

Spectrum and Style Transformation Framework for Omni-Domain COVID-19 Diagnosis

Zhenkun Wang^{ID}, Member, IEEE, Shuangchun Gui^{ID}, Xinpeng Ding^{ID}, Xiaowei Hu^{ID}, Xiaowei Xu^{ID}, Member, IEEE, and Xiaomeng Li^{ID}, Member, IEEE

Abstract—Coronavirus disease 2019 (COVID-19) is an ongoing global pandemic and profoundly affects almost all people around the world. Thus, many automatic diagnosis methods based on computed tomography (CT) images have been proposed to reduce the workload of radiologists. Most of the existing methods focus on the in-domain predictions, i.e., the training and testing have similar distributions, which is impractical in real-world situations, since the CT images can be collected from different devices and in different hospitals. To improve the diagnosis performance of COVID-19 for both in-domain and out-of-domain data, this paper proposes a spectrum and style transformation framework for omni-domain COVID-19 diagnosis. To achieve this, we first present a spectrum transform module, which helps to discover the discriminating features of each domain to recognize the in-domain data. Then, we formulate a cross-domain stylization module, which learns the cross-domain knowledge to enhance the model generalization capability to deal with out-of-domain data. Moreover, our framework is a plug-and-play module that can be easily integrated into existing deep models. We evaluate our framework on four COVID-19 datasets and show our method consistently improves the diagnosis performance of various methods on both in-domain and out-of-domain data.

Index Terms—COVID-19 diagnosis, image classification, data augmentation, and cross-domain learning.

I. INTRODUCTION

COVID-19 is continuing to spread around the world and leads to more than 600 million confirmed cases and more

Manuscript received 17 June 2022; revised 8 September 2022; accepted 30 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62106096, in part by the Shenzhen Technology Plan under Grant JCYJ20220530113013031, in part by the Shenzhen Municipal Central Government Guides Local Science and Technology Development Special Funded Projects under Grant 2021Szvup139, and in part by the Shanghai Committee of Science and Technology under Grant 21DZ1100100. (Zhenkun Wang and Shuangchun Gui contributed equally to this work.) (*Corresponding authors: Xiaowei Hu; Xiaowei Xu*)

Zhenkun Wang and Shuangchun Gui are with the School of System Design and Intelligent Manufacturing (SDIM), Southern University of Science and Technology, Shenzhen 518055, China (e-mail: wangzhenkun90@gmail.com; 12132667@mail.sustech.edu.cn).

Xinpeng Ding and Xiaomeng Li are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China (e-mail: xpding.xidian@gmail.com; xmengli999@gmail.com).

Xiaowei Hu is with the Shanghai AI Laboratory, Shanghai 200232, China (e-mail: xwhu@cse.cuhk.edu.hk).

Xiaowei Xu is with the Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510050, China (e-mail: xiao.wei.xu@foxmail.com).

Digital Object Identifier 10.1109/TETCI.2022.3219858

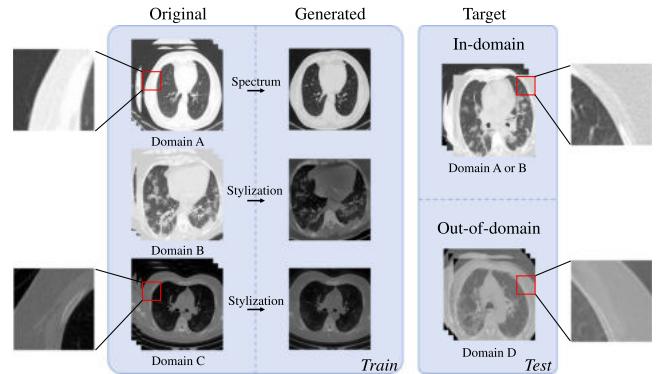


Fig. 1. Example CT images at different domains. “In-domain” indicates the model is trained on Domains A and B, then tested on Domain A or B. “Out-of-domain” indicates the model is trained on Domains A, B and C, then tested on Domain D. The enlarged regions of the red boxes are for better visualization.

than six million deaths until October 2022. To diagnose COVID-19, CT plays a pivotal role and has long been used in the detection of lung diseases. However, the increase in patients leads to a heavy workload for radiologists. To reduce the workload and accumulate the diagnosis process, several methods [1], [2], [3], [4] on automatically CT-based COVID-19 diagnosis have been developed and show remarkable performance, especially using deep learning techniques. The current deep-learning-based approaches assume that the data in both training and testing datasets have a similar distribution. However, in actual applications, CT images are generally collected by different devices and from different hospitals. Fig. 1 shows the images in multiple domains (Domain A-D) with different appearances. Note that we define (i) an *in-domain* task, where the training and testing data are collected from the same domains; (ii) an *out-of-domain* generalization task, where the training and testing data are collected from different domains; and (iii) an *omni-domain* generalization task that aims to improve the performance of deep models for both in-domain and out-of-domain data. In this paper, we focus on the *omni-domain* generalization problem, since in testing time we usually have the CT images collected from the same devices with the training data (in-domain) and also the CT images from other devices (out-of-domain). Hence, it is important to improve the performance of in-domain and out-of-domain data. The existing methods [2], [5], [6], [7] designed for COVID-19 diagnosis have limited performance when the training and testing data have different distributions while

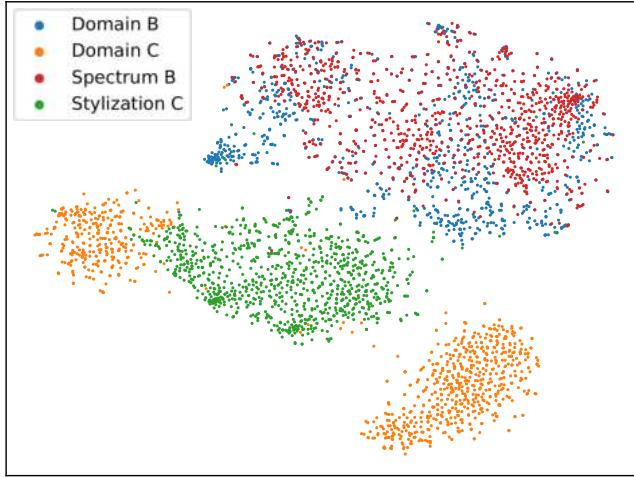


Fig. 2. Visualization of feature maps learned by our deep network. “Domain B” indicates the features of CT images in Domain B; “Domain C” indicates the features of CT images in Domain C; “Spectrum B” denotes the images reconstructed by the low-frequency components of Domain B and the high-frequency components of Domain C; “Stylization C” denotes the images reconstructed by the content of Domain B and the style components of Domain C.

the domain generalization algorithms [8], [9], [10], [11] ignore the individual domain characteristics and the performance on in-domain data may degrade.

In this paper, we aim to improve the diagnosis performance of COVID-19 for omni-domain generalization by discovering the discriminating features of each domain to recognize the in-domain data and learning the cross-domain knowledge to enhance the model generalization capability to deal with out-of-domain data, thus improving the overall performance for COVID-19 diagnosis in omni-domain. To achieve this, we first present a spectrum transform module, which removes the high-frequency information, e.g., edge, from one domain and incorporates that information from other domains as the data augmentation. This is because the high-frequency information mainly contains low-level statistics features, which contribute less to the disease diagnosis. By incorporating the low-level statistics features from other domains to enrich the training data, the deep model will focus on learning the high-level semantic features that help to distinguish COVID-19. Second, to improve the generalization capability of the deep neural network, we design a cross-domain stylization module to transfer the style features, e.g., color and contrast, from one domain to another domain, thus creating more out-of-domain data for network training.

Fig. 2 illustrates the visualization of feature maps learned by our deep network, where we can see that (i) the features in Domain B and that after spectrum operation have similar distributions, showing that the transformed high-frequency spectrum information has fewer effects on changing the semantic features that help to diagnose COVID-19; and (ii) the features in Domain C and that after stylization operation have different distributions, showing that the stylization operation creates more out-of-domain training data, thus improving the generalization capability of the deep model.

