Applied Soft Computing 116 (2022) 108291

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Robust weakly supervised learning for COVID-19 recognition using multi-center CT images

Qinghao Ye [a,b,1], Yuan Gao [c,d,1], Weiping Ding [e,1], Zhangming Niu [d], Chengjia Wang [f], Yinghui Jiang [a,g], Minhao Wang [a,g], Evandro Fei Fang [h], Wade Menpes-Smith [d], Jun Xia [i,*], Guang Yang [j,k,**]

[a] Hangzhou Ocean's Smart Boya Co., Ltd, China
[b] University of California, San Diego, La Jolla, CA, USA
[c] Institute of Biomedical Engineering, University of Oxford, UK
[d] Aladdin Healthcare Technologies Ltd, UK
[e] Nantong University, Nantong 226019, China
[f] BHF Center for Cardiovascular Science, University of Edinburgh, Edinburgh, UK
[g] Mind Rank Ltd, China
[h] Department of Clinical Molecular Biology, University of Oslo, Norway
[i] Radiology Department, Shenzhen Second People's Hospital, Shenzhen, China
[j] Royal Brompton Hospital, London, UK
[k] National Heart and Lung Institute, Imperial College London, London, UK

ARTICLE INFO

Article history:
Received 28 January 2021
Received in revised form 18 October 2021
Accepted 6 December 2021
Available online 13 December 2021

Keywords:
Multicenter data processing
Multi-domain shift
Weakly supervised learning
COVID-19
Medical image analysis

ABSTRACT

The world is currently experiencing an ongoing pandemic of an infectious disease named coronavirus disease 2019 (i.e., COVID-19), which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Computed Tomography (CT) plays an important role in assessing the severity of the infection and can also be used to identify those symptomatic and asymptomatic COVID-19 carriers. With a surge of the cumulative number of COVID-19 patients, radiologists are increasingly stressed to examine the CT scans manually. Therefore, an automated 3D CT scan recognition tool is highly in demand since the manual analysis is time-consuming for radiologists and their fatigue can cause possible misjudgment. However, due to various technical specifications of CT scanners located in different hospitals, the appearance of CT images can be significantly different leading to the failure of many automated image recognition approaches. The multi-domain shift problem for the multi-center and multi-scanner studies is therefore nontrivial that is also crucial for a dependable recognition and critical for reproducible and objective diagnosis and prognosis. In this paper, we proposed a COVID-19 CT scan recognition model namely coronavirus information fusion and diagnosis network (CIFD-Net) that can efficiently handle the multi-domain shift problem via a new robust weakly supervised learning paradigm. Our model can resolve the problem of different appearance in CT scan images reliably and efficiently while attaining higher accuracy compared to other state-of-the-art methods.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The pandemic of coronavirus disease (COVID-19) is spreading all over the world rapidly. The number of infections is growing exponentially in different regions, which has triggered great health concerns in the international communities. One of the effective diagnostic methods confirmed by the World Health Organization is via viral nucleic acid detection using the reverse transcription polymerase chain reaction (RT-PCR) test [1]. However, the RT-PCR test is not sensitive sufficiently in some cases, which may put hurdles for presumptive patients to be identified and treated early.

As a non-invasive imaging technique, computed tomography (CT) can detect those characteristics, e.g., bilateral patchy shadows or ground glass opacity (GGO), manifested commonly in the COVID-19 infected lung. Hence CT may serve as an important tool for COVID-19 patients to be pre-screened and diagnosed early. The quantified imaging biomarkers extracted from CT images can also provide crucial prognostic values. Recently, deep

* Corresponding author.
** National Heart and Lung Institute, Imperial College London, London, UK.
E-mail addresses: xiajun@email.szu.edu.cn (J. Xia), g.yang@imperial.ac.uk (G. Yang).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.asoc.2021.108291
1568-4946/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

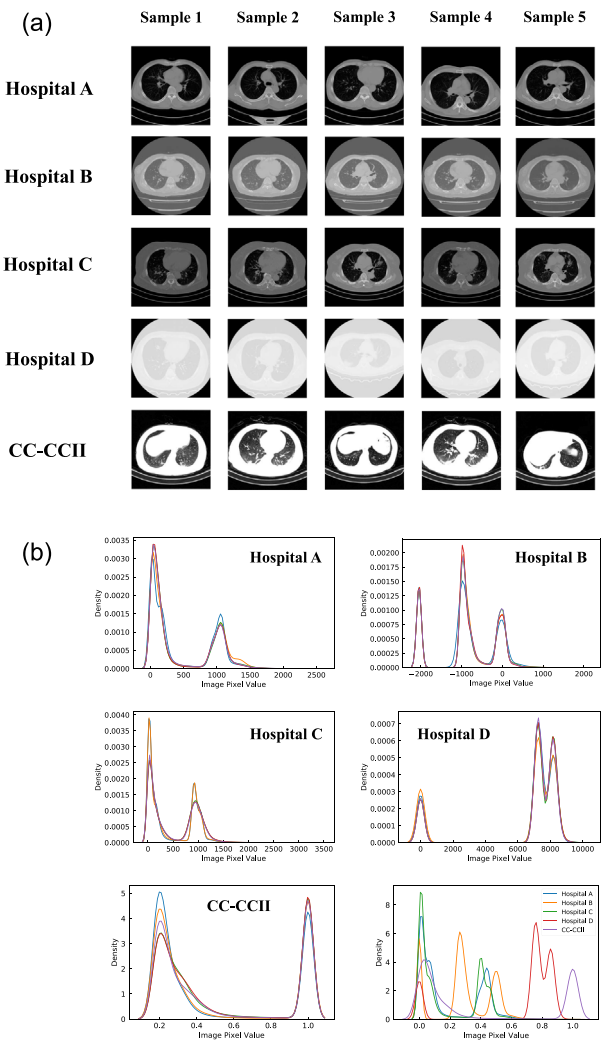


Fig. 1. (a) Samples of CT images are taken from five different hospitals and (b) The histograms of these CT images. Compared with images from Hospital A and Hospital D, it is clear that the brightness levels are distinctive. Moreover, the contrast of the data collected from the China Consortium of Chest CT Image Investigation (CC-CCII) dataset is considerably different from CT images acquired from other hospitals. The right bottom figure demonstrates the distribution of the images from different hospitals after normalization, however, these distributions still behave distinctively. It is of note that there are no visually distinctive features across CT scan images but it is easy for human radiologists to correctly classify despite CT scanner changes. On the contrary, deep learning based automated methods may fail to generalize across CT images acquired from different hospitals.

learning based methods have been developed efficiently for the chest X-ray/CT data analysis and classification [2–4], and these approaches can achieve state-of-the-art performance on X-ray/CT image diagnosis and prognosis.

Nevertheless, most CT scan datasets for COVID-19 only contain CT volumes with a set of CT slices with only patient-level annotations provided (i.e., patient-level class labels available) indicating the patient is infected or not. There is a lack of per-slice labels since annotating each slice is labor-intensive and time-consuming for radiologists. It has been reported that it could take an experienced radiologist about 21.5 min [5] to analyze and label one whole CT volume. Consequently, convolutional neural network (CNN) based deep learning models trained on CT slices with only the patient-level label can perform poorly because some annotations of these CT slices are incorrect (e.g., non-lesion slices of the lung are actually be falsely labeled) leading training data to be noisy.

Yet another challenge when employing deep learning methods to medical image recognition is called *data distribution shift* (a.k.a., **multi-domain shift**). Data distribution shift refers to the phenomenon that a common object or organ collected under various scenarios (e.g., different machine vendors and sequence parameters) can result in vastly different data distributions. Therefore, models trained under the empirical risk minimization (ERM) [6] might cause the failure of model generalization. It is because the ERM assumes that training and testing data are sampled from the same or similar distribution and domains. However, in the data distribution shift scenario, this assumption would be violated.

When a neural network is trained with images from one domain and tested on another domain (i.e., distinct imaging scenarios), the recognition performance often degrades dramatically. Fig. 1 represents images of different CT data collected from different hospitals. In the figure, it can be observed that CT data obtained from different hospitals are visually different although they all present image slices of the lung. It is on the grounds that every hospital uses different protocols and parameters for CT scanners when collecting the images for patients. Therefore, the multi-domain shift problem of the multi-center and multi-scanner studies is nontrivial. It is crucial to solving the multi-domain shift problem to achieve a dependable recognition, which is critical for reproducible diagnosis and prognosis.

