Domain and Content Adaptive Convolution for Domain Generalization in Medical Image Segmentation

Shishuai Hu*, Zehui Liao*, Jianpeng Zhang, Yong Xia✉

[1]School of Computer Science and Engineering, Northwestern Polytechnical University
{sshu, merrical,james.zhang}@mail.nwpu.edu.cn, yxia@nwpu.edu.cn

Abstract

The domain gap caused mainly by variable medical image quality renders a major obstacle on the path between training a segmentation model in the lab and applying the trained model to unseen clinical data. To address this issue, domain generalization methods have been proposed, which however usually use static convolutions and are less flexible. In this paper, we propose a multi-source domain generalization model, namely domain and content adaptive convolution (DCAC), for medical image segmentation. Specifically, we design the domain adaptive convolution (DAC) module and content adaptive convolution (CAC) module and incorporate both into an encoder-decoder backbone. In the DAC module, a dynamic convolutional head is conditioned on the predicted domain code of the input to make our model adapt to the unseen target domain. In the CAC module, a dynamic convolutional head is conditioned on the global image features to make our model adapt to the test image. We evaluated the DCAC model against the baseline and four state-of-the-art domain generalization methods on the prostate segmentation, COVID-19 lesion segmentation, and optic cup/optic disc segmentation tasks. Our results indicate that the proposed DCAC model outperforms all competing methods on each segmentation task, and also demonstrate the effectiveness of the DAC and CAC modules.

Introduction

Medical image segmentation is one of the most critical yet challenging steps in computer-aided diagnosis. Since manual segmentation requires expertise and is time-consuming, expensive, and prone to operator-related bias, automated segmentation approaches are in extremely high demand and have been extensively studied (Litjens et al. 2017; Xie et al. 2021).

Recent years have witnessed the success of deep learning in medical image segmentation (Falk et al. 2019; Zhou et al. 2020; Isensee et al. 2020). As a data-driven technique, deep learning requires a myriad amount of annotated training data to alleviate the risk of over-fitting. However, there is usually a small dataset on medical image segmentation tasks, and this relates to the work required in acquiring the images and then in image annotation. Due to the small data issue, the i.i.d. assumption, *i.e.*, training and test data should be drawn from the same distribution, is less likely to be hold. Indeed, the problem of distribution discrepancy between training and test data is particularly severe on medical image segmentation tasks, since the quality of medical images varies greatly over many factors, including different scanners, imaging protocols, and operators (Wang et al. 2020; Liu et al. 2021a). As a result, a segmentation model learned on a set of training images may over-fit the data, and hence has a poor generalization ability on test images, which are collected in another medical center and follow a different distribution. Such undesired performance drop renders a major obstacle on path between the design and clinical application of medical image segmentation tools.

To address this issue, tremendous research endeavors have recently focused on unsupervised domain adaptation (UDA), test time adaptation (TTA), and domain generalization (DG). UDA attempts to alleviate the decrease of generalization ability caused by the distribution shift between the labeled source domain (training) data and unlabelled target domain (test) data in three ways. At the data level, the image-to-image translation is performed to make the quality of source domain data matches the quality of target domain data, leading to reduced distribution discrepancy (Yang and Soatto 2020). At the feature level, domain adaptation is achieved by using adversarial training or feature normalization to extract domain-irrelevant features (Liu et al. 2020; Shen et al. 2020). At the decision level, various constraints are posed to enforce the consistency between the source domain output and target domain output (Wang et al. 2019). Despite their promising performance, UDA methods have a limited clinical value due to the requirement of accessing target domain data (Tsai et al. 2021; Roth et al. 2021).

To overcome the limitation of UDA, TTA methods have been proposed to train the segmentation model with only the source domain data, while fine-tuning the trained model with the target domain data at the test time. It can be accomplished by adding an additional adaptor network to transform (He et al. 2021) or normalize (Karani et al. 2021) the test data and its features to minimize the domain shift at the test time. Although TTA methods avoid accessing target domain data, they require an extra network to adapt the model to the target data, which increases the spatial and computational complexity.

DG methods target at boosting the generalization abil-

*Equal contribution.

ity of DCNN models and improving their performance in the unseen target domain. An intuitive solution is to extract domain-invariant features via posing domain-invariant constraints to the model or using adversarial training (Wang et al. 2020; Fan et al. 2021; Zhou et al. 2021b). Nevertheless, it is not easy to differentiate domain-invariant features from domain-specific ones, especially when the target data distribution is completely unknown. To increase the diversity of training data, multiple source domains have been increasingly used to replace the single source domain. Multi-source DG methods (Liu, Dou, and Heng 2020; Du et al. 2021; Liu et al. 2021b,a) usually employ meta-learning to minimize the generalization gap between the simulated source domain and the target domain. However, if the simulated domain could not cover the unseen target domain, meta-learning-based methods may not perform well. Alternatively, augmentation-based DG methods (Zhang et al. 2020; Li et al. 2020) attempt to simulate the target data distribution via augmenting either the source data or the features of source data. Despite their advantages, DG methods still suffer from limited performance, which is attributed mainly to their static nature. Specifically, a DG model is frozen after training, and thus it uses the same set of parameters to handle various unseen target data, which have diverse distributions.

