Contrastive Cross-Site Learning With Redesigned Net for COVID-19 CT Classification

Zhao Wang, Quande Liu, and Qi Dou, *Member, IEEE*

***Abstract*—The pandemic of coronavirus disease 2019 (COVID-19) has lead to a global public health crisis spreading hundreds of countries. With the continuous growth of new infections, developing automated tools for COVID-19 identification with CT image is highly desired to assist the clinical diagnosis and reduce the tedious workload of image interpretation. To enlarge the datasets for developing machine learning methods, it is essentially helpful to aggregate the cases from different medical systems for learning robust and generalizable models. This paper proposes a novel joint learning framework to perform accurate COVID-19 identification by effectively learning with heterogeneous datasets with distribution discrepancy. We build a powerful backbone by redesigning the recently proposed COVID-Net in aspects of network architecture and learning strategy to improve the prediction accuracy and learning efficiency. On top of our improved backbone, we further explicitly tackle the cross-site domain shift by conducting separate feature normalization in latent space. Moreover, we propose to use a contrastive training objective to enhance the domain invariance of semantic embeddings for boosting the classification performance on each dataset. We develop and evaluate our method with two public large-scale COVID-19 diagnosis datasets made up of CT images. Extensive experiments show that our approach consistently improves the performanceson both datasets, outperforming the original COVID-Net trained on each dataset by 12.16% and 14.23% in AUC respectively, also exceeding existing state-of-the-art multi-site learning methods.**

***Index Terms*—Contrastive learning, COVID-19 CT diagnosis, multi-site data heterogeneity, network redesign.**

I. Introduction

THE COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has lead to a global public health crisis, and continues to spread worldwide. Medical imaging, especially Computed Tomography (CT), has been playing an important role for clinical diagnosis and monitoring of patients with the disease infections [1]. However, the growth rate of COVID-19 suspicious cases has overloaded the public health service capacity and manifested shortage of trained radiologists. Therefore, developing effective computational methods for automated COVID-19 CT image analysis is highly demanded towards improving the diagnosis outcomes and patient management, as well as helping clinicians on tedious image interpretation workload for releasing their precious time which can otherwise be dedicated to more urgent things on the frontline.

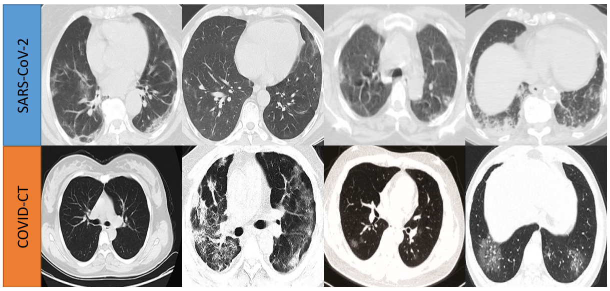


Fig. 1. The CT images of COVID-19 patients from two different clinical centers, showing data heterogeneity on the appearance and contrast.

A considerable amount of data-driven methods have been rapidly developed within this scenario, where the high accuracy is typically attributed to a collected large-scale training database [2]–[4], however, this is difficult to generally achieve in practice. Instead, to mitigate the insufficiency of single-site data amount, aggregating the CT imaging data from different hospitals is desired for establishing a cross-site learning scheme. For instance, Di *et al.* [5] proposed a hypergraph model with multi-site pneumonia data to achieve rapid identification of COVID-19 cases. Wang *et al.* [6] developed COVID-Net using data collected from different repositories to build an accurate deep learning classifier for X-Ray images. However, so far, a major limitation of these works is their negligence of the data heterogeneity across different clinical centers with various imaging conditions (*e.g.*, scanner vendors, imaging protocols, etc). As illustrated in Fig. 1, the CT slices of COVID-19 patients from two different public datasets present apparently different image contrasts. This could potentially affect the model ability to extract robust and general representations as assumed. Previous studies on other medical imaging applications [7]–[9] have frequently observed that straight-forward joint learning with such heterogeneous datasets only brings limited improvement, or even sometimes underperforming individual models trained on a single dataset.

Manuscript received June 20, 2020; revised August 13, 2020; accepted September 2, 2020. Date of publication September 10, 2020; date of current version October 5, 2020. This work was supported by a CUHK start-up research grant and CUHK Shun Hing Institute of Advanced Engineering (project MMT-p5-20). This work was done when Z. Wang did remote internship with CUHK. *(Corresponding author: Qi Dou.)*

Zhao Wang is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: kyfafyd@zju.edu.cn).

Quande Liu and Qi Dou are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China (e-mail: qdliu@cse.cuhk.edu.hk; qdou@cse.cuhk.edu.hk).

Digital Object Identifier 10.1109/JBHI.2020.3023246

To address this real-world challenge, we propose a novel joint learning framework for accurate identification of COVID-19 CT images by effectively combing different data sources with distribution heterogeneity tackled. First, we redesign the recent state-of-the-art COVID-Net [6] from aspects of network architecture and learning strategies to boost its computational efficiency and recognition performance. Moreover, on top of our new backbone, we conduct effective joint learning to fully exploit the benefit of combining multiple datasets. Specifically, our framework employs domain-specific batch normalization layers which enable to conduct the feature normalization and estimate internal feature statistics for each site separately. Importantly, we further propose a contrastive learning objective to explicitly regularize the latent semantic feature space being category sensitive while domain invariant. We evaluate the effectiveness of our approach using two public COVID-19 CT classification datasets. Extensive experiments show that our approach consistently outperforms single-site training models, straight-forward joint learning, as well as existing state-of-the-art multi-site learning methods, on both the datasets. Our main contributions are summarized as follows:

- We redesign the COVID-Net [6] (originally developed for X-Ray) in aspects of network architecture and learning strategy to improve the computation efficiency and prediction accuracy for COVID-19 CT images.
- We propose a novel joint learning framework to improve the COVID-19 diagnosis by effectively learning from heterogeneous datasets, in which we conduct separate feature normalization to tackle the inter-site data discrepancy and propose a contrastive objective to explicitly promote more robust semantic representations.
- Extensive experiments with two public datasets show that our method consistently and significantly improves the classification performance on both datasets. Code is available at: https://github.com/med-air/Contrastive-COVIDNet.

