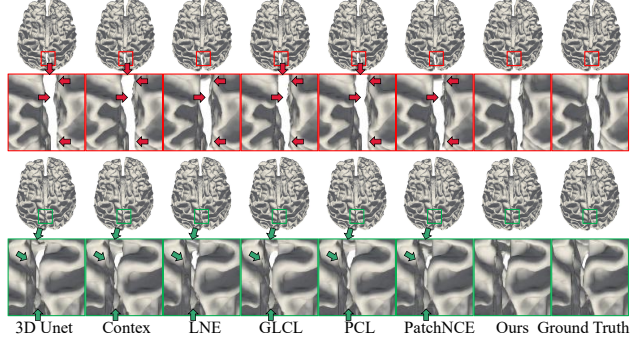


Figure 6. The 3D WM surface views of representative one-shot segmentation results on the ADNI dataset.

ing meta-training, and we froze the feature extractor and only fine-tuned the segmentation head with one 6-month-old subject during meta-test. The quantitative segmentation

results obtained by different competing methods are summarized in Table 1. From Table 1, we can see that our DuMeta outperformed other competing methods by a large margin in learning a longitudinally generalizable segmentation model. As a support to the quantitative evaluations, we also show the qualitative experimental results in Figs. 3 and 4, from which we can see that our method performs better in segmenting detailed structures, e.g., cortical foldings.

One-shot segmentation on ADNI: By leveraging the same meta-training strategy used in iSeg-2019, we further meta-tested the pre-trained 3D U-Net on the ADNI dataset. Correspondingly, the quantitative segmentation results and representative visualization results are summarized in Table 2 and Figs. 5 and 6, respectively. It can be seen that, consistent with the observations on iSeg-2019, our DuMeta also outperformed other competing methods by a large margin in the case of aging brain tissue segmentation. These results suggest the promising generalization capacity of our DuMeta in segmenting brain tissues across the lifespan.

Besides the state-of-the-art segmentation accuracies, it is worth mentioning that the pre-training by other competing methods was based on self-supervision without ground-truth label information. Similarly, our DuMeta does not need the ground truth either during meta-training, as we used pseudo labels efficiently generated by iBEAT, which can be understood as semi-supervision. However, the inclusion of pseudo label information can help avoid mode collapse potentially caused by contrastive self-supervision. For example, because of the U-Net skip connection, the self-supervised model degenerated into identity mappings can still achieve good self-reconstruction, which hampers the model to learn informative representation for downstream segmentation tasks. Moreover, in contrast to contrastive self-supervision that typically needs longitudinally paired training, our DuMeta works on unpaired cross-sectional data, which is much easier to satisfy in practice.

Evaluation of longitudinal consistency: We evaluated the longitudinal consistency of different segmentations on ADNI by using two metrics, i.e., STCS [19] and ASPC [33]. Results in Table 5 suggest the better performance of DuMeta. We also conducted a t-sne visualization of the meta-learned features learned from the ADNI and i-Seg subjects. Examples in Fig. 7 imply that DuMeta can learn reliably time-invariant discriminative representations.

4.4. Ablation Studies

We performed ablation experiments on the iSeg-2019 training set. Specifically, iSeg-2019 provides 10 publicly available labeled training samples. In the meta-test stage, we used one sample for fine-tuning, two subjects for validation, and the remaining two subjects for test.

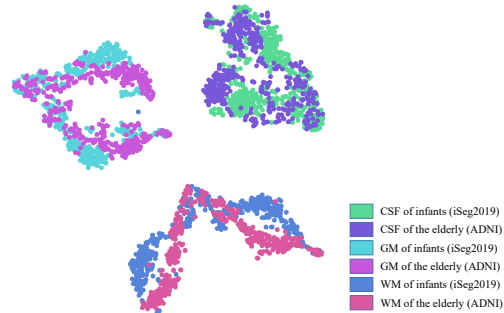


Figure 7. Visualizing meta-learned features for different ages.

Role of DuMeta and class-aware regularizations: To evaluate the efficacy of the DuMeta paradigm and class-aware regularization terms proposed in this paper, we conducted a series of ablation experiments by removing them to check the influence on segmentation accuracies, with the results summarized in Table 3. It can be seen that, compared with the baseline (i.e., A in Table 3), DuMeta led to significantly improved performance, implying its efficacy in learning generalizable segmentation networks across the lifespan. Also, we can see that both the two class-aware regularizations brought respective performance gains. Among them, the improvements by the inter-tissue spatial orthogonality are relatively larger than the intra-tissue temporal similarity. The potential reason could be that the inter-tissue spatial orthogonality encourages the differentiation between different class representations and could be more straightforwardly linked to the segmentation task. The best result was achieved by using both regularizations, which suggest that they are complementary in longitudinally consistent representation learning for segmentation.