In this paper, we propose a multi-source domain generalization model, namely domain and content adaptive convolution (DCAC), for medical image segmentation. We adopt an encoder-decoder structure as the backbone and design the domain adaptive convolution (DAC) and content adaptive convolution (CAC). To adapt our model to the unseen target domain, the DAC module provides a domain-adaptive head, whose parameters are dynamically generated by the domain-aware controller based on the estimated domain code of the input. To adapt our model to each test image, the CAC module has a content-adaptive head, whose parameters are dynamically produced by the content-aware controller based on the global image features. We have evaluated the proposed DCAC model on three medical image segmentation benchmarks, including the prostate segmentation in MRI scans from six domains, COVID-19 lung lesion segmentation in CT scans from four domains, and optic cup (OC)/optic disc (OD) segmentation in fundus images from four domains.

In this work, we have made the following contributions.

- We used the domain-discriminative information embedded in the encoder feature maps to generate the domain code of each input image, which establishes the relationship between multiple source domains and the unseen target domain.
- We designed the dynamic convolution-based DAC module and CAC module to enable our DCAC model to adapt not only to the unseen target domain but also to each test image.
- We presented extensive experiments, which demonstrate the effectiveness of our DCAC model against the state-of-the-art in three medical image segmentation benchmarks with different imaging modalities.

Related Work

DG in Medical Image Segmentation DG methods designed for medical image segmentation can be roughly categorized into augmentation-based, meta-learning-based, and domain-invariant feature learning approaches. **Augmentation-based methods**, such as the deep stacked transformation (Zhang et al. 2020), simulate the distribution of target domain data by augmenting the source domain data. Alternatively, the linear-dependency DG method (Li et al. 2020) performs the augmentation in the feature space, aiming to simulate the distribution of features instead of the distribution of data. With the recent advance of the episodic training strategy for domain generalization in computer vision (Li et al. 2019), many **meta-learning-based methods** have been developed to generalize medical image segmentation models to unseen domains (Liu et al. 2021b; Li et al. 2021). For example, a shape-aware meta-learning scheme (Liu, Dou, and Heng 2020), which takes the incomplete shape and ambiguous boundary of prediction masks into consideration, was proposed to improve the model generalization for prostate MRI segmentation. In another example, the continuous frequency space interpolation was combined with the episodic training strategy to achieve further performance gains in cross-domain retinal fundus image segmentation and prostate MRI segmentation (Liu et al. 2021a). Although these methods work well on specific tasks using elaborately tuned parameters, their performance degrades substantially on the target domain when there are only few source domains. Given this, **domain-invariant feature learning methods** (Onofrey et al. 2019) have been proposed. Zhao *et al.* (Zhao et al. 2021) adopted domain adversarial learning and mix-up to improve white matter hyperintensity prediction on an unseen target domain. Wang *et al.* (Wang et al. 2020) built a domain knowledge pool to store domain-specific prior knowledge and then utilized domain attribute to aggregate features from different domains. Different from these methods, the proposed DCAC model uses dynamic convolutions whose parameters are generated by a controller according to the features of an input image, and thus is able to adapt to the test image from an unknown domain.

Dynamic Convolution Dynamic convolutions, which are more flexible than traditional ones, have been increasingly studied in the field of deep learning (He, Deng, and Qiao 2019; Pang et al. 2020; Zhou et al. 2021a; Han et al. 2021). A dynamic convolutional layer, in which the filters are generated conditioned on the input image, was proposed for short-range weather prediction using radar images (Klein, Wolf, and Afek 2015). The dynamic parameter generation conditioned on each input image was integrated to the mask head for instance segmentation, resulting in improved accuracy and efficiency (Tian, Shen, and Chen 2020). The dynamic filter network was also applied to partially labelled abdominal CT image dataset for multi-organ segmentation (Zhang et al. 2021). In this network, the parameters of dynamic segmentation head are generated for each target organ, conditioned on the input image features and task code. Due to their adaptive nature, dynamic convolutions successfully in-

crease the flexibility and enable the network to have a better representation capacity. The proposed DCAC model employs dynamic convolutions to resolve the domain generalization issue for cross-domain medical image segmentation. In our solution, the parameters of dynamic convolutions are generated on condition of the domain code or global features of the input image.

Method

Problem Definition and Method Overview

Let a set of K source domains be denoted by $D_s = \{(x_{ki}, y_{ki})_{i=1}^{N_k}\}_{k=1}^{K}$, where x_{ki} is the i-th image in the k-th source domain, and y_{ki} is the segmentation mask of x_{ki}. Our goal is to train a segmentation model $F_\theta : x \rightarrow y$ on D_s, which can generalize well to an unseen target domain $D_t = (x_i)_{i=1}^{N_t}$.

The proposed DCAC model is an encoder-decoder structure (Falk et al. 2019) equipped with a domain predictor, a domain-aware controller, a content-aware controller, and a series of domain-adaptive heads and content-adaptive heads. The workflow of this model consists of four steps. First, the feature map produced by each encoder layer is aggregated and fed to the domain predictor. Second, based on the generated domain code, the domain-aware controller predicts the parameters of the domain-adaptive head. Third, the content-aware controller uses the final output of the encoder as its input to generate the parameters of the content-adaptive head. Finally, according to the deep supervision strategy, the output of each decoder layer is fed sequentially to a domain-adaptive head and a content-adaptive head, which predict the segmentation result on a pixel-by-pixel basis. The diagram of our DCAC model is shown in Figure 1. We now delve into its details.

