SODA: Detecting COVID-19 in Chest X-Rays With Semi-Supervised Open Set Domain Adaptation (Invited Paper)

Jieli Zhou, Baoyu Jing, Zeya Wang, Hongyi Xin, and Hanghang Tong

Abstract—Due to the shortage of COVID-19 viral testing kits, radiology imaging is used to complement the screening process. Deep learning based methods are promising in automatically detecting COVID-19 disease in chest x-ray images. Most of these works first train a Convolutional Neural Network (CNN) on an existing large-scale chest x-ray image dataset and then fine-tune the model on the newly collected COVID-19 chest x-ray dataset, often at a much smaller scale. However, simple fine-tuning may lead to poor performance for the CNN model due to two issues, first the large domain shift present in chest x-ray datasets and second the relatively small scale of the COVID-19 chest x-ray dataset. In an attempt to address these two important issues, we formulate the problem of COVID-19 chest x-ray image classification in a semi-supervised open set domain adaptation setting and propose a novel domain adaptation method, Semi-supervised Open set Domain Adversarial network (SODA). SODA is designed to align the data distributions across different domains in the general domain space and also in the common subspace of source and target data. In our experiments, SODA achieves a leading classification performance compared with recent state-of-the-art models in separating COVID-19 with common pneumonia. We also present initial results showing that SODA can produce better pathology localizations in the chest x-rays.

Index Terms—COVID-19, medical image analysis, domain adaptation, open set domain adaptation, semi-supervised learning

1 INTRODUCTION

Since the Coronavirus disease 2019 (COVID-19) was first declared as a Public Emergency of International Concern (PHEIC) on January 30, 2020,1 it has quickly evolved from a local outbreak in Wuhan, China to a global pandemic, taking away millions of lives and causing dire economic loss worldwide. In the US, the total COVID-19 cases grew from just one confirmed on Jan 21, 2020 to over 1 million on April 28, 2020 in a span of 3 months. Despite drastic actions like shelter-in-place and contact tracing, the total cases in US kept increasing at an alarming daily rate of 20000–30000 throughout April, 2020. A key challenge for preventing and controlling COVID-19 right now is the ability to quickly, widely and effectively test for the disease, since testing is usually the first step in a series of actions to break the chains of transmission and curb the spread of the disease. COVID-19 is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).2 By far, the most reliable diagnosis is through Reverse Transcription Polymerase Chain Reaction (RT-PCR)3 in which a sample is taken from the back of throat or nose of the patients and tested for viral RNA. Once the sample is collected, the testing process usually takes several hours and recent study reports that the sensitivity of RT-PCR is around 60–70 percent [1], which suggests that many people tested negative for the virus may actually carry it thus could infect more people without knowing it. On the other hand, the sensitivity of chest radiology imaging for COVID-19 was much higher at 97 percent as reported by [1], [2].

Due to the shortage of viral testing kits, the long period of waiting for results, and low sensitivity rate of RT-PCR, radiology imaging has been used as a complementary screening process to assist the diagnosis of COVID-19. Furthermore, radiology imaging can also provide more detailed information about the patients, e.g., pathology location, lesion size, the severity of lung involvement [3]. These

1. https://www.statnews.com/2020/01/30/who-declares-coronavirus-outbreak-a-global-health-emergency/

2. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

3. https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/how-do-coronavirus-tests-work
insights can help doctors to timely triage patients into different risk levels, bring patients in severe conditions to ICU earlier, and saving more lives.

In recent years, with the rapid advancement in deep learning and computer vision, many breakthroughs have been developed in using Artificial Intelligence (AI) for medical imaging analysis, especially disease detection [4], [5], [6] and report generation [7], [8], [9], [10], and some AI models achieve expert radiologist-level performance [11]. Right now, with healthcare workers busy at front lines saving lives, the scalability advantage of AI-based medical imaging systems stand out more than ever. Some AI-based chest imaging systems have already been deployed in hospitals to quickly inform healthcare workers to take actions accordingly.4

