Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images

Qiangguo Jin [a,c,1], Hui Cui [b], Changming Sun [c], Zhaopeng Meng [a,d], Leyi Wei [e], Ran Su [a,*]

[a] *School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China*
[b] *Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia*
[c] *CSIRO Data61, Sydney, Australia*
[d] *Tianjin University of Traditional Chinese Medicine, Tianjin, China*
[e] *School of Software, Shandong University, Shandong, China*

ABSTRACT

The capability of generalization to unseen domains is crucial for deep learning models when considering real-world scenarios. However, current available medical image datasets, such as those for COVID-19 CT images, have large variations of infections and domain shift problems. To address this issue, we propose a prior knowledge driven domain adaptation and a dual-domain enhanced self-correction learning scheme. Based on the novel learning scheme, a domain adaptation based self-correction model (DASC-Net) is proposed for COVID-19 infection segmentation on CT images. DASC-Net consists of a novel attention and feature domain enhanced domain adaptation model (AFD-DA) to solve the domain shifts and a self-correction learning process to refine segmentation results. The innovations in AFD-DA include an image-level activation feature extractor with attention to lung abnormalities and a multi-level discrimination module for hierarchical feature domain alignment. The proposed self-correction learning process adaptively aggregates the learned model and corresponding pseudo labels for the propagation of aligned source and target domain information to alleviate the overfitting to noises caused by pseudo labels. Extensive experiments over three publicly available COVID-19 CT datasets demonstrate that DASC-Net consistently outperforms state-of-the-art segmentation, domain shift, and coronavirus infection segmentation methods. Ablation analysis further shows the effectiveness of the major components in our model. The DASC-Net enriches the theory of domain adaptation and self-correction learning in medical imaging and can be generalized to multi-site COVID-19 infection segmentation on CT images for clinical deployment.

1. Introduction

Deep learning (DL) methods achieved remarkable success with the fundamental hypothesis that the training data and test data are from an identical distribution (Robinson et al., 2020). However, in real-world scenarios, this hypothesis is often violated in the fields of natural image and medical image processing. A typical situation in the medical field is the usage of various image datasets, which can differ significantly in data distribution due to different hospitals, scanner vendors, imaging protocols, patient populations, etc (Abràmoff, Lavin, Birch, Shah, & Folk, 2018; Dou et al., 2018). As the distribution of domain changes, a well-trained system may fail to produce precise predictions for unseen data with domain shift.

To tackle this issue, domain adaptation (DA) algorithms normally learn to align source and target data in the domain-invariant discriminative feature space (Dou et al., 2018; Chen, Lin, Yang, & Huang, 2019; Zhang, Qiu, Yao, Liu, & Mei, 2018; Saporta, Vu, Cord, & Pérez, 2020; Chen, Dou, Chen, Qin, & Heng, 2020; Zhu, Du, & Yan, 2019). Many of those DA methods embed the source and target data into the latent space to obtain similar distributions. By such embedding, the well-trained model on the source domain can be applied to the target domain (Wang, Du, & Guo, 2019). However, obstacles exist in those types of DA methods. Firstly, most of the DA models focus on narrowing the distribution variations in higher-order latent space while neglecting the influence of hierarchical low-level semantic features in the antecedent. Secondly, few DA models integrate prior knowledge in the feature
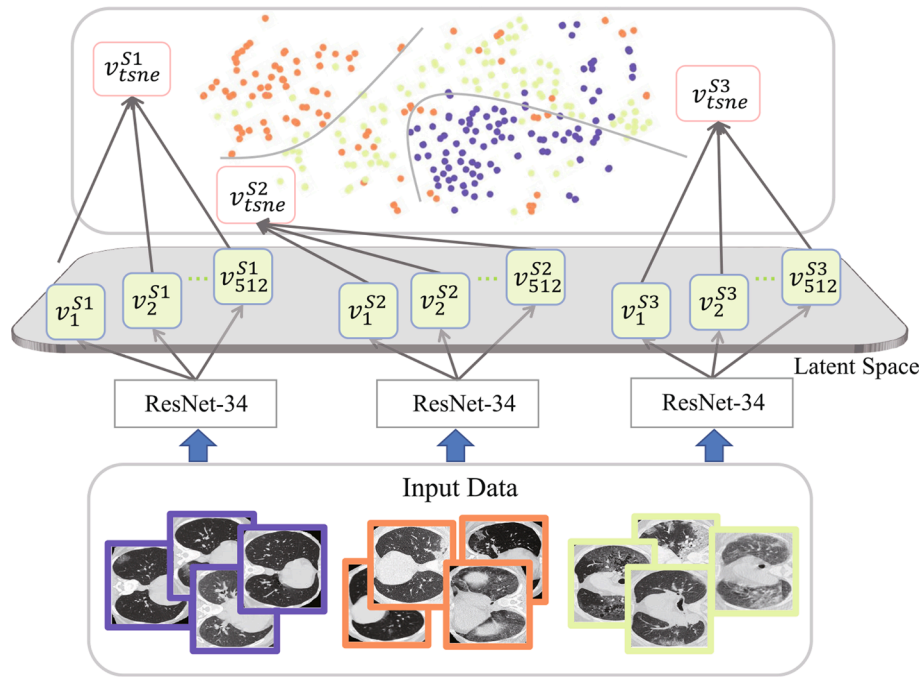
Fig. 1. Illustration of the domain shift problem in the three datasets of COVID-19 used in this work. Latent space features extracted by ResNet-34 are visualized by t-SNE, where each cluster has 100 randomly selected samples from each domain. v_i^{S1} denotes the i-th feature vector extracted from domain $S1$ in the latent space, $S1$, $S2$, and $S3$ denote different domains, and the feature vector v_{tsne}^{S1} is embedded by t-SNE.
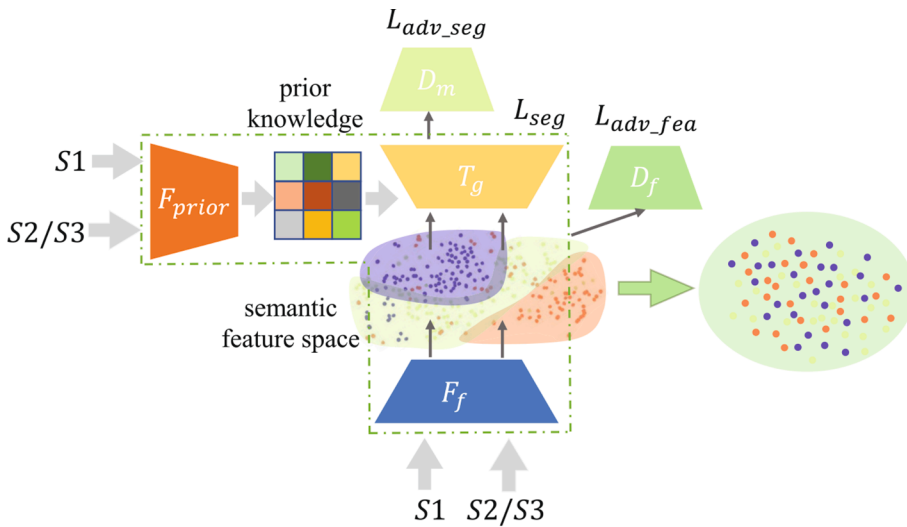
(a) Prior knowledge driven domain adaptation
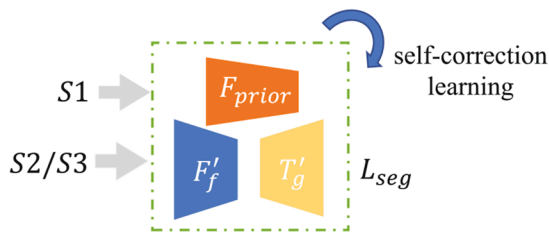
Fig. 2. An overview of the proposed (a) prior knowledge-driven domain adaptation and (b) dual-domain enhanced self-correction learning scheme to address the domain shift problem in semantic feature space, especially when there are limited amounts of data for training. F_f and T_g denote the feature extractor and decoder. D_m and D_f are the mask-level and hierarchical feature level discriminators. F_{prior} is prior knowledge extractor. F_f' and T_g' refer to the adaptively refined F_f and T_g. L_{seg}, L_{adv_seg}, and L_{adv_fea} denote the segmentation loss, mask domain adversarial loss, and feature domain adversarial loss function.

(b) Self-correction learning

extraction and discrimination process. Prior knowledge, however, can provide supervision especially when there is a limited amount of data available for training. Finally, for those DA based segmentation models, the segmentation masks, or the so-called pseudo labels, are obtained based on DA results of the target domain. The source and aligned target domains, however, are not revisited for potential refinement over the pseudo labels.

To address those issues, we propose a novel domain adaptation based self-correction learning (DASC) algorithm. In practice, the novel algorithm is applied to the coronavirus infection segmentation task. Coronavirus (COVID-19) infection segmentation from computed tomography (CT) images is an emerging issue with the outbreak of COVID-19 (Zhu et al., 2020). The automatic segmentation of COVID-19 infections on CT images, however, is challenging because of several reasons. Firstly, the data volume of COVID-19 CT images is limited. Besides, datasets with pixel-level annotations are even harder to obtain when the world, especially experts with domain knowledge, is fighting the COVID-19 pandemic. Secondly, domain shifts exist in the currently available public COVID-19 CT image datasets as illustrated in Fig. 1. Thirdly, large variations exhibit in COVID-19 infections on CT images, such as irregular shapes, indistinct boundaries, and inhomogeneous intensity distributions (Wang et al., 2020). It makes the automated localization and segmentation of abnormalities even challenging.

