Alternative Baselines for Low-Shot 3D Medical Image Segmentation—An Atlas Perspective

Shuxin Wang,[1,2] **Shilei Cao,**[2] **Dong Wei,**[2] **Cong Xie,**[1,2] **Kai Ma,**[2] **Liansheng Wang,**[1,3] **Deyu Meng,**[4] **Yefeng Zheng**[2]

[1] Department of Computer Science, Xiamen University, Xiamen, China
[2] Tencent Jarvis Lab, Shenzhen, China
[3] Department of Digestive Diseases, School of Medicine, Xiamen University, Xiamen, China
[4] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
{sxwang, xiecong}@stu.xmu.edu.cn, {eliasslcao, donwei, kylekma, yefengzheng}@tencent.com, lswang@xmu.edu.cn, dymeng@mail.xjtu.edu.cn

Abstract

Low-shot (one/few-shot) segmentation has attracted increasing attention as it works well with limited annotation. State-of-the-art low-shot segmentation methods on natural images usually focus on *implicit* representation learning for each novel class, such as learning prototypes, deriving guidance features via masked average pooling, and segmenting using cosine similarity in feature space. We argue that low-shot segmentation on medical images should step further to *explicitly* learn dense correspondences between images to utilize the anatomical similarity. The core ideas are inspired by the classical practice of multi-atlas segmentation, where the indispensable parts of atlas-based segmentation, *i.e.*, registration, label propagation, and label fusion are unified into a single framework in our work. Specifically, we propose two alternative baselines, *i.e.,* the Siamese-Baseline and Individual-Difference-Aware Baseline, where the former is targeted at anatomically stable structures (such as brain tissues), and the latter possesses a strong generalization ability to organs suffering large morphological variations (such as abdominal organs). In summary, this work sets up a benchmark for low-shot 3D medical image segmentation and sheds light on further understanding of atlas-based few-shot segmentation.

Introduction

The recent success of deep neural networks in medical image analysis relies on large amounts of labeled training data. However, it is labor-intensive and time-consuming to manually produce 3D annotations for supervised training. To reduce the demand for manual annotations, low-shot segmentation is proposed to solve this problem. Recently, a large body of literature (Nguyen and Todorovic 2019; Shaban et al. 2017; Siam, Oreshkin, and Jagersand 2019; Wang et al. 2019a; Zhang et al. 2018) has been devoted to the development of low-shot segmentation algorithms in the natural image domain, and inspired the development of low-shot algorithms in the medical imaging domain.

S. Wang, S. Cao and D. Wei contributed equally. S. Wang and C. Xie contributed to this work during internships at Tencent. Corresponding authors: L. Wang and Y. Zheng.

Although evolved from the natural image domain, low-shot segmentation in medical imaging develops distinct characteristics. First, in the natural image domain, the segmentor is trained on **base classes**, and the performance is evaluated on **novel classes**. However, in medical imaging, both the training and evaluation focus on the "base" classes due to the relatively fixed human tissue types (Wang et al. 2020; Xu and Niethammer 2019; Zhao et al. 2019), *i.e.*, the low-shot learning is more like semi-supervised learning. Second, human organs/tissue are anatomically similar across individuals, hence geometric and topological priors can be exploited to help segmentation with limited data. Such differences in concepts and levels of domain knowledge lead to essential differences in algorithm design. Low-shot segmentation in natural images emphasizes on representation learning, which is more likely to learn a class-agnostic object "concept" with a large pool of labeled data of base classes, where the features learned on base classes should transfer well to novel classes; whereas low-shot segmentation in medical images makes effort to better utilize the anatomical similarity between subjects, which is served as surrogate supervision to remedy the scarcity of labeled data.

Recently, low-shot segmentation methods in medical images (Wang et al. 2020; Xu and Niethammer 2019; Zhao et al. 2019) resorted to the practice of classical atlas-based segmentation (Jia, Yap, and Shen 2012; Lorenzo-Valdés et al. 2002), and proposed to implement this classical concept under the deep learning (DL) framework, which shifts from *implicit* representation learning in the feature space to *explicit* dense correspondence learning. The main idea of these works can be summarized as follows: (1) utilizing few labeled images (the *atlases*), the algorithms exploit the anatomical priors to establish structural correspondences between the atlases and unlabeled images; and (2) these correspondences are then utilized to propagate the segmentation labels from the atlases to the unlabeled images. Given the encouraging results obtained, the idea of extending atlas-based segmentation for low-shot learning is compelling.

In this paper, we first adopt a simple yet effective baseline, which leverages the cutting edge technologies (Sun et al. 2018) in correspondence learning to boost the performance of pixel/voxel-wise matching. Specifically, it em-

ploys a Siamese structure (Koch, Zemel, and Salakhutdinov 2015) with a principled network design—feature pyramid, warping, and cost volume (Sun et al. 2018). We name this baseline the Siamese-Baseline (Fig. 1). We experimentally demonstrate that such a simple baseline can outperform the state-of-the-art (SOTA) methods (Balakrishnan et al. 2019; Wang et al. 2020; Zhao et al. 2019) in segmentation of brain anatomical structures, which are anatomically stable with small inter-subject morphological variations.

Further, we observe that existing methods and the proposed Siamese-Baseline often generalize poorly on organs/tissue presenting large morphological variations, such as abdominal organs. We experimentally identify that the individual differences between subjects are the main obstacle for learning accurate correspondences in such applications, as pixels/voxels involved in such differences are exclusive to a specific individual and should not contribute to the loss computation. Hence, we further propose an individual-difference-aware (IDA) network (Fig. 2)—which is incrementally built on top of the Siamese-Baseline—as a new baseline named IDA-Baseline to learn the bidirectional correspondences between the atlas and target images, and define individual differences as the large displacement between forward and backward correspondences. To learn the bidirectional correspondence, we employ a forward and a backward branch with the same structure as the Siamese-Baseline to simultaneously model the bidirectional correlations with shared weights, which does not incur any extra parameter but makes the training more stable.

In summary, our contributions are as follows:

- We present the Siamese-Baseline, which innovatively incorporates feature pyramid, warping, and cost volume into a Siamese structure for low-shot segmentation.
- We further propose the IDA-Baseline to segment organs with large morphological variations. As far as we know, this is the first work that involves individual differences in atlas-based low-shot segmentation within a DL framework.
- Our one-shot settings naturally match the core concepts of registration-based segmentation, and innovatively implement registration and label propagation with a single framework. In addition, we further implement label fusion to match the practice of multi-atlas segmentation.