Annotated datasets are required for training AI-based methods, and a small chest x-ray dataset with COVID-19 is collected recently: COVID-ChestXray [12]. Shortly after the COVID-19 outbreak, several works [13], [14], [15], [16] apply Convolutional Neural Networks (CNN) and transfer learning to detect COVID-19 cases from chest x-ray images. They first train a CNN on a large dataset like CheXpert [5] and ChestXray14 [4], and then fine-tune the model on the small COVID-19 dataset. By far, due to the lack of large-scale open COVID-19 chest x-ray imaging datasets, most works only used a very small amount of positive COVID-19 imaging samples [12]. While the reported performance metrics like accuracy and AUC-ROC are high, it is likely that these models overfit on this small dataset and may not achieve the reported performance on a different and larger COVID-19 x-ray dataset. Besides, these methods suffer a lot from label domain shift: these newly trained models lose the ability to detect common thoracic diseases like “Effusion” and “Nodule” since these labels do not appear in the new dataset. Moreover, they also ignored the visual domain shift between the two datasets. On the one hand, the large-scale datasets like ChestXray14 [4] and CheXpert [5] are collected from top U.S. health institutions like National Institutes of Health (NIH) clinical center and Stanford University, which are well-annotated and carefully processed. On the other hand, COVID-ChestXray [12] is collected from a very diverse set of hospitals around the world and they are of very different qualities and follow different standards, such as the viewpoints, aspect ratios and lighting, etc. In addition, COVID-ChestXray contains not only chest x-ray images but also CT scan images.

In order to fully exploit the limited but valuable annotated COVID-19 chest x-ray images and the large-scale chest x-ray image dataset at hand, as well as to prevent the above-mentioned drawbacks of those fine-tuning based methods, we define the problem of learning a x-ray classifier for COVID-19 from the perspective of open set domain adaptation (Definition 1) [17]. Different from traditional unsupervised domain adaptation which requires the label set of both source and target domain to be the same, the open set domain adaptation allows different domains to have different label sets. This is more suitable for our problem because COVID-19 is a new disease which is not included in the ChestXray14 or CheXpert dataset. However, since our task is to train a new classifier for COVID-19 dataset, we have to use some annotated samples. Therefore, we further propose to view the problem as a Semi-Supervised Open Set Domain Adaptation problem (Definition 2).

Under the given problem setting, we propose a novel Semi-supervised Open set Domain Adversarial network (SODA) comprised of four major components: a feature extractor \( G_f \), a multi-label classifier \( G_w \), domain discriminators \( D_g \) and \( D_{dg} \), as well as common label recognizer \( R \). SODA learns the domain-invariant features by a two-level alignment, namely, domain level and common label level. The general domain discriminator \( D_g \) is responsible for guiding the feature extractor \( G_f \) to extract domain-invariant features. However, it has been argued that the general domain discriminator \( D_g \) might lead to false alignment and even negative transfer [18], [19]. For example, it is possible that the feature extractor \( G_f \) maps images with “Pneumonia” in the target domain and images with “Cardiomegaly” in the source domain into similar positions, which might result in the mis-classification of \( G_w \). In order to solve this problem, we propose a novel common label discriminator \( D_s \) to guide the model to align images with common labels across domains. For labeled images, \( D_s \) only activates when the input image is associated with a common label. For unlabeled images, we propose a common label recognizer \( R \) to predict their probabilities of having a common label.

The main contributions of the paper are summarized as follows:

- To the best of our knowledge, we are the first to tackle the problem of COVID-19 chest x-ray image classification from the perspective of domain adaptation.
- We formulate the problem in a novel semi-supervised open set domain adaptation setting.
- We propose a novel two-level alignment model: Semi-supervised Open set Domain Adversarial network (SODA).
- We present a comprehensive evaluation to demonstrate the effectiveness of the proposed SODA.

2 PRELIMINARY

2.1 Problem Definition

**Definition 1. Unsupervised Open Set Domain Adaptation**

Let \( x \) be the input chest x-ray image, \( y \) be the ground-truth disease label. We define \( D^s = \{(x^s_i, y^s_i)\}_{i=1}^{N^s} \) as a source domain with \( N^s \) labeled samples, and \( D^t = \{(x^t_i)\}_{i=1}^{N^t} \) as a target domain with \( N^t \) unlabeled samples, where the underlying label set \( \mathcal{L}^t \) of the target domain might be different from the label set \( \mathcal{L}^s \) of the source domain. Define \( \mathcal{L}^t = \mathcal{L}^s \cap \mathcal{L}^t \) as the set of common labels shared across different domains, \( \mathcal{L}^s = \mathcal{L}^t \setminus \mathcal{L}^s \) and \( \mathcal{L}^t = \mathcal{L}^t \setminus \mathcal{L}^s \) be the sets of domain-specific labels which only appear in the source and the target domain. The task of Unsupervised Open Set Domain Adaptation is to build a model which could accurately assign common labels in \( \mathcal{L}^c \) to samples \( x^t_i \) in the target domain, as well as distinguish those \( x^t_i \) belonging to \( \mathcal{L}^t \).