The reminder of the article is arranged as follows. We review the related works in Section II, describe our proposed method in Section III, and elaborate the extensive experiments in Section IV. We then analyze and discuss our work in Section V and finally draw the conclusion in Section VI.

II. Related Works

Many research works have been intensively and rapidly conducted on developing AI methods in responding to COVID-19 global pandemic [10]. We hereafter briefly review deep learning approaches for the task of image-level classification for diagnosis which are closely relevant to this article.

In the beginning, Butt *et al.* [11] aimed to establish a screening model for distinguishing COVID-19 pneumonia from those Influenza-A viral pneumonia and healthy cases with chest CT images using ResNet18 with a location-attention mechanism. Some following-up methods based on transfer learning have been proposed, and most of which used popular existing network architectures, such as VGG [12], ResNet [13]–[15] and DenseNet [16]. Apostolopoulos *et al.* [17] relied on MobileNet with its interpretability for helping radiologist to understand how the model prediction was produced.

At the same time, there were new network architectures emerging, carefully designed and validated. Representatively, the COVID-Net [6] was tailored for COVID-19 recognition, which achieved a promising accuracy for image-level diagnosis based on chest X-Ray (abbr. CXR). Javaheri *et al.* [2] later designed the CovidCTNet to differentiate positive COVID-19 infections from community-acquired pneumonia and other lung diseases. An alternative redesigned framework was based on Capsule Network [18], aiming to more effectively handle small-scale datasets, which is of valuable significance given the emergency of COVID-19 initial outbreak. The method of Gozes *et al.* [19] presented a system that can utilize robust 2D and 3D deep learning models, relying on modifying and adapting out-of-the-box AI models and combining them with domain-wise clinical understanding. Tang *et al.* [20] tackled automated severity assessment (i.e., differentiating non-severe and severe) for COVID-19 based on chest CT images through designed exploration of those identified severity-related features. Rahimzadeh *et al.* [21] developed a neural network that used concatenation of features from Xception and ResNet50V2 networks, with benefits on recognition performance demonstrated.

With the wide spread of disease, more attentions have been dedicated to joint learning of multiple sites for data sources aggregation. For instance, a hypergraph based model [5] achieved efficient COVID-19 identification with multi-site pneumonia data; Zhang *et al.* [22] developed an AI system for COVID-19 diagnosis based on a very large scale dataset (containing about 0.6 million images) which achieved promising performance on several unseen datasets. DasAdhikari *et al.* [23] combined four datasets based on CT and CXR to study the infection severity of COVID-19. Victor *et al.* [24] used data collected from different repositories [6] for effective COVID-19 screening based on deep learning method. More broadly speaking, the issues of merging multi-site data have been actively investigated in recent literature on medical image analysis. For instance, Nguyen *et al.* [25] proposed a novel multi-site learning algorithm to learn different features and aggregate spatial-temporal features through a weighted regularizer based on an integrated multiple heterogeneous dataset. The deep multi-task learning (MTL) framework [26] could effectively improve the accuracy of skin lesion classification through the additional context information provided by body location. Meanwhile, several previous works [27], [28] studied the construction of effective manually generated features and how to design classifiers for medical image analysis tasks across different domains respectively. The federated learning approach [29] provided private multi-site fMRI analysis through a privacy-preserving pipeline and investigated the federated models communication frequency and privacy-preserving mechanisms from various practical aspects.

III. Methods

An overview of our framework for COVID-19 diagnosis is illustrated in Fig. 2. In this section, we first describe our model redesign from COVID-Net. We then introduce our joint learning

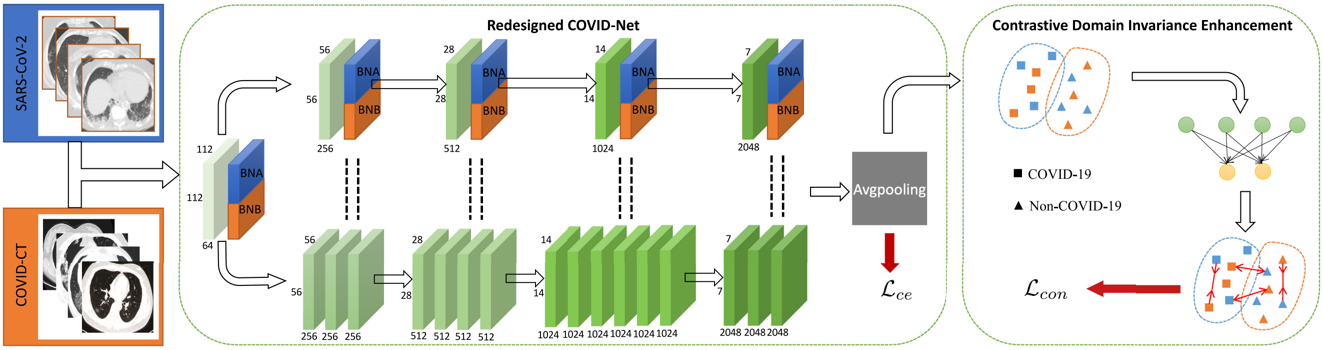


Fig. 2. The overview of our proposed joint learning framework, which redesigns the original COVID-Net as backbone and performs separate feature normalization to tackle the statistical difference of heterogeneous datasets. The proposed contrastive training objective helps to further enhance the domain invariance of semantic embeddings of infected and non-infected cases for boosted diagnosis accuracy on each dataset.