Encoder-decoder Backbone

The backbone used in our DCAC model is a U-shape structure that has an encoder and a decoder, each being composed of $N = 4 \sim 6$ blocks depending on the given segmentation task. Each encoder block contains two convolutional layers with a kernel size of 3, and the first layer has a stride of 2 to downsample the feature map, except for the first encoder block. Each layer is followed by instance normalization and the LeakyReLU activation. In the encoder, the number of filters is set to 32 in the first layer, then doubled in each next block, and finally fixed with 320 when it becomes larger than 256 (Isensee et al. 2020). The computation in each encoder block can be formally expressed as

$$f_E^i = F_E^i(f_E^{i-1}; \theta_E^i), \quad i = 1, 2, \cdots, N \tag{1}$$

where θ_E^i represents the parameters of the i-th encoder block F_E^i, f_E^i is the feature map produced by F_E^i, and $f_E^0 = x^i$ is the input image.

Symmetrically, the decoder upsamples the feature map and refines it gradually. In each decoder block, the transposed convolution with a stride of 2 is used to improve the resolution of input feature maps, and the upsampled feature map is concatenated with the corresponding low-level feature map from the encoder before being further processed by two convolutional layers. The computation in each decoder block can be formally expressed as

$$f_D^i = F_D^i(C(f_E^i, U(f_D^{i+1})); \theta_D^i),\ i = N-1, N-2 \cdots, 1 \tag{2}$$

where $U(\cdot)$ represents upsampling, $C(\cdot)$ represents concatenation, θ_D^i represents the parameters of the i-th decoder block F_D^i, f_D^i is the feature map produced by F_D^i, and $f_D^N = f_E^N$.

With this encoder-decoder architecture, multiscale encoder feature maps $\{f_E^i\}_{i=1}^{N}$ and multiscale decoder feature maps $\{f_D^i\}_{i=1}^{N-1}$ can be generated. It is expected that $\{f_E^i\}_{i=1}^{N}$ are domain-sensitive and can be utilized to calculate the probabilities of belonging to source domains of the input image. Meanwhile, $\{f_D^i\}_{i=1}^{N-1}$ are expected to be rich-semantic and not subjected to a specific domain, *i.e.*, containing the semantic information of domains and target tasks.

Domain Adaptive Convolution

Due to the discrepancy between source domains and the unseen target domain, the encoder-decoder backbone trained with source domain images may not be optimal for target domain images. Therefore, we equipped the backbone with domain-adaptive heads, in which the filters are variable and adaptive to the domain of the input image in the inference stage. For each test image, its probabilities of belonging to source domains, known as a domain code, are calculated by the domain predictor and fed to the domain-aware controller to generate the filters used in the domain-adaptive heads (see Figure 1).

Domain Predictor. Although the target domain is not identical to each source domain, an image in the target domain may similar to those in one or more source domains. And such 'domain attribute' of the image can be used as the clue to guide the adaptive processing of it. Therefore, we design the domain predictor to predict the probability of each target domain image belonging to each source domain.

The domain predictor takes multiscale encoder feature maps $\{f_E^i\}_{i=1}^{N}$ as its input. Each feature map f_E^i is aggregated with global average pooling (GAP), and the aggregated features at all scales are then concatenated into a vector. To predict the domain code of the input image, the vector is fed to a classification module, which is composed of a fully-connected layer $FC(\cdot)$ and a soft-max layer $SM(\cdot)$. The calculation of each domain code can be formally expressed as

$$\mathscr{D}^p = SM((FC(C(GAP(f_E^1), \cdots, GAP(f_E^N))); \theta_{FC})), \tag{3}$$

where θ_{FC} represents the parameters of $FC(\cdot)$. The domain code $\mathscr{D}^p$ is a K-dimensional vector that satisfies $\sum_{k=1}^{K} \mathscr{D}_k^p = 1$. During training, since each input image is sampled from one of K source domains, the ground truth domain code that supervises the training of domain predictor is a one-hot K-dimensional vector. Note that image segmentation and domain prediction are different tasks, though using the same set of features extracted by encoder blocks.