Superior results towards several SOTA methods verify that the proposed baselines not only perform well for anatomically stable structures (such as brain tissues) in one-shot settings but also possess a strong generalization ability for abdominal organs with large morphological variations in low-shot settings.

Related Work

Low-shot Segmentation of Natural Images. Low-shot segmentation can be seen as a natural extension of low-shot classification (Koch, Zemel, and Salakhutdinov 2015; Wang et al. 2019b) to the pixel level on natural images. As aforementioned, the main idea behind existing methods is to learn transferable weights/features with a large pool of training samples of base classes, which are expected to be easily adaptable for novel classes with a limited support set. They approached this task via (i) learning prototypes for each novel classes (Dong and Xing 2018; Wang et al. 2019a) as an extension of the Prototypical Network (Snell, Swersky, and Zemel 2017) from classification to segmentation, (ii) deriving guidance features via masked average pooling (Siam, Oreshkin, and Jagersand 2019; Zhang et al. 2018; Zhao et al. 2020b), segmenting using cosine similarity in feature space (Nguyen and Todorovic 2019; Zhang et al. 2018), *etc*. Different from natural images, low-shot segmentation in medical imaging places more emphasis in utilizing anatomical priors to learn dense correspondence. In the remaining of this paper, the low-shot segmentation is raised in the context of medical image segmentation unless explicitly stated.

Atlas-based Low-shot Segmentation of Medical Images. Existing one/few-shot segmentation methods often approached atlas-based segmentation by learning transformations for data augmentation (Zhao et al. 2019), learning the correspondence for propagating labels (Wang et al. 2020), joint learning of segmentation with registration (Xu and Niethammer 2019), *etc.* For example, Zhao et al. (2019) proposed a data augmentation method for one-shot segmentation, which firstly modeled the spatial and appearance transformations between the atlas and unlabeled images, and then synthesized new labeled images by randomly applying the learned transformations to the atlas. With the enlarged training set, the supervised segmentation network was expected to improve performance. Wang et al. (2020) proposed to directly learn bidirectional correspondences between the atlas and unlabeled images, and incorporate supervision in the image, transformation, and label spaces to push the learning system towards an anatomically meaningful direction, by revisiting the classic forward-backward consistency concepts for supervision. Xu and Niethammer (2019) proposed the joint learning of segmentation and registration with a single DL framework, under the assumption that joint optimization would lead to better performance.

We strengthen the network capacity of Zhao et al. (2019) and Wang et al. (2020) with the Siamese-Baseline to incorporate feature pyramid, warping, and cost volume into a Siamese structure, while avoiding extra overhead for registration (Xu and Niethammer 2019). Besides, these works mainly focused on knee (Xu and Niethammer 2019) or brain (Balakrishnan et al. 2019; Zhao et al. 2019), which in general present relatively small inter-subject variability compared with abdominal organs (Schreibmann, Marcus, and Fox 2014). To account for wide morphological variations due to individual differences, we further propose the IDA-Baseline to zero out the contribution of pixels/voxels exclusive to a single individual.

Correspondence Learning. Learning correspondence is a fundamental computer vision problem closely related to a variety of tasks from object to pixel levels (Li et al. 2019; Wang, Jabri, and Efros 2019), such as tracking (Pan, Porikli, and Schonfeld 2009; Kalal, Mikolajczyk, and Matas 2010),

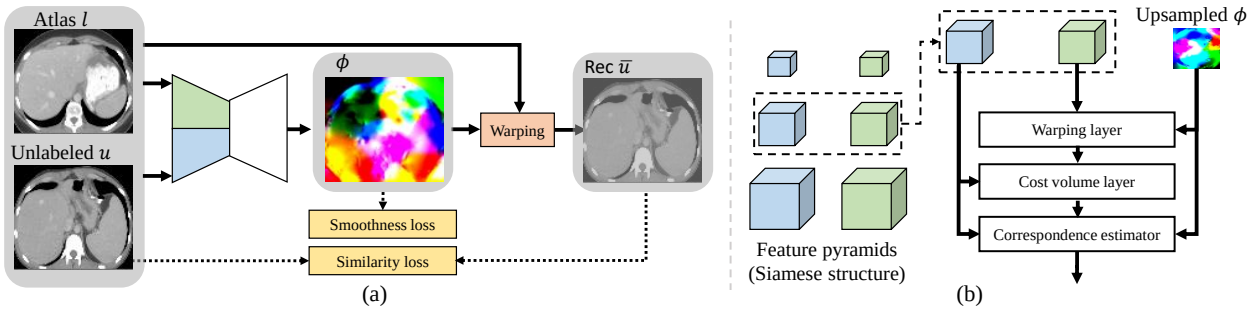

Figure 1: (a) An overview of the proposed Siamese-Baseline. It innovatively incorporates (b) feature pyramid, warping, and cost volume into a Siamese structure for low-shot segmentation. Given the feature pyramids extracted by the encoder, the Siamese-Baseline warps features of the atlas image using the upsampled flow, computes a cost volume, and process the cost volume using CNNs at each pyramid level in the decoder (illustrated with the middle level). The Siamese-Baseline is 3D, but shown here in 2D for simplicity.

patch matching (Bailer, Varanasi, and Stricker 2017), optical flow estimation (Meister, Hur, and Roth 2018; Wang et al. 2018; Sun et al. 2018; Zhao et al. 2020a; Hur and Roth 2017; Liu et al. 2019), stereo matching (Lai, Tsai, and Chiu 2019; Wu et al. 2019), and registration (Balakrishnan et al. 2019; Xu and Niethammer 2019; Haskins, Kruger, and Yan 2020). By treating atlas-based segmentation as a correspondence problem, we could utilize cutting edge technologies from these research areas to guide the design of our framework. For example, the feature pyramid, warping, and cost volume are important building blocks in the network design for dense correspondence learning (Meister, Hur, and Roth 2018; Wang et al. 2018; Sun et al. 2018; Hosni et al. 2012; Lai, Tsai, and Chiu 2019; Wu et al. 2019; Zhao et al. 2020a; Hur and Roth 2017; Liu et al. 2019). Forward-backward consistency has been the evaluation metric (Kalal, Mikolajczyk, and Matas 2010) as well as the measure of uncertainty (Pan, Porikli, and Schonfeld 2009) for tracking, and is a means to define an occluded region, which was excluded for training in optical flow estimation (Meister, Hur, and Roth 2018; Wang et al. 2018; Hur and Roth 2017; Liu et al. 2019).

