Complex & Intelligent Systems (2024) 10:2717–2731
https://doi.org/10.1007/s40747-023-01297-9

ORIGINAL ARTICLE

Learning robust features alignment for cross-domain medical image analysis

Zhen Zheng[1] · Rui Li[1] · Cheng Liu[1]

Received: 9 August 2023 / Accepted: 11 November 2023 / Published online: 14 December 2023

Abstract

Deep learning demonstrates impressive performance in many medical image analysis tasks. However, its reliability builds on the labeled medical datasets and the assumption of the same distributions between the training data (source domain) and the test data (target domain). Therefore, some unsupervised medical domain adaptation networks transfer knowledge from the source domain with rich labeled data to the target domain with only unlabeled data by learning domain-invariant features. We observe that conventional adversarial-training-based methods focus on the global distributions alignment and may overlook the class-level information, which will lead to negative transfer. In this paper, we attempt to learn the robust features alignment for the cross-domain medical image analysis. Specifically, in addition to a discriminator for alleviating the domain shift, we further introduce an auxiliary classifier to achieve robust features alignment with the class-level information. We first detect the unreliable target samples, which are far from the source distribution via diverse training between two classifiers. Next, a cross-classifier consistency regularization is proposed to align these unreliable samples and the negative transfer can be avoided. In addition, for fully exploiting the knowledge of unlabeled target data, we further propose a within-classifier consistency regularization to improve the robustness of the classifiers in the target domain, which enhances the unreliable target samples detection as well. We demonstrate that our proposed dual-consistency regularizations achieve state-of-the-art performance on multiple medical adaptation tasks in terms of both accuracy and Macro-F1-measure. Extensive ablation studies and visualization results are also presented to verify the effectiveness of each proposed module. For the skin adaptation results, our method outperforms the baseline and the second-best method by around 10 and 4 percentage points. Similarly, for the COVID-19 adaptation task, our model achieves consistently the best performance in terms of both accuracy (96.93%) and Macro-F1 (86.52%).

Keywords Adversarial training · Class-level information · Unsupervised domain adaptation · Medical adaptation task

Introduction

Medical image analysis is a crucial task on Computer-aided diagnosis (CAD) systems [23], which aims to assist doctors for detecting abnormal patterns within a medical image. Therefore, an accurate and robust medical image analysis model can significantly decrease the misdiagnosis rate and the diagnostic time, so that improve the entire medical treatment. Recently, deep convolutional neural networks (CNNs) achieve great success on the image classification task [17, 35], which inspires many works to adopt CNNs as the backbone for medical image analysis [20, 22]. For example, Hryniewska et al. [18] presents various deep models for analyzing the COVID-19 (Coronavirus disease 2019) [51]. Dai et al. [9] adopts a novel deep residual architecture to detect different skin lesions.

Despite the impressive performance of the CNNs on the medical image analysis task, the trained deep source models still cannot generalize well to the other domains with different data distributions, and the corresponding performance will be seriously degraded when the domain gap is large. Meanwhile, the different domain distributions are easily resulted from