The contributions of this paper are summarized as follows:

- We design a novel deep neural network framework for COVID-19 diagnosis to improve the performance for both in-domain and out-of-domain data.
- We formulate a spectrum transform module to discover the discriminating features of each domain and a cross-domain stylization module to enhance the model generalization capability to deal with out-of-domain data.
- We evaluate our method on four public COVID-19 datasets collected at different times and in different hospitals and show that our method consistently improves the performance of data in various domains.

II. RELATED WORK

COVID-19 has been an ongoing global pandemic for a long time and lots of works have been developed for COVID-19 diagnosis; please see [1] for a detailed survey. In this section, we will focus on the related studies in the following aspects.

A. Intra-Domain Learning

Intra-domain learning aims to diagnose COVID-19 when the training and testing data are from a single domain and share a similar distribution. To solve this task, Xu et al. [3] established a system for early case screening to distinguish between healthy people, normal pneumonia, and COVID-19 patients. Pham et al. [4] used 16 pre-trained models and performed the two-class classification for COVID-19 diagnosis. Ma et al. [12] developed a multi-receptive field attention module to improve the diagnosis accuracy. Wang et al. [13] provided an easily-trained deep learning algorithm to identify COVID-19 patients and discover lesion regions in chest CT. Wang et al. [14] collected a large X-ray data set on COVID-19 and Afshar et al. [15] leveraged the capsule network to perform COVID-19 diagnosis. To capture the spatial relationship between the region of interest in images, Sitaula et al. [16] proposed an attention-based deep learning model and evaluated the performance on three COVID-19 datasets. El-Kenawy et al. [17] classified COVID-19 diseases into three categories by using a two-stage method based on an advanced squirrel search optimization algorithm. Saha et al. [18] proposed a novel graph isomorphic network based model to detect COVID-19 from CT and X-ray images which pre-processed the image into an undirected graph. Similarly, Yao et al. [19] enhanced the image features using fuzzy stacking and sequentially transferred the trained features between the fractional max-pooling neural network estimators. To extract the rich whole-lung features, Li et al. [20] proposed a novel framework based mask-guided attention mechanism for distant metastasis prediction. However, the number of collected images in a single domain is limited, thus degrading the performance of the diagnosis accuracy of COVID-19.

B. Cross-Domain Learning

The training data in a single domain is limited and we want to collect the data from multiple domains (different devices and hospitals) to train the deep models. However, the data collected

from different domains show different distributions. To learn the knowledge from multiple domains, Wang et al. [2] designed contrastive loss and Ali et al. [5] presented the data augmentation strategies to improve the learning ability of [14]. Di et al. [6] proposed the uncertainty vertex-weighted hypergraph learning to deal with the data collected from five hospitals. Hu et al. [7] leveraged the deep supervised learning with a self-adaptive auxiliary loss to solve the class imbalance issue. Ouyang et al. [21] presented a 3D convolutional network with an online attention module focused on the lesion regions to diagnose COVID-19 from community-acquired pneumonia and developed a dual-sampling strategy to mitigate the imbalanced learning based on the multi-center CT data. Wang et al. [22] proposed a conceptually simple framework with two 3D-ResNets coupled into a single model for the two tasks via a novel prior-attention strategy for COVID-19 diagnosis. Elazab et al. [23] collected a large-scale X-ray dataset from ten resources and designed a multi-site graph convolutional network to perform the disease classification. Unlike the traditional diagnostic models, Zhong et al. [24] learned the optimized embedding space of images and developed a retrieval model based on deep metric learning for COVID-19 diagnosis. Frank et al. [25] integrated domain knowledge into deep neural networks on the task of COVID-19 severity assessment. Also for X-ray images, Sitaula et al. [26], [27] proposed a new multi-scale bag of deep visual words features that can exploit three different scales feature maps from VGG-16 to capture the detailed semantic information in the lungs. Recently, [28] used the lung ultrasound frames to integrate domain knowledge into the deep neural network on COVID-19. However, the above methods only explore to use of the training data across different domains and still have an obvious performance degradation to deal with the testing case where the data is from an unseen domain.

C. Domain Generalization

Domain generalization aims to solve the out-of-domain generalization task, where we train the model on the source domains and test it on the unseen domains. Song et al. [8] developed an augmented multi-center graph convolutional network to diagnose COVID-19 on unseen domains. Li et al. [9] improved the model generalization capability by discovering the knowledge from other medical image datasets. Xu et al. [29] leveraged a stochastic class-balanced boosting sampling strategy to overcome the imbalanced learning problem and improve the generalization ability of the classifier. Jin et al. [30] proposed a prior knowledge-driven domain adaptation and a dual-domain enhanced self-correction learning scheme for COVID-19 segmentation. Liu et al. [31] presented a shape-aware meta-learning scheme to improve the model generalization. To avoid sensitive data-sharing, a novel federated semi-supervised learning technique [32] is proposed for COVID-19 region segmentation by using multi-national data. However, these methods perform limited performance on the in-domain data. In contrast, our work aims to solve the omni-domain generalization issue and consistently improve the performance of both in-domain and out-of-domain data.

III. METHODOLOGY

A. Overall Framework Design

To improve the diagnosis performance of COVID-19 on both in-domain and out-of-domain data, we design a spectrum and style transformation framework to discover the discriminating features of each domain and to learn the cross-domain knowledge that enhances the model generalization capability. Fig. 3 shows the workflow of the proposed framework, where we take the image x_{main} from one domain and the image x_{aux} from another domain as the inputs and output the classification result of COVID-19 in an end-to-end manner. Given the input images x_{main} and x_{aux} , we formulate a spectrum transform module to generate the spectrum-transformed image x_f , and a cross-domain stylization module to produce the cross-domain stylized image x_s . Next, we adopt a selector to randomly select x_f, x_s , and the original as the training data to train a classification network f_θ to predict the probability of COVID-19. In testing, we can adopt the trained framework on any CT images to perform the COVID-19 diagnosis on omni-domain data.

In the next subsections, we will elaborate on the architectures of the proposed spectrum transform module and the cross-domain stylization module in detail.

B. Spectrum Transform Module

In this section, we formulate a spectrum transform module to discover the discriminating features of each domain. The key idea is to transform the low-level statistic features encoded in the high-frequency range from one domain to another domain, thus enriching the training data and forcing the deep models to learn more highly semantic features that contribute to distinguishing COVID-19.

The green block in Fig. 3 shows the architecture of this module. We first transform the x_{main} and x_{aux} from different source domains into the spectrum space $z = \mathcal{F}(x)$, where \mathcal{F} denotes the Fourier transform. Then, we split z into the high-frequency range ($HF = z \odot m$) and the low-frequency range ($LF = z \odot (1 - m)$), where m is a mask obtained by the following equation:

$$m = \begin{cases} 1, & \text{if } (i - \frac{H}{2})^2 + (j - \frac{W}{2})^2 \geq r^2 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $1 \leq i \leq H$, $1 \leq j \leq W$; H and W are the height and width of the image; r is a hyper-parameter and we empirically set it as $(W + H) * (3/32)$. As shown in Fig. 4, the high-frequency components contain the low-level statistic features, e.g., edges and noises. The high-frequency components of the images from different domains show different appearances. Hence, we present to transform the high-frequency component of the image from one domain to the image from another domain to enrich the training data and make the network less sensitive to the low-level statistic features. This is achieved by synthesizing the new data in frequency range z_f :