Methodology

Preliminaries

Let $\{(l^{(i)}, l_s^{(i)})\}_{i=1}^{K}$ denote the atlas and corresponding segmentation pairs, where K is the number of given atlases. Let $\{u^{(i)}\}_{i=1}^{N}$ denote the unlabelled images, where N is the number of unlabelled images. In the following, for an uncluttered notation, we omit the index i and use l, u to denote an atlas image and an unlabelled image, respectively. The goal of atlas-based segmentation in DL is to learn a correspondence map ϕ from l to each $u^{(i)}$ (Fig. 1(a)), which is supervised by an image similarity loss between the warped atlas $\overline{u} = l \circ \phi$, and the original unlabeled image u (where $\circ$ is a warping operation). In addition, a transformation smoothness loss is usually involved to regularize the learned ϕ to be reasonable. When testing, the correspondence map ϕ_s from the atlas l to an unlabeled image u is predicted by the trained network, and the segmentation $\overline{u}_s$ of u can be obtained by $\overline{u}_s = l_s \circ \phi_s$. For example, VoxelMorph (Balakrishnan et al. 2019) employed a 3D U-Net (Çiçek et al. 2016) to learn the correspondence ϕ, and a spatial transformer network (Jaderberg et al. 2015) to implement the warping operation $\circ$.

Siamese-Baseline

A weakness of existing low-shot segmentation works (Wang et al. 2020; Xu and Niethammer 2019; Zhao et al. 2019) is that they do not distinguish between the atlas and target images, and simply fuse them at the first convolutional layer of the network. Such early fusion does not make full use of the multi-level information in the downsampling path. In contrast, our method employs two encoders (one for each image) which interact along the way, mimicking the coarse-to-fine strategy in classical registration approaches. Concretely, we adopt the Siamese structure to separately extract semantics from the atlas and target images (Fig. 1(a)), which has been verified to be effective in low-shot classification problem (Koch, Zemel, and Salakhutdinov 2015) and other computer vision tasks, such as face verification (Taigman et al. 2014), object tracking (Tao, Gavves, and Smeulders 2016), fine-grained classification (Dubey et al. 2018), optical flow estimation (Sun et al. 2018), and visual co-segmentation (Lu et al. 2019). Besides, recent advances in dense correspondence learning practically confirmed the effectiveness of the concepts of feature pyramid, warping, and cost volume (Meister, Hur, and Roth 2018; Wang et al. 2018; Sun et al. 2018; Hosni et al. 2012; Lai, Tsai, and Chiu 2019; Wu et al. 2019; Zhao et al. 2020a; Hur and Roth 2017; Liu et al. 2019; Lai, Tsai, and Chiu 2019; Wu et al. 2019); *e.g.*, in optical flow estimation, all of these concepts are verified useful, such as in the PWC-Net (Sun et al. 2018).

We leverage the above-mentioned advanced technologies to build a strong backbone network for the Siamese-Baseline. Since most of these technologies are implemented in 2D, we reimplement them in 3D. To reduce computational cost of 3D networks, we follow the principle of using fewer channels and downsampling operations. A schematic overview of the network structure and key components is shown in Figs. 1(a) and 1(b), and the detailed network architecture is described in the supplementary material. The image similarity loss $\mathcal{L}_i$ and transformation smoothness loss

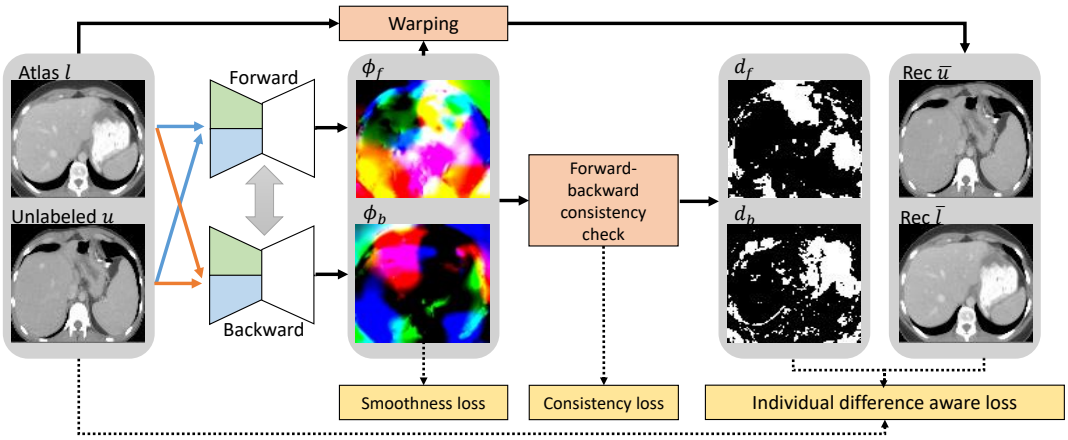

Figure 2: An overview of the proposed IDA-Baseline. Bidirectional training is employed to estimate individual differences between subjects with a forward-backward consistency check. The IDA-Baseline is 3D, but shown here in 2D for simplicity.

$\mathcal{L}_s$ used for the Siamese-Baseline (with a trade-off hyper-parameter λ) are similar to previous methods (Balakrishnan et al. 2019; Zhao et al. 2019; Xu and Niethammer 2019; Wang et al. 2020):

$$\begin{aligned} \mathcal{L}_i(u,\overline{u}) &= 1 - \frac{1}{V}\sum_v \mathcal{CC}(u,\overline{u})^v, \\ \text{and } \mathcal{L}_s(\phi) &= \lambda \sum_v ||\nabla(\phi(v))||_2, \end{aligned} \tag{1}$$

where $\mathcal{CC}$ is the per-voxel map of the local normalized cross-correlation (NCC; Balakrishnan et al. 2019), $V = H \times W \times D$ with H, W, D representing the numbers of voxels along the coronal, sagittal, and axial directions, respectively, and $\mathcal{L}_s$ is formulated with first-order derivatives of ϕ: for each location v of the image space, we approximate $||\nabla(\phi(v))||_2$ with spatial gradient differences between neighboring voxels along x, y, z directions.