To cope with the issues above, in this work, we trained our model on both patient-level and image-level with multiple domain information. In particular, we consider the sequential information within the CT volume when predicting a patient is tested COVID-19 positive or not. To preserve the sequential information, we divide a lung CT volume into individual sections from the upper lobe all the way to the inferior lobe. As illustrated in Fig. 2, our method aggregates these sections as the representation of a patient. When aggregating these sections, we utilize the multiple instance learning method with the k-max selection strategy for images in each section. With the help of the k-max selection, our model can filter out the uncertain and noisy images that can be beneficial to make an accurate prediction. Moreover, multiple instance learning method is incorporated that can enforce our model to mine confident candidates for training and testing [7] while ignoring modeling the joint distribution of sections from the patient rather than a single image, which is rewarding for unseen center prediction.

In summary, our contributions are mainly three-fold:

- We proposed a weakly supervised learning based multi-domain information fusion framework for automated COVID-19 diagnosis from multi-center and multi-scanner CT scans that only requires patient-level annotations for training.
- We propose a novel noisy label correction technique that propagates the patient-level predictions to individual slices and identifies the COVID-19 infected slices accurately.
- We develop a slice aggregation module to alleviate the data distribution shift problem, which is essential for the deployment of the developed model in the real-world scenario.

By validation on the China Consortium of Chest CT Image Investigation (CC-CCII) [8] benchmark dataset, our proposed coronavirus information fusion and diagnosis network achieves superior performance compared to state-of-the-art models on both patient-level and image-level.

2. Related work

Before the COVID-19 pandemic, a huge amount of deep learning based methods has been proposed for lung cancer CT image

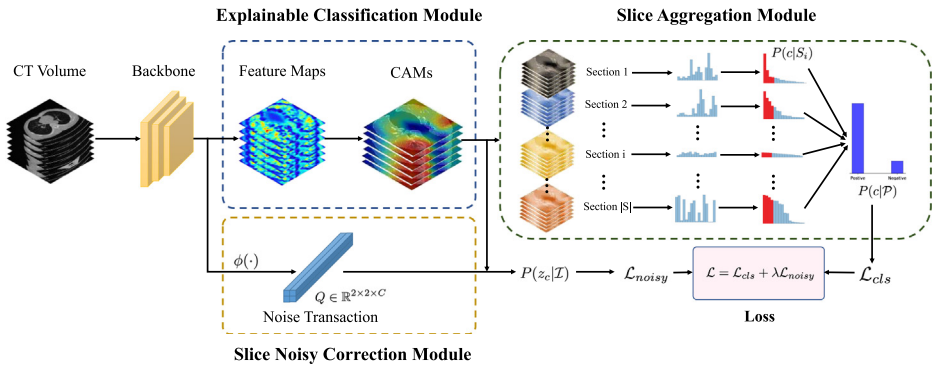


Fig. 2. The architecture of our proposed CIFD-Net. It is of note that $P(c \mid S_i)$ denotes the probability of the Section S_i, and $P(c \mid \mathcal{P})$ represents the probability of the patient who is tested COVID-19 positive or not. $Q \in \mathbb{R}^{2\times2\times C}$ indicates the noise transaction from the probability of the true label $P(y_c \mid \mathcal{I})$ to the probability of the noise label $P(z_c \mid \mathcal{I})$. In addition, $\phi(\cdot)$ is a feature embedding function. In addition, ResNet-50 [9] is adopted for backbone network.

analysis. In this research area, there have been great achievements, culminating in the development of many end-to-end pipelines for lung cancer diagnosis, classification, treatment planning, and prognostic evaluation [10–16]. In the treatment of interstitial lung disease (ILD), deep learning approaches have also been developed [17–20]. In CT scans for COVID-19 patients, image characteristics, e.g., ground glass opacity and/or consolidation, are akin to those observed from lung cancer and ILD patient CT scans. Therefore, in the design of COVID-19 detection algorithms using CT images, insights from research on both lung cancer and ILD are significant and there is a clear translatability to the COVID-19 studies.

CNNs for visual recognition. Convolutional Neural Network (CNN) has been widely used in the medical diagnosis system [3,21,22]. Recently, plenty of COVID-19 recognition algorithms have been proposed, which have adopted artificial intelligence algorithms especially using the CNN. A comprehensive review of artificial intelligence assisted COVID-19 detection and diagnosis can be found elsewhere [23–27], and here we only provided a summary for the most relevant studies.

Jin et al. [28] developed a combined segmentation-classification model for COVID-19 diagnosis. A few pre-trained models were tested, e.g., fully convolutional network (FCN-8s), U-Net, V-Net, and 3D U-Net++, as well as classification models like dual path network (DPN-92), Inception-v3, residual network (ResNet-50), and attention ResNet-50, from which the 3D U-Net++ and ResNet-50 combination achieved the best performance. However, it was unclear which layers were pre-trained and re-trained, the reproducibility of this study is uncertain. Wang and Wong [3] proposed COVID-Net, which stacked multiple convolutional blocks with dilated convolution to recognize chest X-ray images. Li et al. [2] explored the patient label and used max-pooling strategy over features extracted by the CNN from a set of slices to make the prediction. In addition, Ouyang et al. [4] deployed a 3D CNN and used the residual learning mechanism to build the network, which incorporated the depth information of the CT volumes. Shan et al. [29] proposed a human-in-the-loop strategy for infection region quantification, in which a modified V-Net was developed incorporating bottleneck building blocks to reduce training costs. The human-in-the-loop training procedure output a segmentation for subsequent manual corrections performed by radiologists, and then these corrected data were input to re-train the network iteratively. More recently, Hu et al. [30] proposed a weakly supervised multi-scale learning framework for COVID-19 classification and lesions detection, which demonstrated promising results but its performance may be hindered by using the patient-level labels that contain noise labeling.

For automatic prognostication of COVID-19 patients, Huang et al. [31] developed a two-step segmentation model that extracted lung and lobes region followed by pneumonia segmentation. Both steps used separated U-Net and at least two follow-up scans for each patient were analyzed. The authors found significant differences in lung opacification percentage between the initial and the first follow-up scans, but not between the first and the second follow-up scans. Although the study findings are intriguing, there are critiques on lacking important information essential to the reproducibility [32].

Although the aforementioned studies and many others have shown promising results [1,33–43], two major issues can prevent the widespread deployment of these methods: (1) most previously proposed approaches relied on heavily annotated ground truth, e.g., for the infectious areas and slice-based labeling and (2) domain-shift failure for multi-center and multi-scanned datasets and therefore, poor reproducibility was always a concern.

Multiple instance learning. The multiple instance learning (MIL) is a weakly supervised learning problem that has been attempted in several studies including weakly supervised object localization [7], video anomaly detection [44], weakly supervised image segmentation [45] and others. In the MIL framework, a bag can be defined as a set of instances or image slices. Positive bags are assumed to contain at least one instance from a certain category and negative bags do not contain any instances from that category. It is intuitive to consider the classification of CT volumes that contain multiple CT slices as a MIL problem. A few methods have been proposed to solve the MIL problem. For example, Oquab et al. [46] trained a CNN using the max-pooling MIL strategy to classify the object. However, some of the MIL pooling strategies, such as max-pooling and mean-pooling, very often lead to insufficient and unstable training because of gradient vanishing. To fix this problem, Ilse et al. [47] combined the gated attention mechanism with the MIL strategy to solve the medical image classification problem, but it could not predict the instance label accurately. Chen et al. [48] developed a stylized generative method to transfer the knowledge from MRI to CT within unsupervised manner. Xia et al. [49] utilized uncertainties along different volume angles to measure the importance of predicted labels. Chen et al. [50] modeled intra-consistency between two domains to align the feature distributions. However, these methods requires to train the model using both source domain and target domain, which cannot handle the unseen domain scenarios. Our method will provide solutions to these limitations.

Domain adaptation. Domain adaptation refers to the techniques aimed at improving the performance of machine learning tasks, e.g., classification, detection, segmentation, when training the

classifier on the data only from the source domain, but testing it using related samples from a shifted target domain. Some approaches also use domain adaptation to help learn the feature representation. Hoffman et al. [51] proposed a method that learned the difference between classification and detection tasks, and transferred this knowledge from the classifier to detectors using weakly annotated data. In addition, MIL was incorporated for learning feature representation and classifier [52]. Besides, Mahmood et al. [53] utilized transformations such as hue, saturation, contrast, and brightness for RGB images to change the color and texture of the images in the source domain. Existing domain adaptation methods tend to use strongly annotated data in the source domain in order to improve the recognition performance, while our methods will focus on a weakly supervised manner. In other words, our method will require no instance-level annotation or the auxiliary strongly annotated data for recognition.

3. Proposed method

In this section, we introduce the proposed coronavirus information fusion and diagnosis network (CIFD-Net) with their key modules including an explainable classification module (ECM), a slice aggregation module (SAM), and a slice noisy correction module (SNCM), respectively as illustrated in Fig. 2.