**Definition 2. Semi-supervised Open Set Domain Adaptation**

Given a source domain \( D^s = \{(x^s_i, y^s_i)\}_{i=1}^{N^s} \) with \( N^s \) labeled samples, and a target domain \( D^t \cup D^\prime \) consisting of \( D^t = \{(x^t_i)\}_{i=1}^{N^t} \) with \( N^t \) unlabeled samples and \( D^\prime = \{(x^t_i, y^t_i)\}_{i=1}^{N^t} \)
TABLE 1
Notations

| Symbols | Description |
|---------|-------------|
| \(D^r\) | set of labeled samples in the source domain |
| \(D'^r\) | set of unlabeled samples in the target domain |
| \(D^s\) | set of labeled samples in the target domain |
| \(D'^s\) | set of labeled samples in the domain |
| \(\mathcal{L}^c\) | set of common labels across domains |
| \(\mathcal{L}^t\) | set of domain-specific labels in the target domain |
| \(\mathcal{L}^e\) | set of domain-specific labels in the source domain |
| \(L_z\) | set of labels for the target domain |
| \(N_z\) | number of labeled samples in the source domain |
| \(N_e\) | number of labeled samples in the target domain |
| \(G_f\) | feature extractor |
| \(G_g\) | multi-label classifier for \(\mathcal{L}\) |
| \(G_{\text{w}}\) | binary classifier for label \(l\) (part of \(G_g\)) |
| \(R\) | common label recognizer |
| \(D_h\) | domain discriminator for common labels \(\mathcal{L}^c\) |
| \(L_{D_g}\) | general domain discriminator |
| \(L_{G_{\text{c}}}\) | loss of multi-label classification over the entire dataset |
| \(L_R\) | loss of \(R\) over the entire dataset |
| \(L_{D_s}\) | loss of \(D_s\) over the entire dataset |
| \(\lambda\) | the coefficient of losses |
| \(\mathbf{x}\) | input image |
| \(\mathbf{h}\) | hidden features |
| \(y\) | ground-truth label |
| \(\hat{y}\) | predicted probability |
| \(\hat{d}\) | predicted probability that \(x\) belongs to source domain |
| \(\hat{\tau}\) | predicted probability that \(x\) has common labels |

with \(N'^t\) labeled samples. The task of Semi-supervised Open Set Domain Adaptation is to build a model to assign labels from \(\mathcal{L}'\) to unlabeled samples in \(D'^r\).

2.2 Notations

We summarize the symbols used in the paper in Table 1.

3 METHODOLOGY

3.1 Overview

An overview of the proposed Semi-supervised Open Set Domain Adversarial network (SODA) is shown in Fig. 1. Given an input image \(x\), it will be first fed into a feature extractor \(G_f\), which is a Convolutional Neural Network (CNN), to obtain its hidden feature \(h\) (green part). The binary classifier \(G_{\text{w}}\) (part of the multi-label classifier \(G_g\)) takes \(h\) as input, and will predict the probability \(\hat{y}\) for the label \(l \in \mathcal{L}\) (blue part).

We propose a novel two-level alignment strategy for extracting the domain invariant features across the source and target domain. On one hand, we perform domain alignment (Section 3.2), which leverages a general domain discriminator \(D_h\) to minimize the domain-level feature discrepancy. On the other hand, we emphasize the alignment of common labels \(\mathcal{L}'\) (Section 3.3) by introducing another domain discriminator \(D_c\) for images associated with common labels. For labeled images in \(D^r\) and \(D'^r\), we compute loss for \(D_c\) and conduct back-propagation [20] during training only if the input image \(x\) is associated with a common label \(l \in \mathcal{L}'\). As for unlabeled data in \(D'^r\), we propose a common label recognizer \(R\) to predict the probability \(\hat{\tau}\) that an image \(x\) has a common label, and use \(\hat{\tau}\) as a weight in the losses of \(D_c\) and \(D_g\).