To make full use of the limited amount of available datasets and address the domain shift problems, we propose a novel prior knowledge driven domain adaptation and a dual-domain enhanced self-correction learning scheme (Fig. 2). Given data from both source ($S1$) and target ($S2/S3$) domains, we firstly extract features under the supervision of prior knowledge from image-level labels. To solve the domain shift problem in semantic feature space, we introduce hierarchical feature-level alignment and discrimination. Then the initial segmentation, i.e., pseudo labels, is obtained after alignment. Finally, we propose a dual-domain enhanced self-correction learning mechanism to adaptively refine the pseudo labels by learning from both source and target domain samples. In practice, we propose a domain adaption based self-correction model (DASC-Net). The unique contributions of the proposed algorithm are summarized below.

- Theoretically, we propose a novel prior knowledge driven domain adaptation and a dual-domain enhanced self-correction learning scheme. The proposed DA scheme is capable of minimizing the impact of domain shift in the mask-level domain and hierarchical feature-level domain. In the meanwhile, the potential refinement over the pseudo labels is formulated by the self-correction learning scheme.
- Practically, we propose a novel attention and feature domain enhanced DA model (AFD-DA) to improve the segmentation performance, especially when there are variations of infections and domain shifts. The innovations include a class activation map (CAM) (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016) based segmentation to emphasize lung abnormalities, and hierarchical feature-level alignment and discrimination.
- To fully utilize the limited amount of available data for training, we propose a dual-domain enhanced self-correction learning algorithm to refine the segmentation results. The self-correction learning algorithm enables model optimization and pseudo label aggregation during training cycles. Besides, the source and target domains are co-learnt in self-correction to minimize the misleading supervision from noises caused by pseudo labels.
- Extensive experiments using three public COVID-19 CT datasets and comprehensive evaluations demonstrated improved segmentation results over state-of-the-art methods for medical image segmentation and domain adaptation. Besides, the extensive experiments show

that our proposed DASC-Net fills the gap between theoretical and practical implications. The source code will be released publicly at GitHub.[2]

In the following sections, we present related work in Section 2. The methodologies are given in Section 3, where we first introduce the basic DA model, followed by the novel AFD-DA model, self-correction learning, and implementation details. Datasets and pre-processing strategy are given in Section 4. Section 5 presents and discusses experimental results with a conclusion in Section 6.

2. Related works

In the past decades, various computer vision applications have been developed in stereo-vision, 3D reconstruction, object detection, and segmentation (Tang et al., 2019; Jin, Cui, Sun, Meng, & Su, 2020; Zlocha, Dou, & Glocker, 2019; Dou, Liu, Heng, & Glocker, 2020). Among the proposed vision applications, artificial intelligence (AI) models attracted intensive research interests (Chen et al., 2020; Tang et al., 2020; Tang, Wang, Luo, & Zou, 2020; Jin et al., 2020). For instance, Chen et al. (2020) utilized DeepLabV3 + network to segment banana central stocks as a way of image pre-processing. Tang et al. (2020) corrected the relative positions of the two clouds using PointNet++ (Qi, Yi, Su, & Guibas, 2017) for 3D reconstruction of large-scale concrete-filled steel tubes. Recent methods for semantic segmentation are mainly based on deep learning (Lv, Liang, Chen, & Lin, 2020; Zhou, Canu, & Ruan, 2020; Qiu, Liu, & Xu, 2020; Shi et al., 2021), and some of which have achieved state-of-the-art performances in COVID-19 analysis and domain adaptation fields.

2.1. Computer-aided COVID-19 analysis from CT images

Recent applications of AI technologies to COVID-19 CT images fall into three categories: severity assessment, COVID-19 screening from other types of pneumonia, and infection segmentation (Shi et al., 2021).

Severity assessment: He et al. (2020) proposed a multi-instance based architecture to predict the severity of COVID-19. The synergistic learning framework achieved joint segmentation of lung lobes and classification of COVID-19. Tang et al. (2020) trained a random forest model using severity-related features to automatically quantify the severity.

COVID-19 screening: Shi et al. (2020) developed an infection size-aware random forest model where location-specific features were extracted and sorted for classification. Attention-based 3D convolutional neural network (CNN) was developed by Ouyang et al. (2020) to improve the screening performance with the focus on COVID-19 infected regions. Kang et al. (2020) proposed a multi-view representation learning model to extract complementary types of features for COVID-19 diagnosis.

Infection segmentation: To automatically segment COVID-19 infected regions, Zhou et al. (2020) developed a DL model using aggregated residual transformations and attention mechanism. Chen, Yao, and Zhang (2020) extracted contextual features by incorporating spatial and channel attentions to a U-Net architecture. The proposed method was evaluated on a COVID-19 CT segmentation dataset with 473 CT slices. Zhou et al. (2020) proposed a machine-agnostic method that can segment and quantify the infection regions on multi-sites CT scans. A joint classification and segmentation (JCS) system for real-time and interpretable COVID-19 diagnosis was developed by Wu et al. (2020). Because public COVID-19 dataset is limited and hard to obtain, Qiu et al. (2020) developed a lightweight deep learning model for efficient COVID-19 segmentation while avoiding overfitting. Fan et al. (2020) proposed a semi-supervised learning model to tackle the challenge of a

---

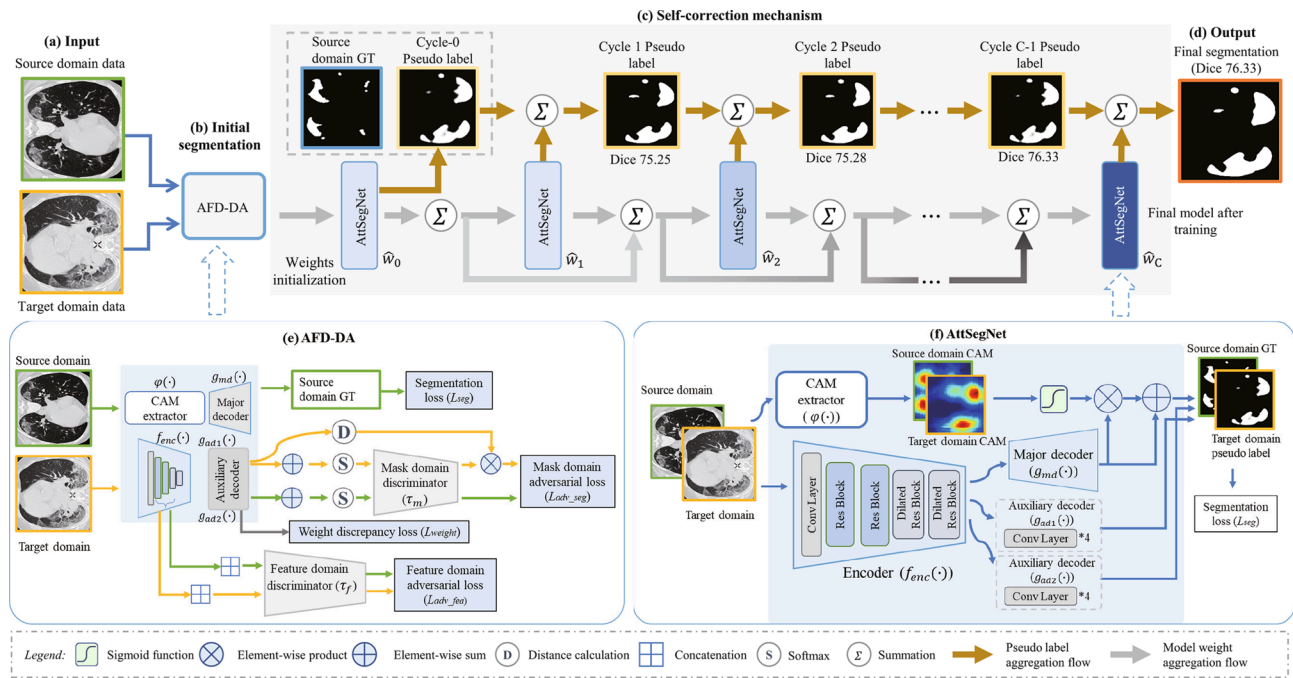[2] https://github.com/qgking/DASC_COVID19.git.

Fig. 3. Overview of the proposed DASC-Net for COVID-19 CT segmentation. Given (a) input data, (b) AFD-DA is firstly trained to solve the domain shift problem and obtain initial segmentation results. The detailed structure of ADF-DA is in (e). (c) The self-correction learning process iteratively aggregates (f) AttSegNet model and pseudo labels until the final cycle to generate (d) final segmentation results. AttSegNet is initialized by the generator of ADF-DA.

limited amount of data for training.

Given a small amount of data for training which may also exhibit domain shifts, the machine learning (ML) methods could have over-fitting problems, especially for data-hungry DL models. Hence, how to make the best use of a limited amount of cross-site data for COVID-19 analysis is an emerging and challenging problem to be solved.