$$z_f = HF(x_{aux}) + LF(x_{main}) \quad (2)$$

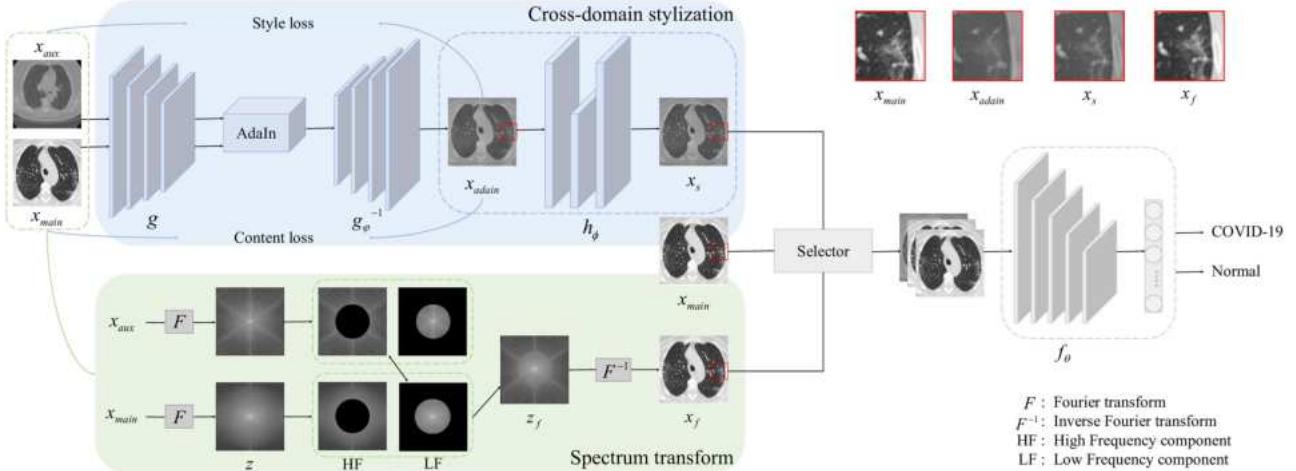


Fig. 3. The overall architecture of the proposed spectrum and style transformation framework. Two major components, the spectrum transform module and the cross-domain stylization module, are designed to improve the model performance for both in-domain and out-of-domain data. x_{main} is the original image. x_{adain} is the corresponding stylized image. x_f is the spectrum-transformed image. x_s is the stylized image after revision. The enlarged regions of the red boxes are for better visualization.

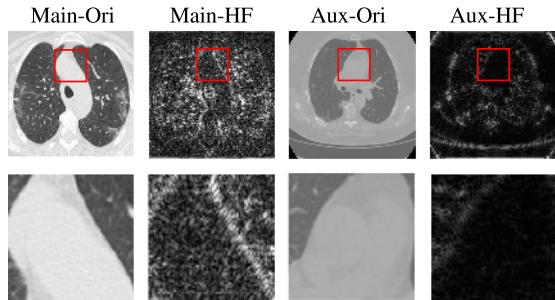


Fig. 4. High-frequency components. “Main-Ori” and “Aux-Ori” indicate x_{main} and x_{aux} from different domains. “Main-HF” and “Aux-HF” denote the high-frequency components of these images. The second row shows the enlarged regions of the red boxes in the first row for better visualization.

where the high-frequency information (HF) is from the auxiliary image while the low-frequency information (LF) is from the main image. After that, we reconstruct the newly synthesized image into the image space by the inverse Fourier transform $x_f = |\mathcal{F}^{-1}(z_f)|$.

C. Cross-Domain Stylization Module

The CT images collected from different domains usually have different styles and different distributions. To enhance the model generalization capability that helps to deal with the out-of-domain data, we design a cross-domain stylization module, which transforms the style from one domain to another domain, thus generating the training data in a new domain. Following [10], [33], [34], we use adaptive instance normalization (AdaIN) [35] fast style transfer. Compared with iterative stylization [36], AdaIN achieves a faster speed and enables arbitrary style transfer in real-time. To enable stylizing the entire COVID-19 dataset, AdaIN would take prohibitively short with an iterative approach.

The blue block in Fig. 3 shows the architecture of the proposed cross-domain stylization module. Given the images x_{main} and x_{aux} as the inputs, we first use several convolutional layers to build an encoder sub-network g , which extracts the features of the input images. After that, we extract the style of the image x by computing the mean $\mu(x)$ and standard deviation $\sigma(x)$:

$$\mu(x) = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W g(x) \quad (3)$$

$$\sigma(x) = \sqrt{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (g(x) - \mu(x))^2 + \epsilon}. \quad (4)$$

Next, we transform the style from the auxiliary image x_{aux} from one domain to the main image x_{main} from another domain through AdaIn:

$$\begin{aligned} \text{AdaIn}(x_{main}, x_{aux}) &= \sigma(x_{aux}) \left(\frac{g(x_{main}) - \mu(x_{main})}{\sigma(x_{main})} \right) \\ &\quad + \mu(x_{aux}). \end{aligned} \quad (5)$$

Further, we reconstruct the image x_{adain} after the cross-domain stylization by a decoder sub-network g_φ^{-1} :

$$x_{adain} = g_\varphi^{-1}(\text{AdaIn}(x_{main}, x_{aux})). \quad (6)$$

During the training process, we initialize the parameters of encoder g by the weights of VGG-19 [37] trained on ImageNet. Then, we train the parameters of decoder g_φ^{-1} by computing the mean squared error (MSE) on both content loss ℓ_c and style loss ℓ_s :

$$\begin{aligned} \ell_{adain} &= \ell_c(g_M(x_{main}), g_M(x_{adain})) \\ &\quad + \sum_{i=1}^M \ell_s(g_i(x_{aux}), g_i(x_{adain})) \end{aligned} \quad (7)$$

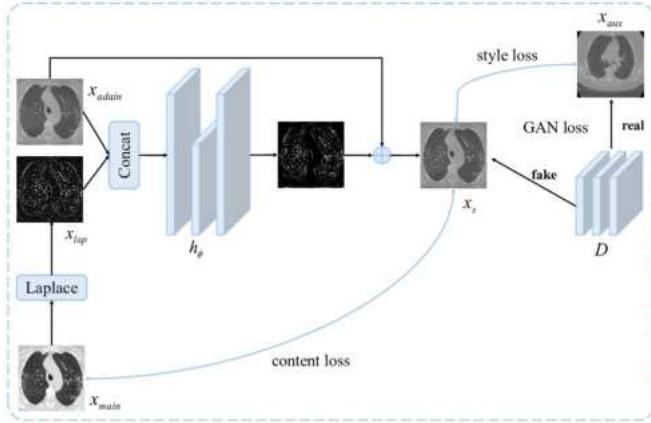


Fig. 5. The detailed structures of the proposed revision network.

where g_i denotes the i -th layer in VGG-19 used to extract the features of style and content; g_M is the features obtained from the last layer; M is the total number of the layers, and we have four layers in total. The content loss $\ell_c(\cdot, \cdot)$ and the style loss $\ell_s(\cdot, \cdot)$ are defined as follows:

$$\ell_s(F_1, F_2) = \|\mu(F_1) - \mu(F_2)\|_2 + \|\sigma(F_1) - \sigma(F_2)\|_2 \quad (8)$$

$$\ell_c(F_1, F_2) = \|F_1 - F_2\|_2 \quad (9)$$

where F_1 and F_2 denote the features extracted by the network g ; please see Eq. (7) for the input features.

After obtaining the stylized image x_{adain} , we formulate a revision network [38] to enhance the detail information of the image. Fig. 5 illustrates the detailed structures of the revision network. Given the original input image x_{main} , we first adopt the Laplace transform to extract the detail information x_{lap} , which is concatenated with the stylized image x_{adain} . Then, we use a revision network h_ϕ to complete the detail information of x_{lap} and generate a new detail image, which is added with x_{adain} to obtain x_s . To optimize the network to enhance the detail information of x_{adain} , we adopt a content loss, a style loss, and a GAN loss to train the parameters of h_ϕ . The style loss is the same as Eq. (8). The content loss is defined as:

$$\ell'_c(F_1, F_2) = \|norm(F_1) - norm(F_2)\|_2 \quad (10)$$

where $norm$ denotes the channel-wise normalization, which helps to better retain content details. The GAN loss is formulated by training a discriminator D to make the generated image x_s be similar to x_{aux} . Concretely, we apply standard adversarial loss [39] to classify each of the generated image and its style image as real or fake. Following [38], we express the gan loss as:

$$\begin{aligned} \ell_{gan} &= \mathbb{E}_{y \sim p_{data}(x_{aux})} [\log D(y)] \\ &+ \mathbb{E}_{x \sim p_{data}(x_{adain})} [\log(1 - D(h_\phi(x)))] \end{aligned} \quad (11)$$