Let $f_u(v)$ and $f_{\bar{u}}(v)$ denote images with local mean intensities: $f_u(v) = \frac{1}{n^3}\sum_{v_i} u(v_i)$, where v_i iterates over a n^3 volume around v. Then $\mathcal{CC}$ can be written as:

$$\mathcal{CC}(u,\bar{u})^v = \frac{\left(\sum_{v_i}\left(u(v_i) - f_u(v)\right)\left(\bar{u}(v_i) - f_{\bar{u}}(v)\right)\right)^2}{\left(\sum_{v_i}\left(u(v_i) - f_u(v)\right)^2\right)\left(\sum_{v_i}\left(\bar{u}(v_i) - f_{\bar{u}}(v)\right)^2\right)}. \tag{2}$$

IDA-Baseline

Bidirectional Framework. Although we will experimentally demonstrate the effectiveness of the Siamese-Baseline over existing methods (Balakrishnan et al. 2019; Wang et al. 2020; Zhao et al. 2019) in segmenting brain anatomical structures, all of these methods generalize poorly on organs/tissues with large intrinsic morphological variations, such as abdominal organs (a noteworthy weakness that was also mentioned in the survey of classical atlas-based segmentation (Iglesias and Sabuncu 2015)). To address this issue, we propose the IDA-Baseline to specifically model the individual differences between subjects to enhance the backbone of the Siamese-Baseline (Balakrishnan et al. 2019) via the addition of the forward and backward branches. Concretely, the addition of the forward-backward branches enables us to identify the individual differences by checking the forward-backward consistency, and avoid enforcing similarity constraint on those regions corresponding to the individual differences. An overview of the IDA-Baseline is shown in Fig. 2.

To compute bidirectional correspondences, the IDA-Baseline contains a forward and a backward branch with shared weights. The forward branch takes the atlas and one unlabeled image (l and u) as input and outputs the forward correspondence ϕ_f from l to u, whereas the backward branch takes the same pair of images u and l as input but outputs the backward correspondence ϕ_b from u to l. After that, ϕ_f is used to warp l to reconstruct $\overline{u}$, and ϕ_b is used to warp u to reconstruct $\overline{l}$. A smoothness loss $\mathcal{L}_s'$ is proposed as a regularization constraint on ϕ_f and ϕ_b as:

$$\mathcal{L}_s'(\phi_f,\phi_b) = \lambda_f\mathcal{L}_s(\phi_f) + \lambda_b\mathcal{L}_s(\phi_b), \tag{3}$$

where λ_f and λ_b are the weights balancing the importance of learning forward and backward correspondences.

Individual Difference Identification. The individual difference problem is analogous to the occlusion problem in optical flow estimation, where the differences between organs can be analogously interpreted as being occluded in optical flow estimation, and should not contribute to the image similarity loss. In this sense, we borrow ideas from recent works on occlusion estimation (Meister, Hur, and Roth 2018; Wang et al. 2018; Hur and Roth 2017; Liu et al. 2019), which incorporate a forward-backward consistency check to identify the occluded pixels based on the assumption that, for non-occluded pixels, the forward correspondences should be the inverse of the backward correspondences. Specifically, in our method, we identify the individual difference by a forward-backward consistency check, that is, a voxel is considered as belonging to a single individual when the displacement between the forward correspondence and backward correspondence is too large. Taking the

individual difference in the forward direction for example, we define a difference flag d_f^v to be one for each location v of the image space in the correspondence if the constraint

$$\begin{aligned} &|\phi_f(v) + \phi_b(v + \phi_f(v))|^2 < \\ &\alpha_1(|\phi_f(v)|^2 + |\phi_b(v + \phi_f(v))|^2) + \alpha_2 \end{aligned} \tag{4}$$

is violated, and zero otherwise. In the backward direction, we define d_b^v in the same way with ϕ_f and ϕ_b exchanged. Following Meister, Hur, and Roth (2018), we set $\alpha_1 = 0.01, \alpha_2 = 0.5$. Then, we mask the voxels flagged as individual differences to avoid learning incorrect deformations, and the IDA image similarity loss is thus defined as

$$\begin{aligned} \mathcal{L}_d = &\lambda_f \Big(1 - \frac{1}{V} \sum_v \big(\mathcal{CC}(u, \overline{u})^v \cdot (1 - d_f^v)\big)\Big) \\ &+ \lambda_b \Big(1 - \frac{1}{V} \sum_v \big(\mathcal{CC}(l, \bar{l})^v) \cdot (1 - d_b^v)\big)\Big) \\ &+ \frac{1}{V} \sum_v \big(\rho(d_f^v) + \rho(d_b^v)\big). \end{aligned} \tag{5}$$

We add a constraint on both d_f^v and d_b^v with the robust generalized Charbonnier penalty function (Sun, Roth, and Black 2014) $\rho(x) = (x^2 + \epsilon^2)^\gamma$, where we set $\gamma = 0.45$, again following Meister, Hur, and Roth (2018).

Along with the smoothness loss and IDA image similarity loss, the consistency loss (Meister, Hur, and Roth 2018; Hur and Roth 2017) is adopted to impose constraints on the network for producing consistent predictions for both forward and backward directions, which is formulated as

$$\begin{aligned} \mathcal{L}_c = \frac{0.2}{V} \Big(&\sum_v \rho\Big(\phi_f(v) + \phi_b\big(v + \phi_f(v)\big)\Big) \\ &+ \sum_v \rho\Big(\phi_b(v) + \phi_f\big(v + \phi_b(v)\big)\Big)\Big). \end{aligned} \tag{6}$$

The final loss for training the IDA-Baseline is thus the sum of $\mathcal{L}_d$, $\mathcal{L}_c$, and $\mathcal{L}_s'$. Note that in the loss function of the IDA-Baseline, we only include two hyper-parameters of λ_f and λ_b to balance the importance of learning forward and backward correspondences to other regularization constraints, avoiding too much parameter tuning.