2.2. DA based semantic segmentation

Most of the ML models for semantic segmentation are based on the hypothesis that the training and test data have an identical distribution (Lv et al., 2020). The hypothesis, however, is not always true in real-world data. When transferring knowledge from a label-rich source domain to a label-scarce target domain, it is common to find the discrepancy from the training to the test stage. The DA based technique aims to rectify this discrepancy and tune the models towards better generalization at the testing step (Patel, Gopalan, Li, & Chellappa, 2015). When the data distribution gap between source and target domains is narrowed, an improved generality can be obtained.

Over the past few years, DA techniques attract intensive interests in semantic segmentation (Lv et al., 2020; Luo, Liu, Guan, Yu, & Yang, 2019; Tsai, Sohn, Schulter, & Chandraker, 2019; Zhu, Park, Isola, & Efros, 2017; Chen, Xu, & Jia, 2020; Li, Yuan, & Vasconcelos, 2019; Li, Xu, Wei, & Yang, 2019; Du et al., 2019; Zheng & Yang, 2019). Generally, DA models can be divided into two categories (Lv et al., 2020): domain-invariant representation learning and pseudo label guided learning.

2.2.1. Domain-invariant representations learning

One of the key motivations behind this category of DA based semantic segmentation is to learn domain invariant representations. Most of the recent studies are based on domain adversarial learning and generative adversarial networks (GANs). For instance, Luo et al. (2019) combined an information bottlenecked strategy and an adversarial learning framework for feature-space DA semantic segmentation. Tsai et al. (2019) proposed an adversarial adaptation scheme, which makes the feature representations of target patches closer to the distributions of source patches in clustered spaces.

CycleGAN (Zhu et al., 2017) provides an alternative solution for DA by translating the source images into target-style images. For instance, Chen et al. (2020) proposed a domain adaptive image-to-image translation architecture for out-of-domain samples. Li et al. (2019) proposed a bidirectional learning framework for self-supervised DA segmentation. Unlike domain adversarial learning based methods, Lv et al. (2020) demonstrated that pivot information could improve the performance of DA based segmentation, which served as a bridge to share common knowledge between source and target domains.

2.2.2. Pseudo label guided learning

The pseudo-labelling method, which is a typical technique in semi-supervised learning (Lee, 2013), can also be used to solve domain shifts. The pseudo label normally refers to the label for target data (Li et al., 2019). For instance, Zheng and Yang (2019) proposed a model to mine the target domain knowledge and fine-tune unlabelled target data in a self-training approach. Du et al. (2019) introduced a progressive confidence strategy taking full advantage of pseudo labels via adversarial learning without global feature alignment. In Saporta et al. (2020), entropy is used as a confidence indicator to improve the quality of pseudo labels in an entropy-guided self-supervised learning (ESL) model.

A recent self-correction training strategy (Li et al., 2019) is proposed to improve the segmentation performance during different training cycles by self-supervision. The apparent drawback of using this approach is that if the pseudo labels of target domain contain noises, the noisy information may mislead the correction of pseudo labels and result in overfitting to the noise. However, our approach leverages both source and target domain information for improved self-correction, especially when there is a limited amount of data with domain shifts.

3. Methods

The network architecture of DASC-Net is given in Fig. 3. The DASC-Net consists of an attention and feature domain enhanced DA model (AFD-DA) for initial segmentation and a self-correction procedure for segmentation correction and improvement. The AFD-DA has three

innovations: a class activation map (CAM) (Zhou et al., 2016) (i.e., prior knowledge) based attentive segmentation branch to drive the network's attention to lung infections, a discrimination module for hierarchical feature-level domain alignment and discrimination, and a hybrid loss function for optimization. The major components in the self-correction learning process include learning cycles where each cycle has an attention enhanced segmentation network (AttSegNet) and a self-correction mechanism to propagate model parameters and pseudo labels.

3.1. Baseline DA model via adversarial learning

3.1.1. Problem formulation

Given an image $I_s \in \mathbb{R}^{H \times W}$ from infected source domain I_S and an image $I_t \in \mathbb{R}^{H \times W}$ from target domain I_T, where H and W are the height and width of the images, S and T are the index set of images in source and target domains respectively, and we use the labelled source domain $\mathscr{D}_S = \{(I_s, Y_s)\}_{s \in S}$ and the unlabelled target domain $\mathscr{D}_T = \{I_t\}_{t \in T}$ to train a segmentation network, where $Y_s \in \{0, 1\}^{H \times W}$ denotes the corresponding label of I_s. The I_s is first passed to the segmentor with $Y_s \in \{0, 1\}^{H \times W}$. Afterwards, the segmentor infers the segmentation output P_t for image I_t. The goal of domain adaptation is to achieve the desirable semantic segmentation performance on the target domain.

3.1.2. Basic DA framework

We exploit the DA model in (Luo, Zheng, Guan, Yu, & Yang, 2019) as the basic network for a better local semantic consistency enforcement during the procedure of global alignment, which is composed of a generator G and a mask domain discriminator $\tau_m(\cdot)$. The generator G is divided into feature encoder $f_{enc}(\cdot)$ and two decoders, termed as $g_{ad1}(\cdot)$ and $g_{ad2}(\cdot)$. In our experiments, the dilated ResNet-34 (He, Zhang, Ren, & Sun, 2016; Yu, Koltun, & Funkhouser, 2017) is adopted as the feature encoder. During the co-training procedure (Zhou & Li, 2005), the weights of g_{ad1} and g_{ad2} are diversified by a cosine distance cost function.

High-level feature maps are extracted from $f_{enc}(\cdot)$ after feeding I_s from the source domain. Those high-level feature maps serve as the input of the two decoders, i.e, g_{ad1} and g_{ad2}, to yield the final predictions P_s^1 and P_s^2. P_s^1 and P_s^2 are used to penalize the network for learning a better prediction under the supervision of Y_s. The segmentation loss function L_{seg} is defined as:

$$L_{seg} = \mathbb{E}\big[L_{ce}\big(P_s^1, Y_s\big) + L_{ce}\big(P_s^2, Y_s\big)\big], \tag{1}$$

where L_{ce} denotes the cross-entropy loss, and P_s^1 and P_s^2 are obtained based on $g_{ad1}(f_{enc}(I_s))$ and $g_{ad2}(f_{enc}(I_s))$ respectively. In the meantime, P_s^1 and P_s^2 are also treated as mask domain inputs to the discriminator τ_m for mask domain adversarial learning.

On the other hand, P_t^1 and P_t^2 can also be obtained by G. Apart from mask domain adversarial loss, pixel-wise discrepancy $Dis(P_t^1, P_t^2)$ between P_t^1 and P_t^2 is measured by a cosine similarity as:

$$Dis\big(P_t^1, P_t^2\big) = 1 - \cos\big(P_t^1, P_t^2\big). \tag{2}$$

In summary, the adversarial learning procedure can be formulated as:

$$L_{adv\_seg} = \mathbb{E}\big[\log\big(\tau_m\big(P_s^1 + P_s^2\big)\big)\big] + \\ \mathbb{E}\big[\big(\lambda_{dis}Dis\big(P_t^1, P_t^2\big)\big)\log\big(1 - \tau_m\big(P_t^1 + P_t^2\big)\big)\big], \tag{3}$$

where λ_{dis} is a weight factor. Each pixel on the segmentation map is differently weighted with respect to the mask domain adversarial loss (Luo et al., 2019).

The basic DA model also provides the weight discrepancy loss on g_{ad1} and g_{ad2} for providing two different views on features (Luo et al., 2019). The L_{weight} is formulated as:
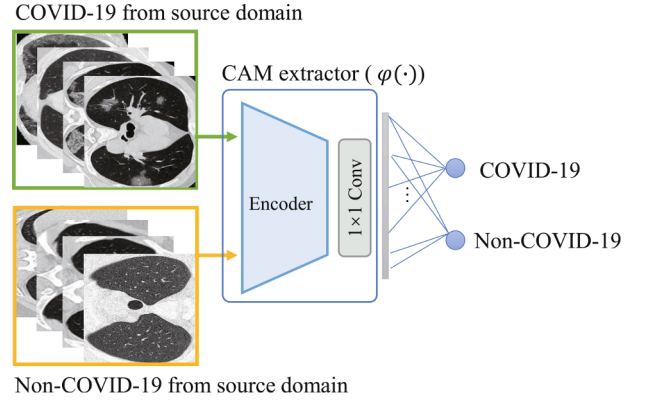

Fig. 4. CAM extractor for attention to lung abnormalities. The classification network is pre-trained using COVID-19/Non-COVID-19 data from the source domain.

$$L_{weight} = \frac{\overrightarrow{w_1} \cdot \overrightarrow{w_2}}{\|\overrightarrow{w_1}\| \|\overrightarrow{w_2}\|}, \tag{4}$$

where $\overrightarrow{w_1}$ and $\overrightarrow{w_2}$ are obtained by flattening and concatenating the weights of the convolutional layers of g_{ad1} and g_{ad2}.

The final training objective L_{BASE} in the base DA model can be summarized as:

$$L_{BASE} = L_{seg} + \lambda_{weight}L_{weight} + \lambda_{adv\_seg}L_{adv\_seg}, \tag{5}$$

where λ_{weight} and λ_{adv_seg} are weight factors.