Note that during the training process of the revision network, the parameters of other parts, e.g., the encoder g and decoder g_φ^{-1} , are fixed. Finally, the overall loss function ℓ_{rev} is defined

as follows:

$$\begin{aligned} \ell_{rev} &= \sum_{i=1}^M \ell'_c(g_i(x_{main}), g_i(x_s)) \\ &+ \sum_{i=1}^M \ell_s(g_i(x_{aux}), g_i(x_s)) + \ell_{gan}. \end{aligned} \quad (12)$$

D. Overall Training

Except for the training losses mentioned above, we further adopted a cross-entropy loss ℓ_{ce} to optimize the classifier f_θ to diagnose COVID-19:

$$\ell_{ce} = \frac{1}{N} \sum_{i=1}^N -y_i \log f_\theta(x_i) \quad (13)$$

where x_i and y_i denote the input image and the ground truth label respectively and N is the number of training samples. Note that we adopt ResNet-18 as the network structure of f_θ and initialize the parameters using the weights trained on ImageNet. In detail, given a set of training data S_{train} , we create the main and auxiliary image pairs S_{pair} , where the ground truth labels of the main and auxiliary images are the same, i.e., $x_{main}, x_{aux} \in S_{train}$, where $S_{train} = \{(x_i, y_i)\}_{i=1}^n$ and $y_{main} = y_{aux}$. Then, in the cross-domain stylization module, we use the parameters of VGG-19 to initialize the encoder g , train a decoder g_φ^{-1} , and then fix the parameters of the decoder and train the parameters of h_ϕ . After that, we use a selector to randomly leverage the generated images x_s , x_f , and the original input image x_{main} with the selector probability p_{sel} to train the classifier f_θ . The overall training procedure of the proposed framework is summarized in Algorithm 1.

IV. EXPERIMENTS

In this section, we first introduce the multi-domain datasets evaluated on the spectrum and style transformation framework. Next, we summarize the details of the implementation from two aspects of the data generator and label classifier. Finally, the experimental results are presented to evaluate the effectiveness of the proposed framework.

A. Datasets and Preprocessing

We select four public COVID-19 CT datasets to evaluate the performance of our method, including COVID-CT [45] (denoted as **Domain A**), SARS-cov [46] (denoted as **Domain B**), COVID-CTset [47] (denoted as **Domain C**) and Mosmed-data [48] (denoted as **Domain D**). For the evaluation of in-domain performance, we follow the method [2], which use the training sets of Domain A and B for training and their test sets for testing for a fair comparison. For the evaluation of out-of-domain performance, we adopt the leave-one-domain-out strategy, for example, we train the model on the training sets of Domain A, B, and C and report the result on the test set of Domain D.

Domain A. COVID-CT [45] includes 349 COVID-19 and 397 non-COVID-19 CT images, mainly derived from case information published on the website. Later in the research process,

Algorithm 1: Training Procedure of the Spectrum and Style Transformation Framework.

```

1: Input: source domain  $S_{train}$ , encoder  $g$ , decoder  $g_\varphi^{-1}$ ,
   revision network  $h_\phi$ , discriminator  $D$ , label classifier  $f_\theta$ ,
   and other hyper-parameters.
2: Output: classifier  $f_\theta$ .
3: Initialize:  $g$ : pre-trained VGG-19;  $f_\theta$ : pre-trained
   ResNet-18.
4: Create image pairs  $S_{pair}$ .                                // prepare data
5: while not end of epoch do
6:    $(x_{main}, x_{aux}) \sim S_{pair}$ 
7:    $x_{adain} = g_\varphi^{-1}(\text{AdaIn}(x_{main}, x_{aux}))$ 
8:    $\varphi = \varphi - \eta \nabla_\varphi \ell_{adain}$ 
9: end while
10: Obtain  $g_\varphi^{-1}$                                          // style transfer
11: while not end of epoch do
12:    $(x_{main}, x_{aux}) \sim S_{pair}$ 
13:    $x_s = h_\phi[x_{adain}]$ 
14:   Update  $D$ 
15:    $\phi = \phi - \eta \nabla_\phi \ell_{rev}$ 
16: end while
17: Obtain  $h_\phi$                                               // revision network
18: while not end of  $S_{pair}$  do
19:   if  $p < p_{spec}$  then
20:     Create  $x_s$ 
21:   else
22:     Create  $x_f$ 
23:   end if
24: end while
25: Obtain  $S_{new}$                                             // generate images
26: while not end of epoch do
27:    $(x_{ori}, x_{new}) \sim (S_{pair}, S_{new})$ 
28:   Replace  $x_{ori}$  to  $x_{new}$  with the selector probability  $p_{sel}$ 
29:    $l = f_\theta(x_{ori})$ 
30:    $\theta = \theta - \eta \nabla_\theta \ell_{ce}$ 
31: end while
32: Obtain  $f_\theta$                                               // classifier

```

we consulted a professional doctor and removed 42 images that were not used to diagnose COVID-19, and this domain finally includes 511 CTs for training and 193 for testing. We will release the removed images for future research.

Domain B. SARS-cov [46] is collected from 120 patients in a hospital in Sao Paulo, Brazil. There are 1252 positive images and 1229 negative ones. The negative images may have other types of lung diseases. The training set contains 1861 images and the testing set contains 620 images.

Domain C. COVID-CTset [47] contains 48260 positive CT images and 15589 negative images, which were obtained in a hospital in Iran between March 5 and April 23, 2020. In order to be consistent with the number of other domains, the final experiment contains 3376 images for training and 1446 images for testing.

Domain D. Mosmeddata [48] contains 3D CT images of human lungs that are positive and negative for COVID-19, and

the positive cases are further subdivided into four levels. The image was acquired at a hospital in Moscow from March 1, 2020, to April 25, 2020. After preprocessing, we convert the 3D images into 2D. The 1689 images in the experiments are used for training and 723 are used for testing.

B. Implementation Details

All models in this paper were built with PyTorch [49] and trained on an NVIDIA Tesla V100 32 GB GPU.

Generator. The trained generator is the cross-domain stylization module, which consists of two main components, i.e., the AdaIn and revision network. In the AdaIn part, its encoder uses the first four layers of the pre-trained VGG-19 [37], and the decoder has a symmetrical structure. The revision network is a simple encoder-decoder architecture that contains six 3×3 convolutional layers and each layer is followed by ReLU. The discriminator is also composed of three convolutional layers, with BN and LeakyRelu inserted after the first two layers. In the process of training all networks, we use the same data augmentation as [35], i.e., RandomCrop and adopt Adam optimizer [50] with a learning rate of 0.0001. The number of the epoch is set to 50,000 and the batch size is set to 8.

Classifier. The architecture of the classifier, i.e., f_θ is the ResNet-18 [51] which is pre-trained by ImageNet [52] with a dropout of 0.5. In the training stage, we use data augmentations including RandomHorizontalFlip, ColorJitter, and RandomGrayscale. f_θ is optimized with SGD, with the initial learning rate of 0.01 and the batch size of 64 for 100 epochs. The learning rate is decayed by 0.1 every 60 epochs.

C. Comparison With the State-of-The-Arts

We evaluate our proposed method on the four domain datasets under different backbones, i.e., ResNet [51] and Swin Transformer [53], and compare it with recently proposed state-of-the-art methods and report the results in Table I. Since current approaches for COVID-19 CT image classification only focus on the in-domain setting. We use their release code to obtain the results on the out-of-domain setting. Note that we remove some images that are not used to diagnose COVID-19 under the suggestion of the doctors. All models are trained in the same setting.

