COVID-19 Screening in Chest X-Ray Images Using Lung Region Priors

Jianpeng An, Qing Cai, Zhiyong Qu, and Zhongke Gao, *Senior Member, IEEE*

***Abstract*—Early screening of COVID-19 is essential for pandemic control, and thus to relieve stress on the health care system. Lung segmentation from chest X-ray (CXR) is a promising method for early diagnoses of pulmonary diseases. Recently, deep learning has achieved great success in supervised lung segmentation. However, how to effectively utilize the lung region in screening COVID-19 still remains a challenge due to domain shift and lack of manual pixel-level annotations. We hereby propose a multi-appearance COVID-19 screening framework by using lung region priors derived from CXR images. Firstly, we propose a multi-scale adversarial domain adaptation network (MS-AdaNet) to boost the cross-domain lung segmentation task as the prior knowledge to the classification network. Then, we construct a multi-appearance network (MA-Net), which is composed of three sub-networks to realize multi-appearance feature extraction and fusion using lung region priors. At last, we can obtain prediction results from normal, viral pneumonia, and COVID-19 using the proposed MA-Net. We extend the proposed MS-AdaNet for lung segmentation task on three different public CXR datasets. The results suggest that the MS-AdaNet outperforms contrastive methods in cross-domain lung segmentation. Moreover, experiments reveal that the proposed MA-Net achieves accuracy of 98.83% and F1-score of 98.71% on COVID-19 screening. The results indicate that the proposed MA-Net can obtain significant performance on COVID-19 screening.**

***Index Terms*—Unsupervised domain adaptation, lung segmentation, multi-appearance, chest X-ray image, COVID-19.**

I. Introduction

PULMONARY disease is one of the leading causes of morbidity and mortality throughout the world. Especially the ongoing global pandemic of COVID-19 [1], caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused more than three million deaths and severe harm to the world's economy. In spite of such high fatality and social burden, the diagnostic tests for COVID-19 are far from satisfactory. The rapid and accurate screening of patients plays an important role in controlling the outbreak and relieving the pressure on the medical system due to the strong infectivity of COVID-19.

Fig. 1. t-SNE visualization of the domain shift problems where the three datasets are collected from different hospitals.

As CXR is cheaper and faster to obtain than computed tomography (CT) scan, CXR has become one of the most regularly used medical image to perform early-stage screening of patients with acute respiratory distress syndrome (ARDS) and currently with COVID-19 symptoms. However, manual reading of CXR images is expertise-required, time-consuming, and error-prone. With the fast development of deep learning (DL) methods in medical image analysis [2] and physiological signals recognition [3], an emerging trend is to employ DL methods to assist radiologists and other physicians in reading and comprehending CXRs.

Lung segmentation is a fundamental task to assist clinical diagnosis owing to the abundant image-based structural information, such as shape irregularities, size measurements, and the volume of the lungs provided by it. The provided information can be used to facilitate disease pattern recognition, segmentation, quantization, and further be employed to analyze biomarkers for lung diseases, including COVID-19 [4].

While lung segmentation is of great clinical importance, two main problems hinder the performance and development of lung segmentation. One is domain shift problem, which is exacerbated by the under-representation of some rare medical conditions, varying imaging configurations, modalities, and other factors. As shown in Fig. 1, the source domain dataset cannot be well generalized to the target domain dataset due to the different data distribution. Another problem is that manual annotations of CXR images by radiologists would impose enormous strains on medical resources. Accordingly, there is a strong motivation to

Manuscript received December 2, 2020; revised June 3, 2021 and July 1, 2021; accepted August 5, 2021. Date of publication August 13, 2021; date of current version November 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61922062 and 61873181. *(Corresponding author: Zhongke Gao.)*

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: anjianpeng@tju.edu.cn; caiqing@tju.edu.cn; 13043267067@163.com; zhongkegao@tju.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3104629

learn a lung segmentation model on the source domain data with pixel-level labels and obtain good performance on the unlabeled target domain data to realize unsupervised adaptation.

In this paper, we propose a framework by leveraging unsupervised lung segmentation results as prior knowledge and further develop a DL network for the early-stage screening of patients who are suspected to be infected with COVID-19. Specifically, we firstly develop a multi-scale adversarial domain adaptation network (MS-AdaNet) for lung segmentation. The proposed MS-AdaNet combines the concepts of unsupervised domain adaptation (UDA) and adversarial learning to force the distribution of the target domain closer to the source domain. Then with the segmentation results, we build a multi-appearance network (MA-Net) ensemble of three models for multi-appearance feature extraction of the Original CXR (OC), Heterogeneity in Lung Shapes (HLS), and Heterogeneity in Pixel Values (HPV), respectively. At last, the network fuses the multi-appearance features at the decision level for screening COVID-19 in CXR images.

To conclude, the main contributions of this paper are:

1) We propose a lung region-based COVID-19 screening framework in CXR. Via the fusion of multi-appearance CXRs, our proposed framework conducts the combined model to perform COVID-19 screening task.

2) To overcome the domain shift and unsupervised problems, a multi-scale feature adversarial learning scheme called MS-AdaNet is developed to mitigate the domain gap for lung segmentation from different datasets without annotations.

3) Through extensive experiments for COVID-19 screening based on multi-appearance CXRs, we demonstrate that lung region prior knowledge can substantially improve the classification accuracy on COVID-19 datasets. And the results further provide an explainable conclusion to assist clinical diagnosis.

In the following sections, we introduce the related works involving image segmentation, domain adaptation and COVID-19 classification in Section II. In Section III we discuss our framework in details. We provide the implementation details and conduct extensive experiments in Section IV and Section V. Finally, we conclude this paper in Section VI.

II. Related Works

In this section, we briefly review three types of works that are related to the proposed lung region-based COVID-19 screening framework: (1) DL in image segmentation, (2) domain adaptation in medical image analysis, (3) DL in COVID-19 classification.