scheme, in which we incorporate separate feature normalization to tackle the cross-site heterogeneity and a contrastive loss to explicitly enhance the domain invariance of latent embeddings for improved classification performance.

A. COVID-Net Redesign for Improved CT Classification

The starting point of our model is COVID-Net [6], a recent new deep learning architecture for COVID-19 CXR image, that has achieved superior performance over several popular classification networks pretrained on ImageNet. As shown in Fig. 2, the network is composed of two branches, in which the upper branch is a light design with four separate convolutional layers, and the lower branch is composed of blocks with heavier dense connections for representation learning. The skip connection between these two branches are employed for long-range multi-level feature fusion. However, the COVID-Net [6] was tailored to meet some specific challenges on CXR images in which the lesions are relatively coarse. Its appropriateness would be changed to a certain extent when applied to CT images where the lesion pattern turns to be more clear, so that presenting richer information to be learned by the model. In this regard, we aim to build upon the strength of this backbone, while further improving its learning efficiency and classification accuracy from two major complementary angles.

1) Network Architecture Redesign: One limitation of the original COVID-Net [6] is the lack of internal feature normalization layers, which is empirically observed to lead to notable variance of the learned representations across different layers and overall branches. As the CT images contain more elaborated patterns, such feature variance will be further amplified if not properly calibrated, which therefore will slow down the the training process and affect the prediction accuracy. To address this problem, we incorporate batch normalization [30] (BN) layers into the specific components of the network to reduce the internal covariate shift and thus helping improve feature discrimination capability and speed up the convergence rate. Importantly, such BN layers are not necessarily beneficial to be naively used as add-on for every single convolution layer. As the computation blocks in the lower branch contain highly dense short-range connections, adding the BN layers there will significantly increase the parameter scale and decrease training speed. As a result, considering the balance between the compution efficiency and stable representation, we add a BN for the initial convolutional layer and a BN after each convolutional layer in the upper branch.

Formally, given M-channel feature maps $\mathbf{x} = \{x_1, \ldots, x_M\}$ of a certain layer, the BN obtains the normalized features $\mathbf{y} = \{y_1, \ldots, y_M\}$ by applying affine transformation on the whitened feature maps along each channel $i \in \{1, \ldots, M\}$:

$$y_i = \gamma \hat{x}_i + \beta, \quad \text{where} \quad \hat{x}_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}}, \tag{1}$$

where μ_i and σ_i^2 refer to the mean and variance of feature x_i; ϵ is an infinitesimal; γ and β are the trainable parameters. Besides, the BN layer collects the moving average values as pair of mean and variance of γ and β during training to capture the global data statistics, and employ these estimated values for feature normalization in the testing phase.

In addition, we have added a global averaging pooling layer after the extracted high-level features for compact semantic embeddings, which helps to significantly decrease the parameters of output dense layers (i.e., by 12 times specific in this network architecture) for alleviating overfitting issues.

2) Learning Strategy Redesign: The CT images used in this study present notable appearance differences for COVID-19 patients across different severity. For examples as shown in Fig. 1, the mild patient may only contain a small lesion while severe patient can be infected almost in whole lung scope. Such large variance within the input space further presents difficulties for the model to explore a robust optimal solution from heterogeneous COVID-19 datasets. To address this problem, we expect a smooth learning process to facilitate the model optimization to reach a relatively robust solution. To this end, we propose to improve the COVID-Net learning strategy by adjusting learning rate more smoothly in a cosine annealing manner [31]. Specifically, denoting the total training epoch as T, the learning rate at a current epoch t is calculated as follows:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta - \eta_{\min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right), \tag{2}$$

where η is the initial learning rate, $\eta_{\min}$ is a predefined threshold of minimum learning rate.

B. Joint Learning Scheme with Redesigned COVID-Net

Given insufficiency of COVID-19 samples from individual hospitals, it is usually desired to aggregate cases from different data sources for deep learning model development. On top of the redesigned COVID-Net backbone, we further propose a joint learning scheme to explicitly tackle the data heterogeneity problem for boosted diagnosis performance.

1) Separate Batch Normalization at Data Heterogeneity: Previous studies have revealed the limited improvement or even performance degradation of simple joint training at severe data heterogeneity [9], [32]. One crucial reason is that the BN layer in joint model will suffer from an inaccurate estimation of moving average values during the training phase due to the statistical difference across datasets (as shown in Fig. 1). During testing phase, the estimated values cannot accurately represent the testing data statistics in each site and hence will lead to performance degradation. In this paper, we employ the domain-specific batch normalization (DSBN) method [9], [33], [34] by assigning an individual BN layer for each site independently to explicitly tackle the statistic discrepancy. As shown in Fig. 2, we replace the BN layers incorporated at redesigned COVID-Net with the DSBN layers. Compared with original BN layer, the DSBN layer enables to capture domain-specific moving values that can accurately represent the statistics of each site, also supplies domain-specific training variables of γ and β to tackle the inter-site variations by performing separate internal feature normalization.