From Table I, it is clear that our proposed method achieves the best performance in terms of both in-domain and out-of-domain data on the COVID-19 CT datasets. Specifically, for the performance of in-domain, our method obtains 89.1% average accuracy on Domain A and Domain B, outperforming 6.2% over the previous SOTA, i.e., COVIDNet [2]. For the performance of out-of-domain, our method outperforms the current methods by a large margin, i.e., 22.8% over COVIDNet [2] on the average accuracy. Compared with 66.8% accuracy of COVIDNet [2], the joint performance of our method, i.e., the average accuracy of in-domain and out-of-domain, achieves 81.2%, which outperforms 14.4% over COVIDNet [2]. In the last four experiments, we evaluate the four backbones: Tiny Swin Transformer, Small Swin transformer, ResNet-18 and ResNet-101. It is clear that our

TABLE I

COMPARISON OF THE IN-DOMAIN AND OUT-OF-DOMAIN ACCURACY OF STATE-OF-THE-ART METHODS ON COVID-19 DATASET. EACH COLUMN DENOTES THE TEST DOMAIN. “AVG.” DENOTES THE AVERAGE ACCURACY UNDER THE CURRENT SETTING. “JOINT AVG.” DENOTES THE AVERAGE ACCURACY OF IN-DOMAIN AND OUT-OF-DOMAIN

Method	#Params	In-domain			Out-of-domain				Joint avg.
		Domain A	Domain B	Avg.	Domain A	Domain B	Domain C	Domain D	
COVID-Net [14]	11.9M	47.5±0.0	79.8±0.0	63.7	53.4±0.0	53.2±0.0	72.7±0.0	58.6±0.0	59.5
COVID-CT-Mask-Net [40]	6.3M	54.9±0.0	82.4±0.0	68.7	51.3±0.0	58.1±0.0	72.9±0.0	68.9±0.0	62.8
COVIDNet [2]	48.7M	75.8±1.8	89.9±1.1	82.9	56.1±4.7	54.8±2.6	49.7±12.5	42.2±1.1	50.7
Ali et.al. [5]	11.2M	83.4±0.0	92.9±0.0	88.2	47.7±0.0	59.5±0.0	66.4±0.0	53.8±0.0	56.8
Ours-SwinTransformer-Tiny	27.5M	77.7±0.0	95.0±0.0	86.4	54.9±0.0	59.7±0.0	95.4±0.0	66.0±0.0	69.0
Ours-SwinTransformer-Small	48.8M	78.8±0.0	96.6±0.0	87.7	54.9±0.0	61.1±0.0	97.2±0.0	62.0±0.0	68.8
Ours-ResNet-18	11.2M	78.4±0.5	97.8±0.2	88.1	56.0±1.5	80.5±0.5	90.6±0.5	66.7±2.4	73.5
Ours-ResNet-101	42.5M	80.4±1.6	97.8±0.5	89.1	54.7±1.7	78.1±2.2	91.3±0.9	68.5±1.6	73.2
									81.2

TABLE II

THE IN-DOMAIN AND OUT-OF-DOMAIN ACCURACY OF DIFFERENT METHODS COMBINED WITH OUR METHOD ON COVID-19 DATASET UNDER RESNET-18 ARCHITECTURE. “*” DENOTES THAT THE STATE-OF-ART METHODS ARE COMBINED WITH OUR METHOD. EACH COLUMN DENOTES THE TEST DOMAIN. “AVG.” DENOTES THE AVERAGE ACCURACY UNDER THE CURRENT SETTING. “JOINT AVG.” DENOTES THE AVERAGE ACCURACY OF IN-DOMAIN AND OUT-OF-DOMAIN. “(+a)” DENOTES THE INCREASED PERFORMANCE ACHIEVED BY OUR METHOD

Method	In-domain			Out-of-domain					Joint avg.
	Domain A	Domain B	Avg.	Domain A	Domain B	Domain C	Domain D	Avg.	
Ali et.al. [5]	83.40.0	92.90.0	88.2	47.70.0	59.50.0	66.40.0	53.80.0	56.8	72.5
Ali et.al. *	85.50.0	94.20.0	89.8 (+1.6)	40.90.0	62.90.0	72.20.0	53.90.0	57.5 (+0.7)	73.7 (+1.2)
SagNet [41]	73.41.7	96.70.7	85.1	53.71.7	69.41.8	89.61.8	71.91.5	71.1	78.1
SagNet *	76.44.8	96.20.3	86.3 (+1.2)	53.61.1	68.91.8	90.81.4	75.00.1	72.1 (+1.0)	79.2 (+1.1)
EISNet [11]	75.52.5	97.20.3	86.3	54.20.3	71.81.0	90.60.8	66.94.8	70.9	78.6
EISNet *	77.20.7	97.10.3	87.2 (+0.9)	54.64.3	74.84.3	90.92.2	67.92.4	72.0 (+1.1)	79.6 (+1.0)
RSC [42]	75.63.2	98.00.2	86.8	56.80.8	75.32.4	93.31.1	71.51.1	74.2	80.5
RSC *	76.90.8	97.50.2	87.2 (+0.4)	57.32.3	75.61.9	93.71.6	73.71.7	75.1 (+0.9)	81.2 (+0.7)

proposed method is generalized to both convolution and transformer backbones. And applying our method to convolution-based backbones achieves better performance than transformer-based backbones. This is because transformer-based backbones are easy to overfit the small-scale labeled data [54], and it is difficult to collect large-scale COVID-19 datasets.

D. Evaluation on Different Methods Plus Our Framework

In order to evaluate that our method is a generalized plug-and-play method that can be easily applied to current methods, we compare the performance of current in-domain and out-of-domain approaches combined with our method in Table II. Specifically, we choose one in-domain methods i.e. Ali et al. [5] and three out-of-domain methods including: SagNet [41], EISNet [11] and RSC [42]. From Table II, we find that our proposed method improves the performance of both in-domain and out-of-domain methods on COVID-19 datasets. For the in-domain method, combined with our approach, Ali et al. [5] achieves the 1.2% improvement on joint average accuracy. Note that, our method improves the accuracy of Ali et al. [5] on both the in-domain and the out-of-domain setting, i.e., 1.6% and 0.7% respectively. For the out-of-domain methods, our method also achieves considerable improvement, i.e., achieving 81.2% accuracy, when integrated into RSC [42]. Hence, our method can consistently improve the performance of COVID-19 diagnosis for omni-domain.

E. Comparison With Data Augmentation Methods

Table III shows the classification accuracies of our methods and other data augmentation methods. Our framework achieves the best joint average accuracy among Mixup [43] and CutMix [44]. Notably, we can see that our method brings a significant improvement from Mixup on in-domain (+1.5%). In an out-of-domain setting, we observe that our method attains slightly lower performance than CutMix (73.5% vs. 73.8%) since CutMix can preserve more lesion details in lung CT images, which could be especially helpful in COCID-19 diagnosis.

F. Ablation Study

We conduct ablation experiments to show how each module in our framework design contributes to the diagnosis performance of COVID-19. The baseline model refers to [51].

1) *Effectiveness of Different Modules:* Table IV reports the effectiveness of each our proposed module, i.e., the spectrum transform module and the cross-domain stylization module. We find that the model trained with spectrum-transformed images outperforms the baseline over 2.3% and 2.7% on in-domain and out-of-domain, respectively. Meanwhile, adding the cross-domain stylization module to the baseline achieves 1.7% and 4.0% improvement on in-domain and out-of-domain, respectively. The results show that the spectrum transform module improves the performance of the in-domain more than the out-of-domain performance while the stylization module contributes more to the out-of-domain data. This is because the

TABLE III
COMPARISON OF THE IN-DOMAIN AND OUT-OF-DOMAIN ACCURACY OF DIFFERENT DATA AUGMENTATION METHODS ON THE COVID-19 DATASET

Method	In-domain			Out-of-domain					Joint avg.
	Domain A	Domain B	Avg.	Domain A	Domain B	Domain C	Domain D	Avg.	
Baseline	72.9±1.2	97.2±0.1	85.0	50.9±0.8	75.0±2.2	88.7±0.5	60.9±1.1	68.9	77.0
Mixup [43]	75.5±2.4	97.8±0.3	86.6	52.0±1.3	77.8±0.5	91.6±0.6	67.4±1.1	72.2	79.4
CutMix [44]	75.3±1.7	97.1±0.3	86.2	56.8±1.8	79.9±1.0	90.3±1.1	68.2±1.5	73.8	80.0
Ours	78.4±0.5	97.8±0.2	88.1	56.0±1.5	80.5±0.5	90.6±0.5	66.7±2.4	73.5	80.8

TABLE IV
ABLATION STUDY ON WHERE TO APPLY OUR FRAMEWORK IN THE RESNET-18 ARCHITECTURE. “SPECTRUMTRANS” DENOTES THE SPECTRUM TRANSFORM MODULE, I.E., THE LOWER BRANCH; “STYLIZATION” DENOTES THE CROSS-DOMAIN STYLIZATION MODULE, I.E., THE UPPER BRANCH