Training and Inference Details

To investigate the k-shot segmentation problem, we assume that only k images have associated labels (k from one to five). Since the atlas is important for learning correspondences and atlas selection is a crucial step in atlas-based segmentation (Iglesias and Sabuncu 2015), in the training phase we select the atlases with a similarity rank by computing the average score of per-voxel local NCC (Zhao et al. 2019) between each of the image pair (one from the training set, and the other from the test set). Specifically, for each image in the training set, we aggregate scores by averaging the scores of the image to all test images, and select the top k examples as the atlases. In the training process, the images in the k atlases are randomly coupled with the unlabeled images in the training set to train the Siamese-Baseline and IDA-Baseline. Note that the segmentation labels are not used during training. In the test phase, to get the segmentation $\overline{u}_s$ of the target image u, we first compute k warped atlases by separately inputting the k atlases each coupled with the target image into the Siamese-Baseline and IDA-Baseline. For the one-shot setting, the segmentation can be obtained with the forward correspondence ϕ_f from the selected atlas to u. For few-shot settings, we employ two different label fusion strategies in accordance with different scenarios. First, for the Siamese-Baseline targeted at anatomically stable structures, we compute per-voxel local NCC between each pair of k warped atlas $\overline{u}$ and the original target image u, and fed the local NCCs to a softmax function to weight the k warped segmentation $\{\overline{u}_s^{(i)}, i = 1, 2, ..., k\}$; the final $\overline{u}_s$ is then obtained by selecting the maximum score among $\overline{u}_s^{(i)}$ voxel by voxel. Second, for the IDA-Baseline proposed to deal with large morphological variations, we use the segmentation of the warped atlas $\overline{u}$ which is most similar to the original target image u (ranked by the average score of per-voxel local NCCs (Zhao et al. 2019)) as the segmentation $\overline{u}_s$ of u.

Correlation with Atlas-based Segmentation. As suggested in a survey (Iglesias and Sabuncu 2015), registration and label propagation are indispensable parts of registration-based segmentation; when further combined with label fusion, they constitute the indispensable parts of multi-atlas segmentation. Our one-shot and few-shot settings naturally implements the core concepts of registration-based segmentation and multi-atlas segmentation within a single framework, and such an end-to-end framework is expected to achieve high performance by Iglesias and Sabuncu (2015).

Experiments

Since the tasks of segmenting brain anatomical structures and abdominal organs are noticeably different, we evaluate the Siamese-Baseline on brain anatomical structures (the CANDI Dataset (Kennedy et al. 2011)), and the IDA-Baseline on abdominal organs (the Multi-organ Dataset (Gibson et al. 2018; Roth et al. 2015; Clark et al. 2013; Landman et al. 2015)). The CANDI dataset consists of 103 brain magnetic resonance imaging (MRI) scans, and we use 28 anatomical structures that were used in VoxelMorph (Balakrishnan et al. 2019) as the segmentation target. The Multi-organ Dataset comprises 90 CT images, and we choose seven organs (listed in Table 3) that exist in most of the images as the segmentation targets. For both datasets, we randomly select 20 volumes as test data, and use the others for training. The details of both datasets can be found in the supplementary material.

Implementation Details

All experiments are implemented with Keras 2.2.0 (Chollet et al. 2015) and TensorFlow 1.10.0 (Abadi et al. 2016). The network is trained with the Adam (Kingma and Ba 2014) optimizer with a learning rate of 0.0002 for the Siamese-Baseline for 600 epochs and 0.0001 for the IDA-Baseline for 2,000 epochs. We train the Siamese-Baseline and IDA-Baseline on one NVIDIA GeForce RTX 2080 Ti GPU with

Table 1: Comparison of our Siamese-Baseline ($\lambda = 1$) with VoxelMorph (Balakrishnan et al. 2019), DataAug (Zhao et al. 2019), LT-Net (Wang et al. 2020), and DeepAtlas (Xu and Niethammer 2019) in low-shot settings on the CANDI dataset; the fully supervised U-Net (Çiçek et al. 2016) trained with 83 samples is served as the upper bound. We show the mean Dice scores (%) with standard deviations. Besides, Min and Max represent the minimum and maximum Dice scores (%).

Method	1-shot			5-shot		
	Mean (std)	Min	Max	Mean (std)	Min	Max
VoxelMorph	76.0 (9.7)	61.7	80.1	83.4 (6.4)	75.9	87.1
DataAug	80.4 (4.3)	73.8	84.0	84.1 (6.2)	78.2	87.5
LT-Net	82.3 (2.5)	75.6	84.2	84.6 (6.2)	77.1	87.8
DeepAtlas	66.5 (1.8)	63.7	70.6	74.1 (2.3)	67.5	77.3
Siamese-Baseline	**83.0 (1.8)**	77.6	85.9	**86.2 (2.1)**	78.1	88.4
U-Net (upper bound)	86.5 (6.3)	83.7	89.2	86.5 (6.3)	83.7	89.2

a single pair of volumes for each batch, on a workstation with Ubuntu 18.04.2 LTS and 251 GB memory. The same as DataAug (Zhao et al. 2019) and LT-Net (Wang et al. 2020), we use the average Dice score as the evaluation metric. In all experiments, the NCC searches an n^3 local cube with $n = 9$ as Balakrishnan et al. (2019). We use random flipping for data augmentation to alleviate the overfitting problem. It is worth mentioning that we also tune the comparison methods for optimal performance.

Siamese-Baseline on Brain Anatomical Structures

We compare the proposed Siamese-Baseline with several SOTA approaches implemented with DL, *i.e.,* VoxelMorph (Balakrishnan et al. 2019), DataAug (Zhao et al. 2019), LT-Net (Wang et al. 2020), and DeepAtlas (Xu and Niethammer 2019). Although VoxelMorph was proposed for medical image registration, we compare to it in low-shot settings as also done in DataAug and LT-Net. Besides, as DataAug and LT-Net were proposed for one-shot setting, we adapt them for the few-shot settings, where more atlases (here, five) are involved for training. The results are shown in Table 1. We can observe that the proposed Siamese-Baseline outperforms the SOTA methods in both the one-shot and few-shot settings. Especially, our Siamese-Baseline using five atlases achieves very competitive results compared to the fully-supervised U-Net trained with 83 samples. The good performance confirms the important roles of feature pyramid, warping, and cost volume in atlas-based segmentation.