2) Contrastive Domain Invariance Enhancement: In addition to tacking the inter-site heterogeneity under joint learning, we further aim to encourage robust semantic embeddings that cluster regardless of the data source domains. This is crucial, as the benefit from aggregating multi-site data would only be partially leveraged if the model fails to project inputs of different sites into a harmonized feature space. In this regard, we propose to explicitly promote the intra-class cohesion and inter-class separation of the semantic embeddings of infected (i.e., positive COVID-19) and non-infected cases across sites.

We adopt the contrastive learning [37] to achieve that goal. Given a pair of samples (m, n), we denote their semantic embeddings extracted after the global average pooling layer of the network as e_m and e_n, which are 8096-dimensional vectors. In the preliminary experiment, we observed that imposing the compactness regularization directly on the semantic features might be a too strict constraint that impede the convergence. We therefore introduce an embedding network H_ϕ to project the embeddings to a lower-dimensional space. The similarity between this pair of samples (m, n) is then computed on the projected features instead of the original features as:

$$sim(m,n) = \frac{H_\phi(e_m) \cdot H_\phi(e_n)}{\| H_\phi(e_m) \|_2 \cdot \| H_\phi(e_n) \|_2}. \tag{3}$$

We denote the pair (m, n) as positive pair if sample m and n are of the same class, otherwise negative pair. In each iteration, we randomly sample a minibatch of K examples from the two sites. The contrastive loss over each positive pair (m, n) within the minibatch is defined as follows:

$$\ell_{contrastive}(m,n) = -log \frac{exp(sim(m,n)/\tau)}{\sum_{k=1}^{K} \mathbb{F}(m,k) \cdot exp(sim(m,k)/\tau)}, \tag{4}$$

where the value of $\mathbb{F}(m, k)$ is 0 and 1 for positive and negative pair, respectively; τ denotes a temperature parameter. The final loss function is computed over all positive pairs in the given mini-batch for both (m, n) and (n, m). Trained in this way, the model will be enhanced to explore the domain invariance of representations such that the semantic embeddings of samples of same class can lie close to each other in angle space regardless of domain, and away from those of different classes.

C. Overall Training Objective and Technical Details

The overall training objective $\mathcal{L}_{\text{overall}}$ composes the cross entropy loss $\mathcal{L}_{ce}$ to assess the classification error and the contrastive loss $\mathcal{L}_{con}$ to regularize latent space:

$$\mathcal{L}_{\text{overall}} = \mathcal{L}_{ce} + \alpha \cdot \mathcal{L}_{con}, \tag{5}$$

where $\mathcal{L}_{ce} = \frac{1}{N} \sum_i -g_i \cdot \log p_i$, in which N is the number of samples, g_i denote the one-hot groundtruth label and p_i is the predicted probability map, and the $\mathcal{L}_{con}$ sums over pairs according to (4). The embedding network H_ϕ has two fully connected layers, with output size of 1024 and 128 using ReLU activation function. This component is only optimized with $\mathcal{L}_{con}$.

The framework is implemented with PyTorch [38] using an Nvidia TITAN Xp GPU. The classification model and embedding network are trained from scratch with the same Adam Optimizer. The learning rate was initialized with 1e-4 and decayed with cosine annealing. We have used grid search with a random small subset of the entire dataset to empirically adjust the hyper-parameters, setting the temperature parameter τ as 0.05 and α is 1.0. For our proposed method and all the comparsion methods, we totally trained 100 epochs with batch size as 32, containing 16 images from each dataset. Considering the imbalance of sample number between the two datasets, we reloaded the smaller dataset by four times. Data augmentation of random crop and random vertical, horizontal flip were used to mitigate the overfitting problem.