Deep Learning in Image Segmentation. DL in image segmentation, focusing on computer vision and medical imaging, has become one of the most popular research in recent years. Litjens *et al.* [5] have surveyed the DL methods applied to medical image segmentation. Specially, the Fully Convolutional Network (FCN) proposed by Long *et al.* [6] has become a powerful and promising framework by replacing the fully connected layers in the classification networks with upscaling layers. Many works [7], [8] developed for organ segmentation in medical images, were inspired by the end-to-end FCN model. Such work has also combined with a convolutional network (e.g. ResNet [9] and DenseNet [10]) for efficient prostate segmentation [11]. With the dramatic development of FCN in medical image processing, Ronneberger *et al.* [12] proposed a novel U-Net framework for semantic segmentation in medical images. U-Net combined the FCN architecture with skip connections between encoder and decoder layers. Many networks were proposed subsequently based on U-Net by introducing nested and dense skip connections [13] or adding densely connected blocks [14]. In this work, we employ U-Net with some careful adjustments as our segmentation network.

Domain Adaptation in Medical Image Analysis. Domain shift is a long-standing problem in medical image analysis due to the under-representation of some rare medical conditions, varying imaging configurations and other factors. Unsupervised domain adaptation has been proved to be effective in addressing domain shift problem in medical image analysis [15], [16], as well as challenging and yet to be explored. Most approaches focus on minimizing measures of distances between the source domain and the target domain features [17]. Recent works follow a similar spirit via adversarial learning, including image-to-image translation [18], feature adaptation [19], and cross-modality segmentation [20] with the mixtures of image-to-image translation and feature adaptation [21]. Moreover, Zhang *et al.* [22] proposed a collaborative UDA method for medical image diagnosis and evaluated on different datasets including COVID-19 to prove its generalization and superiority.

Deep Learning in COVID-19 Classification. Due to the strain on medical system caused by the outbreak of COVID-19 pandemic, many DL methods have been developed to assist radiologists in screening pneumonia patients, including COVID-19 patients on radiology images [23]–[26]. Especially on CXRs, Wang *et al.* [27] proposed a three-stage nodule detection method. They trained the cost-sensitive random forest for lung nodule classification based on the lung segmentation. Minaee *et al.* [28] trained four convolutional networks using transfer learning on CXRs for COVID-19 classification. Nour *et al.* [29] proposed a convolutional neural network which can be used to assist the radiologists in detecting positive COVID-19 cases. Wang *et al.* [30] proposed a computer-aided detection system with two steps: lung feature extraction and local patches training. They claimed that the system can efficiently and accurately screen COVID-19 cases.

Although researchers have made these efforts for the classification of COVID-19 on CXR images, there is a lack of extending the use of lung region and multi-appearance images for pneumonia classification. As suggested by Xie el al. [31], using DL models, the lung nodule characteristics can be learned from multi-views and multi-appearances images. Jaeger *et al.* [32] developed an automated CXR screening system. The system segments the lung region in CXR firstly and then computes a set of shape, edge, and texture features as input to a binary classifier to realize tuberculosis screening. Oh *et al.* [4] analyzed the biomarkers in COVID-19 CXR images. They trained a lung segmentation network and investigated the correspondence between lung contours and categorical classes. A pneumonia classification network was trained with four different types of CXRs.