3.2 Domain Alignment

Domain adversarial training [21] is the most popular method for helping feature extractor \(G_f\) learn domain-invariant features such that the model trained on the source domain can be easily applied to the target domain. The objective function of the domain discriminator \(D_h\) is:

\[
L_{D_h} = -\mathbb{E}_{(x', y') \in D'^r} \left[ \log \hat{d}_{y'} \right] - \mathbb{E}_{(x, y) \in D^r} \left[ \log (1 - \hat{d}) \right],
\]

where \(\hat{d}\) denotes the predicted probability that the input image belongs to the source domain.

We use a Multi-Layer Perceptron (MLP) as the general domain discriminator \(D_g\).

3.3 Common Label Alignment

In the field of adversarial domain adaptation, most of the existing methods only leverage a general domain discriminator \(D_h\) to minimize the discrepancy between the source and target domain. Such a practice ignores the label structure across domains, which will result in false alignment and even negative transfer [18], [19]. If we only use a general domain discriminator \(D_h\) in the open set domain adaptation setting (Definitions 1 and 2), it is possible that the feature extractor \(G_f\) will map the target domain images with a common label \(l \in \mathcal{L}'\), "Pneumonia," and the source domain images with a specific label \(l \in \mathcal{L}\), e.g., "Cardiomegaly," to similar positions in the hidden space, which might lead to the classifier misclassifying a "Pneumonia" image in the target domain as "Cardiomegaly".

To address the problem of the miss-matching between the common and specific label sets, we propose a domain discriminator \(D_c\) to distinguish the domains for the images with a common label. For the labeled data from the source domain \(D^r\) and the target domain \(D'^r\), we know whether an image \(x\) has a common label or not, and we only calculate the loss \(L_{D_c}\) for \(D_c\) on the samples with common labels:

\[
L_{D_c}^{\text{label}} = -\mathbb{E}_{(x', y', y'' \in \mathcal{L}') \in D'^r} \left[ \log \hat{d}_{y''} \right] - \mathbb{E}_{(x, y' \in \mathcal{L}') \in D^r} \left[ \log (1 - \hat{d}) \right],
\]

where \(\hat{d}\) denotes the predicted probability that the input images is associated with a common label.

However, a large number of images in the target domain are unlabeled, and thus extra effort is required for determining whether an unlabeled image is associated with a common label. To address this problem, we propose a novel common label recognizer \(R\) to predict the probability \(\hat{\tau}\) whether an unlabeled image has at least one common label. The probability \(\hat{\tau}\) will be used as a weight in the loss function of \(D_c^{\text{cl}}\):

\[
L_{D_c}^{\text{cl}} = -\mathbb{E}_{(x', y' \in \mathcal{L}') \in D'^r} \left[ \hat{\tau} \log (1 - \hat{d}) \right].
\]

Note that gradients stop at \(\hat{\tau}\) in the training period.
To re-weigh unlabeled samples are not allowed to pass through \( \mathcal{L} \) (Eq. (1)) to further emphasize the alignment of common labels, the probability \( \hat{y} \) that \( x \) is associated with a common label and the probability \( \hat{d} \) that \( x \) belongs to the source domain. \( L_y \) and \( L_d \) denote the losses of image classification, common label classification and domain classification. \( D_c \) is the general domain discriminator, and \( D_l \) is the domain discriminator for images associated with a common label. \( G_y \) denotes the image classifier for the first label in the label set of the entire dataset \( \mathcal{L} = \mathcal{L}' \cup \mathcal{L}'' \). Note that the gradients from \( L_y \) and \( L_d \) are not allowed to pass through \( \hat{r} \) (grey arrows).