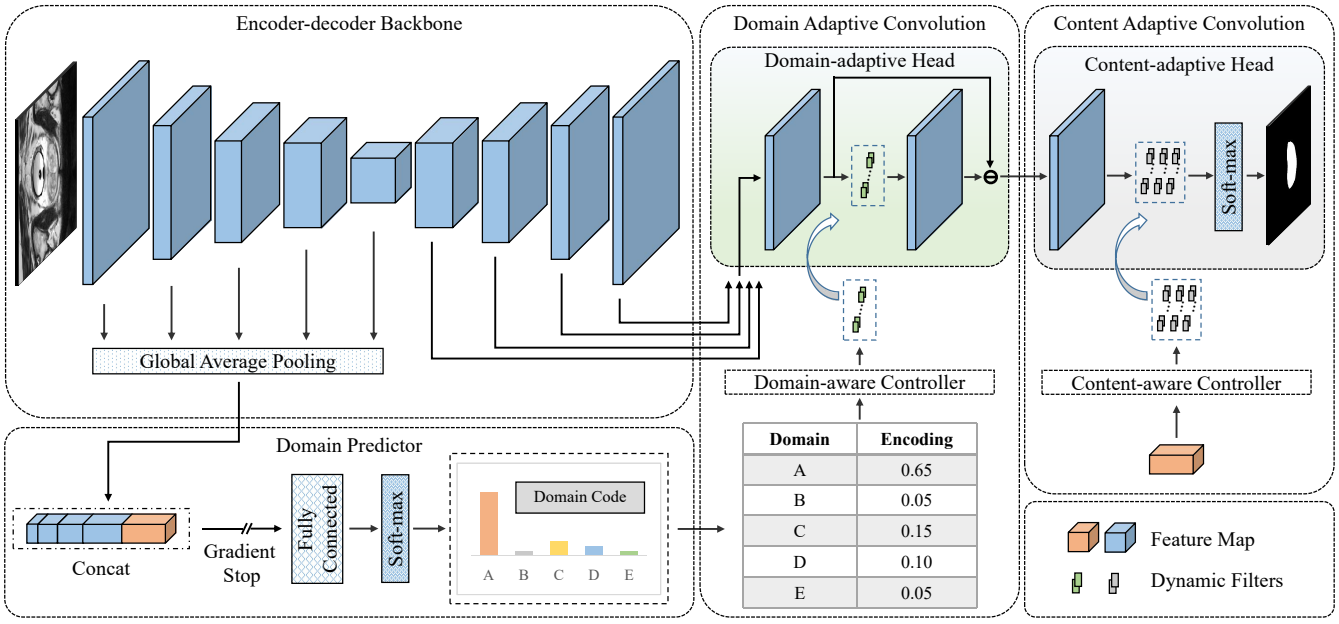

Figure 1: Architecture of the proposed method. The feature map in orange color represents $GAP(f_E^N)$, *i.e.*, the output of the N-th encoder block after global average pooling.

To avoid the interference with the image segmentation performance caused by domain prediction, we adopt the gradients truncation strategy to stop the gradients back propagated from the fully-connected layer in the domain predictor(see Figure 1).

Domain-aware Controller. We use a single convolutional layer as the domain-aware controller $\phi_d(\cdot)$, which maps the domain code to the parameters ω_d of the filters in the domain adaptive head. Such mapping can be formally expressed as

$$\omega_d = \phi_d(\mathscr{D}^p; \theta_\phi^d) \tag{4}$$

where θ_ϕ^d represents the parameters in this controller.

Domain-adaptive Head. A lightweight domain-adaptive head is designed to enable dynamic convolutions, which are responsive to specific domains. This head contains a traditional convolutional layer and a dynamic convolutional layer, both using filters with a kernel size of 1. The traditional layer reduces the channels of the input feature map to $K \times C$, where C is the number of segmentation classes. Since there exists a skip connection to enforce residual learning, the output of the dynamic layer has $K \times C$ channels, too. Therefore, there are totally $(K \times C)^2 + (K \times C)$ parameters in the dynamic layer, which are generated dynamically by the domain-aware controller $\phi_d(\cdot)$ conditioned on the domain code $\mathscr{D}^p$.

To accelerate the convergence of our DCAC model, we adopt the multi-scale supervision strategy. Given the feature map f_D^i generated by the i-th decoder block, the output of the domain-adaptive head is computed as

$$\begin{aligned} f_{DAC}^i = F_O^i(f_D^i) - F_O^i(f_D^i) * \omega_d, \\ i = N-1, N-2 \cdots, 1 \end{aligned} \tag{5}$$

where $*$ represents the convolution, and $F_O^i(\cdot)$ is the traditional convolutional layer.

Content Adaptive Convolution

The proposed DCAC model is expected to adapt not only to the unseen test domain but also to each test image. Therefore, we equipped our segmentation backbone with content adaptive convolutions, which are implemented using a content-adaptive head whose parameters are generated dynamically by a content-aware controller.

Content-aware Controller. The content-aware controller is a convolutional layer, denoted by ϕ_c. The input of this controller is the global image representation, which is the feature map generated by the encoder (*i.e.*, the output f_E^N of the N-th encoder block) and aggregated by global average pooling. The output is the ensemble of parameters of the content-adaptive head, which can be formally expressed as

$$\omega_c = \phi_c(GAP(f_E^N); \theta_\phi^c) \tag{6}$$

where θ_ϕ^c represents the parameters of the controller ϕ_c.

Content-adaptive Head. The content-adaptive head, which is placed after the domain-adaptive head, contains three stacked convolutional layers using filters with a kernel size of 1. The first two layers have $K \times C$ channels, and the last layer has C channels. Thus there are totally $2 \times ((K \times C)^2 + (K \times C)) + ((K \times C) \times C + C)$ dynamic parameters in this head. These parameters, denoted by $\omega_c = \{\omega_{c1}, \omega_{c2}, \omega_{c3}\}$, are generated by the controller ϕ_c according to the globally aggregated image feature map f_E^N.

The content-adaptive head uses the output of domain-adaptive head f_{DAC}^i as its input. This head acts as a pixel classifier, performing image segmentation via predicting

class labels on a pixel-by-pixel basis. The computation of segmentation result p^i can be formally expressed as

$$p^i = SM(((f^i_{DAC} * \omega_{c1}) * \omega_{c2}) * \omega_{c3}), \quad i = N-1, N-2, \cdots, 1 \tag{7}$$

where $SM(\cdot)$ represents the soft-max operation.

Training and Test

Besides image segmentation, the proposed DCAC model also performs domain classification using the domain predictor. For the classification task, the objective is the cross-entropy loss, which can be calculated as

$$\mathcal{L}_{cls} = -\sum_{k=1}^{K} d_k \log(d^p_k) \tag{8}$$

where d_k is the domain label, and d^p_k is the soft-max probability of belonging to the k-th domain.

For the segmentation task, the Dice loss and cross-entropy loss are used jointly as the objective. The segmentation loss at each scale can be calculated as

$$\begin{aligned}\mathcal{L}^i_{seg} =& 1 - \frac{2\sum_{v=1}^{V} p^i_v y^i_v}{\sum_{v=1}^{V}(p^i_v + y^i_v + \epsilon)} \\ &- \sum_{v=1}^{V}\left(y^i_v \log p^i_v + \left(1 - y^i_v\right)\log\left(1 - p^i_v\right)\right)\end{aligned} \tag{9}$$

where p^i_v and y^i_v denote the prediction and ground truth of the v-th voxel in the output of the i-th decoder block, V represents the number of voxels, and ϵ is a smooth factor to avoid dividing by 0.