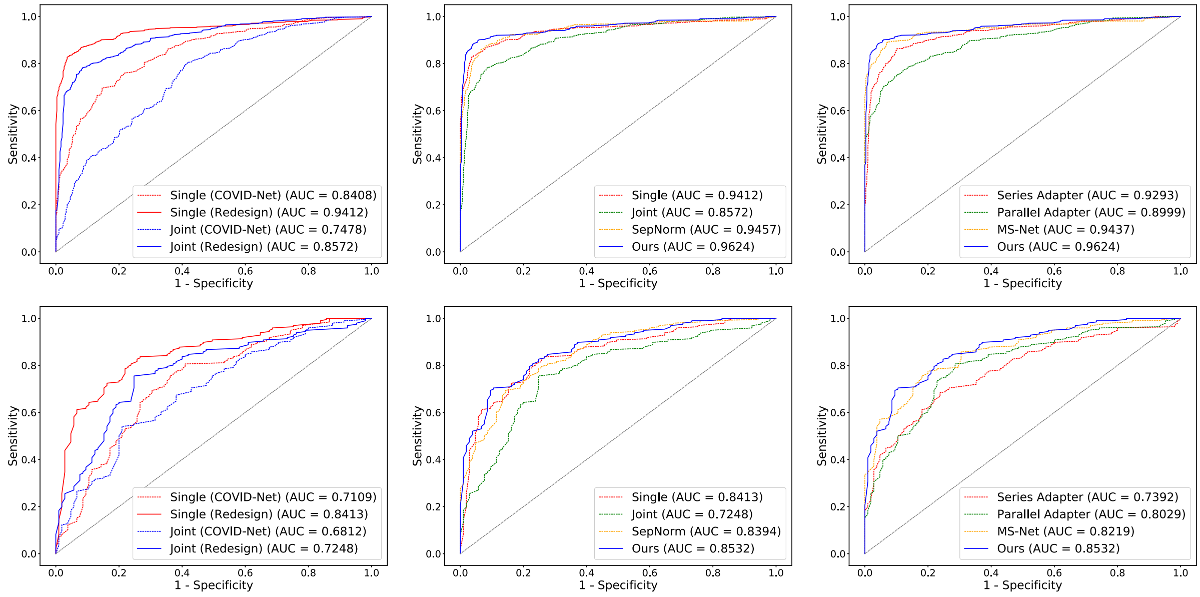
IV. Experiments

A. Datasets and Evaluation Metrics

We adopt two public COVID-19 CT datasets to evaluate our joint learning framework, including *SARS-CoV-2* [39] and *COVID-CT* [40]. To the best of our knowledge, these two datasets are the only relatively large-scale high-quality COVID-19 datasets which are currently publicly available for research. Among the two datasets, the *SARS-CoV-2* (denoted as Site A) consists of 2482 CT images from 120 patients, in which 1252 are positive with COVID-19 and 1230 are non-COVID but with other types of lung disease manifestations. The spatial sizes of these images range from 119 × 104 to 416 × 512. The *COVID-CT* dataset (denoted as Site B) includes 349 CT images from

TABLE I
RESULTS OF DIFFERENT METHODS ON THE TWO DATASETS FOR COVID-19 CT IMAGE CLASSIFICATION (MEAN±STD)

Methods	Site A					Site B				
	Accuracy	F1	Recall	Precision	AUC	Accuracy	F1	Recall	Precision	AUC
Single (COVID-Net [6])	77.12±0.98	76.03±1.13	70.97±2.37	80.04±2.87	84.08±0.92	63.12±2.09	61.09±1.28	57.73±2.94	64.03±3.91	71.09±2.18
Single (Redesign)	89.09±1.08	88.97±0.91	83.78±0.62	94.58±2.07	94.12±0.87	77.07±1.92	77.04±2.17	74.69±3.91	79.48±0.96	84.13±0.82
Joint (COVID-Net [6])	68.72±1.94	69.17±1.93	69.41±3.91	68.27±1.21	74.78±2.91	63.27±2.82	59.78±3.12	54.19±4.17	64.27±3.81	68.12±2.11
Joint (Redesign)	78.42±2.19	77.86±2.01	74.07±3.16	80.82±1.05	85.72±3.54	69.67±0.92	66.89±4.91	66.94±5.86	64.98±3.17	72.48±2.17
Series Adapter [35]	85.73±0.71	86.19±1.65	81.91±2.61	90.98±0.79	92.93±1.42	70.01±3.82	67.08±3.09	74.91±1.89	63.04±4.87	73.92±2.36
Parallel Adapter [36]	82.13±1.91	82.39±1.78	80.02±2.47	83.51±1.87	89.99±0.97	74.93±1.83	73.46±1.68	71.81±2.47	79.84±1.75	80.29±1.76
MS-Net [9]	87.98±1.31	88.73±1.20	84.91±2.83	93.78±2.76	94.37±0.79	76.23±1.81	76.54±1.73	74.07±1.29	79.29±1.48	82.19±1.47
SepNorm	88.76±0.78	87.88±0.81	82.97±1.66	95.46±0.74	94.57±0.77	76.89±0.65	75.02±1.14	70.34±3.76	**80.74±2.98**	83.94±0.43
+ Contrastive (**Ours**)	**90.83±0.93**	**90.87±1.29**	**85.89±1.05**	**95.75±0.43**	**96.24±0.35**	**78.69±1.54**	**78.83±1.43**	**79.71±1.42**	78.02±1.34	**85.32±0.32**



(a) (b) (c)

Fig. 3. (a) ROC curves of Single and Joint approaches with redesigned and original backbone of COVID-Net on Site A (upper) and Site B (lower); (b) ROC curves of our approaches and baseline approaches (Single and Joint) on Site A (upper) and Site B (lower), using redesigned backbone; (c) ROC curves of our approach and other comparison methods on Site A (upper) and Site B (lower), using redesigned backbone.

216 patients containing clinical findings of COVID-19 and 397 CT images from 171 patients without COVID-19. Resolutions of these images range from 102 × 137 to 1853 × 1485. For the preprocessing of the two datasets, all images are first resized to 224 × 224 in axial plane, and then normalized into zero mean and unit variance for intensity values along channel dimension.

Our experiment conducted four-fold cross-validation on the two datasets. Following the literature of COVID-19 diagnosis [39], we adopt five metrics to provide comprehensive evaluation for the models, including: (1) Accuracy (%), (2) F1 score (%), (3) Sensitivity (%), (4) Precision (%) and (5) AUC (%). We report the results in form of average and standard deviation over three independent runs.

B. Effectiveness of Network Redesign on COVID-Net

We first compare our redesigned backbone with the original COVID-Net to validate the effectiveness of network redesign. The comparisons are conducted on two different experimental settings, including 1) *Single* setting which trains a model for each single site; and 2) *Joint* setting which trains a model jointly using two datasets with naive aggregation. From the results in Table I, we see that our *Redesign* model outperforms the original COVID-Net [6] in *Single* setting on both two sites by a large margin, with consistent increase on all five evaluation metrics. Similar observations are shown in *Joint* setting, except the slightly marginal improvement of precision in Site B. These results highlight the superior representation learning ability of our redesigned backbone for COVID-19 diagnosis. Fig. 3(a) further displays the receiver operating characteristic (ROC) curves of the *Single* and *Joint* settings on the two sites, with our redesigned model and the original COVID-Net as backbone respectively. The benefits of our architecture and learning strategy redesign can be further observed from the overwhelming advantage in ROC curves.