The proposed ECM integrates the generation of class activation mapping into the forward propagation of the CIFD-Net, enabling CAMs generation during training and testing, which provides explainable results for the prediction of our model.

Besides, instead of training on image-level (slice-wise) labels, which requires a significant amount of labor for manual labeling, we propose the SAM to train on patient-level labels. We model the joint probability of slices for each patient by which slices are divided into several consecutive sections with equal length. We then compute the probability of each section by adopting a k-max selection strategy, which can ignore some slice with large uncertainty, thus reduce the noise during modeling the joint probability at the patient level. With the help of modeling the joint probability, our model pays more attention to modeling the distribution of affected sections leading to better generalization on multiple domains.

Moreover, in order to improve the accuracy on the image-level, we further proposed the SNCM, which models the transaction between the true label and noisy label since the labels at the patient-level are considered to be noisy with respect to slice-wise labels.

3.1. Problem formulation

The ultimate goal of our model is to diagnose whether a patient is tested positive or negative given a 3D volumetric CT lung scan. Let $\mathcal{P} = [\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n]$ denotes the lung CT volume for a patient with n CT slices, where $\mathcal{I}_i$ is a 2D CT slice image. Let $Y \in \mathbb{R}^{\{0,1\}}$ denotes whether a patient is tested to be COVID-19 positive or not. $Y = 1$ when the patient gets COVID-19, while $Y = 0$ stands for the patient is not COVID-19 infected. During the training stage, we only have patient-level labels, and the number of CT lung slices can vary significantly.

3.2. Explainable classification module

As the predicting process of CNN is in a black box. Several techniques [54,55] have been proposed to shed light on how CNN makes the prediction and obtains the remarkable localization ability without any supervision of localization maps. As an explainable auxiliary diagnosis tool for radiologists, we employ the class activation mapping (CAM) [54], which can generate the localization maps for the prediction from the output of the backbone networks, e.g., ResNet [9], VGG [56], GoogLeNet [57], etc. However, the process of generating CAM is a two-step process, in which the backbone network is trained on the dataset and utilizes the weights of the final fully connected layer to compute the weighted sum of feature maps of the last convolutional layer. Suppose $F^k \in \mathbb{R}^{H\times W}$ is the kth feature map with height H and width W from the last convolutional layer, and $W^{fc} \in \mathbb{R}^{K\times C}$ is the weight of the last fully connected layer, where C is the number of classes and K is the number of feature maps from the last convolutional layer. Therefore, the class score s_c of the class c can be calculated by

$$s_c = \sum_{k=1}^{K} W_{k,c}^{fc} \left(\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j}^{k} \right). \tag{1}$$

Therefore, the localization map for the class c proposed in [54] is defined by

$$A_c^{fc} = \sum_{k=1}^{K} W_{k,c}^{fc} F^k, \tag{2}$$

and we can visualize the object localization maps via A_c^{fc}.

Although CAM is a useful way to locate the region, it requires a post-processing procedure to generate. In our method, we plug the generation of CAM into the network with only one forward pass. Instead of directly applying global average pooling after the last convolutional layer, we replace the fully connected layer using a 1×1 convolutional layer with the stride of 1 before the global average pooling operation. Suppose the weight of the convolutional layer is $W^{conv} \in \mathbb{R}^{K\times C}$ which is the same mathematical form as the weight of the fully connected layer, i.e., W^{fc}, we tweak Eq. (1) as follows,

$$s_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\sum_{k=1}^{K} W_{k,c}^{conv} F_{i,j}^{k} \right), \tag{3}$$

which results in the same output with Eq. (1). Thus, the modified CAM for the class c is computed as

$$A_c^{conv} = \sum_{k=1}^{K} W_{k,c}^{conv} F^k. \tag{4}$$

The modified activation mapping can accurately indicate the importance of the activation from CT images and locate the infected areas of the COVID-19 patients, providing the explainable and reliable results for prediction. The region with higher activation score indicates more importance the activation contributed to the prediction. The modified activation mapping can also offer the auxiliary diagnostic information for radiologists. The differences between the original CAM and our ECM strategy are demonstrated in Fig. 3.

3.3. Slice aggregation module

In some mild COVID-19 cases, there might be only part of the CT volume that has an infection, and very often the lesions are quite small. If we simply treat all slices as COVID-19 positive and train a classifier with the image-level label, it could lead to a noisy learning and poor results as the consequence. To overcome this problem, we propose the SAM and use the joint distribution to model the probability of patient is COVID-19 positive or negative. We assume that lesions are consecutive and only affect adjacent slices, consequently, we use a section based strategy to tackle the problem. The intuition of using the section based strategy is that

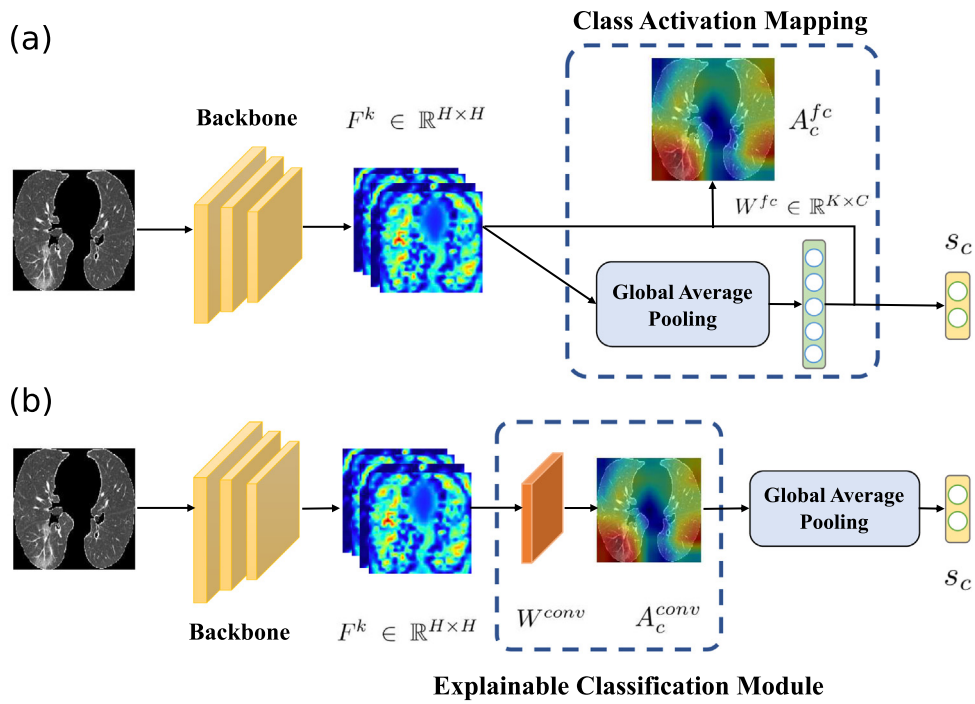


Fig. 3. (a) The workflow of the class activation mapping (CAM) scheme and (b) The proposed explainable classification module (ECM). It shows that our ECM can generate the CAM using only one forward pass, but the original method proposed by Zhou et al. [54] needs a post-processing procedure to generate the CAM. F^k is the kth feature map from the backbone network. Besides, W^{fc} and W^{conv} are the weights for the fully connected layer and the convolutional layer. A_c^{fc} and A_c^{conv} are the class activation maps for class c. s_c is the class score for class c.

it can be directly mapped to the problem of multiple instance learning (MIL) [58]. In MIL, samples are divided into two bags classified as positive and negative bags. A positive bag contains at least one positive instance and a negative bag only has the negative instance. In the problem, only bag labels (patient annotations) are provided, and sections can be treated as instances in the corresponding bags.

Given a patient $\mathcal{P} = [\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n]$ with n CT slices, we divide these slices into disjoint sections, which can be considered as a set that contains an equal number of consecutive CT slices, i.e., $\mathcal{P} = \{S_i\}_{i=1}^{|S|}$, where $|S|$ is the amount of sections for patient $\mathcal{P}$ as defined as follows,

$$|S| = \max\left(1, \left\lfloor \frac{n}{l_s} \right\rfloor\right), \tag{5}$$

where l_s is an empirically designed parameter named as section length.