In addition, we also use \( \hat{r} \) to re-weigh unlabeled samples in \( D_g \) (Eq. (1)) to further emphasize the alignment of common labels:

\[
L_{D_g} = -\mathbb{E}_{(x' \in \mathcal{D}')} [\log \hat{d}] - \mathbb{E}_{(x' \in \mathcal{D}')} [\log (1 - \hat{d})] - \mathbb{E}_{(x' \in \mathcal{D}')} [\log (1 - \hat{y})].
\]

(4)

Finally, the recognizer \( R \) is trained on the labeled set \( \mathcal{D}' \cup \mathcal{L}' \) via cross-entropy loss:

\[
L_R = -\mathbb{E}_{(x \in \mathcal{D}'\cup\mathcal{L}' \mid y \in \mathcal{L}')} [\log \hat{r}] - \mathbb{E}_{(x \in \mathcal{D}'\cup\mathcal{L}' \mid y \notin \mathcal{L}')} [\log (1 - \hat{r})].
\]

(5)

### 3.4 Overall Objective Function

The overall objective function of SODA is a min-max game between classifiers \( G_y, R \) and discriminators \( D_y, D_c \):

\[
\min_{G_y} \max_{R, D_y, D_c} L_{G_y} + \lambda_R L_R - \lambda_D L_{D_g} - \lambda_{label} L_{label} - \lambda_{un} L_{un}.
\]

(6)

where \( L_{G_y}, L_R, L_{D_g}, L_{label} \) and \( L_{un} \) are defined in Eqs. (5), (4), (2) and (3); \( L_{G_y} \) denotes the cross-entropy loss for multi-label classification; \( \lambda \) denotes the coefficient of different loss functions.

### 4 EXPERIMENTS

#### 4.1 Experiment Setup

**4.1.1 Dataset**

**Source Domain.** We use ChestXray-14 [4] as the source domain dataset. This dataset is comprised of 112 120 anonymized chest x-ray images from the National Institutes of Health (NIH) clinical center. The dataset contains 14 common thoracic disease labels: “Atelectasis,” “Consolidation,” “Infiltration,” “Pneumothorax,” “Edema,” “Emphysema,” “Fibrosis,” “Effusion,” “Pneumonia,” “Pleural thickening,” “Cardiomegaly,” “Nodule,” “Mass” and “Hernia”. **Target Domain** The newly collected COVID-ChestXray [12] is adopted as the target domain dataset, which contains images collected from various public sources and different hospitals around the world. This dataset (by the time of this writing) contains 328 chest x-ray images in which 253 are labeled positive as the new disease “COVID-19,” whereas 61 are labeled as other well-studied “Pneumonia”.

**4.1.2 Evaluation Metrics**

We evaluate our model from four different perspectives. First, to test the classification performance, following the semisupervised protocol, we randomly split the 328 x-ray images in COVID-ChestXray into 40 percent labeled set, and 60 percent unlabeled set. We report the AUC-ROC score for each label in the target domain. Second, we compute the Proxy-A Distance (PAD) [22] to evaluate models’ ability for minimizing the feature discrepancy across domains. Third, we use t-SNE to visualize the feature distributions of the target domain. Finally, we also qualitatively evaluate the models by visualizing their saliency maps.

**4.1.3 Baseline Methods**

We compare SODA with two types of baselines methods: fine-tuning based transfer learning models and domain adaptation models. For fine-tuning based models, we select the two most popular CNN models DenseNet121 [23] and ResNet50 [24] as our baselines. These models are first trained on the ChestXray-14 dataset and then fine-tuned on the COVID-ChestXray dataset. For domain adaptation models, we compare our model with two classic models, Domain Adversarial Neural Networks (DANN) [21] and Partial Adversarial Domain Adaptation (PADA) [25]. Note that DANN and PADA were designed for unsupervised domain adaptation, and we implement a semi-supervised version of them.

**4.1.4 Implementation Details**

We use DenseNet121 [23], which is pretrained on the ChestXray-14 dataset [4], as the feature extractor \( G_f \) for...
SODA. The multi-label classifier $G_y$ is a one layer neural network and its activation is the sigmoid function. We use the same architecture for $D_g$, $D_c$ and $R$: a MLP containing two hidden layers with ReLU activation and an output layer. The hidden dimension for all of the modules: $G_y$, $D_g$, $D_c$ and $R$ is 1024. For fair comparison, we use the same setting of $G_f$, $G_y$ and $D_g$ for DANN [21] and PADA [25]. All of the models are trained by Adam optimizer [27], and the learning rate is $10^{-4}$.