Since deep supervision is used, the total loss is defined as follows

$$\mathcal{L} = \mathcal{L}_{cls} + \sum_{i=1}^{N-1} \omega^i \mathcal{L}^i_{seg} \tag{10}$$

where ω^i is a weighting vector that enables higher resolution output to contribute more to the total loss (Isensee et al. 2020).

During inference, given a test image x, the multiscale encoder feature maps $\{f^i_E\}^N_{i=1}$ and multiscale decoder feature maps $\{f^i_D\}^{N-1}_{i=1}$ can be produced by the trained encoder-decoder backbone. Based on $\{f^i_E\}^N_{i=1}$, the trained domain predictor can generate a K-dimensional domain code. Based on the code, the trained domain-aware controller can generate the parameters for the domain-adaptive head. Meanwhile, based on the feature map produced by the last encoder block (*i.e.*, f^N_E), the content-aware controller can generate the parameters for the content-adaptive head. Finally, the feature map produced by the decoder is fed sequentially to the domain-adaptive dynamic head and content-adaptive head to generate the segmentation result. Note that deep supervision is carried out only in the training stage and the segmentation is not performed at course scales in the test stage.

Experiments and Discussions

We evaluated the proposed DCAC model against the baseline and state-of-the-art DG models on three tasks, including the prostate segmentation using MRI, COVID-19 lesion segmentation using CT, and OC/OD segmentation using fundus imaging. These tasks cover different image modalities and represent variable domain shifts in cross-domain medical image segmentation problems.

Datasets

Three datasets were used for this study. For prostate segmentation, the dataset contains 116 T2-weighted MRI cases from six domains (Liu, Dou, and Heng 2020). Following (Liu, Dou, and Heng 2020; Liu et al. 2021a), we preprocessed the MRI data and only preserved the slices with the prostate region for consistent and objective segmentation evaluation. For COVID-19 lesion segmentation, the dataset consists of 120 RT-PCR positive CT scans with pixel-level lesion annotations, collected from the first multi-institutional, multi-national expert annotated COVID-19 image database (Tsai et al. 2021). For OC/OD segmentation, the dataset contains 789 cases for training and 281 cases for test, which are collected from four public fundus image datasets and have inconsistent statistical characteristics (Wang et al. 2020). The statistics of three datasets were summarized in Table 1.

Implementation Details

The images in each segmentation task were normalized by subtracting the mean and dividing by the standard deviation. To make a compromise between the network complexity and input image size, the mini-batch size was set to 32 for 2D prostate segmentation with a patch size of 256×256, set to 16 for 2D OC/OD segmentation with a patch size of 512×512, and set to 2 for 3D COVID-19 lesion segmentation with a patch size of $128 \times 196 \times 196$. To expand the training set, several data augmentation techniques were used, including random cropping, rotation, scaling, flipping, adding Gaussian noise, and elastic deformation. The SGD algorithm with a momentum of 0.99 was adopted as the optimizer. The initial learning rate lr_0 was set to 0.01 and decayed according to the polynomial rule $lr = lr_0 \times (1 - t/T)^{0.9}$, where t is the current epoch and T is the maximum epoch. The maximum epoch T was set to 200 for 2D prostate segmentation, 500 for 2D OC/OD segmentation, and 1000 for 3D COVID-19 lesion segmentation. The whole framework was implemented using the PyTorch framework on NVIDIA 2080Ti.

Comparative Experiments

We compared the proposed DCAC model with the 'Intra-domain' setting (*i.e.*, training and testing on the data from the same domain), 'DeepAll' baseline (*i.e.*, training on the data aggregated from all source domains and testing directly on the unseen target domain), and four DG methods, including (1) BigAug: a data-augmentation based method (Zhang et al. 2020), (2) SAML (Liu, Dou, and Heng 2020) and FedDG (Liu et al. 2021a): two meta-learning-based methods, and (3) DoFE: a domain-invariant feature learning approach (Wang et al. 2020). For each segmentation task,

Table 1: Statistics of three datasets used for this study.

Task	Modality	Number of Domains	Cases in each Domain	Total Cases
Prostate Segmentation	MRI	6	30; 30; 19; 13; 12; 12	116
COVID-19 Segmentation	CT	4	28; 19; 58; 15	120
OC/OD Segmentation	Fundus Imaging	4	50/51; 99/60; 320/80; 320/80*	789/281*

* Data split (training/test cases) was provided by (Wang et al. 2020).

the leave-one-domain-out strategy was used to evaluate the performance of each DG method, *i.e.*, training on $K-1$ source domains and evaluating on the left unseen target domain. Each domain is chosen as the target domain in turn. The segmentation performance was measured by the Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD), which characterize the accuracy of predicted masks and boundaries, respectively.