Influence of different base/meta-learner splits: We have carried out experiments to check the influence brought by different base/meta-learner (i.e., feature extractor/segmentation head) splits of the 3D U-Net, with the results shown in Table 4. It can be seen from Table 4 that, compared with the pre-trained model (i.e., w/o ft), fine-tuning all parameters (i.e., ft all) achieved much better results, suggesting that there are indeed domain gaps across the lifespan. On the other hand, when only a small number of parameters were fine-tuned (e.g., ft 1 upsample layer), the performance decreased compared with ft all. The reason could be that there is a gap between the local and global optimum due to too few parameters available for fine-tuning. In contrast, when fine-tuning the last three upsample layers (i.e., ft 3 upsample layers) in the meta-test stage, our DuMeta led to the best performance. These results suggest that there is a trade-off between convergence difficulty and convergence quality in model fine-tuning, considering that

Table 3. Ablation study of different components on the held-out test set of iSeg-2019.

Exp	L_{inter}	L_{intra}	DuMeta	Dice \uparrow			ASD \downarrow		
				CSF	GM	WM	CSF	GM	WM
A				0.9155 \pm 0.0147	0.8724 \pm 0.0114	0.8375 \pm 0.0094	0.1113 \pm 0.0194	0.2954 \pm 0.0498	0.4082 \pm 0.0955
B			\checkmark	0.9358 \pm 0.0060	0.8955 \pm 0.0011	0.8679 \pm 0.0235	0.0781 \pm 0.0106	0.2176 \pm 0.0421	0.2997 \pm 0.0847
C	\checkmark		\checkmark	0.9431 \pm 0.0057	0.9015 \pm 0.0029	0.8742 \pm 0.0247	0.0669 \pm 0.0093	0.2009 \pm 0.0428	0.2775 \pm 0.0826
D		\checkmark	\checkmark	0.9414 \pm 0.0051	0.8977 \pm 0.0025	0.8690 \pm 0.0235	0.0701 \pm 0.0087	0.2127 \pm 0.0447	0.2935 \pm 0.0830
E	\checkmark	\checkmark	\checkmark	0.9492\pm0.0078	0.9091\pm0.0001	0.8826\pm0.0205	0.0597\pm0.0104	0.1824\pm0.0328	0.2488\pm0.0664

Table 4. Ablation study of different base/meta-learner splits on the held-out test set of iSeg-2019 (ft means fine-tune).

Exp	Dice \uparrow			ASD \downarrow		
	CSF	GM	WM	CSF	GM	WM
w/o ft	0.8450 \pm 0.0074	0.8082 \pm 0.0089	0.7564 \pm 0.0091	0.2487 \pm 0.0320	0.5298 \pm 0.0652	0.7460 \pm 0.2265
ft all	0.9454 \pm 0.0069	0.9076 \pm 0.0008	0.8823 \pm 0.0220	0.0653 \pm 0.0076	0.1857 \pm 0.0369	0.2493 \pm 0.0715
ft 5 upsample layers	0.9484 \pm 0.0087	0.9083 \pm 0.0003	0.8820 \pm 0.0220	0.0617 \pm 0.0130	0.1842 \pm 0.0322	0.2487\pm0.0669
ft 4 upsample layers	0.9491 \pm 0.0081	0.9096\pm0.0005	0.8829\pm0.0194	0.0618 \pm 0.0124	0.1841 \pm 0.0331	0.2502 \pm 0.0644
ft 3 upsample layers	0.9492\pm0.0078	0.9091 \pm 0.0001	0.8826 \pm 0.0205	0.0597\pm0.0104	0.1824\pm0.0328	0.2488 \pm 0.0664
ft 2 upsample layers	0.9451 \pm 0.0076	0.9058 \pm 0.0002	0.8793 \pm 0.0211	0.0654 \pm 0.0101	0.1908 \pm 0.0373	0.2600 \pm 0.0730
ft 1 upsample layer	0.9450 \pm 0.0066	0.9049 \pm 0.0013	0.8782 \pm 0.0229	0.0661 \pm 0.0103	0.1920 \pm 0.0394	0.2620 \pm 0.0775

Table 5. Evaluation of longitudinal consistency on the ADNI dataset.