Method	In-domain			Out-of-domain					Joint avg.
	Domain A	Domain B	Avg.	Domain A	Domain B	Domain C	Domain D	Avg.	
Baseline	72.9±1.2	97.2±0.1	85.0	50.9±0.8	75.0±2.2	88.7±0.5	60.9±1.1	68.9	77.0
+ SpectrumTrans.	77.0±1.6	97.6±0.3	87.3	57.0±1.0	76.8±0.4	90.1±1.3	62.4±0.8	71.6	79.5
+ Stylization	76.3±0.8	97.2±0.2	86.7	55.4±1.0	81.1±1.2	91.3±0.2	63.7±1.4	72.9	79.8
Ours	78.4±0.5	97.8±0.2	88.1	56.0±1.5	80.5±0.5	90.6±0.5	66.7±2.4	73.5	80.8

TABLE V
ABLATION STUDY ON WHERE TO APPLY REVISION NETWORK MODULE IN THE RESNET-18 ARCHITECTURE

Method	In-domain			Out-of-domain					Joint avg.
	Domain A	Domain B	Avg.	Domain A	Domain B	Domain C	Domain D	Avg.	
AdaIn	75.1±1.4	97.3±0.3	86.2	53.7±2.7	78.4±1.1	91.5±0.8	62.1±0.1	71.4	78.8
AdaIn + Revision	76.3±0.8	97.2±0.2	86.7	55.4±1.0	81.1±1.2	91.3±0.2	63.7±1.4	72.9	79.8

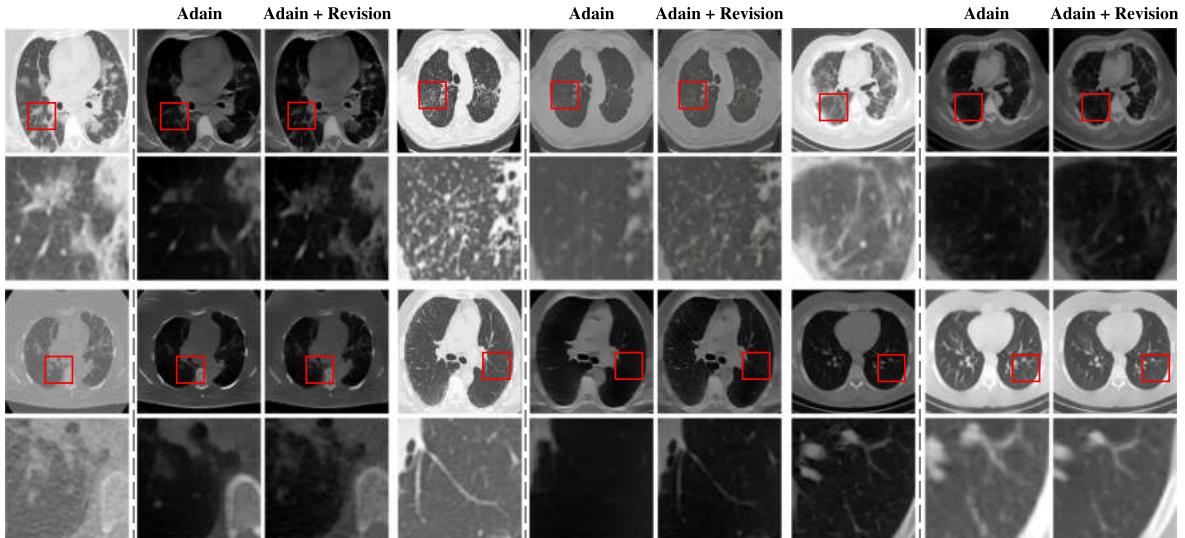


Fig. 6. Generated images based on different stylization methods. “AdaIn” denotes the original AdaIn method; “AdaIn + Revision” denotes the proposed cross-domain stylization module. The enlarged regions of the red boxes are for better visualization.

spectrum transform module aims to discover the discriminating features of each domain to enhance the in-domain performance, while the stylization module transfers the style features of the cross-domain to enhance the model generalization capability for out-of-domain. The visualization analysis is shown in Fig. 2.

2) *Effectiveness of the Revision Network:* To evaluate the effectiveness of the revision network for stylization, we conduct the ablation study in Table V. The results show that the original AdaIn only achieves 71.4% average accuracy on out-of-domain.

Using the revision module improves the performance of the basic AdaIn on the out-of-domain, i.e., achieving 72.9% average accuracy, 1.5% over the basic AdaIn. We also compare the visualization results of the basic AdaIn and the revision module (shown in Fig. 6). It is clear that the images stylized by AdaIn would reduce the lesion details, which can be seen in red boxes in Fig. 6. After adding the revision module, the model generates more accurate stylized images, i.e., preserving the lesions on the original images. Furthermore, combined with

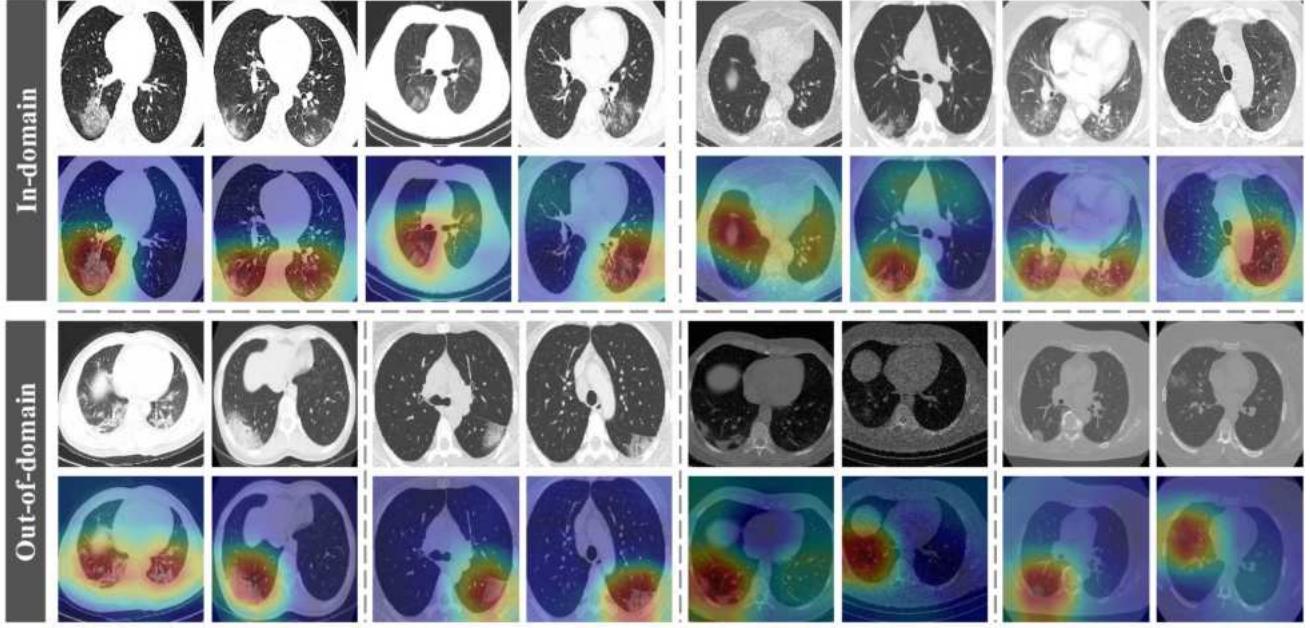


Fig. 7. Grand-CAM visualization of the results produced from our framework.

TABLE VI
ABLATION STUDY OF THE SPECTRUM RADIUS r

r	1/32	2/32	3/32	4/32	5/32
In-domain Acc.	87.7	87.5	88.1	87.8	87.4
Out-of-domain Acc.	70.7	72.8	73.5	71.7	71.3
Joint Acc.	79.2	80.2	80.8	79.8	79.3

the revision module, the model can generate images that contain vessel details. In contrast, the stylized images generated by AdaIn contain fewer vessel details.

3) *Ablation Study on Different Parameters*: In this section, we explore the effectiveness of different hyper-parameters for the diagnosis of COVID-19.