3.2. Prior knowledge driven domain adaptation

As there are large variations between the COVID-19 source domain and target domain, the baseline DA model may fail to identify infected lung regions which are tiny or have high similarities with healthy tissues. Besides, the basic DA only penalizes mask-domain alignment. However, there still exists domain misalignment between features in antecedent. Thus, we propose an AFD-DA model to improve the performance of the basic DA.

We hypothesize that a DA model with attention to lung abnormities and hierarchical feature domain alignment and discrimination can enhance the segmentation performance. Fig. 3(e) gives the architecture of AFD-DA. Compared with the basic DA, the AFD-DA includes an attentive segmentation branch with a CAM extractor and a major decoder, and a feature-level domain alignment and discrimination module where there are a feature encoder, two auxiliary decoders, and a feature-level discriminator.

3.2.1. Prior knowledge driven segmentation branch

In the segmentation branch, we firstly design a CAM enhanced encoder with attention to lung abnormalities. This is achieved by pre-training a classification network using image-level labels as shown in Fig. 4. Since CAM can highlight image regions associated with image-level labels, we use positive samples (CT slices with infections, i.e., I_S) and negative samples (CT slices without infections) from the source domain as the training dataset. The classification network consists of a dilated ResNet-34 encoder and a 1×1 convolutional layer. The pre-trained network is referred to as CAM extractor φ.

As CAM may contain non-infected regions which can mislead the segmentation (Xie, Zhang, Xia, & Shen, 2020), we propose an online CAM attention based decoder to improve the conventional decoders g_{ad1} and g_{ad2} in the basic DA. Inspired by residual learning (He et al., 2016), if soft CAM can be constructed as the attention map, the performance shall be at least better than a decoder without attention.

Given the original decoder of DeepLab V3+ (Chen, Papandreou,

Kokkinos, Murphy, & Yuille, 2017) as the major segmentor g_{md}, we modify output P_s^0 of major segmentor g_{md} as

$$P_s^0 = (1 + \sigma(Up(\varphi(I_s))))g_{md}(f_{enc}(I_s)), \tag{6}$$

where $\sigma(\cdot)$ denotes the sigmoid function, $Up(\cdot)$ is a bilinear upsample function to interpolate the CAM to the input resolution, $\sigma(Up(\varphi(I_s)))$ is within the range of [0, 1]. After combining Eqs. (1) and (6), we can rewrite the final segmentation loss function as:

$$L_{seg} = \mathbb{E}\left[\lambda_{seg}L_{ce}(P_s^0, Y_s) + L_{ce}(P_s^1, Y_s) + L_{ce}(P_s^2, Y_s)\right], \tag{7}$$

where λ_{seg} is a weight factor to magnify the impact of the major decoder, the two auxiliary decoders.

3.2.2. Hierarchical feature-level alignment of semantic feature

According to (Luo et al. (2019)), lower $Dis(P_t^1, P_t^2)$ values indicate less severe domain shifts and larger overlap between source and target domain distributions. It seems that the domain shift problem is solved in the latent space. However, due to the hierarchical architecture of the convolutional layers, the features extracted from a specific layer heavily rely on the activations from its previous layers and vice versa (Dou et al., 2018). Thus, if a model only focuses on the high-level latent space alignment, domain shift problems may still exist during the lower-level feature extractions. The problem may be magnified when the misaligned low-level features are propagated to the deeper layers for final integration.

To overcome this problem, and inspired by deep supervision (Lee, Xie, Gallagher, Zhang, & Tu, 2015), we propose hierarchical feature-level domain alignment discrimination. As shown in Fig. 3(e), there are a feature extractor f_{enc} for high-level and low-level knowledge encoding and feature-level discriminator τ_f to facilitate the conventional mask-level discriminator. We aggregate and concatenate the outputs, which are the input of the feature-level discriminator τ_f, from every residual block in f_{enc} except the last one. Given f_{enc}, the feature-level domain alignment can be supervised by the gradients from τ_f to f_{enc} as

$$
\begin{aligned}
L_{adv\_fea} = \ & \mathbb{E}\left[\log\left(\tau_f\left(f_{enc}^1(I_s), f_{enc}^2(I_s), f_{enc}^3(I_s)\right)\right)\right] \\
+ \ & \mathbb{E}\left[\log\left(1 - \tau_f\left(f_{enc}^1(I_t), f_{enc}^2(I_t), f_{enc}^3(I_t)\right)\right)\right],
\end{aligned} \tag{8}
$$

where $f_{enc}^i(I_s)$ denotes the output of the i-th residual block in feature extractor f_{enc}.

3.3. Pseudo label refinement via self-correction learning

To fully utilize the available information for training and decrease the impact of noise in the self-correction process, we propose a dual-domain enhanced self-correction learning mechanism. Because the AFD-DA model addressed the domain shift problem, our motivation is that the label-rich source domain can be blended with the target domain to correct the noise caused by pseudo labels. As shown in Fig. 3, the process consists of C training cycles where each cycle has an AttSegNet. The self-correction mechanism aggregates the model and propagates the updated labels.

Algorithm 1. Self-correction learning mechanism

Input: $\widehat{\omega}_0$: initialized model weight by f_{enc}, g_{md}, g_{ad1}, g_{ad2}, and φ after domain adaptation; $\widehat{y}_0^a$: augmented initial pseudo label generated by $\widehat{\omega}_0$; Ep: epoch in each cycle; C: total number of cycles;

Output: Optimal $\widehat{\omega}_C$

1: Construct a training dataset $\mathscr{D}_{training}$ using $\mathscr{D}_S = (I_S, Y_S)$ and $\mathscr{D}_T = \left(I_T, \widehat{Y}_0^a\right)$

2: **for** each $c \in [1, C]$ **do**

3:    **for** each $e \in [1, Ep]$ **do**

4:       Update the learning rate;

5:       Calculate loss L_{seg} by Eq. (7);

6:    **end for**

7:    Update weight $\widehat{\omega}_{c+1}$ for initializing the weight of the next cycle $c+1$ by Eq. (9);

(continued)

8:    Generate pseudo label $\widehat{Y}_{c+1}$ using $\widehat{\omega}_c$;

9:    Augment pseudo label $\widehat{Y}_{c+1}^a$ using $\widehat{\omega}_c$;

10:    Pseudo label refinement by Eq. (10);

11:    Re-construct a training dataset $\mathscr{D}_{training}$ using $\mathscr{D}_S = (I_S, Y_S)$ and $\mathscr{D}_T = \left(I_T, \widehat{Y}_{c+1}^a\right)$

12: **end for**

3.3.1. Self-correction mechanism

Given C training cycles and a set of optimal segmentation models $S = \{\widehat{\omega}_0, \widehat{\omega}_1, ..., \widehat{\omega}_C\}$ where each of them is obtained after a training cycle, $\widehat{\omega}_0$ denotes the well-trained generator in our AFD-DA model. At the beginning of the c-th cycle, we initialize model $\widehat{\omega}_c$ by aggregating model $\widehat{\omega}_{c-1}$ with $\widehat{\omega}_0$. The aggregation is formulated as:

$$\widehat{\omega}_c = \frac{c}{c+1}\widehat{\omega}_{c-1} + \frac{1}{c+1}\widehat{\omega}_0, \tag{9}$$

where c is in $\{1, ..., C\}$ and C is set as 9 in our experiments. The pseudo-code of the self-learning algorithm is given in Algorithm 1. During the self-correction integration, the pseudo labels are progressively improved in terms of performance and generability.

3.3.2. Pseudo label aggregation and AttSegNet

As shown in Fig. 3(f), AttSegNet consists of a CAM extractor φ, feature extractor f_{enc}, CAM enhanced decoder g_{md}, and two auxiliary decoders g_{ad1} and g_{ad2}, which are structurally identical to the generator in AFD-DA. In each training cycle, the segmentation model takes mixed source data with ground-truth and target data with pseudo labels as input. In practice, we update the pseudo label of the target domain at the end of each cycle. To further alleviate the incorrect segmentation in pseudo labels from the target domain, we augment the target domain by horizontal and vertical flipping, and then aggregate those augmented predictions to increase the reliability of pseudo labels.

With $Y^a = \left\{\widehat{Y}_0^a, \widehat{Y}_1^a, ..., \widehat{Y}_C^a\right\}$ denoting the augmented set of all the pseudo labels after each cycle, pseudo labels are updated as follows based on our model aggregation strategy:

$$\widehat{Y}_c^a = \frac{c}{c+1}\widehat{Y}_{c-1}^a + \frac{1}{c+1}\widehat{Y}_0^a, \tag{10}$$

where $\widehat{Y}_0^a$ is generated by the initial model $\widehat{\omega}_0$.

3.4. Network architecture and loss function

In our DASC-Net, there are three networks requiring optimization: a CAM extractor for online CAM generation, AFD-DA for domain adaptation training, and AttSegNet for pseudo label and model aggregation in self-correction.