Exp	STCS \uparrow			ASPC \downarrow		
	CSF	GM	WM	CSF	GM	WM
RandInitUnet.2D	0.8809	0.8584	0.8998	6.35	8.32	5.16
RandInitUnet.3D	0.9063	0.8862	0.9177	5.06	6.02	4.20
Context Restore [5]	0.8386	0.8180	0.8580	9.15	11.07	8.08
LNE [29]	0.9015	0.8796	0.9132	4.78	6.81	4.49
GLCL [3]	0.8956	0.8706	0.9057	5.18	7.57	4.75
PCL [40]	0.9155	0.8968	0.9275	3.98	5.63	3.13
PatchNCE [30]	0.9250	0.9087	0.9356	3.80	5.05	2.94
Ours	0.9373	0.9222	0.9468	2.66	3.40	2.20

Table 6. Ablation of hyperparameters on the ADNI dataset.

beta	gama	Dice \uparrow			ASD \downarrow		
		CSF	GM	WM	CSF	GM	WM
1e-1	1e-1	0.9714	0.9516	0.9684	0.0407	0.0600	0.0647
1e-3	1e-3	0.9710	0.9509	0.9679	0.0417	0.0617	0.0652
1e-3	1e-1	0.9697	0.9490	0.9671	0.0448	0.0656	0.0675
1e-1	1e-3	0.9809	0.9678	0.9796	0.0222	0.0315	0.0322

fine-tuning the whole network is hard to converge, while fine-tuning few network layers is prone to suboptimal solutions. Our method found a good compromise, leading to accurate and generalizable segmentation results.

Hyperparameter sensitivity evaluation: We performed a hyperparameter sensitivity evaluation. Results in Table 6 suggest the inter-class term Eq. (12) does have a larger influence than the intra-class term Eq. (13). This may be due to the fact that the inter-class term has a more direct effect on the segmentation task.

5. Conclusion

In this paper, we have proposed a dual meta-learning (DuMeta) paradigm coupled with dedicated class-aware regularizations to learn longitudinally consistent representations from brain MRIs for accurate brain tissue segmentation across the lifespan. Our DuMeta unifies the advantages

of both meta-feature learning and meta-initialization learning to jointly meta-learn an age-agnostic plug-and-play feature extractor and a well-initialized segmentation head. In the meta-test, only one labeled data is needed by DuMeta to adopt the segmentation head to unseen age groups. Experiments carried out on the ADNI and iSeg2019 datasets show that our method significantly outperforms existing longitudinally consistent representation learning methods.

Acknowledgement

This work was supported in part by STI 2030—Major Projects (No. 2022ZD0209000), NSFC 62101431, and NSFC 62101430.

References

- [1] Atilim Gunes Baydin and Barak A Pearlmutter. Automatic differentiation of algorithms for machine learning. *arXiv preprint arXiv:1404.7456*, 2014. 2
- [2] Luca Calatroni, Chung Cao, Juan Carlos De Los Reyes, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Bilevel approaches for learning of variational imaging models. *Variational Methods: In Imaging and Geometric Control*, 18(252):2, 2017. 2
- [3] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in Neural Information Processing Systems*, 33:12546–12558, 2020. 3, 7, 9
- [4] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, and Sayan Mukherjee. Choosing multiple parameters for support vector machines. *Machine learning*, 46:131–159, 2002. 2
- [5] Liang Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. Self-supervised learning for medical image analysis using image context