4.2 Classification Results

To investigate the effects of domain adaptation and demonstrate the performance improvement of the proposed SODA, we present the average AUC-ROC scores for all models in Table 2. Comparing the results for ResNet50 and DenseNet121, we observe that deeper and more complex models achieve better classification performance. For the effects of domain adaptation, it is obvious that the domain adaptation methods (DANN, PADA, and SODA) outperform those fine-tuning based transfer learning methods (ResNet50 and DenseNet121). Furthermore, the proposed SODA achieves higher AUC scores on both COVID-19 and Pneumonia than DANN and PADA, demonstrating the effectiveness of the proposed two-level alignment.

4.3 Feature Visualization

We use t-SNE to project the high dimensional hidden features $h$ extracted by DANN, PADA, and SODA to low dimensional space. The 2-dimensional visualization of the features in the target domain is presented in Fig. 2, where the red data points are image features of “Pneumonia” and the blue data points are image features of “COVID-19”. It can be observed from Fig. 2 that SODA performs the best for separating “COVID-19” from “Pneumonia,” which demonstrates the effectiveness of the proposed common label recognizer $R$ as well as the domain discriminator for common labels $D_c$.

4.4 Proxy $\mathcal{A}$-Distance

Proxy $\mathcal{A}$-Distance (PAD) [22] has been widely used in domain adaptation for measuring the feature distribution discrepancy between the source and target domains:

$$\text{PAD} = 2(1 - \min(C)),$$

where $C$ is the domain classification error (e.g., mean absolute error) of a classifier (e.g., linear SVM [28]).

Following [21], we train SVM models with different $C$ and use the minimum error to calculate PAD. In general, a lower PAD means a better ability for extracting domain invariant features. As shown in Fig. 3, SODA has a lower PAD compared with the baseline methods, which indicates the effectiveness of the proposed two-level alignment strategy.

5 RELATED WORK

5.1 Domain Adaptation

Domain adaptation is an important application of transfer learning that attempts to generalize the models from source domains to unseen target domains [19], [21], [30], [31], [32], [33], [34], [35]. Adversarial training, inspired by the success of generative adversarial modeling [36], has been widely

| Model       | COVID-19 | Pneumonia |
|-------------|----------|-----------|
| ResNet50 [24] | 0.8143   | 0.8342    |
| DenseNet121 [23] | 0.8202   | 0.8414    |
| DANN [21]     | 0.8785   | 0.8961    |
| PADA [25]     | 0.8822   | 0.9038    |
| SODA          | 0.9006   | 0.9082    |

Table 2: Target Domain Average AUC-ROC Score

Fig. 2. t-SNE visualization for DANN, PADA and SODA on the target domain.
applied for promoting the learning of transfer features in image classification. It takes advantage of a domain discriminator to classify whether an image is from the source or target domains. Recently, researchers have started to study the open set domain adaptation problem, where the target domain has images that do not come from the classes in the source domain [17], [33]. Universal domain adaptation is the latest method that is proposed through using an adversarial domain discriminator and a non-adversarial domain discriminator to successfully solve this problem [33]. Although domain adaptation has been well explored, its application in medical imaging analysis, such as domain adaptation for chest x-ray images, is still under-explored.

![Proxy A-Distance](image1.png)

**Fig. 3.** Proxy A-Distance.

![Grad-CAM visualization](image2.png)

**Fig. 4.** Grad-CAM [29] visualization for ResNet50, DenseNet121, DANN, PADA, and SODA.
5.2 Chest X-Ray Image Analysis

There has been substantial progress in constructing publicly available databases for chest x-ray images as well as a related line of works to identify lung diseases using these images. The largest public datasets of chest x-ray images are CheXpert [5] and ChestXray14 [4], which respectively include more than 200,000 and 100,000 chest x-ray images collected by Stanford University and National Institute of Healthcare. The creation of these datasets have also motivated and promoted the multi-label chest x-ray classification for helping the screening and diagnosis of various lung diseases. The problems of disease detection [4], [5], [6] and report generation using chest x-rays [7], [8], [9], [10] are investigated recently and have achieved much-improved results upon recently. However, there are very few attempt for studying the domain adaptation problems with the multi-label image classification problem using chest x-rays.