The effectiveness of different spectrum radius r . We also ablate the effectiveness of different spectrum radius r (defined in Eq. 1) in Table VI. We can observe that too large and too small r would hurt the performance. This is because the generated images with too large r lose the high-level details, and too small r reserves the original low-level features, which contributes less to COVID-19 diagnosis. In the experiment, we find that setting r as $(H + W) * (3/32)$ achieves the best performance. We also visualize the generated images with different spectral radius in Fig. 8. It is clear that the middle image retains the original details and the new low-level features, i.e., edges and noises.

The effectiveness of different spectrum probabilities p_{spec} . Table VII reports the results of different spectrum probabilities p_{spec} (defined in Algorithm 1, line: 19). We observe that small p_{spec} , e.g., 0.0 and 0.2, obtains lower in-domain performance and higher out-of-domain accuracy compared with the large one, e.g., 0.8 and 1.0. This indicates that the spectrum-transformed images contribute more to in-domain diagnosis, while the stylized image contributes more to the out-of-domain setting. To obtain the best trade-off between the in-domain and

TABLE VII
ABLATION STUDY OF THE SPECTRUM PROBABILITIES p_{spec}

p_{spec}	0.0	0.2	0.4	0.5	0.6	0.8	1.0
In-domain Acc.	86.7	86.5	88.2	88.4	87.9	87.8	87.3
Out-of-domain Acc.	72.7	72.8	71.3	72.2	71.7	71.9	71.6
Joint Acc.	79.8	79.6	80.0	80.3	79.6	79.9	79.5

TABLE VIII
ABLATION STUDY OF THE SELECTORS. COMB. $a\%$ DENOTES THAT EACH IMAGE FROM THE SOURCE DOMAIN IS COMBINED WITH A GENERATED IMAGE WITH A CERTAIN PROBABILITY OF $a\%$. REP. $a\%$ DENOTES THE SOURCE IMAGES ARE REPLACED BY THE GENERATED IMAGES, WHICH IS USED IN OUR FRAMEWORK

Type	Baseline	Comb. 100%	Comb. 10%	Rep. 10%
In-domain Acc.	85.0	86.0	87.2	88.1
Out-of-domain Acc.	68.9	72.1	72.6	73.5
Joint Acc.	77.0	79.1	79.9	80.8

the out-of-domain, we set p_{spec} as 0.5, which achieves the best joint accuracy of 80.3%.

The effectiveness of different selectors. Table VIII includes an ablation of selector types and the effect it has on COVID-19 dataset accuracy. Equipped with 100% generated images, the accuracy improves by 2.1%. By selecting fewer samples from the generated images, we obtain the joint accuracy of 79.9%. Following [55], we replace each original image with a generated image with a certain probability of 0.1, and achieve the best performance both on in-domain and out-of-domain, demonstrating the effectiveness of the proposed selector. This is likely due to it can not only boost predictions but also provide their relevant uncertainty.

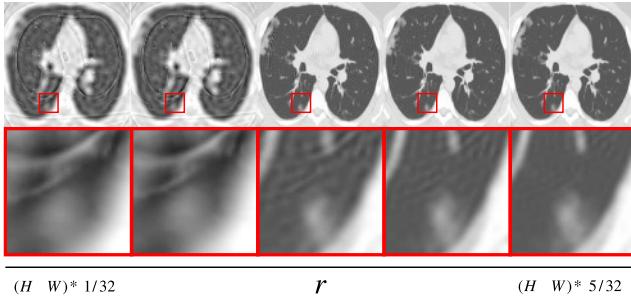


Fig. 8. Generated images based on different spectrum radius. The spectrum radius r is $(H \cdot W) * 1/32$ to $(H \cdot W) * 5/32$ in Eq. (1). The enlarged regions of the red boxes are for better visualization.

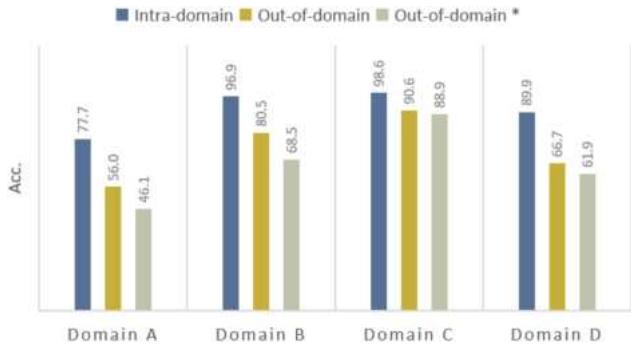


Fig. 9. Results obtained with our classifiers trained with different settings. Intra-Domain denotes that the source domain and target domain share the same data distribution and vice versa. * denotes that we train the source domains with the same data number.

V. DISCUSSION

In this section, we will explore three issues: 1) whether our model can capture the lesion regions, 2) data imbalance issues between the source domains, and 3) the limitations of our work.

Grand-CAM visualization. We generate Grand-CAM [56] visualization images in Fig. 7 to demonstrate that our method can focus on the lesion regions of the images. Furthermore, our model successfully finds the lesions in different regions of images. Take the fifth column and the sixth column of the in-domain row as an example, our model can focus on the lesions in both the upper and lower parts of the lungs. We also observe that our approach can localize multiple lesions on the same image; see examples in the second column of the in-domain row and the first column of the out-of-domain row.

Data imbalance issue. It is interesting to note that the performance of out-of-domain could be rather sensitive to the imbalanced sample numbers between different domains. To get more insights into this issue, we train the model using the same samples from each source domain and test on the left domain. Fig. 9 reports numerical results obtained. We can still observe that Domain A reaches the worst accuracy. This is due to the differences between test domains, as Domain A has the worst image quality, which is collected from published websites and papers.

Limitations. Although the proposed spectrum and style transformation framework achieves satisfactory performance in the experiments, there is still a limitation in our framework design. The branch is dependent on the domain characteristics, e.g.,

styles and noises, which show different appearances between domains. If the domains have divergence features out of the introduced style and noise, we should revise our data augmentation methods. In our future work, we will explore an adaptive strategy to automatically learn the proper data augmentation methods.

VI. CONCLUSION

This paper formulates a spectrum and style transformation framework for omni-domain COVID-19 diagnosis, which improves the diagnosis performance of both in-domain and out-of-domain data. Two new techniques, the spectrum transform module, and the cross-domain stylization module, are presented to improve the capability of deep models to deal with in-domain and out-of-domain data, respectively. The spectrum transform module incorporates the high-frequency information of other domains to augment training data, thus benefiting to discover the discriminating features of each domain to recognize the in-domain data. The cross-domain stylization module transfers the style features from one domain to another domain, thus creating more out-of-domain data to learn the cross-domain knowledge and enhance the model generalization capability. In the end, we compare our method with various methods for COVID-19 diagnosis on four datasets and show the effectiveness of the proposed method on both in-domain and out-of-domain data. In the future, we will explore our method for other CT-based disease diagnosis tasks and integrate our framework into the medical diagnostics systems.

REFERENCES

- [1] S. Bharati, P. Podder, M. Mondal, and V. Prasath, “Medical imaging with deep learning for COVID-19 diagnosis: A comprehensive review,” *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, pp. 91–112, 2021, doi: [10.48550/arXiv.2107.09602](https://arxiv.org/abs/2107.09602).
- [2] Z. Wang, Q. Liu, and Q. Dou, “Contrastive cross-site learning with redesigned net for COVID-19 CT classification,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2806–2813, Oct. 2020.
- [3] X. Xu et al., “A deep learning system to screen novel coronavirus disease 2019 pneumonia,” *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.
- [4] T. D. Pham, “A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.
- [5] A. Ali, T. Shaharabany, and L. Wolf, “Explainability guided multi-site COVID-19 CT classification,” 2021, [arXiv:2103.13677](https://arxiv.org/abs/2103.13677).
- [6] D. Di et al., “Hypergraph learning for identification of COVID-19 with CT imaging,” *Med. Image Anal.*, vol. 68, 2021, Art. no. 101910.
- [7] K. Hu et al., “Deep supervised learning using self-adaptive auxiliary loss for COVID-19 diagnosis from imbalanced CT images,” *Neurocomputing*, vol. 458, pp. 232–245, 2021.
- [8] X. Song et al., “Augmented multicenter graph convolutional network for COVID-19 diagnosis,” *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 6499–6509, Sep. 2021.
- [9] C. Li et al., “Domain generalization on medical imaging classification using episodic training with task augmentation,” *Comput. Biol. Med.*, vol. 141, 2022, Art. no. 105144.
- [10] O. Nuriel, S. Benaim, and L. Wolf, “Permuted adain: Reducing the bias towards global statistics in image classification,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9482–9491.
- [11] S. Wang, L. Yu, C. Li, C.-W. Fu, and P.-A. Heng, “Learning from extrinsic and intrinsic supervisions for domain generalization,” in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 159–176.
- [12] X. Ma, B. Zheng, Y. Zhu, F. Yu, R. Zhang, and B. Chen, “COVID-19 lesion discrimination and localization network based on multi-receptive field attention module on CT images,” *Optik*, vol. 241, 2021, Art. no. 167100.
- [13] X. Wang et al., “A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT,” *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2615–2625, Aug. 2020.