Z. Zheng and R. Li have contributed equally.

✉ Cheng Liu
cliu@stu.edu.cn

Zhen Zheng
20zzheng1@gmail.com

Rui Li
ruili@stu.edu.cn

[1] Department of Computer Science, Shantou University, Shantou, China

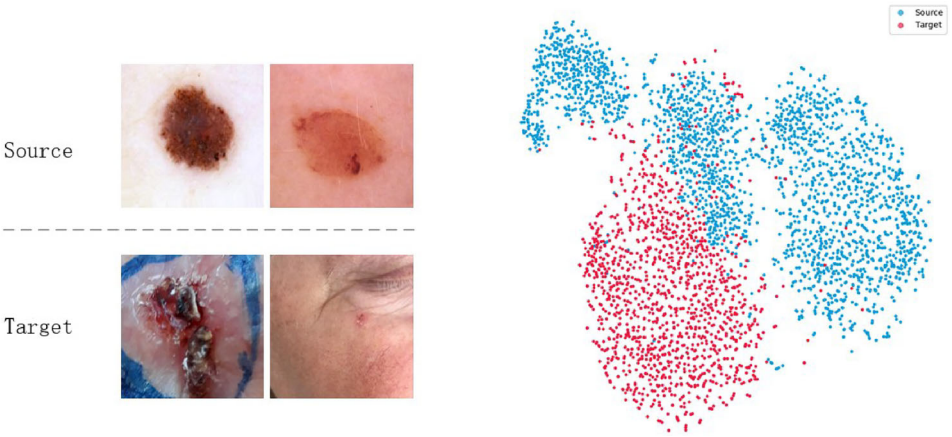

Fig. 1 Cross-domain medical image analysis tasks, source domain samples (top left), target domain samples (bottom left). t-SNE visualization of two domains (right side)

various data-collection devices, institutions, or new types of diseases, which are wildly observed in medical fields. As shown in Fig. 1, the skin lesion images collected by the dermoscopy and the smartphone follow different distributions. Similarly, both COVID-19 and typical pneumonia can lead to abnormality on lungs, while different types of viruses result in distribution discrepancy between the corresponding medical images. All these scenarios can result in performance drop. This is referred to as domain shift [2, 48], which is one of the key factors that prevent the deep neural networks from large-scale usages in the real-world applications, and becomes a hot research topic in recent years.

An intuitive way to address the domain shift is to fine-tune the source model with newly collected labeled target dataset [30, 45]. However, manually annotating large amount of medical data in the target domain requires the expertise of the doctors and is very expensive, or even impossible for the new diseases, i.e., COVID-19. Unsupervised domain adaptation [4, 27] provides another alternative, which can transfer knowledge from rich labeled source domain to the unlabeled target domain, so that the adapted model can achieve reliable performance in the target environment. Therefore, there is a strong motivation to develop a robust medical image analysis model that can generalize well to different domain with few or no doctor annotations [15], especially with the rapid development of health care system.

Most unsupervised domain adaptation methods focus on learning domain-invariant features [11, 27, 37], where the source features and the target features should be aligned or follow the same distributions, thus the source classifier can be directly applied for the target features. This is often realized by minimizing a specific distribution distance between two domains, such as maximum mean discrepancy (MMD) [13], second-order statistics difference (CORAL) [37], and so on. Meanwhile, inspired from generative adversarial networks (GAN) [12], some recent works use a discriminator to deliver more advanced results by adversarial training [11]. On one hand, the discriminator aims to distinguish the source and the target features; On the other hand, the features extractor is trained to fool the discriminator, which makes the extracted features indistinguishable so that the similar features distributions can be achieved. However, these global alignment methods do not consider the class-level information of the features, and implicitly assume that the features belonging to the same class will be related and aligned together. This assumption may not be reasonable for the medical domain, due to the various characteristics of the benign or malignant data from different patients, and the target features can be near the class boundary. Thus, forcing two domain features matching via conventional adversarial training is likely to deteriorate its discriminative information [25], so that the features from different classes will be aligned and result in negative transfer.

To address the above issues, we propose dual-consistency regularizations on the classifiers to achieve more robust features alignment. First, an auxiliary classifier is introduced in our medical adaptation framework, and we use diverse training to detect the unreliable (less-aligned) target features. Then, the *cross-classifier consistency regularization* is proposed to minimize the discrepancy of the two classifiers so that the unreliable target features will be aligned with more attentions, and the reliable (already-aligned) target features is less affected during alignment. With the class-level information from the classifiers, the corresponding features discriminative information can be preserved. Second, since the unreliable target features detection is highly related to performance of the corresponding classifiers, we further propose a *within-classifier consistency regularization* to enforce the classifier predictions invariant to the perturbations of the target data to increase its robustness. Specifically, we use several data augmentations to perturb the target medical data, and we demonstrate that the classifiers which are insensitive to these perturbations can deliver more superior results in terms of both accuracy and Macro-F1-measure during adaptation, since they can in turn facilitate the distribution alignment with more accurate class-level information.