C. Effectiveness of Our Joint Learning Framework

We then study the effectiveness of our proposed joint learning framework. Specifically, we first conduct comparison with the

TABLE II
THE P-VALUE WITH PAIRED T-TEST OF OUR METHOD WITH SINGLE, JOINT AND SEPNORM LEARNING SCHEMES

Methods	Single	Joint	SepNorm
Site A	0.002	0.007	0.023
Site B	0.004	0.005	0.009

TABLE III
THE P-VALUE WITH PAIRED T-TEST OF OUR METHOD WITH THE STATE-OF-THE-ART COMPARISON METHODS

Methods	Series-Adapter	Parallel-Adapter	MS-Net
Site A	0.009	0.012	0.021
Site B	1e-5	0.014	0.008

two baseline settings, *i.e.*, *Single* and *Joint*, and then compare with state-of-the-art joint learning approaches. Note that all these comparisons are based on the same backbone of redesigned COVID-Net for fair comparison.

1) Comparison With Baseline Settings: From the results in Table I, the *Joint* approach underperforms the *Single* approach in both Site A and Site B, with 8.40% and 11.65% decrease of AUC score respectively. Such performance degradation reveals the severe statistical discrepancy between the two datasets, and also highlights the urgency and clinical significance to design effective ways for improving the joint learning outcomes from heterogeneous datasets. It is worthy to point out that when conducting separate feature normalization for the two datasets, the joint learning model, *i.e.*, *SepNorm*, outperforms the *Joint* approach on both two sites consistently, which indicates the effectiveness of separate feature normalization scheme in solving the data heterogeneity problem. Notably, by further leveraging the proposed contrastive training objective, the model gains additional improvements on both Site A and Site B, achieving the AUC score of 96.24% and 85.32%, respectively. Such results demonstrate the effectiveness of the contrastive objective to promote more robust semantic embeddings from heterogeneous datasets. Our final results outperforms the *Single* approach in 9 out of 10 metrics on the two sites, which further endorses the practical values of our approach to maximize the data utility of different datasets for boosting diagnosis accuracy. Fig. 3(b) displays the ROC curves of our approach and the two baseline approaches for reference.

We conduct paired t-test to analyze the significance of the improvements of our method over the *Joint*, *Single* and *SepNorm* approaches. The detailed results are shown in Table II. We see that all paired t-tests present p-value smaller than 0.05, indicating the statistically significant improvements of our method on both two sites.

2) Comparison With State-of-the-Art Methods: We then compare our approach with state-of-the-art joint learning methods in both medical image analysis and natural imaging domain, including:

Series-Adapter [35]: This study proposes series domain adapter for joint learning from multiple datasets, in which domain-adaptive layers are incorporated into residual block to mitigate the cross-domain visual discrepancy in natural image processing.

Parallel-Adapter [36]: They develop parallel domain adapter where the domain-adaptive convolutional layer is inserted into residual block in parallel with filter banks to tackle the visual domain gap. This method achieves the state-of-the-art performance for the joint learning task from 10 different natural imaging classification datasets.

MS-Net [9]: This work constructs a multi-site model that incorporates domain-specific auxiliary branches to improve the feature learning capacity and an online knowledge transfer strategy to explore the robust knowledge from multiple heterogeneous prostate MRI datasets for boosted segmentation.

The *Joint* approach serves as a reference to evaluate these joint learning methods. As shown in Table I, the *Series Adapter* achieves higher performance than *Joint* model in both Site A and Site B, while its improvements are highly imbalanced across the two sites and still underperforms the *Single* approach. Compared with *Series Adapter*, the *Parallel Adapter* presents relatively balanced improvements over the *Joint* model, with 4.27% and 7.81% increase of AUC score in Site A and Site B respectively. Improvements of the two approaches over *Joint* model indicate that the domain-specific parameters in domain adapter are beneficial for handling the problem of data heterogeneity. The *MS-Net* is superior to the two domain-adaptive approaches, demonstrating the benefits of the knowledge transfer process in this framework. Notably, our method considerably outperforms all three state-of-the-art joint learning methods on both two sites, demonstrating the superiority of our approach to exploit more robust representations from heterogeneous datasets. The advantage of our method can also be reflected from the ROC curves in Fig. 3(c). Results of paired t-test in Table III indicate the statistical significance of our improvements over the state-of-the-art methods.

V. DISCUSSIONS

With the rapid growth rate of COVID-19 suspection all over the world, designing effective automated tools for COVID-19 diagnosis from CT imaging is highly demanded to improve the clinical diagnosis efficiency and release the tedious workload of clinicians and radiologists. However, accurate diagnosis of COVID-19 from CT images is a non-trival problem, mainly due to the highly similar patterns of COVID-19 and other pneumonia types, as well as the large appearance variance of COVID-19 lesions of patients in different severity level [42]. Recently, a variety of data-driven models have been proposed to solve this problem [4], [19], [43], [44], leading to considerable progress in the field of automated COVID-19 diagnosis in the past few months.