6 Conclusion

In this paper, in order to assist and complement the screening and diagnosing of COVID-19, we formulate the problem of COVID-19 chest x-ray image classification within a semi-supervised open set domain adaptation framework. We propose a novel deep domain adversarial neural network, Semi-supervised Open set Domain Adversarial network (SODA), which is able to align the data distributions across different domains at both domain level and common label level. Through evaluations of the classification accuracy, we show that SODA achieves better AUC-ROC scores than the recent state-of-the-art models. We further demonstrate that the features extracted by SODA is more tightly related to the lung pathology locations, and get initial positive feedback from an experienced radiologist. In practice, SODA can be generalized to any semi-supervised open set domain adaptation settings where there are a large well-annotated dataset and a small newly available database.

Acknowledgments

Jieli Zhou and Baoyu Jing contributed equally to this work.

References

[1] T. Ai et al., “Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases,” Radiology, vol. 296, no. 2, E32–E40, 2020.
[2] Y. Fang et al., “Sensitivity of chest CT for COVID-19: Comparison to RT-PCR,” Radiology, vol. 296, no. 2, E115–E117, 2020.
[3] 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, 1423–1433, 2020.
[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2097–2106.
[5] J. Irvin et al., “CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in Proc. AAAI Conf. on Artif. Intell., vol. 33, 2019, pp. 590–597.
[6] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, “Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 9049–9058.
[7] B. Jing, P. Xie, and E. Xing, “On the automatic generation of medical imaging reports,” 2017, arXiv:1711.08195.
[8] Y. Li, X. Liang, Z. Hu, and E. P. Xing, “Hybrid retrieval-generation reinforced agent for medical image report generation,” in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 1530–1540.
[9] B. Jing, Z. Wang, and E. Xing, “Show, describe and conclude: On exploiting the structure information of chest x-ray reports,” in Proc. Thirty-Second AAAI Conf. Comput. Linguistics, 2019, pp. 6750–6758.
[10] S. Biswal, C. Xiao, L. Gleave, J. Verstower, and J. Sun, “Clinical report auto-completion,” in Proc. Web Conf., 2020, pp. 541–550.
[11] P. Lakhani and B. Sundaram, “Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks,” Radiology, vol. 284, no. 2, pp. 574–582, 2017.
[12] J. P. Cohen, P. Morrison, and L. Dao, “COVID-19 image data collection,” 2020, arXiv:2003.11597. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset
[13] 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,” Sci Rep., vol. 10, 2020, Art. no. 19549.
[14] L. Li et al., “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT,” Radiology, vol. 296, no. 2, E65–E71, 2020.
[15] I. D. Apostolopoulos and T. A. Mpiesiana, “COVID-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks,” Phys. Eng. Sci. Med., vol. 43, no. 2, pp. 635–640, 2020.
[16] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, “DeepCOVID: Predictive tool to diagnose COVID-19 from chest x-ray images using deep transfer learning,” 2020, arXiv:2004.09363.
[17] P. P. Busto and J. Gall, “Open set domain adaptation,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 754–763.
[18] Z. Pei, Z. Cao, M. Long, and J. Wang, “Multi-adversarial domain adaptation,” in Thirty-Second AAAI Conf. Artif. Intell., 2018, 3934–3941.
[19] Z. Wang, B. Jing, Y. Ni, N. Dong, P. Xie, and E. P. Xing, “Adversarial domain adaptation for disease diagnosis: being aware of class relationships,” 2019, arXiv:1905.11931.
[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” Inst. for Cognitive Science, Univ. California San Diego, La Jolla, CA, USA, Tech. Rep. 8506, 1985.
[21] Y. Ganin et al., “Domain-adversarial training of neural networks,” J. Mach. Learn. Res., vol. 17, no. 1, pp. 2096–2030, 2016.
[22] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, “Analysis of representations for domain adaptation,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 137–144.
[23] G. Huang, Z. Liu, L. Van DerMaaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4700–4708.
[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.
[25] Z. Cao, L. Ma, M. Long, and J. Wang, “Partial adversarial domain adaptation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 135–150.
[26] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in Proc. 27th Int. Conf. Mach. Learn. (ICML-10), 2010, pp. 807–814.
[27] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.
[28] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
[29] 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.
[30] Y. Ganin and V. Lempitsky, “Unsupervised domain adaptation by backpropagation,” in Proc. Int. Conf. Mach. Learn., 2015, pp. 1180–1189.
[31] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2893–2902.
[32] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep domain confusion: Maximizing for domain invariance,” 2014, arXiv:1412.3474.
[33] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, “Universal domain adaptation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 2720–2729.
B. Jing, C. Lu, D. Wang, F. Zhuang, and C. Niu, “Cross-domain labeled LDA for cross-domain text classification,” in Proc. IEEE Int. Conf. Data Mining (ICDM), 2018, pp. 187–196.