We further study the performance with different combinations of warping and cost volume at different levels of feature pyramids, and the sensitivity of the Siamese-Baseline with different settings of λ in Eq. (1). Due to the page limit, we put the results in the supplementary material.

IDA-Baseline on Abdominal Organs

Ablation Study on the Number of Atlases: We present experimental results of the proposed IDA-Baseline with k-shot segmentation across various organs. The results with k equaling to one to five are shown in Table 2. The results indicate that, as the number of atlases increases, better results are achieved, which is in accordance with our intuition. Besides, the consistent improvements verify the flexibility and expansibility of our IDA-Baseline, which should benefit the clinical practice, that is, with more data being labeled, IDA-Baseline should present a better performance.

Ablation Study on Key Components: To investigate the effectiveness of the key components of our proposed method—forward-backward correspondence consistency and individual difference estimation, we perform ablation studies on the training set with five atlases. We first adopt the Siamese-Baseline as the initial performance baseline. Specifically, the network is trained to learn forward correspondence only, and propagates the labels of atlases to the segmentation of unlabeled images. Based on the basic framework, we then add a backward correspondence learning path to form a complete cycle. As the forward correspondence should be the inverse of the backward correspondence at the corresponding pixels, we introduce forward-backward correspondence consistency loss defined as Eq. (6) in the learning process. After that, we further replace the image similarity loss with the proposed IDA loss to model wide variations in organ morphology as defined in Eq. (5).

The results are listed in Table 3. As shown, compared to the baseline, the proposed IDA-Baseline achieves 2.14% and 4.97% improvements in average Dice score by gradually adding the forward-backward correspondence consistency and IDA loss. It is worth noting that the IDA loss achieves a further 2.83% improvement in average Dice score over only imposing the forward-backward correspondence consistency. Our IDA-Baseline takes into account the wide variations in organ morphology, which may damage the performance in correspondence learning. Unsurprisingly, by removing the contribution of voxels belonging to a single individual in the loss computation, the network can be trained better under more uncluttered supervision.

Comparison with Other SOTA Methods: We first compare our IDA-Baseline with two classical multi-atlas/registration methods: PICSL_MALF (Wang and Yushkevich 2013) with the joint label fusion and corrective learning techniques, and DEEDS (Heinrich et al. 2013) the SOTA for classical registration-based abdominal organ segmentation (Rohlfing et al. 2004). All methods are trained with five atlases. The results are tabulated in Table 4. We can observe that our IDA-Baseline outperforms PICSL_MALF by a large margin of 10.91%, and outperforms DEEDS by a noticeable margin of 3.38%. In addition, we set up the comparison experiment with two recent SOTA approaches implemented with DL, *i.e.,* DataAug (Zhao et al. 2019) and LT-Net (Wang et al. 2020)[1]. Comparatively, we also present the results by DataAug and LT-Net in one-shot settings in Table 2. As shown in Table 2 and Table 4, the

[1]Despite our efforts, we are unable to tune VoxelMorph (Balakrishnan et al. 2019) or DeepAtlas (Xu and Niethammer 2019) to yield comparable results with DataAug and LT-Net on abdominal organs. Therefore, their results are not included.

Table 2: Dice scores (%) of IDA-Baseline ($\lambda_f = 3, \lambda_b = 1$) with k-shot segmentation across various organs. One-shot results with DataAug (Zhao et al. 2019), and LT-Net (Wang et al. 2020) are also presented.

Num. of Atlases	Spleen	Left Kidney	Liver	Stomach	Pancreas	Duodenum	Esophagus	Mean
1 (DataAug)	71.00 ± 16.29	81.22 ± 9.67	86.55 ± 5.59	60.57 ± 15.00	35.59 ± 12.21	$\mathbf{38.12 \pm 10.72}$	40.09 ± 18.17	59.02
1 (LT-Net)	74.72 ± 11.24	78.91 ± 11.41	82.76 ± 4.78	47.69 ± 13.12	$\mathbf{43.13 \pm 9.88}$	30.35 ± 8.31	$\mathbf{48.75 \pm 14.41}$	58.05
1 (IDA-Baseline)	$\mathbf{75.03 \pm 16.64}$	$\mathbf{83.94 \pm 8.45}$	$\mathbf{87.58 \pm 5.31}$	$\mathbf{61.66 \pm 14.55}$	35.94 ± 13.34	38.09 ± 11.10	46.73 ± 17.18	**61.28**
2 (IDA-Baseline)	76.91 ± 17.73	84.55 ± 8.93	89.05 ± 5.03	64.08 ± 16.92	40.47 ± 14.35	41.55 ± 13.16	45.52 ± 16.07	63.16
3 (IDA-Baseline)	76.99 ± 15.54	86.25 ± 7.16	88.6 ± 5.54	61.97 ± 19.33	42.81 ± 16.55	42.68 ± 15.11	51.55 ± 13.75	64.41
4 (IDA-Baseline)	77.36 ± 16.41	85.72 ± 10.02	90.07 ± 3.76	64.65 ± 16.93	44.12 ± 17.48	44.00 ± 14.30	51.83 ± 15.21	65.39
5 (IDA-Baseline)	$\mathbf{80.34 \pm 15.61}$	$\mathbf{88.44 \pm 5.33}$	$\mathbf{91.55 \pm 2.47}$	$\mathbf{68.40 \pm 15.18}$	$\mathbf{51.14 \pm 12.94}$	$\mathbf{48.50 \pm 12.34}$	$\mathbf{51.91 \pm 14.33}$	**68.61**

Table 3: Ablation study on the proposed IDA-Baseline ($\lambda_f = 3, \lambda_b = 1$) equipped with different modules using five atlases (5-shot). We show the Dice scores (%) across various organs.