3.4.1. Network architecture

Even though performance can be improved with a deeper depth of network architecture, a deeper network may hamper the training process and cause problems of vanishing and exploding gradients (He et al., 2016). To address this issue, He et al. (2016) proposed a deep residual learning scheme, where the deep residual unit makes the deep network easy to train and alleviates the problem of vanishing gradients (He et al., 2016). Because of these advantages, we use ResNet as the backbone encoders in our models. Since there are a limited amount of available COVID-19 data and computational resources, we choose the 34-layer ResNet. Besides, atrous convolutions with different dilation rates are employed within the last two residual blocks to increase the receptive field of each layer for better and multi-scale feature representation. The encoders in all three models in our work are identical to the dilated ResNet-34 (He et al., 2016; Yu et al., 2017) and initialized with the weight pre-trained on ImageNet (Deng et al., 2009).

The mask domain discriminator τ_m consists of 5 convolutional layers with 4×4 kernels, channel numbers of $\{64, 128, 256, 512, 1\}$, and stride of 2. Each convolutional layer is followed by a Leaky-ReLU with a negative slope of 0.2 except the last layer. The last layer is an up-sampling layer to restore the output to the size of the input domain. The structure of feature domain discriminator τ_f is the same as τ_m but without the last up-sampling layer, and the channel numbers are set as $\{256, 128, 64, 64, 1\}$.

3.4.2. Loss function

We use the cross-entropy loss as the objective function to train the CAM extractor:

$$\text{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{m=1}^{M} p_i^m \log \widehat{p}_i^m, \qquad (11)$$

where $\widehat{p}_i^m$ denotes the probability of an image sample i belonging to class m, p_i^m denotes the ground-truth class, N indicates the total number of samples, and M represents the total class type of a slice, i.e., $M = 2$.

In adversarial learning, the G of AFD-DA is trained against an adversary with the two discriminators τ_m and τ_f. The training objective of AFD-DA is formulated as:

$$
\begin{aligned}
L_{da} = \; & L_{seg} + \lambda_{weight} L_{weight} + \\
& \lambda_{adv\_seg} L_{adv\_seg} + \lambda_{adv\_fea} L_{adv\_fea},
\end{aligned} \qquad (12)
$$

where $\lambda_{weight}, \lambda_{adv\_seg}$, and λ_{adv_fea} are weight factors, L_{seg} is calculated by Eq. (7), L_{weight} is defined by Eq. (4), L_{adv_seg} and L_{adv_fea} are formulated in Eq. (3) and Eq. (8), respectively.

AttSegNet is initialized by the generator of AFD-DA. We penalize the network using Eq. (7) with source and target domain data.

4. Dataset and preprocessing

4.1. Dataset

We collect COVID-19 CT datasets from three sites to evaluate the performance of DASC-Net.

4.1.1. Coronacases Initiative and Radiopaedia (CIR)

CIR released 20 public 3D CT scans without manual segmentations. Ma et al. (2020) provided manually segmented lung and infections by radiologists. Among the images, 10 scans are 630×630 in-plane resolution, while the other 10 scans are of 512×512. In total, there are 3,520 axial CT slices, including 1,844 slices with annotated COVID-19 infections (considered as positive 2D samples) and 1,676 slices without annotations (considered as negative 2D samples).

4.1.2. Italian Society of Medical and Interventional Radiology (ISMIR)

ISMIR provided 100 axial CT images from 60 patients with COVID-19.[3] The manual segmentation of COVID-19 infected areas, including ground-glass opacity (GGO), consolidation, and pleural effusion, was performed by a radiologist. The image size is 512×512 pixels and has been greyscaled and compiled into a single NIFTI file.

4.1.3. MOSMEDDATA

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department (Morozov et al., 2020) provided 50 CT volumetric scans and corresponding annotations by experts. For each scan, the annotations of GGO and consolidation are recorded by non-zero values in a binary mask. In total, there are 2,049 axial CT slices, including 785 COVID-19 positive and 1,264 negative samples. It is noted that the proportion of the samples

containing GGO and consolidation is less than 25%. The in-plane resolution of all the images is 512×512.

In the experiments, we use the CIR data as the source domain, termed as COVID-19-S. The ISMIR and MOSMEDDATA data are used as target domains which are denoted by COVID-19-T1 and COVID-19-T2 respectively.

4.2. Data pre-processing and augmentation

Pre-processing includes image normalization and patch generation. All the CT images were normalized by windowing operation of [-1,250, 250] Hounsfield Units (HU) values, and then linearly scaled to [0, 1] by min–max normalization. Secondly, we cropped the images by localizing lung regions. Manual-segmented lungs were provided in COVID-19-S and COVID-19-T1. To obtain the lung mask in COVID-19-T2, we trained a 3D U-Net (Ronneberger, Fischer, & Brox, 2015) using COVID-19-S. Finally, the image patches were obtained by cropping the original CT image using a bounding box outside the lung.

Data augmentation transformations included random flippings in vertical and horizontal directions, affine transformations using translation within [0.01,0.01], scaling within [0.8, 1.2], shearing within [−10, 10], and rotating by up to 90°. Finally, we down-sampled the infected 2D CT slices and infection masks into 320×320 pixels for efficient computation.

5. Experimental results and discussion

5.1. Implementation and parameter settings

The DASC-Net model was implemented using the PyTorch (Paszke et al., 2017) library and trained on an NVIDIA 1080Ti GPU. For the loss function, the hyper-parameters λ_{seg}, λ_{weight}, λ_{adv_seg}, λ_{adv_fea}, and λ_{dis} are set to 3, 0.01, 0.001, 0.001, and 10 respectively. Adam optimizer, with a batch size of 4, is applied to minimize the objectives in DASC-Net. The learning rate for the discriminators is set to 1×10^{-4}, while the others are set as 2.5×10^{-4}. All the optimizers are decayed by a polynomial learning rate policy, where the initial learning rate is multiplied by $\left(1 - \frac{iter}{total\_iter}\right)^{power}$ with *power* at 0.9. The total number of training iterations is set to $100 * (iter\ per\ epoch)$, i.e., 100 epoch, for AFD-DA. For AttSegNet, the total number of cycles C is set to 9 and the total number of epochs in each cycle is set to 2. The training of the CAM extractor stopped after one epoch obtains a map which can totally cover the infections.

5.2. Comparison methods and evaluation metrics

The segmentation models are validated using a comprehensive list of evaluation metrics including sensitivity (SEN), specificity (SPC), Jaccard (JA), Dice coefficient (Dice), and Hausdorff distance (HD) (Taha & Hanbury, 2015).

To demonstrate the effectiveness of our DASC-Net, we compare with several segmentation models, which can be categorized as U-Net based model for medical image segmentation, DA based model for semantic segmentation, and most recent segmentation models published for COVID-19 infection segmentation.

U-Net based model: The U-Net based models are most widely used in medical image analysis. As there are few published peer-reviewed methods for COVID-19 infection segmentation, we firstly compared with U-Net based state-of-the-art models. U-Net (Ronneberger et al., 2015) is one of the most widely used CNN models for medical image segmentation. The architecture captures contextual features and utilizes

---

[3] http://medicalsegmentation.com/covid19/.

Table 1

Results on the COVID-19-T1 dataset for DASC-Net and SOTA methods. The best results are shown in bold font.

Method	Dice (%)	SEN (%)	SPC (%)	HD	JA (%)
U-Net (Ronneberger et al., 2015)	70.30	71.83	98.27	80.80	55.72
U2-Net (Qin et al., 2020)	69.58	77.16	97.26	87.82	55.03
U-Net++ (Zhou et al., 2018)	67.46	75.67	96.73	89.63	52.69
U-Net 3+ (Huang et al., 2020)	69.60	78.00	97.09	91.52	54.84
AdaptSegNet* (Tsai et al., 2018)	71.29	76.62	97.64	83.53	56.83
ADVENT* (Vu et al., 2019)	72.01	78.39	97.66	77.49	57.82
Ours (AFD-DA)*	74.33	80.92	97.72	71.41	60.52
Ours (DASC-Net)*	**76.33**	**83.24**	**97.82**	**65.31**	**62.97**

* DA based method.

Table 2

Results on the COVID-19-T2 dataset for DASC-Net and SOTA methods. The best results are shown in bold font.

Method	Dice (%)	SEN (%)	SPC (%)	HD	JA (%)
U-Net (Ronneberger et al., 2015)	51.95	61.99	99.76	95.09	39.56
U2-Net (Qin et al., 2020)	56.91	66.52	99.76	71.55	44.12
U-Net++ (Zhou et al., 2018)	54.00	73.88	99.58	95.72	40.41
U-Net 3+ (Huang et al., 2020)	55.46	73.80	99.61	78.58	42.18
AdaptSegNet* (Tsai et al., 2018)	56.18	69.87	99.70	**64.97**	43.15
ADVENT* (Vu et al., 2019)	56.40	69.00	99.70	66.78	43.36
Ours (AFD-DA)*	59.04	**75.17**	99.74	74.14	45.42
Ours (DASC-Net)*	**60.66**	72.44	**99.78**	75.62	**46.96**

* DA based method.

a symmetric expanding path for precise localization. Compared to U-Net, U2-Net (Qin et al., 2020)[4] is able to capture more contextual information from different scales with the receptive fields of different sizes in the Residual U-blocks. The residual U-blocks increase the depth of the network with few computational cost and shows promising results for segmentation. U-Net++ (Zhou, Siddiquee, Tajbakhsh, & Liang, 2018)[5] is a nested architecture, which is designed for medical image segmentation, and has been evaluated across multiple tasks. Taking advantage of full-scale skip connections and deep supervisions, the U-Net 3+ (Huang et al., 2020)[6] is proposed, which shows desirable results on liver segmentation.