Then the probability of patient $\mathcal{P}$ belonging to the class c can be represented as

$$P(c \mid \mathcal{P}) = 1 - \prod_{i=1}^{|S|} (1 - P(c \mid S_i)), \tag{6}$$

where $P(c \mid S_i)$ is the probability of the ith section S_i that belongs to the class c. Instead of taking the average of each probability of the slice in that section, we take the k-max probability for each class to compute the section probability. This is because some slices may contain few infection regions which can confound the prediction. To alleviate this problem, we adopt the k-max selection method which can be formulated as

$$P(c \mid S_i) = \sigma\left(\frac{1}{k} \max_{\substack{s^{(j)} \in M \\ s.t. \quad M \subset S_i, |M| = k.}} \sum_{j=1}^{k} s_c^{(j)}\right), \tag{7}$$

where $s_c^{(j)}$ is the top jth class score of the slice in the ith section for the class c, and $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function. Then we use the patient-level annotations $\mathbf{y}$ as the ground-truth during the training. The classification loss can be formulated as

$$\mathcal{L}_{cls} = \sum_{c=0}^{1} [y_c \log P(c \mid \mathcal{P}) + (1 - y_c) \log(1 - P(c \mid \mathcal{P}))]. \tag{8}$$

3.4. Slice noisy correction module

To further alleviate the negative impact of the image-level noises, we propose the SNCM, which is loosely inspired by [59], to model the hidden distribution $P(z_c = i \mid y_c = j, \mathcal{I})$ between the noisy label and the true label. Let $P(y_c \mid \mathcal{I})$ denotes the true posterior distribution, given an image $\mathcal{I}$. The distribution of noisy label, $P(z_c \mid \mathcal{I})$, can be modeled as

$$P(z_c = i \mid \mathcal{I}) = \sum_j P(z_c = i \mid y_c = j, \mathcal{I}) P(y_c = j \mid \mathcal{I}). \tag{9}$$

We estimate the noise transaction $Q_{ij}^c = P(z_c = i \mid y_c = j, \mathcal{I})$ for the class c as follows

$$Q_{ij}^c = P(z_c = i \mid y_c = j, \mathcal{I}) = \frac{\exp(w_{ij}^c \phi(\mathcal{I}) + b_{ij}^c)}{\sum_i \exp(w_{ij}^c \phi(\mathcal{I}) + b_{ij}^c)}, \tag{10}$$

where $i, j \in \{0, 1\}$; $\phi(\cdot)$ is a nonlinear mapping function; w_{ij}^c and b_{ij}^c are trainable parameters for the class c between the status i and j. Transaction score $T_{ij}^c = w_{ij}^c \phi(\mathcal{I}) + b_{ij}^c$ can be regarded as the score of the transaction from the true label i to the noisy label j with respect to the class c. As a result, the estimated probability of noisy label for the class c is represented as

$$P(z_c = i \mid \mathcal{I}) = \sum_j Q_{ij}^c P(y_c = j \mid \mathcal{I}). \tag{11}$$

Algorithm 1 Training procedure of CIFD-Net

Input:
Set of M CT volumes $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_M\}$.
Learning rate: η.
Output:
Learned model parameters of CIFD-Net: Θ.
Initialize all parameters denoted Θ using Xavier.
repeat
  for $m = 1$ **to** M **do**
    Use the backbone network to compute slice features for CT volume $\mathcal{P}_m$ with its slices $\{\mathcal{I}_i\}_{i=1}^n$.
    Calculate class scores $s_c \in \mathbb{R}$ using Eq. (3).
    Compute the probability of patient $P(c \mid \mathcal{P}_m)$ via SAM in Eqs. (6) and (7), and obtain classification loss $\mathcal{L}_{cls}$ via Eq. (8).
    Formulate the noise transaction Q_{ij}^c between the probability of noise label and that of the true label by Eqs. (9) and (10).
    Obtain the estimated probability of noisy label $P(z_c = i|\mathcal{I})$ by Eq. (11), and calculate noisy classification loss $\mathcal{L}_{noisy}$ using Eq. (12).
    Compute the final loss $\mathcal{L}$ with Eq. (13).
    Update parameters: $\Theta \leftarrow \Theta - \eta \nabla_\Theta \mathcal{L}(\Theta)$.
  end for
until convergence
return Θ.

Table 1
The number of CT samples used for training for each class collected by four different hospitals A, B, C, and D. Besides, details of the CC-CCII dataset are also listed, which was used in the independent testing stage. The ratio of positive and negative samples in training set is approximately 1:1, and 2:1 in test dataset.

Dataset	Number of patients			Number of CT images			Subset
	Total	Positive	Negative	Total	Positive	Negative	
Hospital A	424	0	424	24,670	0	24,670	Train
Hospital B	58	58	0	5,512	5,512	0	Train
Hospital C	17	17	0	2,611	2,611	0	Train
Hospital D	305	305	0	12,374	12,374	0	Train
CC-CCII [8]	2,034	1,320	714	130,511	84,629	45,882	Test

Finally, with the help of the estimated noisy probability, for the patient $\mathcal{P}$, the noisy classification loss is computed by

$$\mathcal{L}_{noisy} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=0}^{1}[y_c \log P(z_c = 1 \mid \mathcal{I}) + (1 - y_c)\log P(z_c = 0 \mid \mathcal{I})]. \quad (12)$$

By combining Eqs. (8) and (12), we can obtain the total loss function that we need to optimize for our model that is calculated as follows,

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{noisy}, \quad (13)$$

where λ is a hyper-parameter to balance the loss terms.

During the model training, the above loss functions are optimized iteratively. By incorporating the SAM, we can build a unified end-to-end deep neural network architecture for the COVID-19 diagnosis. The whole training procedure is summarized in Algorithm 1.

4. Experiments and discussions

In this section, the effectiveness of our method is validated and the results are quantified. First, we provide some statistics of the datasets and describe the implementation details as well as the experimental settings, which are followed by the reported results, the ablation studies, and further discussions on the qualitative and quantitative results.

4.1. Datasets

In order to verify the effectiveness of proposed model on the data from an independent hospital, we use data from several hospital, then test the model on an independent dataset. The datasets used in our study are summarized in Table 1. We collect CT datasets from four different local hospitals and anonymize the data by removing all the patient identity information. In total, there are 804 CT scan volumes with 45,167 CT images, 380 of which are COVID-19 positive and the other 424 are negative cases. All COVID-19 positive cases are confirmed by the RT-PCR tests. We train on the cross-domain datasets collected from hospitals A, B, C, and D and test on an open public CC-CCII dataset [8]. CC-CCII dataset consists of 2034 3D CT volumes with 130,511 CT images, which have been acquired by the CT scanner from a different manufacturer representing another image domain.

4.2. Data standardization, pre-processing

Following the protocol used in [8], we first normalized images with *z-score* normalization, then we used the U-Net segmentation network [60] to segment the CT images. After that, we randomly cropped a rectangular region whose aspect ratio is randomly sample in [3/4, 4/3] and area randomly sampled in [90%, 100%], then resized the region into 224 × 224 shape. Meanwhile, we randomly flipped the input volumes horizontally with 0.5 probability. The input data would be a set of CT volumes which are composed by consecutive CT slice images.

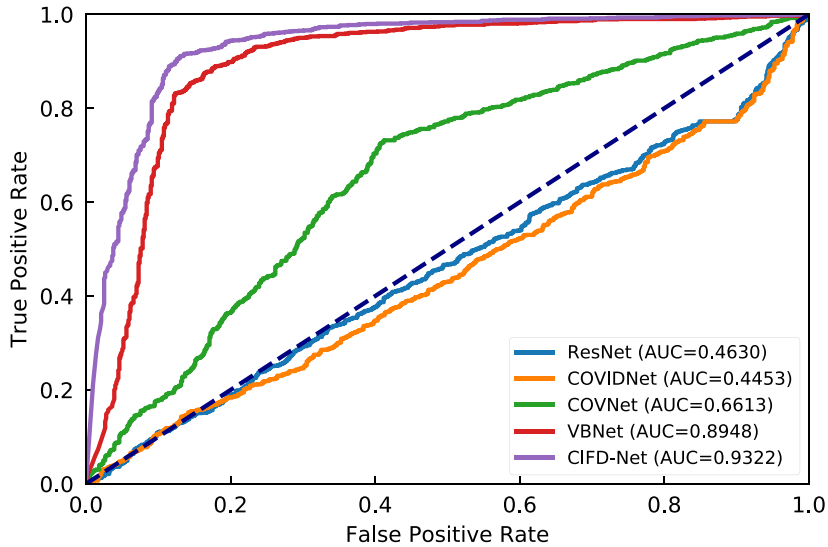
4.3. Implementation details

We use ResNet-50 [9] as the backbone network pre-trained on ImageNet [61]. For data augmentation, we use random horizontal flipping for the input CT volume in the spatial dimension. Each image in a CT volume is randomly horizontal flipped with a probability of 0.5. Then, we resize them into the size of 224 × 224. In addition, brightness and contrast are randomly changed within the range [0.9, 1.1]. The dropout rate is set to 0.7, λ is set to 0.0001, and the L_2 weight decay coefficient is set to 10^{-5}. During the training and testing stage, we set $l_s = 16$ and $k = 8$ to compute the patient probability. We train our model using the Adam optimizer [62] with the initial learning rate $\eta = 1 \times 10^{-3}$, and training is terminated after 4000 iterations with a batch size 10. All experiments have been conducted on a workstation with 4 NVIDIA Tesla V100 GPUs using PyTorch.