Method	Spleen	Left Kidney	Liver	Stomach	Pancreas	Duodenum	Esophagus	Mean
Siamese-Baseline	78.88 ± 13.47	83.79 ± 8.17	88.92 ± 3.66	61.61 ± 16.13	41.13 ± 15.78	43.67 ± 12.07	47.47 ± 14.01	63.64
+cycle-consistency	78.29 ± 15.87	85.81 ± 7.32	89.93 ± 3.85	65.26 ± 17.85	46.26 ± 16.20	44.84 ± 13.42	50.08 ± 14.64	65.78
+Individual modeling	$\mathbf{80.34 \pm 15.61}$	$\mathbf{88.44 \pm 5.33}$	$\mathbf{91.55 \pm 2.47}$	$\mathbf{68.40 \pm 15.18}$	$\mathbf{51.14 \pm 12.94}$	$\mathbf{48.50 \pm 12.34}$	$\mathbf{51.91 \pm 14.33}$	**68.61**

Table 4: Comparison of IDA-Baseline ($\lambda_f = 3, \lambda_b = 1$) with PICSL_MALF (Wang and Yushkevich 2013), DEEDS (Heinrich et al. 2013), DataAug (Zhao et al. 2019), and LT-Net (Wang et al. 2020) trained with five atlases (5-shot). The results with fully supervised U-Net (Çiçek et al. 2016) serve as the upper bound. We show the Dice scores (%) across various organs.

Method	Spleen	Left Kidney	Liver	Stomach	Pancreas	Duodenum	Esophagus	Mean
PICSL_MALF	76.62 ± 14.90	79.22 ± 14.05	85.50 ± 5.06	59.92 ± 16.10	37.93 ± 19.89	32.96 ± 17.12	31.73 ± 21.21	57.70
DEEDS	76.11 ± 11.53	76.17 ± 19.02	89.59 ± 2.89	68.20 ± 12.33	50.53 ± 12.10	45.00 ± 9.66	51.02 ± 17.97	65.23
DataAug	74.05 ± 17.93	82.92 ± 9.68	88.09 ± 4.07	61.57 ± 19.04	40.32 ± 14.68	40.65 ± 13.65	43.80 ± 14.39	61.63
LT-Net	76.79 ± 16.48	83.92 ± 9.10	88.70 ± 4.68	63.30 ± 16.08	38.47 ± 14.57	41.25 ± 13.51	46.20 ± 15.60	62.66
IDA-Baseline	$\mathbf{80.34 \pm 15.61}$	$\mathbf{88.44 \pm 5.33}$	$\mathbf{91.55 \pm 2.47}$	$\mathbf{68.40 \pm 15.18}$	$\mathbf{51.14 \pm 12.94}$	$\mathbf{48.50 \pm 12.34}$	$\mathbf{51.91 \pm 14.33}$	**68.61**
U-Net (upper bound)	93.04 ± 5.27	92.96 ± 2.30	94.31 ± 1.13	81.56 ± 9.88	71.88 ± 10.74	63.10 ± 11.42	62.18 ± 10.98	79.86

IDA-Baseline achieves the best segmentation performance in both one-shot and k-shot settings. Besides, although the margins of improvements over DataAug and LT-Net in one-shot settings are small, our IDA-Baseline in five-shot settings outperforms these two methods by large margins of 6.98% and 5.95%, respectively. This can be attributed to the individual difference modeling in abdominal CT images of the IDA-Baseline. In addition, with more atlases, the IDA-Baseline can select the most similar atlases for warping. We visualize some example slices of segmentation results in the supplementary material.

We further study how the hyper-parameters of λ_f and λ_b in Eqs. (3) and (5) affect the performance, and put the results in the supplementary material due to the limited space.

Conclusion and Future Work

In this paper, we proposed the Siamese-Baseline and IDA-Baseline, which are served as alternative baselines for low-shot segmentation of data suffering slight and considerable individual differences, respectively. Experimental results verified the effectiveness of these two baselines in two distinctive situations, *i.e.,* segmentation of brain structures with relatively stable anatomy and abdominal organs suffering large morphological variations between individuals.

We believe the correlation of atlas-based low-shot segmentation with correspondence learning in natural images would further inspire the research community of medical image analysis to rethink critically the design of neural networks and loss functions. It should be noted that there still remains an apparent gap between low-shot segmentation and fully supervised learning in segmenting organs suffering large morphological variations, which should be further addressed by joint efforts of the research community. The main practice of atlas-based segmentation is on tissues/organs, and due attention should be paid to how it can contribute to the segmentation of tumors/lesions. For example, it is possible to learn the tumor/lesion segmentation by comparing the images of tumor patients to those of normal subjects to explicitly taking into account the difference in appearance, and we plan to explore this direction in future work. Lastly, although our one-shot and few-shot settings naturally match the core concepts of classical atlas-based segmentation with an end-to-end framework, an important next step is how to implement such concepts more effectively and efficiently. We hope this work would inspire the development of low-shot learning for medical image segmentation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61671399), the

Fundamental Research Funds for the Central Universities (Grant No. 20720190012), the Key-Area Research and Development Program of Guangdong Province, China (No. 2018B010111001), the National Key R&D Program of China (2018YFC2000702), and the Scientific and Technical Innovation 2030-"New Generation Artificial Intelligence" Project (No. 2020AAA0104100).

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; et al. 2016. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* .

Bailer, C.; Varanasi, K.; and Stricker, D. 2017. CNN-based patch matching for optical flow with thresholded hinge embedding loss. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 3250–3259.

Balakrishnan, G.; Zhao, A.; Sabuncu, M. R.; Guttag, J.; and Dalca, A. V. 2019. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging* 38(8): 1788–1800.

Chollet, F.; et al. 2015. Keras. https://keras.io.

Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; and Ronneberger, O. 2016. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, 424–432. Springer.

Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; et al. 2013. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging* 26(6): 1045–1057.

Dong, N.; and Xing, E. 2018. Few-Shot Semantic Segmentation with Prototype Learning. In *The British Machine Vision Conference*, volume 3.

Dubey, A.; Gupta, O.; Guo, P.; Raskar, R.; Farrell, R.; and Naik, N. 2018. Pairwise confusion for fine-grained visual classification. In *Proc. European Conf. Computer Vision*, 70–86.

Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; et al. 2018. Automatic multi-organ segmentation on abdominal CT with dense V-Networks. *IEEE Transactions on Medical Imaging* 37(8): 1822–1834.

Haskins, G.; Kruger, U.; and Yan, P. 2020. Deep learning in medical image registration: A survey. *Machine Vision and Applications* 31(1-2).

Heinrich, M. P.; Jenkinson, M.; Brady, M.; and Schnabel, J. A. 2013. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Transactions on Medical Imaging* 32(7): 1239–1248.

Hosni, A.; Rhemann, C.; Bleyer, M.; Rother, C.; and Gelautz, M. 2012. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(2): 504–511.