DA based model: We also compare with recent DA based models including AdaptSegNet (Tsai et al., 2018), ADVENT (Vu, Jain, Bucher, Cord, & Pérez, 2019). Different from our AFD-DA model, those models lack strict hierarchical feature alignment and discrimination. AdaptSegNet (Tsai et al., 2018)[7] is an adversarial learning method for domain adaptation in the context of semantic segmentation. A multi-level adversarial algorithm is proposed to perform output space domain adaptation. Entropy of the pixel-wise predictions is fully utilized by ADVENT (Vu et al., 2019),[8] which provides an alternative solution for DA.

COVID-19 infection segmentation model: Semi-Inf-Net (Fan et al., 2020) is compared as it is a most recent segmentation model proposed for COVID-19 infection segmentation. The model uses a semi-supervised

[4] https://github.com/NathanUA/U-2-Net
[5] https://github.com/4uiiurz1/pytorch-nested-unet.
[6] https://github.com/ZJUGiveLab/UNet-Version
[7] https://github.com/wasidennis/AdaptSegNet
[8] https://github.com/valeoai/ADVENT.

Table 3

Results on the 50 slices from COVID-19-T1 dataset for DASC-Net and Semi-Inf-Net. The best results are shown in bold font.

Method	Dice (%)	SEN (%)	SPC (%)
Semi-Inf-Net (Fan et al., 2020)	76.4	79.7	96.3
Ours	**77.0**	**81.2**	**98.0**

approach to address the problem of training on a limited amount of data. The model was only validated on COVID-19-T1.

5.3. Comparison results

5.3.1. Quantitative results

The evaluation results over COVID-19-T1 using our DASC-Net, DASC-Net without self-correction, referred to as AFD-DA, and all the other models are given in Table 1. As shown in the table, our AFD-DA outperformed all the other methods for all the evaluation metrics especially HD, Dice, and SEN. The results demonstrated the effectiveness of AFD-DA in addressing domain shifts. The self-correction mechanism further boosted the performance of AFD-DA by 2% and 6.1 in terms of Dice and HD, and 2.32%, 0.1%, and 2.45% with respect to SEN, SPC, and JA.

The comparison results over COVID-19-T2 are summarized in Table 2. As shown in this table, the DASC-Net achieved the best results, followed by AFD-DA. We also found that the overall experimental results by all the models on COVID-19-T2 were worse than those on COVID-19-T1. This may be due to the fact that the infections have large variations between COVID-19-T2 and COVID-19-S. For instance, in the COVID-19-T2 samples, image regions with GGO and pulmonary parenchymal typically cover less than 25% of the whole lung. In the source domain, however, there are not too many CT slices with less than 25% infected areas. Besides, our empirical finding is that the less the portion of the infections, the more influence on the segmentation results. Nevertheless, our model outperformed all the methods in comparison over the cases with smaller infection regions.

When comparing with Semi-Inf-Net (Fan et al., 2020), we found that the experimental results in (Fan et al., 2020) were using 50 samples from COVID-19-T1 only. Thus, we selected the same 50 slices as reported in (Fan et al., 2020)[9] and re-calculated the segmentation results for a fair comparison. The comparison results are given in Table 3.

Our DASC-Net achieved a Dice of 77.0%, a SEN of 81.2%, and a SPC of 98.0% which outperformed the published results of Semi-Inf-Net by 0.6%, 1.5%, and 1.7% respectively. As shown in Fig. 5, the proportion with red color indicates false predictions. It reveals that the DASC-Net provides a better segmentation with higher confidence. It is also noted that the Semi-Inf-Net was trained using the rest 50 slices from COVID-19-T1. Our DASC-Net, however, was trained using source domain data without fine-tuning on the 50 slices from the target domain. Compared with Semi-Inf-Net, our model is more applicable to real-world scenarios, especially when the training and testing datasets are from different sites.

Our primary finding is that our DASC-Net substantially alleviates the domain shift problem and achieves superior performance in segmenting COVID-19 infection when compared to the U-Net based models (i.e., U-Net (Ronneberger et al., 2015), U2-Net (Qin et al., 2020), U-Net++ (Zhou et al., 2018), and U-Net 3+ (Huang et al., 2020)). Secondly, the DA based models (i.e., AdaptSegNet (Tsai et al., 2018) and ADVENT (Vu et al., 2019)) gain better quantitative results than the U-Net based models. Finally, our DASC-Net outperforms the COVID-19 infection segmentation model (i.e., Semi-Inf-Net (Fan et al., 2020)) without fine-tuning on extra data, which is extremely important at the early stage of COVID-19 outbreak when limited well-annotated data samples can be acquired.

[9] https://github.com/DengPingFan/InfNet.

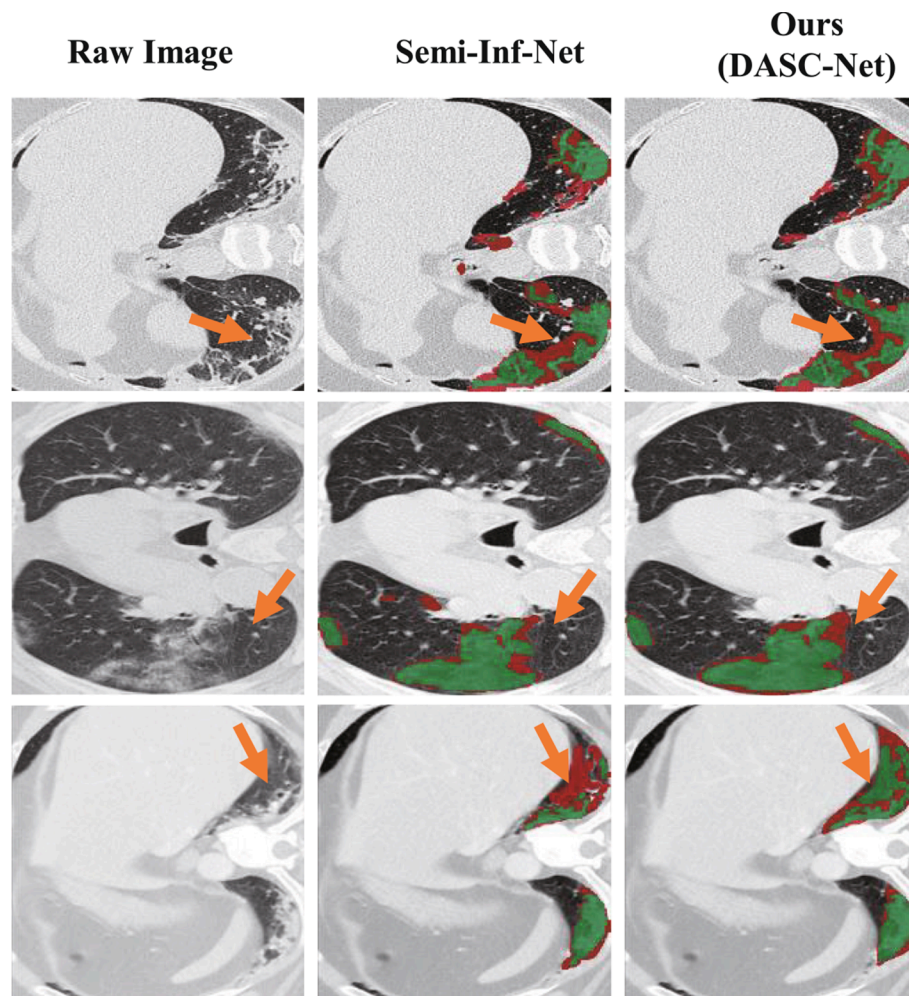## Raw Image   Semi-Inf-Net   Ours (DASC-Net)

Fig. 5. Three examples of COVID-19 infection segmentation against Semi-Inf-Net. The segmentation results from Semi-Inf-Net were downloaded from the authors' GitHub repository. The false predictions, i.e., false-positive and false-negative, are shown in red while the correct predictions are in green. The significant improvement by our model is marked with orange arrows.

5.3.2. Qualitative results

Six segmentation results on COVID-19-T1 and COVID-19-T2 are given in Fig. 6. The segmentation results of DASC-Net are the closest to manual segmentations whether the regions to be segmented are large or small, as shown by the top three and bottom three rows. It is also found that the DA based models, including AdaptSegNet (Tsai et al., 2018) and ADVENT (Vu et al., 2019), outperformed those U-Net based models without domain adaptation.

5.4. Ablation studies for effectiveness validation

Ablation study was performed to evaluate the contributions of major components in DASC-Net. We first obtain the performance of the baseline DA model (Luo et al., 2019). Afterwards, we gradually added the newly proposed components to the baseline model. The experimental results on COVID-19-T1 are presented in Table 4.

5.4.1. Contributions of prior knowledge driven segmentation branch and hierarchical feature-level alignment of semantic features in AFD-DA

The baseline DA model (Luo et al., 2019) achieved a Dice of 73.65%. When an online CAM based attentive segmentation branch was added to the baseline, the experimental results were improved. The newly proposed feature-level domain alignment also contributed to the experimental results where the Dice value further increased to 73.84%. The SEN and HD metrics were improved by 0.99% and 0.71 respectively.