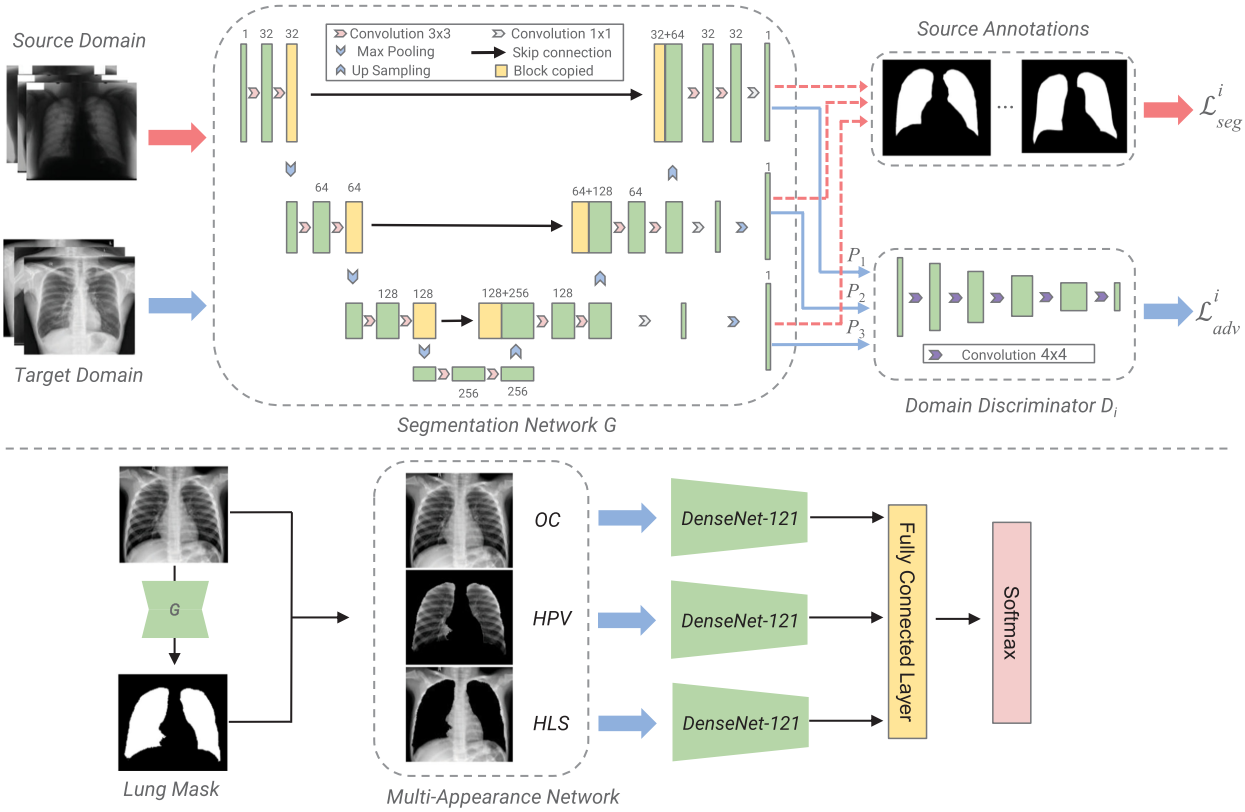

Fig. 2. Overview of our proposed lung region-based COVID-19 screening framework. The framework architecture consists two stages. Step-1: MS-AdaNet. The supervised loss L_{seg} is employed to update the segmentation network as the sum of three segmentation losses. The three discriminators $\{D_1, D_2, D_3\}$ differentiate the input of the three scales respectively to obtain the adversarial loss L_{adv}. Step-2: MA-Net. The lung masks are obtained based on the well-trained network G in step-1, and the multi-appearance CXR images are further generated as the input of the corresponding sub-network to get the final classification results.

Inspired by the above works, we adapt the U-Net with some careful adjustments and introduce a multi-scale adversarial learning strategy for UDA lung segmentation, and then utilize the segmentation results as prior knowledge to generate multi-appearance images for improving the performance on COVID-19 screening.

III. Methods

A. Overview

Our framework aims to attain reliable classification results via a lung region-based multi-appearance network. As illustrated in Fig. 2, the framework consists of two steps: a UDA lung segmentation network, and a multi-appearance pneumonia classification network. In the first step, we apply a multi-scale adversarial learning method to obtain a well-trained lung segmentation network G on source and target domain datasets. Specially, we denote the source dataset as X_s with pixel-level annotations Y_s and the target dataset as X_t without any annotations. In the second step, the lung masks can be attained by the segmentation network G; then the multi-appearance CXR images OC, HLS, and HPV can then be generated from the lung masks; eventually, we train three DL models with three appearances respectively and fuse them at the decision level for COVID-19 screening.

B. Adversarial Learning for Unsupervised Segmentation

Lung segmentation in CXRs is a basic task for computer-aided diagnosis system. However, the domain shift problem in different CXR datasets has caused an enormous challenge in improving lung segmentation results. Some methods [33], [34] applied adversarial learning in high-dimensional feature space to solve the widely existed domain shift problem. Nonetheless, the high-dimensional features, encoding complex representations, are difficult to be adapted directly cross domain for pixel-wise segmentation task. To fully exploit information and achieve better performance in different domains, we utilize the features produced in semantic prediction space of the different decoder levels to exploit object features via an adversarial learning scheme.

Single-scale Training: Firstly, for the segmentation network G training part, it is trained with X_s and the corresponding label Y_s by minimizing a hybrid loss from source domain:

$$L_{seg} = CE\left(Y_s, P_s\right) + Dice\left(Y_s, P_s\right) \tag{1}$$

where $CE(\cdot)$ represents cross-entropy loss, $Dice(\cdot)$ represents dice loss and $P_s = G(X_s)$ is the segmentation output.

Subsequently, in order to reduce the domain gap between P_s and P_t, where $P_t = G(X_t)$, we use the adversarial loss L_{adv} with an inverted domain label ($c = 1$). When the parameters of

discriminator D are frozen, L_{adv} encourages the segmentation network G for learning common feature distributions between P_t and P_s by fooling the discriminator D. It can be defined as:

$$L_{adv}\left(P_t\right) = -\sum \log\left(\mathbf{D}\left(P_t\right)\right) \tag{2}$$

Sencodly, for the discriminator D training part, the fully-convolutional discriminator D with CE loss L_d to classify the segmentation outputs corresponding to P_s or P_t. The discriminator loss can be defined as:

$$L_d(P_s, P_t) = -\sum \left[(1-c)\log\left(\mathbf{D}(P_t)\right) + c\log\left(\mathbf{D}(P_s)\right)\right] \tag{3}$$

where $c \in \{0, 1\}$ specifies the domain label from the target domain or the source domain.

The semantic prediction space contains the information of lung structures. Specially, the distribution between different lung regions should be consistent with typical lung regions. If the lung features extracted from source CXRs are close to target CXRs, the discriminator D would fail to differentiate which domain that the feature belongs to, as the lung regions are consistent.

Multi-Scale Training: The final output layer of the decoder may not be adapted high-dimensional complex representation, for the reason that it is far away from the layer in latent space. Thus, we incorporate multi-level adversarial learning method [35] into our framework to boost the gradients back-propagation to the compact spaces. Accordingly, we extend the hybrid loss in Equation (1) and the adversarial loss in Equation (2) as L^i_{seg} and L^i_{adv}, where i indicates the i-th scale of the decoder used for predicting the segmentation output.

Taken them together, we optimize the segmentation network G in following min-max criterion:

$$\min_{G} \max_{D_1, D_2, D_3} \lambda^i L^i_{seg} + \beta^i L^i_{adv} \tag{4}$$

where the λ^i and β^i with $i = 1, 2, 3$ are trade-off weights. Due to that high-dimensional feature outputs carry less information to predict the segmentation, we empirically set λ^i and β^i as $\{1.0, 0.1, 0.1\}$ and $\{0.1, 0.1, 0.01\}$ respectively.

C. Multi-Appearance Classification

We characterize the original CXR image as the OC appearance. Based on the segmented lung region in OC image, the HLS appearance and the HPV appearance can be obtained by setting pixel values of the lung region to 0 and non-lung region to 0, respectively.

For the MA-Net, the n-th input can be denoted by $X^m_n = S(G(X_n))_m$, where $S(\cdot)$ is the operation to generate the appearance images and $m \in \{OC, HPV, HLS\}$. Then the corresponding output can be denoted by O^m_n. The inference procedure can be formulated as:

$$M_n = f_s\left(\sum_m \left(O^m_n, W^m_{cls} \mid X^m_n\right), W_{fc}\right) \tag{5}$$

where W^m_{cls} and W_{fc} indicate the learned convolutional weights of m-th appearance classification sub-network and the learned weighs of fully-connected layers, respectively. $f_s(\cdot)$ represents the softmax function.