D. Wang et al., “Coarse alignment of topic and sentiment: A unified model for cross-lingual sentiment classification,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 2, pp. 736–747, Feb. 2021.

I. Goodfellow et al., “Generative adversarial nets,” in Proc. 27th Int. Conf. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

Jieli Zhou received the B.S. degree in mathematical sciences and the M.S. degree in computational data science from Carnegie Mellon University, Pittsburgh, PA, USA, in 2017 and 2018, respectively. He is currently working toward the PhD degree with the University of Michigan–Shanghai Jiao Tong University (UM-SJTU) Joint Institute, Shanghai Jiao Tong University, Shanghai, China. After graduation, he joined C3.ai as a data scientist and worked on a series of high-dimensional time-series modeling projects, such as predictive maintenance and anomaly detection. His research interests mainly include computational biology, medical image analysis, and time series analysis.

Baoyu Jing received the bachelor’s degree from Beihang University, Beijing, China, in 2016 and the master’s degree from the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, in 2018. He is currently working toward the PhD degree with Computer Science Department, University of Illinois at Urbana-Champaign. His research interests mainly include data mining, graph mining, and transfer learning.

Zeya Wang received the PhD degree in statistics from Rice University, Houston, TX, USA, in 2017. He received postdoctoral training with the University of Texas MD Anderson Cancer Center, Houston, TX, USA, where he conducted research to build statistical methods for biomedical studies. Since October 2017, he has been a scientist with Petuum Inc., Pittsburgh, PA, USA, for designing models to power large-scale industrial machine learning applications. His research interests include applied mathematics, statistical machine learning, biostatistics, and computer vision focusing on biomedical image analysis.

Hongyi Xin received the PhD degree from Computer Science Department, Carnegie Mellon University, where he worked on developing novel algorithms to improve the speed and the sensitivity of read mappers. He is currently a tenure-track assistant professor with Shanghai Jiao Tong University. He is jointly appointed by both the UM-SJTU Joint Institute and the Department of Automation. After graduation, he joined the School of Medicine, University of Pittsburgh, as a postdoc and switched focus to single-cell multiomics. He led the development of several single-cell multiomics analytical methods with multiple submissions both published in progress with UPMC Children’s Hospital. His research interests include computer architecture, immunology, and cancer research.

Hanghang Tong received the MSc and PhD degrees majored in machine learning from Carnegie Mellon University in 2008 and 2009, respectively. Since August 2018, he has been an associate professor with Computer Science Department, University of Illinois at Urbana-Champaign. In August 2014, he was an assistant professor with the School of Computing, Informatics, and Decision Systems Engineering (CIDSE), Arizona State University. He was an assistant professor with Computer Science Department, City College, City University of New York, a research staff member with IBM T.J. Watson Research Center, and a postdoctoral fellow with Carnegie Mellon University. He has authored or coauthored more than 100 referred articles. His research focuses on large scale data mining for graphs and multimedia. He was the recipient of several awards, including the NSF CAREER Award (2017), the ICDM 2015 Highest-Impact Paper Award, four best paper awards (the TUP’14, the CIKM’12, the SDM’08, and the ICDM’06), five ‘bests of conference’ (the KDD’16, the SDM’15, the ICDM’15, the SDM’11, and the ICDM’10), and one best demo, honorable mention (the SIGMOD’17). He is an associate editor for the SIGKDD Explorations (ACM), an action editor of Data Mining and Knowledge Discovery (Springer), and an associate editor for the Neurocomputing (Elsevier), and was a program committee member in multiple data mining, databases, and artificial intelligence venues, including SIGKDD, SIGMOD, AAAI, WWW, and CIKM.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.