When considering both CAM based attentive segmentation branch and feature domain alignment module, the performance was further improved to a Dice of 74.33%.

The experimental results verified our hypothesis that although domain shift problems are tackled in a high-level latent space by the basic DA, the problem still exists in the low-level feature space. The proposed feature-level domain alignment module enabled our model to resolve the feature-level domain shifts, as demonstrated by the improved segmentation results.

5.4.2. Effectiveness of self-correction learning

When examining the contributions of the self-correction algorithm, Table 4 shows that the most noticeable improvement was in HD, followed by JA, Dice, and SEN.

We also investigate the segmentation performance in different cycles during the self-correction learning process. The detailed segmentation results and the corresponding Dice scores in each cycle are given in Fig. 7. As shown in this figure, with the aggregation of the model and pseudo labels, the segmentation results gradually increased from a Dice of 75.25% in Cycle 1 to a Dice of 76.33% in the final cycle. More importantly, the segmentation results on small or isolated infected regions were improved as indicated by the orange arrows in Fig. 7.
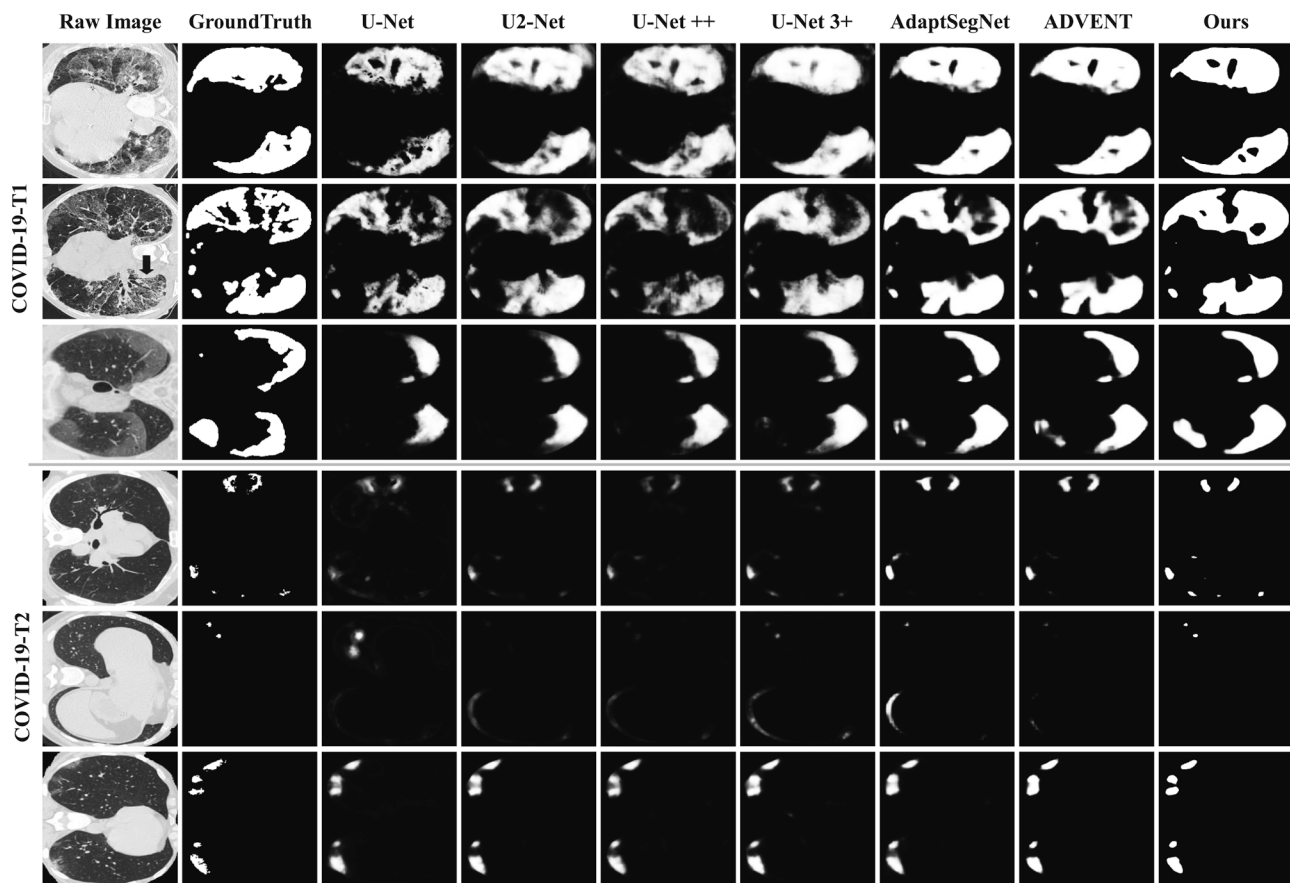
Fig. 6. Visual comparison of COVID-19 infection segmentation against other methods on two target datasets.

Table 4
Ablation studies of our DASC-Net on the COVID-19-T1 dataset. A component marked by √ means that the model contains this component. The best results are shown in bold font.

Basic DA	CAM attention enhanced	Feature-level alignment	Self-correction	Dice (%)	SEN (%)	SPC (%)	HD	JA (%)
√				73.65	79.97	97.76	69.70	59.64
√	√			74.36	80.02	97.87	68.46	60.51
√		√		73.84	80.96	97.61	68.99	59.79
√	√	√		74.33	80.92	97.72	71.41	60.52
√	√	√	√	**76.33**	**83.24**	**97.82**	**65.31**	**62.97**

6. Conclusion

We propose a novel domain adaptation based self-correction learning (DASC-Net) model. The DASC-Net utilizes the prior domain knowledge and enhances the distribution similarity of the semantic features by hierarchical feature-level discrimination, and adaptively refines the pseudo labels by self-correction learning mechanism. In practice, CAM drives the segmentation network to emphasize lung abnormalities. The new hierarchical feature-level discriminator complements mask-level discrimination, which enhances feature domain alignment. The self-correction learning mechanism alleviates the misleading supervision caused by noises in pseudo labels. Extensive experiments on multi-sites public COVID-19 datasets demonstrate that our model outperforms the state-of-the-art methods. Our model and framework can be generally applied to other tasks when there are domain shift problems and lack of sufficient dataset for training, which bridges the theory–practice gap. The limitation of our method is that we assume that all of the source data samples are fully annotated. However, in the early stage of the COVID-19 outbreak, the amount of well-annotated data samples is still limited, and the performance of DA methods can degrade substantially with fewer labeled samples. Our future work includes the exploration of DASC-Net for few-shot learning tasks.

CRediT authorship contribution statement

Qiangguo Jin: Writing - original draft, Investigation, Methodology, Formal analysis, Resources, Software. **Hui Cui:** Formal analysis, Writing - review & editing. **Changming Sun:** Investigation, Supervision. **Zhaopeng Meng:** Investigation, Supervision. **Leyi Wei:** Investigation, Supervision. **Ran Su:** Formal analysis, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
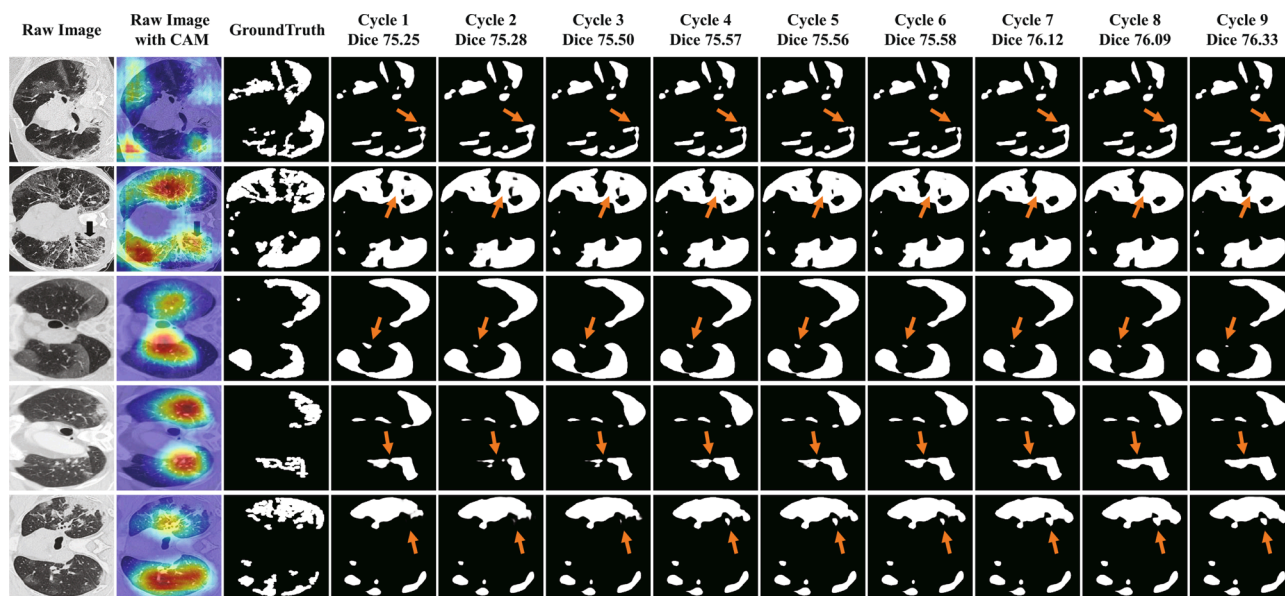
Fig. 7. Illustration of segmentation map in 9 cycles on COVID-19-T1. The refinement is marked with orange arrows.