TABLE I
SEGMENTATION DATASETS

Dataset	Class	Number of Images
JSRT	Non-Nodule/Nodule	93/154
MC	Normal/Tuberculosis	80/58
SC	Normal/Tuberculosis	326/336

As COVID-19 infection has similar features with other viral pneumonia diseases, it is difficult for clinicians to diagnose on CXR images [4]. In other words, MA-Net should have strong abilities to deal with the hard samples between 'COVID-19' and 'viral pneumonia'. Motivated by the outstanding performance of focal loss in solving hard samples and class imbalance problems in medical diagnosis, we employ focal loss to balance the weights of different classes. Specially, the focal loss L_{cls} can be defined as:

$$L_{cls} = \sum \alpha \cdot \left(1 - e^{-CE(M_n, y_n)}\right)^{\gamma} \cdot CE\left(M_n, y_n\right) \tag{6}$$

where y_n is the class label, α is a weighting factor and γ is the tunable focusing parameter.

D. Network Architecture

Segmentation Network: To achieve better prediction performance, we employ an encoder-decoder network based on the U-Net [12] model with some careful adjustment as the backbone network. Each encoder stage consists of two convolutional layers, batch normalization, leaky-ReLU activation, and a max-pooling layer. Each decoder stage is constructed with an upsampling layer, two convolution layers, batch normalization and leaky-ReLU activation. Through every stage of the decoder, the output features are collected at the resolutions of $x/4$ and $x/2$ before a 1×1 convolutional layer followed by an upsampling layer to recover to the resolution x. Moreover, multiple discriminators enable the model to receive feedback at semantic prediction space via multi-scale features. The discriminator consists of five 4×4 convolutional layers with stride of 2 and the numbers of feature maps are $\{32, 64, 128, 256, 1\}$.

Classification Network: The multi-appearance images are used to train MA-Net. The MA-Net contains three sub-networks, and each sub-network is associated with one appearance. DenseNet is employed in this work as the backbone model, which including a 7×7 convolutional layer with a 3×3 max-pooling, and 4 dense blocks with a transition layer between two contiguous dense blocks subsequently. We concatenate the output of the last average pooling layer in each sub-networks and then add two extra fully connected layer for three categories classification task.

IV. EXPERIMENTS

A. Datasets

To validate the effectiveness of MS-AdaNet for lung segmentation, we employ three public datasets, summarized in Table I, for cross domain segmentation. And a public dataset is presented in Table II for classification.

TABLE II
CLASSIFICATION DATASET

Dataset	Class	Number of Images
Chowdhury et al. [36]	Normal	1341
	Viral Pneumonia	1345
	COVID-19	219

Lung Segmentation Datasets: We validate our MS-AdaNet using three publicly available datasets released from different sources including lung mask labels: the Japanese Society of Radiological Technology dataset (JSRT, denoted by J) [37] [38], the Montgomery County chest X-Ray set (MC, denoted by M) [32], and the Shenzhen chest X-ray set (SC, denoted by S) [39]. The input CXR images are resized to 256×256 for both source domain and target domain.

Six cross-domain combinations are constructed, i.e., $J \rightarrow M$, $J \rightarrow C$,..., $M \rightarrow C$. We randomly split each dataset into training and testing sets according to an 80%/20% division.

COVID-19 Classification Dataset: We evaluate our MA-Net on a public dataset database [36], which was created to combine several public databases [40]–[42] of normal, viral pneumonia and COVID-19 CXRs. The appearance images of OC and HPV are resized to 1024×1024, and HLS are resized to 256×256. We use 5-fold cross validation for model evaluation.

B. Environment Setup

Evaluation metric: We use the Dice Similarity Coefficient (DSC) to evaluate segmentation performance. DSC is a common choice in unsupervised domain adaptation medical segmentation tasks [15], [21], [43], which can be defined as:

$$DSC(\mathrm{A},\mathrm{B}) = \frac{2 \times |\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A}| + |\mathrm{B}|} \tag{7}$$

where A and B represent the lung region of prediction and ground truth, respectively.

For classification evaluation, accuracy, F1-score, recall and precision are employed, which can be written as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Implementation details: We conduct experiments using the PyTorch framework and all the experiments are performed on two NVIDIA RTX 2080Ti GPUs.

1) **Segmentation Network:** Each cross-domain combination model is trained 500 epochs with a batch size of 4, and all the modules are trained using Adam optimizer with initial learning rate of 2×10^{-4} and decreased using the polynomial decay with power as 0.9 [14].
2) **Classification Network:** We set the batch size as 2 and the maximum epoch number is 100. The min-batch stochastic gradient descent with initial learning rate of 1×10^{-4} is chosen for training. In addition, the hyper-parameters are the same for each appearance sub-network. To avoid over-fitting, we set the patience of early stop scheme is set as 20.

C. Results

Effectiveness of MS-AdaNet: Two comparative experiments are designed to validate the effect of domain adaptation on lung segmentation. We firstly conduct supervised training for the target domain to obtain the 'Supervised' results as the upper bound. Then the 'No-adaptation' lower bound results are obtained by training the segmentation model directly on the source domain and testing it on the target domain. Notably, the segmentation network in MS-AdaNet is employed to obtain the upper and lower bound results. Table III shows the results of lung segmentation in cases of different combinations of source and target domains. It can be observed that the lower bound DSCs of $S \rightarrow M$ and $M \rightarrow S$ are significantly higher than DSCs of other combinations. The reason for the higher DSC may be the similar imaging methods applied in *M* and *S*, as well as more similar domain distributions of *M* and *S*. The average DSCs of the upper and lower bound obtained by the trained models are 95.2% and 35.3% respectively. Our MS-AdaNet achieved average DSC of 90.6%, which is slightly below the 'Supervised' results. The visual comparison in Fig. 3 shows that our method is able to generate reliable semantic lung segmentation.