- restoration. *Medical image analysis*, 58:101539, 2019. 3, 7, 9
- [6] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021. 3
- [7] Xiaoyang Chen, Jinjian Wu, Wenjiao Lyu, Yicheng Zou, Kim-Han Thung, Siyuan Liu, Ye Wu, Sahar Ahmad, and Pew-Thian Yap. Brain tissue segmentation across the human lifespan via supervised contrastive learning. *arXiv preprint arXiv:2301.01369*, 2023. 3
- [8] Justin Domke. Generic methods for optimization-based modeling. In *Artificial Intelligence and Statistics*, pages 318–326. PMLR, 2012. 2
- [9] Xunli Fan, Shixi Shan, Xianjun Li, Jinhang Li, Jizong Mi, Jian Yang, and Yongqin Zhang. Attention-modulated multi-branch convolutional neural networks for neonatal brain tissue segmentation. *Computers in Biology and Medicine*, 146:105522, 2022. 1
- [10] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017. 2, 7
- [11] Chuan-sheng Foo, Andrew Ng, et al. Efficient multiple hyperparameter learning for log-linear models. *Advances in neural information processing systems*, 20, 2007. 2
- [12] Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil. A bridge between hyperparameter optimization and learning-to-learn. *arXiv preprint arXiv:1712.06283*, 2017. 2
- [13] Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *International Conference on Machine Learning*, pages 1568–1577. PMLR, 2018. 2
- [14] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169, 2021. 2
- [15] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 6
- [16] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *MedRxiv*, 2019. 6
- [17] Jaein Lee, Eunsong Kang, Eunjin Jeon, and Heung-II Suk. Meta-modulation network for domain generalization in multi-site fmri classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 500–509. Springer, 2021. 3
- [18] Jiyeon Lee, Minhoo Lee, Jongseung Lee, Regina EY Kim, Seong Hoon Lim, and Donghyeon Kim. Fine-grained brain tissue segmentation for brain modeling of stroke patient. *Computers in Biology and Medicine*, 153:106472, 2023. 1
- [19] Bo Li, Wiro J Niessen, Stefan Klein, Marius de Groot, M Arfan Ikram, Meike W Vernooij, and Esther E Bron. Longitudinal diffusion mri analysis using segis-net: a single-step deep-learning framework for simultaneous segmentation and registration. *NeuroImage*, 235:118004, 2021. 8
- [20] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017. 3
- [21] Quande Liu, Qi Dou, and Pheng-Ann Heng. Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, pages 475–485. Springer, 2020. 3
- [22] Risheng Liu, Jiaxin Gao, Jin Zhang, Deyu Meng, and Zhouchen Lin. Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10045–10067, 2021. 2
- [23] Xiao Liu, Spyridon Thermos, Alison O’Neil, and Sotirios A Tsaftaris. Semi-supervised meta-learning with disentanglement for domain-generalised medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 307–317. Springer, 2021. 3
- [24] Jonathan Lorraine and David Duvenaud. Stochastic hyperparameter optimization through hypernetworks. *arXiv preprint arXiv:1802.09419*, 2018. 2
- [25] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 1
- [26] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer’s disease neuroimaging initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005. 6
- [27] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018. 3, 5, 7
- [28] Takayuki Okuno, Akiko Takeda, Akihiro Kawana, and Motokazu Watanabe. Hyperparameter learning via bilevel non-smooth optimization. *arXiv preprint arXiv:1806.01520*, 2018. 2
- [29] Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Edith V Sullivan, Adolf Pfefferbaum, Greg Zaharchuk, and Kilian M Pohl. Self-supervised longitudinal neighbourhood embedding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 80–89. Springer, 2021. 1, 3, 5, 7, 9
- [30] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pages 319–345. Springer, 2020. 3, 7, 9

- [31] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019. [3](#)
- [32] Mengwei Ren, Neel Dey, Martin A Styner, Kelly Botteron, and Guido Gerig. Local spatiotemporal representation learning for longitudinally-consistent neuroimage analysis. *arXiv preprint arXiv:2206.04281*, 2022. [1](#), [3](#), [5](#)
- [33] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4):1402–1418, 2012. [8](#)
- [34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. [6](#)
- [35] Matthias W Seeger. Cross-validation optimization for large scale structured classification kernel methods. *Journal of Machine Learning Research*, 9(6), 2008. [2](#)
- [36] Xingyou Song, Wenbo Gao, Yuxiang Yang, Krzysztof Choromanski, Aldo Pacchiano, and Yunhao Tang. Es-maml: Simple hessian-free meta learning. *arXiv preprint arXiv:1910.01215*, 2019. [3](#)
- [37] Yue Sun, Kun Gao, Zhengwang Wu, Guannan Li, Xiaopeng Zong, Zhihao Lei, Ying Wei, Jun Ma, Xiaoping Yang, Xue Feng, Li Zhao, Trung Le Phan, Jitae Shin, Tao Zhong, Yu Zhang, Lequan Yu, Caizi Li, Ramesh Basnet, M. Omair Ahmad, M. N. S. Swamy, Wenao Ma, Qi Dou, Toan Duc Bui, Camilo Bermudez Noguera, Bennett Landman, Ian H. Gotlib, Kathryn L. Humphreys, Sarah Shultz, Longchuan Li, Sijie Niu, Weili Lin, Valerie Jewells, Dinggang Shen, Gang Li, and Li Wang. Multi-site infant brain segmentation algorithms: The iseg-2019 challenge. *IEEE Transactions on Medical Imaging*, 40(5):1363–1376, 2021. [1](#), [6](#)
- [38] Li Wang, Zhengwang Wu, Liangjun Chen, Yue Sun, Weili Lin, and Gang Li. ibeat v2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. *Nature Protocols*, pages 1–32, 2023. [6](#)
- [39] Liang Wu, Shunbo Hu, and Changchun Liu. Mr brain segmentation based on de-resunet combining texture features and background knowledge. *Biomedical Signal Processing and Control*, 75:103541, 2022. [1](#)
- [40] Dewen Zeng, Yawen Wu, Xinrong Hu, Xiaowei Xu, Haiyun Yuan, Meiping Huang, Jian Zhuang, Jingtong Hu, and Yiyu Shi. Positional contrastive learning for volumetric medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 221–230. Springer, 2021. [3](#), [7](#), [9](#)