Acknowledgment

References

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine, 1*, 1–8.

Chen, C., Dou, Q., Chen, H., Qin, J., & Heng, P. A. (2020). Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Transactions on Medical Imaging, 39*, 2494–2505.

Chen, X., Yao, L., & Zhang, Y. (2020). Residual attention U-Net for automated multi-class segmentation of COVID-19 Chest CT Images. arXiv preprint arXiv:2004.05645.

Chen, Y.-C., Lin, Y.-Y., Yang, M.-H., & Huang, J.-B. (2019). CrDoCo: Pixel-level domain transfer with cross-domain consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1791–1800).

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*, 834–848.

Chen, M., Tang, Y., Zou, X., Huang, K., Huang, Z., Zhou, H., Wang, C., & Lian, G. (2020). Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology. *Computers and Electronics in Agriculture, 174*, Article 105508.

Chen, Y.-C., Xu, X., & Jia, J. (2020). Domain adaptive image-to-image translation. In *The IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 5274–5283).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., & Heng, P.-A. (2018). *PnP-AdaNet: Plug-and-Play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation*. arXiv preprint arXiv:1812.07907.

Dou, Q., Liu, Q., Heng, P. A., & Glocker, B. (2020). Unpaired multi-modal segmentation via knowledge distillation. *IEEE Transactions on Medical Imaging, 39*, 2415–2425.

Du, L., Tan, J., Yang, H., Feng, J., Xue, X., Zheng, Q., Ye, X., & Zhang, X. (2019). SSF-DAN: Separated semantic feature based domain adaptation network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 982–991).

Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., Shen, J., & Shao, L. (2020). Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. *IEEE Transactions on Medical Imaging, 39*, 2626–2637.

He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., Shi, F., Gao, Y., Liu, J., Zhang, J. et al. (2020). *Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT Images*. arXiv preprint arXiv:2005.03832.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., & Wu, J. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1055–1059). IEEE.

Jin, Q., Cui, H., Sun, C., Meng, Z., & Su, R. (2020). Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. *Applied Soft Computing, 99*, Article 106881.

Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., Jiang, H., Wu, D., Sui, H., Zhang, C., et al. (2020). Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. *IEEE Transactions on Medical Imaging, 39*, 2606–2614.

Lee, D.-H. (2013). Pseudo-Label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*. Vol. 3.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015). Deeply-supervised nets. In *Artificial Intelligence and Statistics* (pp. 562–570).

Li, P., Xu, Y., Wei, Y., & Yang, Y. (2019). *Self-correction for human parsing*. arXiv preprint arXiv:1910.09777.

Li, Y., Yuan, L., & Vasconcelos, N. (2019). Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6936–6945).

Luo, Y., Liu, P., Guan, T., Yu, J., & Yang, Y. (2019). Significance-aware information bottleneck for domain adaptive semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 6778–6787).

Luo, Y., Zheng, L., Guan, T., Yu, J., & Yang, Y. (2019). Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2507–2516).

Lv, F., Liang, T., Chen, X., & Lin, G. (2020). Cross-domain semantic segmentation via domain-invariant interactive relation transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4334–4343).

Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z. et al. (2020). *Towards efficient COVID-19 CT Annotation: A benchmark for lung and infection segmentation*. arXiv preprint arXiv:2004.12537.

Morozov, S., Andreychenko, A., Pavlov, N., Vladzymyrskyy, A., Ledikhova, N., Gombolevskiy, V., Blokhin, I. A., Gelezhe, P., Gonchar, A., & Chernina, V. Y. (2020). *MosMedData: Chest CT Scans With COVID-19 related findings dataset*. arXiv preprint arXiv:2005.06465.

Ouyang, X., Huo, J., Xia, L., Shan, F., Liu, J., Mo, Z., Yan, F., Ding, Z., Yang, Q., Song, B., et al. (2020). Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Transactions on Medical Imaging, 39*, 2595–2605.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in PyTorch. In *NIPS-W*.

Patel, V. M., Gopalan, R., Li, R., & Chellappa, R. (2015). Visual domain adaptation: A survey of recent advances. *IEEE Signal Processing Magazine, 32*, 53–69.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems* (pp. 5099–5108).

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., & Jagersand, M. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition, 106,* Article 107404.

Qiu, Y., Liu, Y., & Xu, J. (2020). *MiniSeg: An extremely minimum network for efficient COVID-19 segmentation.* arXiv preprint arXiv:2004.09750.

Robinson, R., Dou, Q., Castro, D., Kamnitsas, K., de Groot, M., Summers, R., Rueckert, D., & Glocker, B. (2020). *Image-level harmonization of multi-site data using image-and-spatial transformer networks.* arXiv preprint arXiv:2006.16741.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Saporta, A., Vu, T.-H., Cord, M., & Pérez, P. (2020). *ESL: Entropy-guided self-supervised learning for domain adaptation in semantic segmentation.* arXiv preprint arXiv:2006.08658.

Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Jiang, H., Gao, Y., Sui, H., & Shen, D. (2020). *Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification.* arXiv preprint arXiv:2003.09860.

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., & Shen, D. (2021). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering, 14,* 4–15.

Taha, A. A., & Hanbury, A. (2015). An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37,* 2153–2163.

Tang, Y., Chen, M., Lin, Y., Huang, X., Huang, K., He, Y., & Li, L. (2020). Vision-based three-dimensional reconstruction and monitoring of large-scale steel tubular structures. *Advances in Civil Engineering, 2020.*

Tang, Z., Zhao, W., Xie, X., Zhong, Z., Shi, F., Liu, J., & Shen, D. (2020). *Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images.* arXiv preprint arXiv:2003.11988.

Tang, Y., Li, L., Wang, C., Chen, M., Feng, W., Zou, X., & Huang, K. (2019). Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision. *Robotics and Computer-Integrated Manufacturing, 59,* 36–46.

Tang, Y.-C., Wang, C., Luo, L., Zou, X., et al. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. *Frontiers in Plant Science, 11,* 510.

Tsai, Y.-H., Hung, W.-C., Schulter, S., Sohn, K., Yang, M.-H., & Chandraker, M. (2018). Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7472–7481).

Tsai, Y.-H., Sohn, K., Schulter, S., & Chandraker, M. (2019). Domain adaptation for structured output via discriminative patch representations. In *Proceedings of the IEEE international conference on computer vision* (pp. 1456–1465).

Vu, T.-H., Jain, H., Bucher, M., Cord, M., & Pérez, P. (2019). Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2517–2526).

Wang, Z., Du, B., & Guo, Y. (2019). Domain adaptation with neural embedding matching. *IEEE Transactions on Neural Networks and Learning Systems, 31,* 2387–2397.

Wang, G., Liu, X., Li, C., Xu, Z., Ruan, J., Zhu, H., Meng, T., Li, K., Huang, N., & Zhang, S. (2020). A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Transactions on Medical Imaging, 39,* 2653–2663.

Wu, Y.-H., Gao, S.-H., Mei, J., Xu, J., Fan, D.-P., Zhao, C.-W., & Cheng, M.-M. (2020). *JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation.* arXiv preprint arXiv:2004.07054.

Xie, Y., Zhang, J., Xia, Y., & Shen, C. (2020). A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Transactions on Medical Imaging, 39,* 2482–2493.

Yu, F., Koltun, V., & Funkhouser, T. (2017). Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 472–480).

Zhang, Y., Qiu, Z., Yao, T., Liu, D., & Mei, T. (2018). Fully convolutional adaptation networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6810–6818).

Zheng, Z., & Yang, Y. (2019). *Unsupervised scene adaptation with memory regularization in vivo.* arXiv preprint arXiv:1912.11164.

Zhou, T., Canu, S., & Ruan, S. (2020). *An automatic COVID-19 CT segmentation network using spatial and channel attention mechanism.* arXiv preprint arXiv:2004.06673.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).

Zhou, Z.-H., & Li, M. (2005). Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on Knowledge and Data Engineering, 17,* 1529–1541.

Zhou, L., Li, Z., Zhou, J., Li, H., Chen, Y., Huang, Y., Xie, D., Zhao, L., Fan, M., Hashmi, S., et al. (2020). A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. *IEEE Transactions on Medical Imaging, 39,* 2638–2652.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.

Zhu, Q., Du, B., & Yan, P. (2019). Boundary-weighted domain adaptive neural network for prostate MR image segmentation. *IEEE Transactions on Medical Imaging, 39,* 753–763.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223–2232).

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine, 382,* 727–733.

Zlocha, M., Dou, Q., & Glocker, B. (2019). Improving retinanet for ct lesion detection with dense masks from weak recist labels. In *International conference on medical image computing and computer-assisted intervention* (pp. 402–410). Springer.

Qiangguo Jin received his B.E. degree in School of Computer Software from Tianjin University, China, in 2014. He received his M. Eng. degree in School of Computer Software from Tianjin University, China in 2017. Currently, he is completing a Ph.D. in Software Engineering in Tianjin University. His research interests include medical image processing and artificial intelligence.