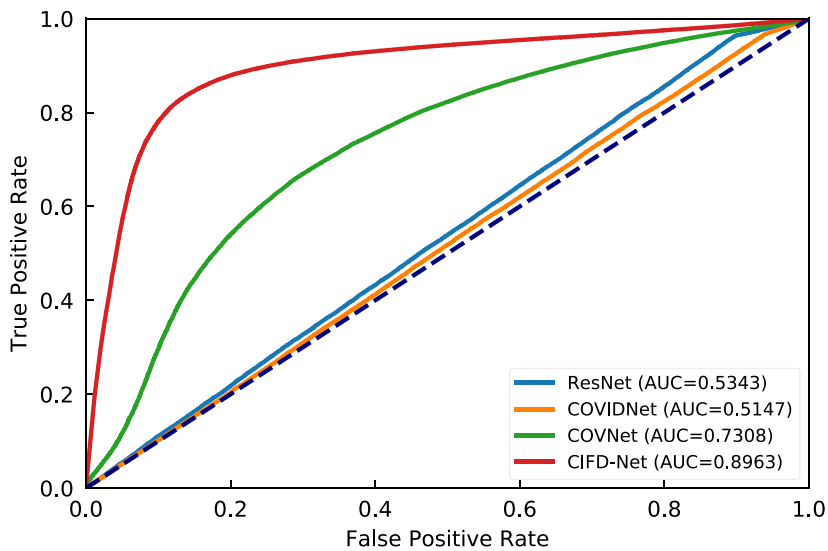
4.4. Quantitative results

We reproduce and compare with four state-of-the-art methods [2–4,9] on the COVID-19 CT classification. The results are shown in Table 2. For image-level supervision, COVID-Net [3] and ResNet-50 [9] employ the patient-level annotations as image annotations. Different to the methods proposed by Wang and Wong [3] and He et al. and [9], VBNet [4] adopts a 3D residual convolutional neural network (3D-ResNet) to train on CT volumes with patient labels. Moreover, COVNet [2] also trains on the patient-level label that they feed a patient-specific set of CT images into a 2D ResNet and simply aggregate the image-level feature descriptors with a max-pooling operator.

From Table 2, several interesting observations can be summarized as follows.



(a) Patient-level Annotation



(b) Image-level Annotation

Fig. 4. Receiver Operating Characteristic (ROC) curves and area under ROC curves (AUC) of different models trained using patient-level annotation (a) and image-level annotation (b) on the CC-CCII dataset.

Table 2
Comparison results of our CIFD-Net method vs. state-of-the-art architectures on the CC-CCII dataset.

Annotation	Method	Patient Acc. (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F_1-score (%)	AUC (%)
Patient-level	ResNet-50 [9]	$53.70^{**}_{\pm 0.02}$	$61.42^{**}_{\pm 0.08}$	$77.13^{**}_{\pm 0.10}$	$10.37^{**}_{\pm 0.17}$	$68.38^{**}_{\pm 0.05}$	$46.30^{**}_{\pm 0.10}$
	COVID-Net [3]	$53.62^{**}_{\pm 0.03}$	$61.35^{**}_{\pm 0.01}$	$77.18^{**}_{\pm 0.05}$	$10.06^{*}_{\pm 0.25}$	$68.36^{**}_{\pm 0.02}$	$44.53^{*}_{\pm 0.18}$
	COVNet [2]	$67.64^{**}_{\pm 0.04}$	$76.03^{*}_{\pm 0.08}$	$73.17^{**}_{\pm 0.07}$	$57.34^{*}_{\pm 0.18}$	$74.57^{**}_{\pm 0.04}$	$66.13^{*}_{\pm 0.15}$
	VB-Net [4]	$76.75^{*}_{\pm 0.04}$	$85.25^{**}_{\pm 0.10}$	$77.61^{**}_{\pm 0.07}$	$75.22^{*}_{\pm 0.19}$	$81.25^{**}_{\pm 0.05}$	$89.48^{*}_{\pm 0.16}$
	CIFD-Net (Ours)	$\mathbf{89.25}^{**}_{\pm 0.02}$	$\mathbf{89.98}^{*}_{\pm 0.13}$	$\mathbf{93.86}^{**}_{\pm 0.06}$	$\mathbf{80.67}^{*}_{\pm 0.13}$	$\mathbf{91.91}^{**}_{\pm 0.07}$	$\mathbf{93.22}^{**}_{\pm 0.06}$
Image-level	ResNet-50 [9]	$67.29^{**}_{\pm 0.04}$	$68.23^{**}_{\pm 0.06}$	$92.95^{**}_{\pm 0.05}$	$20.40^{**}_{\pm 0.16}$	$78.71^{**}_{\pm 0.02}$	$53.43^{**}_{\pm 0.11}$
	COVID-Net [3]	$64.83^{*}_{\pm 0.08}$	$66.28^{*}_{\pm 0.07}$	$\mathbf{93.18}^{**}_{\pm 0.02}$	$12.48^{**}_{\pm 0.04}$	$77.46^{**}_{\pm 0.03}$	$51.47^{**}_{\pm 0.09}$
	COVNet [2]	$70.79^{**}_{\pm 0.03}$	$83.09^{*}_{\pm 0.07}$	$68.95^{**}_{\pm 0.11}$	$74.10^{**}_{\pm 0.08}$	$75.37^{**}_{\pm 0.05}$	$73.08^{**}_{\pm 0.07}$
	CIFD-Net (Ours)	$\mathbf{84.83}^{**}_{\pm 0.02}$	$\mathbf{91.19}^{**}_{\pm 0.03}$	$84.74^{**}_{\pm 0.07}$	$\mathbf{84.99}^{**}_{\pm 0.11}$	$\mathbf{87.86}^{**}_{\pm 0.04}$	$\mathbf{89.63}^{**}_{\pm 0.08}$

* indicates the p-value < 0.05, and ** represents the p-value < 0.01.

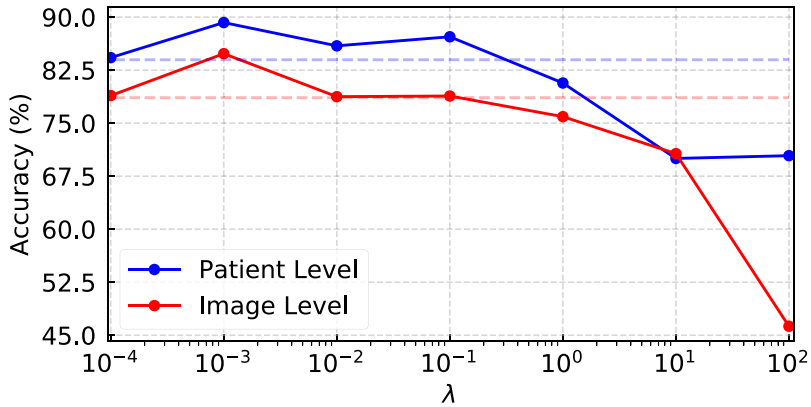


Fig. 5. Variations in classification results by changing the hyper-parameter λ. The light dash line represents the case when $\lambda = 0$. It shows that our model achieves the best performance with $\lambda = 1 \times 10^{-3}$.

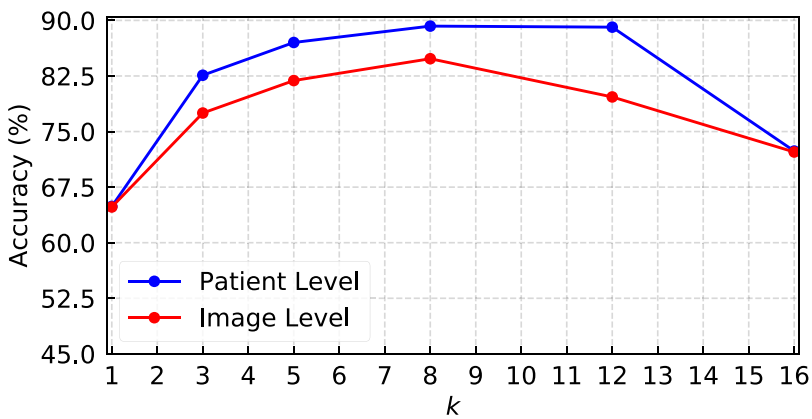


Fig. 6. Variations in classification results by changing the hyper-parameter k. Our model achieves the best performance with $k = 8$ with section size $l_s = 16$.