Table 2 gives the DSC and ASD values obtained by our DCAC model and six competing models in each target domain and the average performance over six domains. As expected, the performance of DeepAll seems to be worse on average than that of Intra-domain, due to the distribution discrepancy between the source (training) data and target (test) data. Meanwhile, it shows that the augmentation-based method BigAug performs worse than meta-learning-based methods (*i.e.*, SAML and FedDG), indicating that simply augmenting training data is insufficient to simulate the data distribution of the target domain. It also shows that DoFE is superior to FedDG but slightly inferior to SAML, suggesting that the domain-invariant feature learning approach (*i.e.*, DoFE) can disentangle domain-sensitive features, but it can hardly adapt to different domain discrepancies automatically. It is worth noting that, meta-learning-based methods are designed to directly minimize the generalization gap between the simulated source domain and the target domain in the training stage, which can be effective when the source training domains are adequate to simulate various generalization gaps (see the performance of SAML and FedDG in Domain 5 and 6 in Table 2). More important, it reveals that the proposed DCAC mode not only beats Intra-domain and DeepAll but also outperforms four state-of-the-art DG methods. We believe the superior performance can be attributed to the fact that, with dynamic convolution, our model is capable of adapting to both the predicted domain code and extracted global features of the input image.

The average segmentation performance of our DCAC model and five competing models on the COVID-19 lesion segmentation task and OC/OD segmentation task was given in Table 3 and Table 4, respectively. In both experiments, the average performance of Intra-domain is surprisingly worse than that of DeepAll. A possible reason is that the amount of training data in a single domain (see Table 1) is far from sufficient for training a DCNN model, leading to serious overfitting of the small training dataset. By contrast, aggregating the data in multiple domains can benefit model training and thus results in improved performance. Meanwhile, it seems that the domain-invariant feature learning method is relatively better than meta-learning methods, indicating

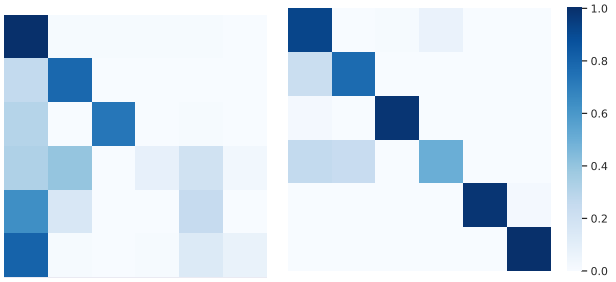

Figure 2: Confusion matrix of the approaches, which use (left) single-scale global features or (right) multiscale global features for domain classification.

the sensitivity of meta-learning-based methods to the number of source domains. When there are less source domains, the diversity of the generalization gap simulated by meta-learning is highly restricted. Comparing to these methods, our model is less susceptible to the number of source domains and achieves stable performance gain on both segmentation tasks. This observation is consistent with what we observed in Table 2.

Ablation Analysis

The prostate segmentation task was chosen as a case study, and ablation studies were conducted on this task to investigate the domain-discriminatory ability of extracted features and the effectiveness of newly designed DAC and CAC modules.

Domain-discriminatory power of extracted features was assessed using the domain classification accuracy. In our DCAC model, we aggregated the feature map produced by each encoder block using GAP and concatenated the multiscale global features to form the input for domain classification. For comparison, we also attempted to apply GAP to only the feature map produced by the last encoder block and use the single-scale global features for domain classification. The confusion matrices of these two approaches were visualized in Figure 2. It shows that the domain attributions can be largely discriminated even using only the deepest semantic enriched global image features. However, using multiscale global features can substantially improve the accuracy of domain classification (see the right part of Figure 2). It suggests that the multiscale feature maps produced by encoder blocks contain domain-specific information, which can be used for reliable domain prediction (*i.e.*, generating the domain code) and should be somehow depressed for domain generalization -based medical image segmentation.

Table 2: Performance (DSC↑/ASD↓) of our DCAC model and six segmentation models in prostate segmentation.

Models	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Domain 6	Average
Intra-domain	89.27/1.41	88.17/1.35	88.29/1.56	83.23/3.21	83.67/2.93	85.43/1.91	86.34/2.06
DeepAll (baseline)	89.27/1.37	87.34/1.20	73.74/2.42	89.19/**1.06**	83.81/2.40	89.41/0.87	85.46/1.55
BigAug (Zhang et al. 2020)	88.62/1.70	86.22/1.56	83.76/2.72	87.35/1.98	85.53/1.90	85.83/1.75	86.21/1.93
SAML (Liu, Dou, and Heng 2020)	89.66/1.38	87.53/1.46	84.43/2.07	88.67/1.56	**87.37/1.77**	88.34/1.22	87.67/1.58
FedDG (Liu et al. 2021a)	90.19/-	87.17/-	85.26/-	88.23/-	83.02/-	90.47/-	87.39/-*
DoFE (Wang et al. 2020)	89.79/1.33	87.42/1.57	84.90/2.13	88.56/1.52	86.47/1.93	87.72/1.33	87.48/1.64
D-CAC	91.24/1.37	89.94/0.92	**86.72/1.67**	89.23/1.34	79.51/3.54	89.90/0.96	87.74/1.70
Ours w/o DAC	91.13/1.12	89.62/1.01	84.75/2.17	**89.31**/1.48	80.79/2.11	89.93/0.93	87.59/1.47
Ours w/o CAC	91.69/1.01	89.96/0.97	85.27/1.89	89.19/1.33	78.44/2.35	**90.65**/0.90	87.53/1.41
Ours (DCAC)	**91.76/0.98**	**90.51/0.89**	86.30/1.77	89.13/1.53	83.39/2.46	90.56/**0.85**	**88.61/1.41**

* ASD was not reported in (Liu et al. 2021a).

Table 3: Average performance (over four domains) of our DCAC model and six models in COVID-19 lesion segmentation.