Appropriate network redesign is commonly required to adapt a well-established model onto a specific task. Our work employs the COVID-Net [6] as backbone, which achieves superior performance in COVID-19 diagnosis with X-ray images than several popular classification networks. Considering that the CT images present more detailed and complex patterns of lesions than the X-rays, we redesign the COVID-Net in terms of network architecture and learning strategies to better capture the

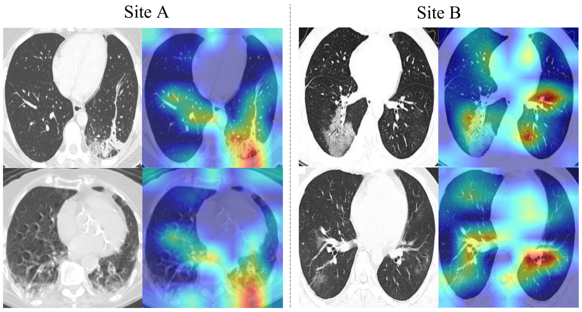


Fig. 4. Visualization of color maps using Grad-CAM [41].

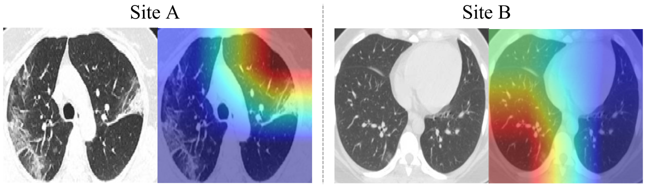


Fig. 5. Visualization of color maps of failure cases with Grad-CAM [41].

semantic representations and facilitate smooth learning process for boosted recognition performance and learning efficiency on COVID-19 diagnosis from CT images.

Given the large appearance variance of COVID-19 lesions and the highly similar patterns with other pneumonia types, the data-driven machine learning models certainly require a large-scale database for training to capture a widespread sample and lesion distribution to attain high accuracy [45]. To mitigate the insufficiency of available COVID-19 CT scans from a certain hospital, it is meaningful and essential to collect the joint data efforts from different clinical centers for robust model development. Some previous studies have also highlighted the importance of learning from multi-site data for rapid and accurate model development in COVID-19 diagnosis [2], [5], but most of them naively mix the data from different sources while ignoring the data heterogeneity that will affect the model to explore the general and robust knowledge for this task. Our experiment reveal that the separate feature normalization can effectively solve the problem of data discrepancy and the benefits of collaborative data efforts can be better explored by explicitly promoting the domain-invariant knowledge during training process.

To understand the behavior of our framework, we observe the Grad-CAM [41] visualization results on the two heterogeneous sites, as saliency maps (shown in Fig. 4). It is consistently observed on both datasets that the suspicious lesion regions are successfully localized across various abnormality patterns (e.g., bilateral and peripheral ground-glass, and consolidative pulmonary opacity), even with quite mild lesions. This analysis reveals promising interpretability of our classification model trained with image-level labels, demonstrating potential clinical relevance for COVID-19 image-based computer-assisted diagnosis. In addition, we present typical failure cases in Fig. 5. We see that the method would mis-classify samples due to wrongly attended regions, and fail to distinguish images with unobvious lesions.

Although promising performance has been achieved as a preliminary study of multi-site learning with COVID-19 data, the limitation of our method still exists. Our method is limited to these two sites used in our paper, which is suboptimal to be directly applied on other unseen sites. This still cannot solve the challenge for wider cross-site deployment thoroughly. Meanwhile, as the lack of computational resources and development urgency, we cannot pretrain our redesigned model on large-scale datasets such as ImageNet. Some previous works in the literature demonstrated that fine-tuning transferred models will bring performance improvement and speed up the training process [46]. As a near future work, we are interested to explore how to connect the carefully redesigned network architectures with model transfer learning from large-scale datasets, by trading off their respective benefits at balance. In addition, we also plan to extend our method to more sites with different environments for wider multi-site learning to validate the generalization capability of AI models in the context of COVID-19 CT image diagnosis.

VI. Conclusion

In this article, we aim to develop a highly-accurate model for COVID-19 CT diagnosis by exploring the benefits of joint learning from heterogeneous datasets of different data sources. We propose a novel joint learning framework through redesigning the recently proposed COVID-Net from architecture and learning strategy as a strong backbone. Our joint learning framework explicitly mitigates the inter-site data heterogeneity by conducting separate feature normalization for each site. A contrastive training objective is further explored to enhance the learning of domain-invariant semantic features to improve the identification performance on each dataset. Experiments on two large-scale public datasets demonstrates the effectiveness and clinical significance of our approach. The future works include improving the generalization capacity of our model, extending it into a wider multi-site setting, as well as employing transfer learning from other large-scale datasets to further enhance the diagnosis accuracy.

References

[1] X. Mei *et al.*, “Artificial intelligence–enabled rapid diagnosis of patients with covid-19,” *Nat. Med.*, vol. 26, pp. 1224–1228, 2020.

[2] T. Javaheri *et al.*, “Covidctnet: An open-source deep learning approach to identify covid-19 using CT image,” 2020, *arXiv:2005.03059*.

[3] J. Zhang *et al.*, “Viral pneumonia screening on chest x-ray images using confidence-aware anomaly detection,” 2020 *arXiv:2003.12338*.

[4] D.-P. Fan *et al.*, “Inf-net: Automatic covid-19 lung infection segmentation from CT scans,” in *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[5] D. Di *et al.*, “Hypergraph learning for identification of covid-19 with CT imaging,” 2020, *arXiv:2005.04043*.

[6] Z. Q. L. Linda Wang and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images,” 2020, *arXiv:2003.09871*.