- The CIFD-Net outperforms most of the competing models by a large margin on the independent testing dataset, which can be attributed to the successful multi-domain shift problem proffered by our model. For the patient-level classification, our model is performed better than other compared methods by at least 12.5% on accuracy. Moreover, our model yields also the best performance on the image-level classification, which outperforms COVNet [2] by 14.1%. In addition, receiver operating characteristic (ROC) analysis and area under curves (AUC) results are obtained to quantify the classification performance. Our CIFD-Net achieves higher AUC value at both patient-level and image-level annotation compared to other state-of-the-art methods. Meanwhile, it is worth noticing that at the patient-level our method significantly outperforms other methods by at least 16.3% with respect to the sensitivity, which is an important indication for diagnosing COVID-19 positive cases.
- Models trained on patient-level, such as [2,4] and ours, achieve significant performance improvement than those trained on the image-level, i.e., [3,9], especially on the patient-level accuracy. This reflects that the image-level noise is non-trivial and can have a negative impact that these models can be overfitted because of the noise. Moreover, the models trained on the image-level may rely on learning the image textures [63], which are highly discriminative between domains. As a consequence, the models are prone to be overfitted and biased toward different textures while predicting, which may explain why these methods, e.g., methods proposed in [3,9], are poorly generalized to the unseen domains.
- Although methods proposed by Li et al. and Ouyang et al. [2, 4] also trained on the patient-level labels, our proposed CIFD-Net is superior to these methods, especially on the patient-level classification. The method proposed by Li et al. [2] performed the worst and this may because it has been trained on randomly selected CT images extracted from each 3D volume that may impede the encoding of lesions (often appearing adjacently between slices). In contrast, Ouyang et al. [4] preserved the sequential information among the CT slices because their method was trained on the whole CT volumes. In contrast, we take the full 3D volume into account and preserve the sequential information by dividing the volume into sections [2,4]. Besides, VB-Net achieves better performance than COVNet because VB-Net is trained with stronger supervision that is additional to the image level classifier. It also employs an auxiliary pixel-wise classifier trained with pixel-level infection annotation (i.e., infection segmentation mask). In comparison, our proposed model achieves better overall classification performance than VB-Net with weak supervision only.

We carried out the ROC analysis and the AUC results were used to quantify the classification performances as shown in Fig. 4. From Fig. 4(a), we can observe that the models trained only on image-level annotations (i.e., ResNet-50 and COVID-Net) are not reliable since their AUCs are less than 50%. In addition, we found that overall our CIFD-Net remains the best performed algorithm with an AUC of 93.22%. It is of note that the overall results at the patient-level are higher than those at the image-level. This could be correlated with our findings in the classification that some CT slices with few lesion parts are hard to diagnose and classify.

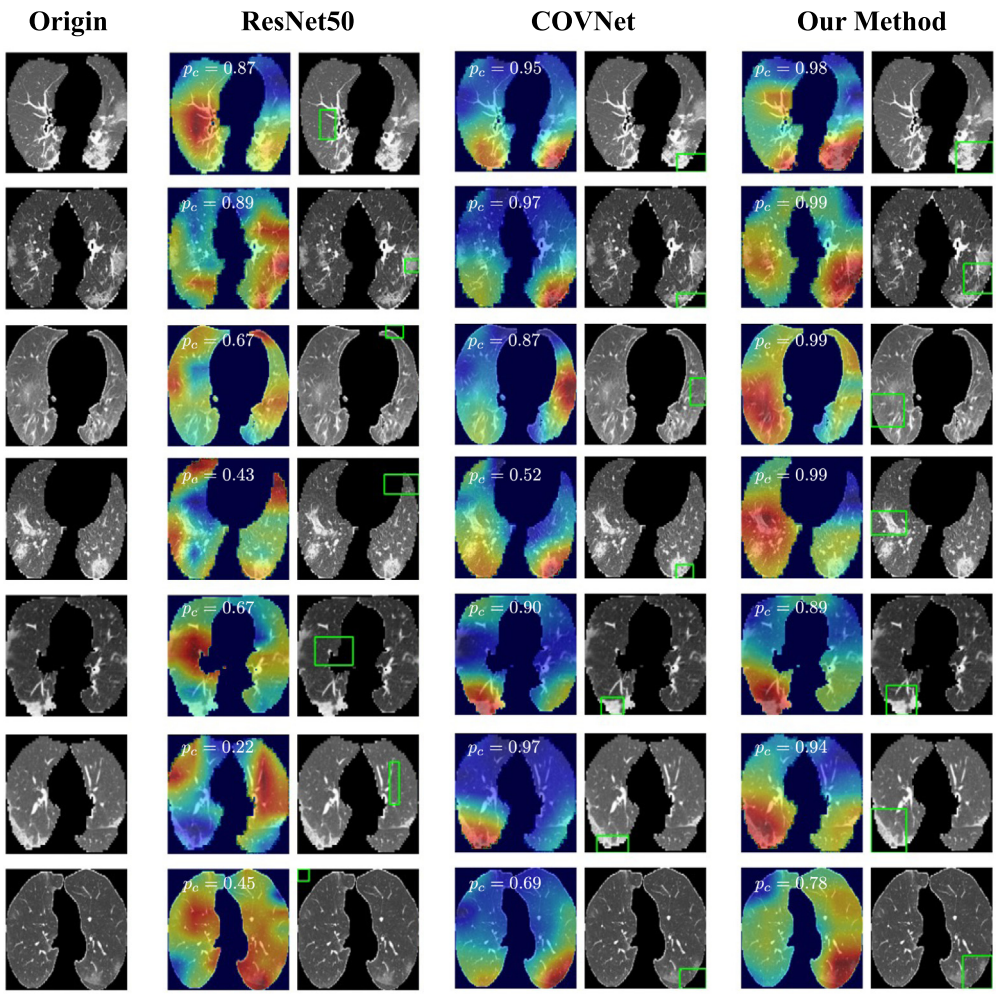


Fig. 7. Visualization of the CAMs and bounding boxes generated by different methods on the CC-CCII dataset. The region with a deeper red color indicates discriminative regions for the prediction by the model. p_c is the probability for being predicted as COVID-19 positive.

Table 3
Accuracy (%) of all the cases where each proposed component is applied.

Exp.	ResNet-50	$\mathcal{L}_{cls}$	$\mathcal{L}_{noisy}$	Patient Acc. (%)	Image Acc. (%)
1	√			53.72	67.31
2	√	√		83.97	78.60
3	√		√	35.10	35.16
4	√	√	√	**89.23**	**84.83**

To examine the influence of different loss terms, we conduct ablation studies on the proposed model and the results are reported in Table 3. As seen in the table, the model with the SNCM slightly outperforms the model without the SNCM on the patient-level. However, the SNCM advances the prediction at the image-level with significant improvement by 6.2% for the image accuracy. However, when only use the SNCM, the model would still be biased to predicting CT images tested negative because we only require our model to correct those CT images wrongly labeled as COVID-19 positive providing strong prior information to the training procedure.

Furthermore, we have examined the sensitivity of the choice of the hyper-parameters λ and k for our model. Fig. 5 shows the effect of the patient-level accuracy and the image-level accuracy while tuning the hyper-parameter λ. We can see that if λ is too large, the model would be biased and the performance would drop significantly since it acts as the regularization terms in the model training. The best results are obtained when $\lambda = 1 \times 10^{-3}$. In addition, for the selection of the hyper-parameter k, we can observe that when k is too large or too small, the performance degrades dramatically. This is because that if k is too large, the uncertainty of the section would increase and cause the noisy prediction. On the contrary, if k is too small (e.g., $k = 1$), some important slice information would be neglected, which leads to inaccurate results (see Fig. 6).

4.5. Qualitative results

For qualitative studies, we use the trained models (e.g., ResNet-50, COVNet, and others) to visualize the CAMs and bounding boxes on the test set. Fig. 7 presents the visualization of CAMs using our ECM. We can clearly see that the model trained on the slice-level (ResNet-50) tend to discard the lesions and focus on non-infected regions, and this also explains why it makes inaccurate and unreliable diagnosis decision causing trouble for radiologist use. On the contrary, models trained on patient-level, COVNet for instance, are able to detect some of the lesions occasionally but mostly failed in estimating the extent of the lesions reliably. In contrast, our model is not only precise in terms of lesion localization but also precise in estimating the extent of the infectious areas.