We compare the performance of our proposed MS-AdaNet with unsupervised domain adaptation methods including CycleGAN [44], MUNIT [45], AdaptSegNet [35], SeUDA [43]. These methods utilize either image-to-image translation, feature adaptation or their mixtures. The first three methods have been validated in natural images. CycleGAN and MUNIT adapted image appearance, and AdaptSegNet employed output space adaptation. SeUDA dedicated to cross-domain lung segmentation using feature adaptation and image-to-image translation. We list the results of different methods in Table III, and the qualitative comparison are shown in Fig. 3. We noticed that the DSCs of CycleGAN and MUNIT for $S \rightarrow M$ and $M \rightarrow S$ were significantly lower than other combinations, indicating that image-to-image translation fails to produce photorealistic results between these two datasets, which is a tedious process. Different from AdaptSegNet, MS-AdaNet maps the high-dimensional features to the output in different scales by the same decoder, which turns out to be more effective for learning the common features between domains. Therefore, our MS-AdaNet achieves the best average DSC, which shows that reducing the domain gap with a simpler adversarial learning strategy is very effective. Furthermore, t-test has indicated the statistical significance of the difference between our method and other methods.

MA-Net for classification: To maximize the mapping of the abnormal lung regions in different pneumonia cases to the segmentation results, we set normal cases in Table I and the training set in Table II as the source domain and target domain in MS-AdaNet for training, respectively. We can then apply the

TABLE III
COMPARISON OF DSC (MEAN ± STANDARD DEVIATION) PERFORMANCE WITH DIFFERENT UDA METHODS FOR LUNG SEGMENTATION. THE ASTERISK INDICATES STATISTICALLY SIGNIFICANT DIFFERENCE ($p < 0.05$)

Method	$J \rightarrow S$	$J \rightarrow M$	$S \rightarrow J$	$S \rightarrow M$	$M \rightarrow S$	$M \rightarrow J$	AVG
Supervised	94.9 ± 3.3*	95.3 ± 5.1*	94.2 ± 4.0*	96.7 ± 3.6*	96.3 ± 4.2*	93.8 ± 3.2*	95.2
No-Adaptation	17.1 ± 11.5*	11.4 ± 5.6*	4.3 ± 6.6*	61.2 ± 9.2*	80.4 ± 5.5*	37.5 ± 4.1*	35.3
CycleGAN [44]	48.3 ± 6.2*	64.1 ± 7.9*	59.1 ± 13.9*	23.4 ± 8.9*	7.7 ± 7.2*	66.9 ± 4.8*	44.9
MUNIT [45]	20.5 ± 9.7*	73.1 ± 3.7*	50.6 ± 10.6*	7.9 ± 8.1*	1.3 ± 4.9*	68.4 ± 11.9*	36.9
AdaptSegNet [35]	78.9 ± 15.3*	58.8 ± 12.2*	79.1 ± 14.0*	81.8 ± 9.5*	86.2 ± 10.6*	66.4 ± 8.1*	75.2
SeUDA [43]	85.7 ± 3.7*	**90.7 ± 5.0***	87.2 ± 6.2*	76.7 ± 5.4*	93.5 ± 4.0*	82.7 ± 8.8*	86.1
MS-AdaNet (Ours)	**88.3 ± 5.3**	88.5 ± 7.4	**87.9 ± 4.9**	**95.8 ± 3.3**	**94.1 ± 6.3**	**89.4 ± 4.7**	**90.6**