- [14] L. Wang, Z. Q. Lin, and A. Wong, "COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.
- [15] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of COVID-19 cases from x-ray images," *Pattern Recognit. Lett.*, vol. 138, pp. 638–643, 2020.
- [16] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest x-ray image classification," *Appl. Intell.*, vol. 51, no. 5, pp. 2850–2863, 2021.
- [17] E.-S. M. El-Kenawy et al., "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 x-ray chest image classification," *IEEE Access*, vol. 9, pp. 36019–36037, 2021.
- [18] P. Saha, D. Mukherjee, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, "Retracted article: Graphcovidnet: A graph neural network based model for detecting COVID-19 from CT scans and X-rays of chest," *Sci. Rep.*, vol. 11, no. 1, pp. 1–16, 2021.
- [19] X. Yao, Z. Zhu, C. Kang, S.-H. Wang, J. M. Gorri, and Y.-D. Zhang, "AdaD-FNN for chest CT-based COVID-19 diagnosis," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Jun. 01, 2022, doi: [10.1109/TETCI2022.3174868](https://doi.org/10.1109/TETCI2022.3174868).
- [20] Z. Li et al., "A novel deep learning framework based mask-guided attention mechanism for distant metastasis prediction of lung cancer," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, May 17, 2022, doi: [10.1109/TETCI2022.3171311](https://doi.org/10.1109/TETCI2022.3171311).
- [21] X. Ouyang et al., "Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2595–2605, Aug. 2020.
- [22] J. Wang et al., "Prior-attention residual learning for more discriminative COVID-19 screening in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2572–2583, Aug. 2020.
- [23] A. Elazab, M. Abd Elfattah, and Y. Zhang, "Novel multi-site graph convolutional network with supervision mechanism for COVID-19 diagnosis from X-ray radiographs," *Appl. Soft Comput.*, vol. 114, 2022, Art. no. 108041.
- [24] A. Zhong et al., "Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101993.
- [25] O. Frank et al., "Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19," *IEEE Trans. Med. Imag.*, vol. 41, no. 3, pp. 571–581, Mar. 2022.
- [26] C. Sitaula, T. B. Shahi, S. Aryal, and F. Marzbanrad, "Fusion of multi-scale bag of deep visual words features of chest X-ray images to detect COVID-19 infection," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021.
- [27] C. Sitaula and S. Aryal, "New bag of deep visual words based features to classify chest x-ray images for COVID-19 diagnosis," *Health Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–12, 2021.
- [28] O. Frank et al., "Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19," *IEEE Trans. Med. Imag.*, vol. 41, no. 3, pp. 571–581, Mar. 2022.
- [29] G.-X. Xu et al., "Cross-site severity assessment of COVID-19 from CT images via domain adaptation," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 88–102, Jan. 2022.
- [30] Q. Jin, H. Cui, C. Sun, Z. Meng, L. Wei, and R. Su, "Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images," *Expert Syst. with Appl.*, vol. 176, 2021, Art. no. 114848.
- [31] Q. Liu, Q. Dou, and P.-A. Heng, "Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 475–485.
- [32] D. Yang et al., "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101992.
- [33] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," 2018, [arXiv:1811.12231](https://arxiv.org/abs/1811.12231).
- [34] K. Hermann, T. Chen, and S. Kornblith, "The origins and prevalence of texture bias in convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 19000–19015, 2020.
- [35] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1501–1510.
- [36] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2414–2423.
- [37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [38] T. Lin et al., "Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 5141–5150.
- [39] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [40] A. Ter-Sarkisov, "COVID-CT-mask-net: Prediction of COVID-19 from CT scans using regional features," *Appl. Intell.*, vol. 52, pp. 9664–9675, 2022.
- [41] H. Nam, H. Lee, J. Park, W. Yoon, and D. Yoo, "Reducing domain gap by reducing style bias," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8690–8699.
- [42] Z. Huang, H. Wang, E. P. Xing, and D. Huang, "Self-challenging improves cross-domain generalization," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 124–140.
- [43] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "MIXUP: Beyond empirical risk minimization," 2017, [arXiv:1710.09412](https://arxiv.org/abs/1710.09412).
- [44] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6023–6032.
- [45] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-CT-dataset: A CT scan dataset about COVID-19," 2020, [arXiv:2003.13865](https://arxiv.org/abs/2003.13865).
- [46] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, "SARS-COV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-COV-2 identification," *MedRxiv*, 2020, doi: [10.1101/2020.04.24.20078584](https://doi.org/10.1101/2020.04.24.20078584).
- [47] M. Rahimzadeh, A. Attar, and S. M. Sakhaei, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset," *Biomed. Signal Process. Control*, vol. 68, 2021, Art. no. 102588.
- [48] S. Morozov et al., "Mosmeddata: Chest CT scans with COVID-19 related findings dataset," 2020, [arXiv:2005.06465](https://arxiv.org/abs/2005.06465).
- [49] A. Paszke, S. Gross, and S. Chintala, "Pytorch deep learning framework," Web, 2017. [Online]. Available: <http://pytorch.org/>
- [50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [53] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.
- [54] S. Wang, A. Raju, and J. Huang, "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 620–623.
- [55] N. Somavarapu, C.-Y. Ma, and Z. Kira, "Frustratingly simple domain generalization via image stylization," 2020, [arXiv:2006.11207](https://arxiv.org/abs/2006.11207).
- [56] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.

Zhenkun Wang (Member, IEEE) received the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2016. He is currently an Assistant Professor with the School of System Design and Intelligent Manufacturing, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. His research interests include computational intelligence, optimization, machine learning, and their applications.



Shuangchun Gui received the B.S. degree from the Anhui Normal University, Wuhu, China, in 2021. She is currently a Postgraduate Student with the School of System Design and Intelligent Manufacturing, Southern University of Science and Technology, Shenzhen, China. Her research interests include computer vision, medical image analysis, machine learning, and their applications.





Xinpeng Ding received the B.E. degree in software engineering and the M.S. degree in information and telecommunications engineering from Xidian University, Xi'an, China, in 2018 and 2021, respectively. He is currently working toward the Ph.D. degree with the Hong Kong University of Science and Technology, Hong Kong. His research interests include computer vision, pattern recognition, medical image analysis, and machine learning.



VLSI journal in 2017.

Xiaowei Xu (Member, IEEE) received the B.S. and Ph.D. degrees in electronic science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2016 respectively. He is currently an Associate Professor with Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangzhou, China. His research interests include deep learning and medical image segmentation. He was the recipient of DAC System Design Contest Special Service Recognition Reward in 2018 and outstanding contribution in reviewing the



Xiaowei Hu received the B.Eng. degree in computer science and technology from the South China University of Technology, Guangzhou, China, in 2016 and the Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong, Hong Kong, in 2020. He is currently a Research Scientist with the Shanghai AI Laboratory. His research interests include computer vision, deep learning, and low-level vision.



Xiaomeng Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong, Hong Kong, in 2019. She is currently an Assistant Professor with the Hong Kong University of Science and Technology, Hong Kong. Her research interests include medical image analysis, label-efficient learning, and robust deep learning algorithms.