Moreover, based on the results of the CAMs, we extracted the bounding boxes using each method. It can be found that our CIFD-Net is able to yield more accurate bounding boxes on the

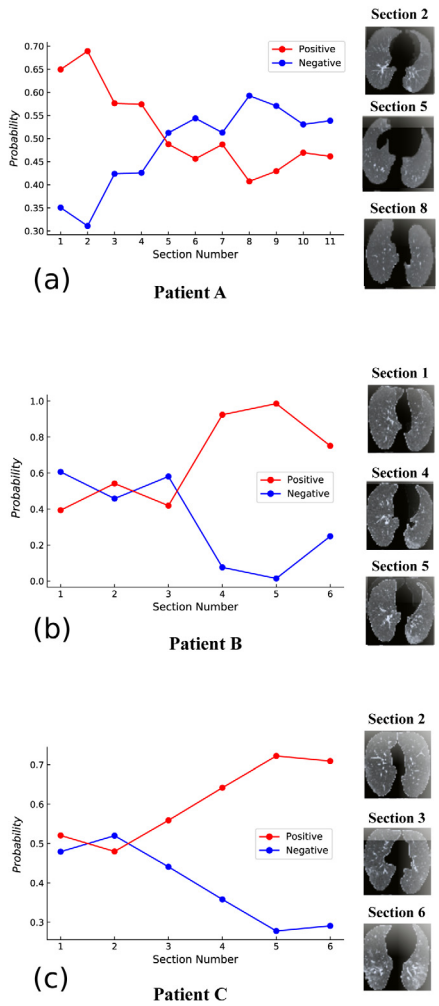


Fig. 8. Visualizations of infected/non-infected probabilities of each section for the patients. The *x*-axis of the plot is the section index of the patient. The right sub-figures of the probability plot are the picture sampled from the section listed above. (a) The first few sections are recognized as COVID-19 positive with high probability and when approaching the last few sections, no obvious lesions are found thus the positive probability drops drastically. (b) It shows that the probabilities of the first three sections are close to 0.5 indicating uncertainty for these sections. (c) For the last few sections, the lesions are gradually showing up in the left and right lower lobes together with increased infected probability.

salient part of the CAMs (Fig. 7) comparing to other methods, which indicates that our methods can be more applicable to perform auxiliary diagnosis. For instance, in diffusive cases (Fig. 7 rows 1 to 4), our CIFD-Net method has produced more accurate saliency maps compared to ResNet-50 and COVNet with less false positives and false negatives. Therefore, more precise localization (bounding boxes) have been generated. For the lesions distributed peripherally and subpleurally, both our CIFD-Net and COVNet approaches have performed better than the ResNet-50 (Fig. 7 rows 5 and 6). However, our CIFD-Net is more sensitive to the infectious regions that are not obvious in the images (Fig. 7 row 7).

In addition, we visualize the infection probability of lung sections for patients and sample the CT slices from corresponding sections. As illustrated in Fig. 8, the red curve depicts the infection probability varying along different lung sections, and the blue curve, on the opposite, depicts the non-infection probability for each section. Overall, it can be seen that the infected lung sections are distributed adjacently and the transition between the sections is smooth. Besides, we found our model is capable and robust of localizing where the infected lung sections are, regardless of the scale and the types of lesions. For example, for patient A Section 2, despite there is a very small lesion (GGO) peripherally, our model is still quite sensitive and is able to identify the infected section. Our model reaches around a saddle point, i.e., 0.5, when there are no apparent lesions detected, for instance, Section 1 for patient B and Section 2 for patient C.

4.6. Discussions

Our proposed CIFD-Net sequentially aggregates image-level features within a CT volume to alleviate the multi-domain shift problems, which turns out to be very effective and we have demonstrated that our CIFD-Net can be better generalized to unseen data domain compared to other state-of-the-art works. This may be attributed to (1) the *k*-max selection strategy: when optimizing the joint probability, only top-*k* probabilities within each section have been considered. Besides, those confounded images are not considered, which can result in a robust prediction; (2) our loss function is designed for modeling the joint probability of the patient instead of the individual image slice. Compared with the naive models, e.g., plain ResNet-50 trained on single image slice, our model is less likely to overfit on varied image styles and appearance, e.g., due to assorted textures and contrasts of the images, because our model takes into account the relationship between sections and the correlation between images in each section.

In addition, we integrated a novel slice noise correction module, i.e., SNCM, in the proposed CIFD-Net, which adds additional regularization to the optimization. Besides, we argue that this not only contributes to boosting the classification performance on the image-level prediction but also leads to more precise localization of lesions. However, since we trained the CIFD-Net under the assumption that CT slices are consecutive and lung segments (sections) are ordered, it may be difficult to handle disordered CT slices by using the slice aggregation, i.e., SAM and as a consequence, it may result in less accurate classification.

5. Conclusion

In this study, we have proposed a robust COVID-19 recognition model named CIFD-Net, which exploits the ECM to assist radiologists for auxiliary diagnosis. To handle the volume information, the model adopts the SAM to combine different sections for the sake of modeling the joint probability of the patient is COVID-19 positive or not. In addition, we extend our CIFD-Net incorporating the SNCM to predict a single CT slice without any image-level annotations. To investigate the prediction performance of the proposed model, we conducted comprehensive experiments on publicly available CT datasets. Experimental results have verified the superiority of our model, which can solve the multi-domain shift problem efficiently and effectively, compared to other state-of-the-art methods.

CRediT authorship contribution statement

Qinghao Ye: Conceived and designed the study, Literature search, Data analysis, Data interpretation, Contributed to the tables and figures, Writing of the report, Writing – review & editig. **Yuan Gao:** Literature search, Data analysis, Contributed to the tables and figures, Writing of the report. **Weiping Ding:** Literature search, Contributed to the tables and figures, Writing of the report. **Zhangming Niu:** Literature search. **Chengjia Wang:** Literature search. **Yinghui Jiang:** Literature search, Data analysis. **Minhao Wang:** Literature search, Data analysis. **Evandro Fei Fang:** Literature search. **Wade Menpes-Smith:** Literature search.

Jun Xia: Conceived and designed the study, Data collection, Data analysis, Data interpretation, Contributed to the tables and figures, Writing of the report. **Guang Yang:** Conceived and designed the study, Data analysis, Data interpretation, Contributed to the tables and figures, Writing of the report, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: QY, YJ, MW, is employed by Hangzhou Ocean's Smart Boya Co., Ltd., China. YG, ZN, WS, is employed by Aladdin Healthcare Technologies, Ltd., UK. YJ, MW, is employed by Mind Rank Ltd, Hongkong. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgment

This work was supported in part by the European Research Council Innovative Medicines Initiative on Development of Therapeutics and Diagnostics Combatting Coronavirus Infections Award 'DRAGON: rapiD and secuRe AI imaging based diaGnosis, stratification, fOllow-up, and preparedness for coronavirus paNdemics' [H2020-JTI-IMI2 101005122], the AI for Health Imaging Award 'CHAIMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management' [H2020-SC1-FA-DTS-2019-1 952 172], the Hangzhou Economic and Technological Development Area Strategical Grant [Imperial Institute of Advanced Technology], China, the British Heart Foundation, UK [TG/18/5/34111, PG/16/78/32402], the UK Research and Innovation Future Leaders Fellowship (MR/V023799/1), the SABER project, UK supported by Boehringer Ingelheim Ltd, the Project of Shenzhen International Cooperation Foundation, China (GJHZ20180926165402083), the Clinical Research Project of Shenzhen Health and Family Planning Commission, China (SZLY2018018), the National Natural Science Foundation of China, China (61976120), the Natural Science Foundation of Jiangsu Province, China (BK20191445), and the Natural Science Key Foundation of Jiangsu Education Department, China (21KJA510004). All authors approved the version of the manuscript to be published.

References

[1] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using CT images to screen for corona virus disease (COVID-19), MedRxiv (2020).

[2] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology (2020) 200905.

[3] L. Wang, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images, 2020, ArXiv preprint arXiv:2003.09871.

[4] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song, et al., Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia, 2020, ArXiv preprint arXiv:2005.02690.

[5] Y. Lin, S. Dong, Y. Yeh, Y. Wu, G. Lan, C. Liu, T.C. Chu, Emergency management and infection control in a radiology department during an outbreak of severe acute respiratory syndrome, British J. Radiol. 78 (931) (2005) 606–611.

[6] V. Vapnik, Principles of risk minimization for learning theory, in: Advances in Neural Information Processing Systems, 1992, pp. 831–838.

[7] D. Li, J.-B. Huang, Y. Li, S. Wang, M.-H. Yang, Weakly supervised object localization with progressive domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3512–3520.

[8] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang, et al., Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, Cell (2020).

[9] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[10] D. Ardila, A.P. Kiraly, S. Bharadwaj, B. Choi, J.J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, et al., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nature Med. 25 (6) (2019) 954–961.