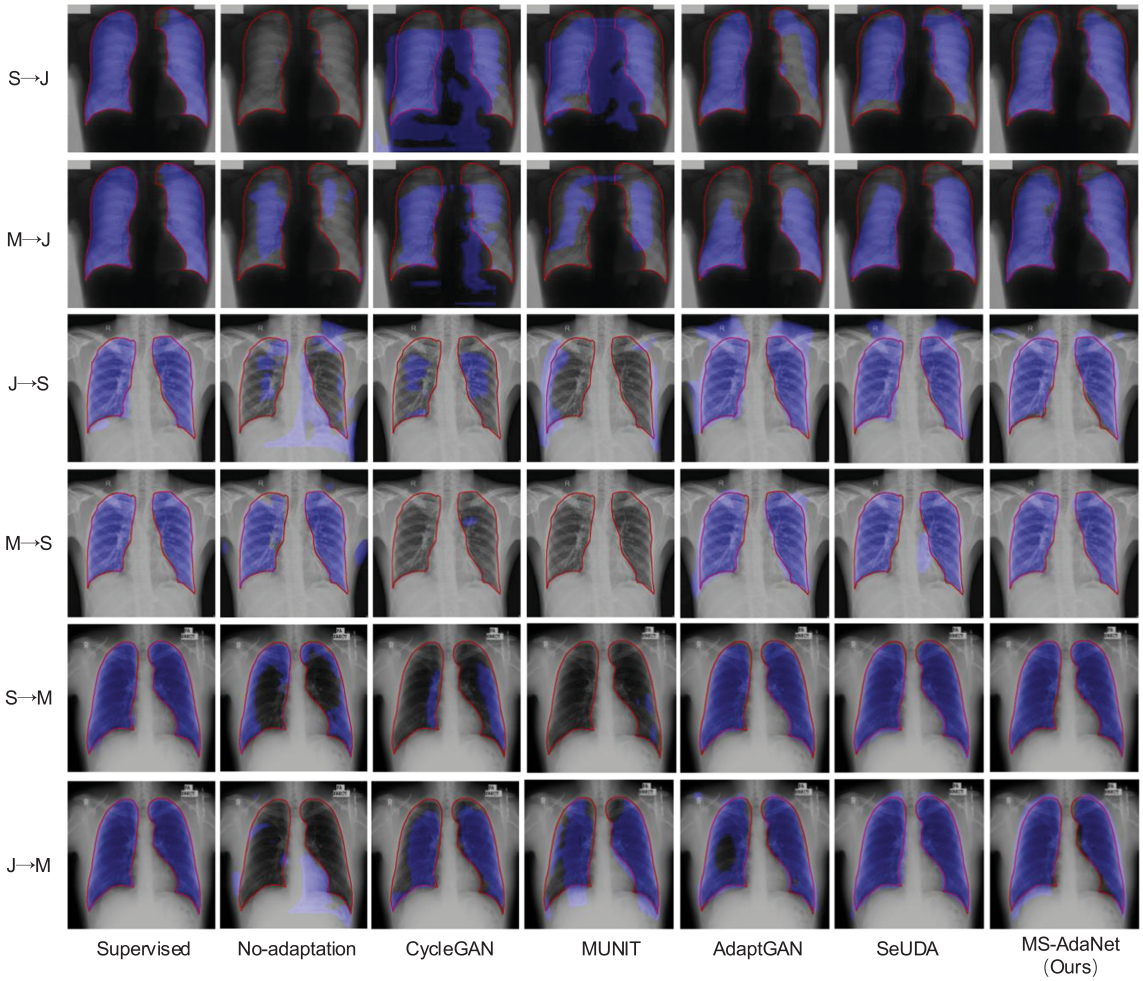

Fig. 3. Visualization of lung segmentation results with advanced approaches for different domain combinations. From left to right are 'Supervised' training upper bound (the first column), 'No-adaptation' lower bound (the second column), results of other methods (the third to sixth column), results of our MS-AdaNet (the last column). The red contours represent the ground truth. The blue-colored fill-in parts are the predictions by different methods. Each row corresponds to an exemplar sample from the target domain test set.

well-trained segmentation network on classification datasets to obtain lung masks. In order to demonstrate the performance of our approach for screening COVID-19, the multi-appearance performance and quantitative results are shown in Fig. 4 and Table IV, respectively. The appearance changes in lung mask caused by different pneumonia diseases are subtle. Due to the opacity of the lungs caused by pneumonia, the lung contours are not obvious in CXRs, as shown in cases of viral pneumonia and COVID-19 in the second and third columns in Fig. 4. More specifically, bilateral consolidations in viral pneumonia results in a deformed lung region [4], while more location of the abnormalities are in the lower lobe in COVID-19 cases [46]. The lung contours are more clearly defined in normal cases in the first column, resulting in better segmentation result. Thus we utilize

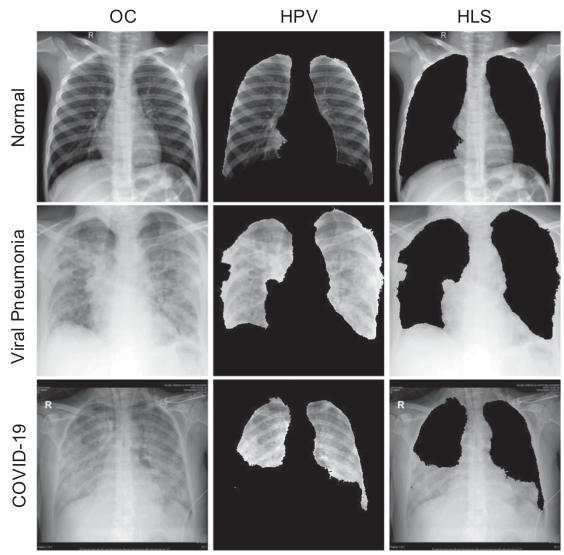

Fig. 4. Exemplar samples of different classes of OC (the first column), HPV (the second column), and HLS (the third columns) selected from the dataset.

TABLE IV
COMPARISON OF CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS. (ACC- ACCURACY, PRE.-PRECISION, REC.-RECALL (%))

Methods	ACC	F1	REC.	PRE.
VGG-19 [47]	94.92 ± 1.18	94.07 ± 1.30	92.78 ± 1.87	95.41 ± 0.99
InceptionV3 [48]	94.59 ± 0.73	94.02 ± 0.45	93.64 ± 0.95	95.02 ± 0.44
ResNet-50 [9]	95.51 ± 0.81	95.43 ± 1.28	94.85 ± 1.23	96.00 ± 1.54
DenseNet-121 [10]	95.68 ± 0.57	95.62 ± 0.98	95.28 ± 0.99	96.34 ± 1.30
Ours(ResNet-50)	97.43 ± 0.59	97.19 ± 0.83	96.75 ± 0.95	97.64 ± 0.86
Ours(DenseNet-121)	98.83 ± 0.50	98.71 ± 0.56	98.53 ± 0.75	98.91 ± 0.42

lung regions as prior knowledge to generate multi-appearance images for improving the classification accuracy. The comparative results of different methods are shown in Table IV. It can be noticed that different methods obtain similar results by using OC images merely as the input. Significantly, our proposed method, which utilizes the multi-appearance of lung regions on CXR, can achieve better evaluation results compared with other methods. In addition, the last two rows in Table IV show that using DenseNet-121 as the backbone in MA-Net achieves the best performance.

V. DISCUSSION

A. Ablation Study for Segmentation

We firstly conduct ablation experiments on different combinations of source and target domains for domain adaptation to compare results by using the semantic prediction space and the feature space adaptation. For feature space adaptation, we directly use the output features of the first two stages in decoder as inputs to the discriminators D_2 and D_3. Table V shows that the proposed adaptation method performs better than the one in the feature space. This indicates that with semantic prediction space adaptation, we can achieve better performance, and the feature in different stages of decoder could encourage domain invariance from integral aspects.