Hur, J.; and Roth, S. 2017. MirrorFlow: Exploiting symmetries in joint optical flow and occlusion estimation. In *Proc. Int'l Conf. Computer Vision*, 312–321.

Iglesias, J. E.; and Sabuncu, M. R. 2015. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis* 24(1): 205–219.

Jaderberg, M.; Simonyan, K.; Zisserman, A.; et al. 2015. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, 2017–2025.

Jia, H.; Yap, P.-T.; and Shen, D. 2012. Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage* 59(1): 422–430.

Kalal, Z.; Mikolajczyk, K.; and Matas, J. 2010. Forward-backward error: Automatic detection of tracking failures. In *20th International Conference on Pattern Recognition*, 2756–2759. IEEE.

Kennedy, D. N.; Haselgrove, C.; Hodge, S. M.; Rane, P. S.; Makris, N.; et al. 2011. CANDIShare: A Resource for Pediatric Neuroimaging Data. *Neuroinformatics* 10(3): 319–322.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Koch, G.; Zemel, R.; and Salakhutdinov, R. 2015. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2. Lille.

Lai, H.-Y.; Tsai, Y.-H.; and Chiu, W.-C. 2019. Bridging Stereo Matching and Optical Flow via Spatiotemporal Correspondence. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1890–1899.

Landman, B.; Xu, Z.; Eugenio Igelsias, J.; Styner, M.; Langerak, T.; and Klein, A. 2015. MICCAI Multi-Atlas Labeling Beyond the Cranial Vault–Workshop and Challenge.

Li, X.; Liu, S.; De Mello, S.; Wang, X.; Kautz, J.; and Yang, M.-H. 2019. Joint-task self-supervised learning for temporal correspondence. In *Advances in Neural Information Processing Systems*, 317–327.

Liu, P.; Lyu, M.; King, I.; and Xu, J. 2019. SelFlow: Self-supervised learning of optical flow. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 4571–4580.

Lorenzo-Valdés, M.; Sanchez-Ortiz, G. I.; Mohiaddin, R.; and Rueckert, D. 2002. Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, 642–650. Springer.

Lu, X.; Wang, W.; Ma, C.; Shen, J.; Shao, L.; and Porikli, F. 2019. See More, Know More: Unsupervised Video Object Segmentation with Co-Attention Siamese Networks. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 3623–3632.

Meister, S.; Hur, J.; and Roth, S. 2018. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. In *Proc. AAAI Conf. Artificial Intelligence*.

Nguyen, K.; and Todorovic, S. 2019. Feature weighting and boosting for few-shot segmentation. In *Proc. Int'l Conf. Computer Vision*, 622–631.

Pan, P.; Porikli, F.; and Schonfeld, D. 2009. Recurrent tracking using multifold consistency. In *Proceedings of the Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*.

Rohlfing, T.; Brandt, R.; Menzel, R.; and Maurer Jr, C. R. 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* 21(4): 1428–1442.

Roth, H. R.; Lu, L.; Farag, A.; Shin, H.-C.; Liu, J.; Turkbey, E. B.; et al. 2015. DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, 556–564. Springer.

Schreibmann, E.; Marcus, D. M.; and Fox, T. 2014. Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search. *Journal of Applied Clinical Medical Physics* 15(4): 22–38.

Shaban, A.; Bansal, S.; Liu, Z.; Essa, I.; and Boots, B. 2017. One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410* .

Siam, M.; Oreshkin, B. N.; and Jagersand, M. 2019. AMP: Adaptive masked proxies for few-shot segmentation. In *Proc. Int'l Conf. Computer Vision*, 5249–5258.

Snell, J.; Swersky, K.; and Zemel, R. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, 4077–4087.

Sun, D.; Roth, S.; and Black, M. J. 2014. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision* 106(2): 115–137.

Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 8934–8943.

Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. 2014. DeepFace: Closing the gap to human-level performance in face verification. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1701–1708.

Tao, R.; Gavves, E.; and Smeulders, A. W. 2016. Siamese instance search for tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1420–1429.

Wang, H.; and Yushkevich, P. 2013. Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. *Frontiers in NeuroInformatics* 7: 27.

Wang, K.; Liew, J. H.; Zou, Y.; Zhou, D.; and Feng, J. 2019a. PANet: Few-shot image semantic segmentation with prototype alignment. In *Proc. Int'l Conf. Computer Vision*, 9197–9206.

Wang, S.; Cao, S.; Wei, D.; Wang, R.; Ma, K.; Wang, L.; et al. 2020. LT-Net: Label Transfer by Learning Reversible Voxel-wise Correspondence for One-shot Medical Image Segmentation. *arXiv preprint arXiv:2003.07072* .

Wang, X.; Jabri, A.; and Efros, A. A. 2019. Learning correspondence from the cycle-consistency of time. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2566–2576.

Wang, Y.; Yang, Y.; Yang, Z.; Zhao, L.; Wang, P.; and Xu, W. 2018. Occlusion aware unsupervised learning of optical flow. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 4884–4893.

Wang, Y.; Yao, Q.; Kwok, J.; and Ni, L. M. 2019b. Generalizing from a few examples: A survey on few-shot learning. In *arXiv preprint arXiv: 1904.05046*.

Wu, Z.; Wu, X.; Zhang, X.; Wang, S.; and Ju, L. 2019. Semantic Stereo Matching with Pyramid Cost Volumes. In *Proc. Int'l Conf. Computer Vision*, 7484–7493.

Xu, Z.; and Niethammer, M. 2019. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, 420–429. Springer.

Zhang, X.; Wei, Y.; Yang, Y.; and Huang, T. 2018. SG-One: Similarity guidance network for one-shot semantic segmentation. *arXiv preprint arXiv:1810.09091* .

Zhao, A.; Balakrishnan, G.; Durand, F.; Guttag, J. V.; and Dalca, A. V. 2019. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 8543–8553.

Zhao, S.; Sheng, Y.; Dong, Y.; Chang, E. I.; Xu, Y.; et al. 2020a. MaskFlownet: Asymmetric Feature Matching with Learnable Occlusion Mask. *arXiv preprint arXiv:2003.10955* .

Zhao, Y.; Price, B.; Cohen, S.; and Gurari, D. 2020b. Objectness-Aware One-Shot Semantic Segmentation. *arXiv preprint arXiv:2004.02945* .