Model	Average DSC↑/ASD↓
Intra-domain	61.58/14.07
DeepAll (baseline)	63.91/**10.39**
BigAug (Zhang et al. 2020)	64.08/13.33
SAML (Liu, Dou, and Heng 2020)	64.41/13.57
FedDG (Liu et al. 2021a)	63.65/15.36
DoFE (Wang et al. 2020)	64.45/13.45
Ours (DCAC)	**65.10**/12.76

Table 4: Average performance (over four domains) of our DCAC model and six models in OC/OD segmentation.

Model	Average DSC↑/ASD↓
Intra-domain	86.91/13.11
DeepAll (baseline)	87.14/12.89
BigAug (Zhang et al. 2020)	88.19/11.56
SAML (Liu, Dou, and Heng 2020)	87.85/12.31
FedDG (Liu et al. 2021a)	87.03/-*
DoFE (Wang et al. 2020)	88.44/11.33
Ours (DCAC)	**88.47/11.32**

* ASD was not reported in (Liu et al. 2021a).

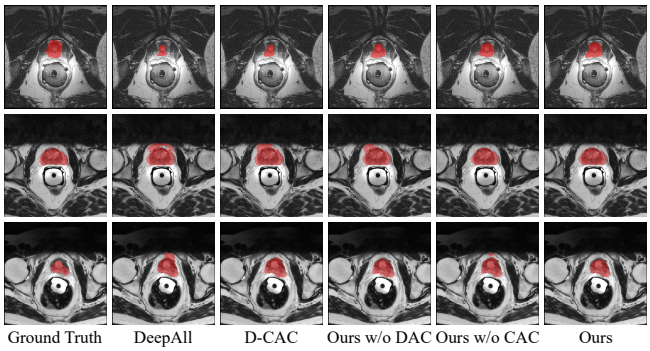

Figure 3: Visualization of the results obtained by applying DeepAll, our DCAC, and three variants of DCAC to three prostate MRI slices, together with the ground truth.

Contributions of DAC and CAC. In this work, we designed the DAC module and CAC module to make our model capable of adapting to the unseen test domain and test image, respectively. To evaluate the contributions of these two modules, we compared our model with its variant that uses only one module. We also compared to a variant, denoted by D-CAC, that uses the concatenation of domain code and global image features to generate one and only one unified dynamic head. The performance of our DCAC model and its three variants was given in Table 2. It reveals that our DCAC model outperforms not only D-CAC but also the variant without either DAC or CAC. The results confirm that either DAC or CAC contributes to the final results and the two-dynamic-head strategy is superior to the unified dynamic head.

We visualized the segmentation results of DCAC and three variants in Figure 3. We also display the results of DeepAll and ground truth for reference. It shows that our DCAC model can produce more accurate segmentation results of unseen test images, particularly in the boundary region.

Conclusion

This paper proposes a multi-source domain generalization model called DCAC, which uses two dynamic convolutional heads. One dynamic head is conditioned on the predicted domain code of the input to make the DCAC model adapt to the target domain, while the other dynamic head is conditioned on global image features to make the model adapt to the input image. Our results on the prostate segmentation, COVID-19 lesion segmentation, and OC/OD segmentation tasks suggest that, after training on the data from multiple source domains, the proposed DCAC model can generalize well on an unseen target domain, achieving substantially improved average performance over the baseline and four state-of-the-art domain generalization methods. In our future work, we will extend the proposed DCAC model to multi-source, multi-modality, and multi-task scenarios, aiming to provide large-scale pre-trained segmentation model for various downstream medical image segmentation applications.

References

Du, Y.; Zhen, X.; Shao, L.; and Snoek, C. G. M. 2021. MetaNorm: Learning to Normalize Few-Shot Batches Across Domains. In *International Conference on Learning Representations (ICLR)*. OpenReview.net.

Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. 2019. U-Net: Deep Learning for Cell Counting, Detection, and Morphometry. *Nature Methods*, 16(1): 67–70.

Fan, X.; Wang, Q.; Ke, J.; Yang, F.; Gong, B.; and Zhou, M. 2021. Adversarially Adaptive Normalization for Single Domain Generalization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8208–8217.

Han, Y.; Huang, G.; Song, S.; Yang, L.; Wang, H.; and Wang, Y. 2021. Dynamic Neural Networks: A Survey. *arXiv:2102.04906 [cs]*.

He, J.; Deng, Z.; and Qiao, Y. 2019. Dynamic Multi-Scale Filters for Semantic Segmentation. In *International Conference on Computer Vision (ICCV)*, 3561–3571.

He, Y.; Carass, A.; Zuo, L.; Dewey, B. E.; and Prince, J. L. 2021. Autoencoder Based Self-Supervised Test-Time Adaptation for Medical Image Analysis. *Medical Image Analysis*, 72: 102136.

Isensee, F.; Jaeger, P. F.; Kohl, S. A. A.; Petersen, J.; and Maier-Hein, K. H. 2020. nnU-Net: A Self-configuring Method for Deep Learning-based Biomedical Image Segmentation. *Nature Methods*, 18(2): 203–211.

Karani, N.; Erdil, E.; Chaitanya, K.; and Konukoglu, E. 2021. Test-time Adaptable Neural Networks for Robust Medical Image Segmentation. *Medical Image Analysis*, 68: 101907.

Klein, B.; Wolf, L.; and Afek, Y. 2015. A Dynamic Convolutional Layer for Short Range Weather Prediction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4840–4848.

Li, C.; Qi, Q.; Ding, X.; Huang, Y.; Liang, D.; and Yu, Y. 2021. Domain Generalization on Medical Imaging Classification using Episodic Training with Task Augmentation. *arXiv:2106.06908 [cs]*.