TABLE V
DSC (MEAN ± STANDARD DEVIATION) PERFORMANCE OF ABLATION STUDY

Method	$J \rightarrow S$	$J \rightarrow M$	$S \rightarrow J$	$M \rightarrow J$	AVG
Feature space	75.1 ± 12.2	78.7 ± 11.0	83.2 ± 6.8	90.7 ± 5.5	83.5
Single-scale	85.4 ± 7.9	83.3 ± 6.4	77.3 ± 14.5	92.2 ± 7.1	84.4
Ours	88.3 ± 5.3	88.5 ± 7.4	87.9 ± 4.9	89.4 ± 4.7	90.6

TABLE VI
COMPARISON OF CLASSIFICATION PERFORMANCE OF DIFFERENT UDA METHODS. (ACC- ACCURACY, PRE.-PRECISION, REC.-RECALL (%))

Methods	ACC	F1	REC.	PRE.
No-adaptation	93.79 ± 1.45	89.66 ± 4.38	87.37 ± 5.96	92.19 ± 3.20
CycleGAN [44]	94.33 ± 0.95	88.82 ± 2.59	84.09 ± 4.85	94.39 ± 1.63
MUNIT [45]	94.19 ± 1.46	90.25 ± 4.95	87.44 ± 8.59	93.75 ± 2.24
AdaptSegNet [35]	96.26 ± 0.94	95.36 ± 0.98	94.38 ± 1.31	96.37 ± 0.67
SeUDA [43]	98.30 ± 0.52	98.07 ± 0.84	97.76 ± 0.93	98.39 ± 0.88
Ours (Single-scale)	97.00 ± 0.76	96.40 ± 0.85	96.56 ± 0.79	96.24 ± 1.06
Ours (Multi-scale)	98.83 ± 0.50	98.71 ± 0.56	98.53 ± 0.75	98.91 ± 0.42

In addition, we further evaluate the multi-scale adversarial learning introduced in this work by comparing to single-scale adversarial learning. The results are presented in Table V. We only use the final output of the proposed network to construct it by removing the segmentation loss $\{L^2_{seg}, L^3_{seg}\}$and the adversarial loss $\{L^2_{adv}, L^3_{adv}\}$ when training the network. The evaluation results show that the multi-scale adversarial learning further improves the DSC of lung segmentation.

B. Impact of UDA Methods on Classification

To demonstrate the effectiveness of the proposed MS-AdaNet in our framework, we compare the performance of MA-Net based on MS-AdaNet with that of MA-Net based on other UDA segmentation methods list in Table III. The evaluation results are presented in Table VI. As MS-AdaNet leads to a better performance in segmentation, it promotes classification network better compared with other UDA methods. Moreover, the performance of SeUDA is also very encouraging, just slightly smaller than that of the MS-AdaNet method, demonstrating that based on the correct lung region prior, multi-appearance could effectively improve the classification results. In addition, the performance of MA-Net based on the no-adaptation method is inferior to the performance of only using OC appearance image shown in Table IV. The above comparative results show that our proposed framework is quite important for improving the performance of COVID-19 screening.

C. Impact of Transfer Learning

Due to the small size of medical image datasets, especially in the COVID-19 screening task, the pre-trained model is a simple yet effective method to improve classification accuracy [49]. To assess the impact of transfer learning on the performance of COVID-19 classification in our framework, we compare the

TABLE VII
PERFORMANCE OF PRE-TRAINED MODEL AND MODEL TRAIN FROM SCRATCH. (ACC- ACCURACY, PRE.-PRECISION, REC.-RECALL (%))

Methods	ACC	F1	REC.	PRE.
Only Pre-trained	96.08 ± 0.58	96.10 ± 0.75	95.78 ± 0.84	96.44 ± 1.09
Ours (Pre-trained)	98.89 ± 0.48	98.75 ± 0.63	98.57 ± 0.66	98.95 ± 0.64
Ours (Scratch)	98.83 ± 0.50	98.71 ± 0.56	98.53 ± 0.75	98.91 ± 0.42

TABLE VIII
PERFORMANCE OF PROPOSED METHOD AND THREE SINGLE APPEARANCE MODELS. (ACC- ACCURACY, PRE.-PRECISION, REC.-RECALL (%))

Appearance	ACC	F1	REC.	PRE.
OC	95.68 ± 0.57	95.62 ± 0.98	95.28 ± 0.99	96.34 ± 1.30
HPV	94.89 ± 0.69	94.39 ± 0.99	93.39 ± 1.92	95.43 ± 0.49
HLS	90.54 ± 1.59	83.22 ± 3.54	81.62 ± 6.00	85.17 ± 2.84
ALL	98.83 ± 0.50	98.71 ± 0.56	98.53 ± 0.75	98.91 ± 0.42

TABLE IX
PERFORMANCE OF ONE AND THREE SUB-NETWORKS. (ACC- ACCURACY, PRE.-PRECISION, REC.-RECALL (%))

Methods	ACC	F1	REC.	PRE.
1 DenseNet-121	97.91 ± 0.37	97.46 ± 0.39	96.96 ± 0.35	97.97 ± 0.49
3 DenseNet-121	98.83 ± 0.50	98.71 ± 0.56	98.53 ± 0.75	98.91 ± 0.42

performance of MA-Net based on the DenseNet-121 trained from scratch with a deep model that has the same architecture but based on pre-trained DenseNet-121. As shown in Table VII, because of the small dataset and the prior knowledge of lung region, the results of the model trained from scratch and the pre-trained model based on our framework are similar. In addition, there is a performance improvement in the pre-trained model without using our framework compared with the model trained from scratch in Table IV.

D. Analysis for Multi-Appearance Classification

We build MA-Net to integrate the morphological characteristics of lung region into the comprehensive analysis of pulmonary diseases, including COVID-19. To prove the effectiveness of multi-appearance fusion, we test the sub-networks individually with the corresponding appearance image. The results as depicted in Table VIII, comparing MA-Net to any single appearance network, the accuracy and other evaluation metrics have certain improvement. This experiment quantitatively revealed the role of the multi-appearance images based on lung region in classifying pneumonia diseases including COVID-19.

Furthermore, we conduct an experiment to demonstrate the effectiveness of the three sub-networks in classification. We concatenate OC, HPV, and HLS appearance image as a three-dimensional tensor, then take the tensor as the input to a single network. The results in Table IX show that our proposed framework with three sub-networks can achieve better performance than a single network with a concatenate multi-appearance tensor.