[7] E. Gibson *et al.*, “Inter-site variability in prostate segmentation accuracy using deep learning,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 506–514.

[8] R. Zech *et al.*, “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study,” *PLoS Med.*, vol. 15, no. 11, 2018.

[9] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "MS-NET: Multi-site network for improving prostate segmentation with heterogeneous MRI data," in *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2713–2724, Sep. 2020.

[10] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Rev. Biomed. Eng.*, 2020. [Online]. Available: http://dx.doi.org/10.1109/RBME.2020.2987975

[11] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Appl. Intell.*, Apr. 2020. [Online]. Available: http://dx.doi.org/10.1007/S10489-020-01714-3

[12] L. Hall, D. Goldgof, R. Paul, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," May 2020. [Online]. Available: http://dx.doi.org/10.36227/techrxiv.12083964

[13] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," 2020, *arXiv:2003.10849*.

[14] A. Abbas, M. Abdelsamea, and M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," Apr. 2020. [Online]. Available: http://dx.doi.org/10.1101/2020.03.30.20047456

[15] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," 2020, *arXiv:2003.14395*.

[16] X. Li, C. Li, and D. Zhu, "Covid-mobilexpert: On-device covid-19 patient triage and follow-up using chest x-rays," 2020, *arXiv:2004.03042*.

[17] I. D. Apostolopoulos, S. I. Aznaouridis, and M. A. Tzani, "Extracting possibly representative covid-19 biomarkers from x-ray images with deep learning approach and image data related to pulmonary diseases," *J. Med. Biol. Eng.*, vol. 40, no. 3, pp. 462–469, May 2020. [Online]. Available: http://dx.doi.org/10.1007/s40846-020-00529-4

[18] 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. Letters*, Sep. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2020.09.010

[19] O. Gozes *et al.*, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," 2020, *arXiv:2003.05037*.

[20] Z. Tang *et al.*, "Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest CT images," 2020, *arXiv:2003.11988*.

[21] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2," *Inform. Med. Unlocked*, vol. 19, 2020, Art. no. 100360. [Online]. Available: http://dx.doi.org/10.1016/j.imu.2020.100360

[22] K. Zhang *et al.*, "Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography," *Cell*, vol. 181, no. 6, pp. 1423–1433.e11, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092867420305511

[23] N. C. Das Adhikari, "Infection severity detection of covid19 from x-rays and ct scans using artificial intelligence," *Int. J. Comput.*, vol. 38, no. 1, pp. 73–92, May 2020. [Online]. Available: https://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/1638

[24] U. Victor, X. Dong, X. Li, P. Obiomon, and L. Qian, "Effective covid-19 screening using chest radiography images via deep learning," 2020.

[25] L. H. Nguyen *et al.*, "Spatial-temporal multi-task learning for within-field cotton yield prediction," *Lecture Notes Comput. Sci.*, pp. 343–354, 2019. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-16148-4_27

[26] L. Haofu and J. Luo, "A deep multi-task learning approach to skin lesion classification," presented at the 31st AAAI Conf. Artif. Intell. Workshops, 2017.

[27] B. Glocker, R. Robinson, D. C. Castro, Q. Dou, and E. Konukoglu, "Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects," presented at the Medical Imaging Meets NeurIPS Workshop, 2019.

[28] A. Van Opbroek, M. A. Ikram, M. W. Vernooij, and M. De Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1018–1030, May 2015.

[29] X. Li, Y. Gu, N. Dvornek, L. Staib, P. Ventola, and J. S. Duncan, "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results," 2020, *arXiv:2001.05647*.

[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[31] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with restarts," in *Proc. Int. Conf. Learn. Representations*, 2017.

[32] L. Yao, J. Prosky, B. Covington, and K. Lyman, "A strong baseline for domain adaptation and generalization in medical imaging," in *Proc. Med. Imag. Deep Learn.-MIDL*, 2019.

[33] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han, "Domain-specific batch normalization for unsupervised domain adaptation," in *Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7346–7354.

[34] Q. Dou, Q. Liu, P. Heng, and B. Glocker, "Unpaired multi-modal segmentation via knowledge distillation," *IEEE Trans. Med. Imag.*, 2020.

[35] S.-A. Rebuffi, H. Bilen, and A. Vedaldi, "Learning multiple visual domains with residual adapters," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 506–516.

[36] S.-A. Rebuffi, H. Bilen, and A. Vedaldi, "Efficient parametrization of multi-domain deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8119–8127.

[37] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020, *arXiv:2002.05709*.

[38] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Advances Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.

[39] E. Soares, P. Angelov, S. Biaso, M. Higa Froes, and D. Kanda Abe, "Sars-cov-2 CT-scan dataset: A large dataset of real patients CT scans for sars-cov-2 identification," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584

[40] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "Covid-CT-dataset: A CT scan dataset about covid-19," 2020, *arXiv:2003.13865*.

[41] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[42] 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. [Online]. Available: http://dx.doi.org/10.1109/tmi.2020.2995508

[43] A. Mobiny *et al.*, "Radiologist-level covid-19 detection using CT scans with detail-oriented capsule networks," 2020, *arXiv:2004.07407*.

[44] S. Chaganti *et al.*, "Quantification of tomographic patterns associated with covid-19 from chest CT," 2020, *arXiv:2004.01279*.

[45] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (covid-19)," 2020, [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.02.14.20023028v5

[46] X. He *et al.*, "Sample-efficient deep learning for covid-19 diagnosis based on CT scans," *medRxiv*, 2020.