Li, D.; Zhang, J.; Yang, Y.; Liu, C.; Song, Y.-Z.; and Hospedales, T. M. 2019. Episodic Training for Domain Generalization. In *International Conference on Computer Vision (ICCV)*.

Li, H.; Wang, Y.; Wan, R.; Wang, S.; Li, T.-Q.; and Kot, A. 2020. Domain Generalization for Medical Imaging Classification with Linear-Dependency Regularization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 3115–3126.

Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J. A.; van Ginneken, B.; and Sánchez, C. I. 2017. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42: 60–88.

Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021a. FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1013–1023.

Liu, Q.; Dou, Q.; and Heng, P.-A. 2020. Shape-Aware Meta-learning for Generalizing Prostate MRI Segmentation to Unseen Domains. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 475–485.

Liu, Q.; Dou, Q.; Yu, L.; and Heng, P. A. 2020. MS-Net: Multi-Site Network for Improving Prostate Segmentation With Heterogeneous MRI Data. *IEEE Transactions on Medical Imaging*, 39(9): 2713–2724.

Liu, X.; Thermos, S.; O'Neil, A.; and Tsaftaris, S. 2021b. Semi-supervised Meta-learning with Disentanglement for Domain-generalised Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*.

Onofrey, J. A.; Casetti-Dinescu, D. I.; Lauritzen, A. D.; Sarkar, S.; Venkataraman, R.; Fan, R. E.; Sonn, G. A.; Sprenkle, P. C.; Staib, L. H.; and Papademetris, X. 2019. Generalizable Multi-Site Training and Testing Of Deep Neural Networks Using Image Normalization. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 348–351.

Pang, Y.; Zhang, L.; Zhao, X.; and Lu, H. 2020. Hierarchical Dynamic Filtering Network for RGB-D Salient Object Detection. In *European Conference on Computer Vision (ECCV)*, 235–252.

Roth, H.; Xu, Z.; Diez, C. T.; Jacob, R. S.; Zember, J.; Molto, J.; Li, W.; Xu, S.; Turkbey, B.; Turkbey, E.; et al. 2021. Rapid Artificial Intelligence Solutions in a Pandemic-The COVID-19-20 Lung CT Lesion Segmentation Challenge. *Research Square*.

Shen, Y.; Sheng, B.; Fang, R.; Li, H.; Dai, L.; Stolte, S.; Qin, J.; Jia, W.; and Shen, D. 2020. Domain-invariant Interpretable Fundus Image Quality Assessment. *Medical Image Analysis*, 61: 101654.

Tian, Z.; Shen, C.; and Chen, H. 2020. Conditional Convolutions for Instance Segmentation. In *European Conference on Computer Vision (ECCV)*, 282–298.

Tsai, E. B.; Simpson, S.; Lungren, M.; Hershman, M.; Roshkovan, L.; Colak, E.; Erickson, B. J.; Shih, G.; Stein, A.; Kalpathy-Cramer, J.; Shen, J.; Hafez, M.; John, S.; Rajiah, P.; Pogatchnik, B. P.; Mongan, J.; Altinmakas, E.; Ranschaert, E. R.; Kitamura, F. C.; Topff, L.; Moy, L.; Kanne, J. P.; and Wu, C. C. 2021. The RSNA International COVID-19 Open Annotated Radiology Database (RICORD). *Radiology*, 299(1): E204–E213.

Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.-W.; and Heng, P.-A. 2019. Boundary and Entropy-driven Adversarial Learning for Fundus Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 102–110.

Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.-W.; and Heng, P.-A. 2020. DoFE: Domain-Oriented Feature Embedding for Generalizable Fundus Image Segmentation on Unseen

Datasets. *IEEE Transactions on Medical Imaging*, 39(12): 4237–4248.

Xie, X.; Niu, J.; Liu, X.; Chen, Z.; Tang, S.; and Yu, S. 2021. A Survey on Incorporating Domain Knowledge into Deep Learning for Medical Image Analysis. *Medical Image Analysis*, 69: 101985.

Yang, Y.; and Soatto, S. 2020. FDA: Fourier Domain Adaptation for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4085–4095.

Zhang, J.; Xie, Y.; Xia, Y.; and Shen, C. 2021. DoDNet: Learning To Segment Multi-Organ and Tumors From Multiple Partially Labeled Datasets. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1195–1204.

Zhang, L.; Wang, X.; Yang, D.; Sanford, T.; Harmon, S.; Turkbey, B.; Wood, B. J.; Roth, H.; Myronenko, A.; Xu, D.; and Xu, Z. 2020. Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains via Deep Stacked Transformation. *IEEE Transactions on Medical Imaging*, 39(7): 2531–2540.

Zhao, X.; Sicilia, A.; Minhas, D. S.; O'Connor, E. E.; Aizenstein, H. J.; Klunk, W. E.; Tudorascu, D. L.; and Hwang, S. J. 2021. Robust White Matter Hyperintensity Segmentation on Unseen Domain. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 1047–1051.

Zhou, J.; Jampani, V.; Pi, Z.; Liu, Q.; and Yang, M.-H. 2021a. Decoupled Dynamic Filter Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6647–6656.

Zhou, Q.; Zhou, W.; Wang, S.; and Xing, Y. 2021b. Duplex Adversarial Networks for Multiple-source Domain Adaptation. *Knowledge-Based Systems*, 211: 106569.

Zhou, Z.; Siddiquee, M. M. R.; Tajbakhsh, N.; and Liang, J. 2020. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, 39(6): 1856–1867.