VI. CONCLUSION

In this paper, our main contribution focused on screening COVID-19 in CXR images based on lung region as prior knowledge. Specifically, we firstly built MS-AdaNet, which utilizes multi-scale features in semantic prediction space to boost cross-domain lung segmentation task. We further investigated the differences in lung regions in different pneumonia including COVID-19. Then the multi-appearance CXR images were generated from MS-AdaNet for classification. MA-Net aggregated across the three appearances performed substantially better than any of the appearances on their own. In particular, we validated MS-AdaNet for CXR domain adaptation on three public challenging lung segmentation datasets and achieved satisfactory results in different combinations of source and target domains. Experimental results demonstrated that integrating OC, HLS, and HPV images can improve the performance of COVID-19 screening tasks on CXR.

REFERENCES

[1] C. Sohrabi *et al.*, "World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, 2020.
[2] A. S. Panayides *et al.*, "AI in medical imaging informatics: Current challenges and future directions," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020.
[3] Z. Gao, Y. Li, Y. Yang, N. Dong, X. Yang, and C. Grebogi, "A coincidence filtering-based approach for CNNs in EEG-based recognition," *IEEE Trans. Ind. Informat.*, vol. 16, no. 11, pp. 7159–7167, Nov. 2020.
[4] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.
[5] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
[7] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, "Pelvic organ segmentation using distinctive curve guided fully convolutional networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 585–595, Feb. 2019.
[8] M. Drozdzal *et al.*, "Learning normalized inputs for iterative estimation in medical image segmentation," *Med. Image Anal.*, vol. 44, pp. 1–13, 2018.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[10] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, and K. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019, doi: 10.1109/TPAMI.2019.2918284.
[11] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. P. Xing, "SCAN: Structure correcting adversarial network for organ segmentation in chest X-rays," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 263–273.
[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Cham, Switzerland: Springer, 2015, pp. 234–241.
[13] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++ : Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
[14] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 11–19.
[15] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2494–2505, Jul. 2020.

[16] R. Shen, J. Yao, K. Yan, K. Tian, C. Jiang, and K. Zhou, "Unsupervised domain adaptation with adversarial learning for mass detection in mammogram," *Neurocomputing*, vol. 393, pp. 27–37, 2020.
[17] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 97–105.
[18] Y. Huo *et al.*, "Synseg-Net: Synthetic segmentation without target modality ground truth," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 1016–1025, Apr. 2019.
[19] S. Wang, L. Yu, X. Yang, C.-W. Fu, and P.-A. Heng, "Patch-based output space adversarial learning for joint optic disc and cup segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2485–2495, Nov. 2019.
[20] G. van Tulder and M. de Bruijne, "Learning cross-modality representations from multi-modal images," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 638–648, Feb. 2019.
[21] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Unsupervised X-ray image segmentation with task driven generative adversarial networks," *Med. Image Anal.*, vol. 62, 2020, Art no. 101664.
[22] Y. Zhang *et al.*, "Collaborative unsupervised domain adaptation for medical image diagnosis," *IEEE Trans. Image Process.*, vol. 29, pp. 7834–7844, Jul. 2020.
[23] D. Wu *et al.*, "Severity and consolidation quantification of COVID-19 from CT images using deep learning based on hybrid weak labels," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3529-3538, Dec. 2020.
[24] Y. Li *et al.*, "Efficient and effective training of COVID-19 classification networks with self-supervised dual-track learning to rank," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2787–2797, Oct. 2020.
[25] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Syst. Appl.*, vol. 164, 2020, Art no. 114054.
[26] W. Shi, L. Tong, Y. Zhuang, Y. Zhu, and M. D. Wang, "Exam: An explainable attention-based model for COVID-19 automatic diagnosis," in *Proc. 11th ACM Int. Conf. Bioinf., Comput. Biol. Health Informat.*, 2020, pp. 1–6.
[27] C. Wang, A. Elazab, J. Wu, and Q. Hu, "Lung nodule classification using deep feature fusion in chest radiography," *Comput. Med. Imag. Graph.*, vol. 57, pp. 10–18, 2017.
[28] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, 2020, Art. no. 101794.
[29] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *Appl. Soft. Comput.*, vol. 97, 2020, Art. no. 106580.
[30] Z. Wang *et al.*, "Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays," *Pattern Recognit.*, vol. 110, 2020, Art. no. 107613.
[31] Y. Xie *et al.*, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019.
[32] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.
[33] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
[34] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7167–7176.
[35] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7472–7481.
[36] M. E. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020.
[37] J. Shiraishi *et al.*, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, 2000.
[38] B. Van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, 2006.
[39] S. Stirenko *et al.*, "Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation," in *Proc. IEEE 38th Int. Conf. Electron. Nanotechnol.*, 2018, pp. 422–428.
[40] P. C. Joseph, M. Paul, and D. Lan, *COVID-19 Database*, 2020. [Online]. Avaliable: https://github.com/ieee8023/covid-chestxray-dataset
[41] S. I. S. o. M. a. I. Radiology, *COVID-19 Database*, 2020. [Online]. Avaliable: https://www.sirm.org/category/senza-categoria/covid-19/
[42] P. Mooney, *Chest X-Ray Images (Pneumonia)*, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
[43] C. Chen, Q. Dou, H. Chen, and P.-A. Heng, "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest X-ray segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, Cham, Switzerland: Springer, 2018, pp. 143–151.
[44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
[45] A. Riesco, "MUnit: A unit framework for Maude," in *Proc. Int. Workshop Rewriting Log. Appl.* Cham: Springer, 2018, pp. 45–58.
[46] M. B. Weinstock, A. Echenique, J. W. R. Dabr, A. Leib, and F. A. Illuzzi, "Chest X-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: A normal chest X-ray is no guarantee," *J. Urgent Care Med.*, vol. 14, no. 7, pp. 13–8, 2020.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
[49] T. D. Pham, "Classification of COVID-19 chest X-rays with deep learning: New models or fine tuning?," *Health Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–11, 2021.