[11] S. Lakshmanaprabu, S.N. Mohanty, K. Shankar, N. Arunkumar, G. Ramirez, Optimal deep learning model for classification of lung cancer on CT images, Future Gener. Comput. Syst. 92 (2019) 374–382.

[12] K.-L. Hua, C.-H. Hsu, S.C. Hidayati, W.-H. Cheng, Y.-J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, OncoTargets Therapy 8 (2015).

[13] H. Jiang, H. Ma, W. Qian, M. Gao, Y. Li, An automatic detection system of lung nodule based on multigroup patch-based deep learning network, IEEE J. Biomed. Health Inf. 22 (4) (2017) 1227–1237.

[14] A. Hosny, C. Parmar, T.P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, J. Bussink, R.J. Gillies, R.H. Mak, H.J. Aerts, Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study, PLoS Med. 15 (11) (2018) e1002711.

[15] A.A.A. Setio, A. Traverso, T. De Bel, M.S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M.E. Fantacci, B. Geurts, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge, Med. Image Anal. 42 (2017) 1–13.

[16] M. Nishio, O. Sugiyama, M. Yakami, S. Ueno, T. Kubo, T. Kuroda, K. Togashi, Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning, PLoS One 13 (7) (2018) e0200721.

[17] S.L. Walsh, L. Calandriello, M. Silva, N. Sverzellati, Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study, Lancet Respir. Med. 6 (11) (2018) 837–845.

[18] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, IEEE Trans. Med. Imaging 35 (5) (2016) 1207–1216.

[19] T. Pang, S. Guo, X. Zhang, L. Zhao, Automatic lung segmentation based on texture and deep features of hrct images with interstitial lung disease, BioMed Res. Int. 2019 (2019).

[20] B. Park, H. Park, S.M. Lee, J.B. Seo, N. Kim, Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks, J. Digital Imag. 32 (6) (2019) 1019–1026.

[21] Q. Ye, D. Tu, F. Qin, Z. Wu, Y. Peng, S. Shen, Dual attention based fine-grained leukocyte recognition for imbalanced microscopic images, J. Intell. Fuzzy Systems 37 (5) (2019) 6971–6982.

[22] T.-C. Pham, C.-M. Luong, M. Visani, V.-D. Hoang, Deep CNN and data augmentation for skin lesion classification, in: Asian Conference on Intelligent Information and Database Systems, Springer, 2018, pp. 573–582.

[23] L. Wynants, B.V. Calster, M.M. Bonten, G.S. Collins, T.P. Debray, M.D. Vos, M.C. Haller, G. Heinze, K.G. Moons, R.D. Riley, E. Schuit, L. Smits, K.I. Snell, E.W. Steyerberg, C. Wallisch, M.v. Smeden, Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection, 2020.03.24.20041020, http://dx.doi.org/10.1101/2020.03.24.20041020, URL https://www.medrxiv.org/content/10.1101/2020.03.24.20041020v1.

[24] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19, p. 1, http://dx.doi.org/10.1109/RBME.2020.2987975, Conference Name: IEEE Reviews in Biomedical Engineering.

[25] Q. Ye, J. Xia, G. Yang, Explainable AI for COVID-19 CT classifiers: An initial comparison study, in: J. ao Rafael Almeida, A.R. González, L. Shen, B. Kane, A. Traina, P. Soda, J.L. Oliveira (Eds.), 34th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2021, Aveiro, Portugal, June 7-9, 2021, IEEE, 2021, pp. 521–526, http://dx.doi.org/10.1109/CBMS52027.2021.00103.

[26] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Inf. Fusion 77 (2022) 29–52, http://dx.doi.org/10.1016/j.inffus.2021.07.016.

[27] H. Ma, Q. Ye, W. Ding, Y. Jiang, M. Wang, Z. Niu, X. Zhou, Y. Gao, C. Wang, W. Menpes-Smith, et al., Can clinical symptoms and laboratory results predict CT abnormality? initial findings using novel machine learning techniques in children with COVID-19 infections, Front. Med. 8 (2021) 855.

[28] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, et al., AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks, MedRxiv (2020).

[29] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, Y. Shi, Lung infection quantification of covid-19 in ct images with deep learning, 2020, ArXiv preprint arXiv:2003.04655.
[30] S. Hu, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, E.F. Fang, W. Menpes-Smith, J. Xia, et al., Weakly supervised deep learning for covid-19 infection detection and classification from ct images, IEEE Access (2020).
[31] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest ct assessment of covid-19: Deep-learning approach, Radiol. Cardiothoracic Imag. 2 (2) (2020) e200075.
[32] M.S. Takahashi, M. Ribeiro Furtado de Mendonça, I. Pan, R.Z. Pinetti, F.C. Kitamura, Regarding "serial quantitative chest CT assessment of COVID-19: Deep-learning approach", Radiol. Cardiothoracic Imag. 2 (3) (2020) e200242.
[33] K. Murphy, H. Smits, A.J. Knoops, M.B. Korst, T. Samson, E.T. Scholten, S. Schalekamp, C.M. Schaefer-Prokop, R.H. Philipsen, A. Meijers, et al., COVID-19 on the chest radiograph: A multi-reader evaluation of an AI system, Radiology (2020) 201874.
[34] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. (2020) 103792.
[35] Y. Oh, S. Park, J.C. Ye, Deep learning covid-19 features on cxr using limited training data sets, IEEE Trans. Med. Imaging (2020).
[36] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. (2020) 1.
[37] C. Butt, J. Gill, D. Chun, B.A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia, Appl. Intell. (2020) 1.
[38] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, MedRxiv (2020).
[39] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, MedRxiv (2020).
[40] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for COVID-19 from chest CT using weak label, MedRxiv (2020).
[41] X. Mei, H.-C. Lee, K.-y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P.M. Robson, M. Chung, et al., Artificial intelligence–enabled rapid diagnosis of patients with COVID-19, Nature Med. (2020) 1–5.
[42] D. Driggs, I. Selby, M. Roberts, E. Gkrania-Klotsas, J.H. Rudd, G. Yang, J. Babar, E. Sala, C.-B. Schönlieb, A.-C. collaboration, Machine Learning for COVID-19 Diagnosis and Prognostication: Lessons for Amplifying the Signal While Reducing the Noise, Radiological Society of North America, 2021.
[43] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A.I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, et al., Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nat. Mach. Intell. 3 (3) (2021) 199–217.
[44] W. Sultani, C. Chen, M. Shah, Real-world anomaly detection in surveillance videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479–6488.
[45] Y. Xu, J.-Y. Zhu, I. Eric, C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Med. Image Anal. 18 (3) (2014) 591–604.
[46] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Is object localization for free?-weakly-supervised learning with convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 685–694.
[47] M. Ilse, J.M. Tomczak, M. Welling, Attention-based deep multiple instance learning, in: 35th International Conference on Machine Learning, ICML 2018, International Machine Learning Society (IMLS), 2018, pp. 3376–3391.
[48] C. Chen, Q. Dou, H. Chen, J. Qin, P.-A. Heng, Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33(01), 2019, pp. 865–872.
[49] Y. Xia, D. Yang, Z. Yu, F. Liu, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, H. Roth, Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation, Med. Image Anal. 65 (2020) 101766.
[50] J. Chen, H. Zhang, R. Mohiaddin, T. Wong, D. Firmin, J. Keegan, G. Yang, Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data, 2021, ArXiv preprint arXiv:2109.08311.
[51] J. Hoffman, S. Guadarrama, E.S. Tzeng, R. Hu, J. Donahue, R. Girshick, T. Darrell, K. Saenko, LSDA: Large scale detection through adaptation, in: Advances in Neural Information Processing Systems, 2014, pp. 3536–3544.
[52] J. Hoffman, D. Pathak, T. Darrell, K. Saenko, Detector discovery in the wild: Joint multiple instance and representation learning, in: Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition, 2015, pp. 2883–2891.
[53] F. Mahmood, R. Chen, S. Sudarsky, D. Yu, N.J. Durr, Deep learning with cinematic rendering: fine-tuning deep neural networks using photorealistic medical images, Phys. Med. Biol. 63 (18) (2018) 185012.
[54] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[55] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[56] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, ArXiv preprint arXiv:1409.1556.
[57] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[58] Z.-H. Zhou, Multi-instance learning: A survey, Tech. Rep 2, Department of Computer Science & Technology, Nanjing University, 2004.
[59] A.J. Bekker, J. Goldberger, Training deep neural-networks based on unreliable labels, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2016, pp. 2682–2686.
[60] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[61] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.
[62] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, ArXiv preprint arXiv:1412.6980.
[63] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F.A. Wichmann, W. Brendel, Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, 2018, ArXiv preprint arXiv:1811.12231.