To evaluate the superiority of our method, we conduct extensive experiments on multiple medical image adaptation benchmarks, including the lung X-ray images adaptation from the typical pneumonia to the COVID-19 and skin lesion images adaptation from the dermoscopy device to the smartphone device. We show that our proposed method can significantly outperform previous state-of-the-art methods. Furthermore, the ablation studies and several visualization results are presented to verify the effectiveness of each proposed module. In summary, the main contributions of this paper are outlined as follows:

To summarize, the contributions of this paper are as follows:

1. We propose a new framework for cross-domain medical image analysis, which avoids medical experts annotations for various medical domains. This is beneficial for the the real-world computer-aided diagnosis applications.
2. Compared with conventional global alignment methods, we introduce an auxiliary classifier for achieving robust features alignment with the class-level information, which includes the cross-classifier and within-classifier consistency regularizations.
3. We confirm the superiority of our proposed method through extensive experiments on several medical benchmarks with state-of-the-art performances.

The remainder of the paper is organized as follows. The next section introduces related work about cross-domain medical image analysis. The third section presents the details of our proposed framework, followed by the extensive experiment results in the fourth section. Finally, the conclusion is drawn in the last section.

Related work

Medical image analysis In recent years, deep learning techniques have received increasing attention in the field of medical image analysis. Using computers to analyze medical images and perform assisted diagnosis of diseases, doctors can improve the efficiency of diagnosis and make early detection and treatment, thus avoid delaying the best treatment time for the patients. For example, the identification of skin lesions, especially for melanoma cases, plays an important role in assisting clinical diagnosis. Since damaged skin has various characteristics within the same diseases, which leads to difficulties in distinguishing the lesion types [46]. Yu et al. propose to use very deep residual networks [17] to distinguish melanoma from non-melanoma, Liu et al. [26] proposes a consistency-based semi-supervised method to diagnose skin lesions, [14] introduces a two-step progressive transfer adversarial learning technique to transfer from the source domain to the target domain by discovering the invariant properties. In addition, chest radiography (e.g., x-ray or CT) images are often used to understand the abnormality on lungs, such as the COVID-19 disease [34]. On one hand, deep CNN-based approaches [38] are widely used to predict the novel COVID-19 disease with the chest x-ray images. On the other hand, Minaee et al. [31] adopts deep migration learning for improved detection. However, all these methods require labeled medical data for training which is not always feasible in the real-world applications, since medical annotations are very expensive and need expertise of specific doctors. In this work, we tend to use unsupervised domain adaptation techniques to achieve model generalization on the unlabeled target domain.

Domain adaptation To solve the problem of scarcity of labeled data in the target domain, domain adaptation learning is used to migrate knowledge from different yet relevant source domains to the target domain [6, 24]. There are three types of existing deep domain adaptation methods. The first category is the distance-minimization-based methods, which aim to reduce the distribution distance between domains by minimizing a distance between two domains. For example, maximum mean discrepancy (MMD) used in DDC [42] inspires a series of MMD-based methods, such as [39, 44]. DDDA [47] uses central alignment and correlation alignment strategies to jointly generate domain-invariant and discriminative features. The second category is adversarial-training-based methods that learn domain-invariant features to reduce the domain discrepancy with a discriminator. Zhang et al. [50] use a dual features extractor and classifier to detect the COVID-19 with the domain adversarial training. In addition to the conventional global alignment methods [11, 14], some recent works also adopt class-level information [21, 33] during alignments, which adopts adversarial training between features extractor and the classifiers. The third category is based on reconstruction methods. For example, deep separation networks (DSN) [5] adopt a reconstruction loss to learn domain-specific features and domain-invariant features for interpretability. We focus on unsupervised domain adaptation problem, where there is only unlabeled data in the target domain. Beyond global alignment, we consider to achieve the robust features alignment with class-level information and the dual-consistency regularizations, thus the improved adaptation performance can be expected.

Consistency training utilizes unlabeled data for augmented prediction and enforces consistent outputs under different perturbations. Berthelot et al. [3] introduce distribution alignment and augmented anchoring to use the ensembled output of the network as the consistency target. Pseudo-labeling [16] can be regarded as a hard case of consistency learning. FixMatch [36] combines pseudo-labeling and consistency regularization to achieve great success in the semi-supervised

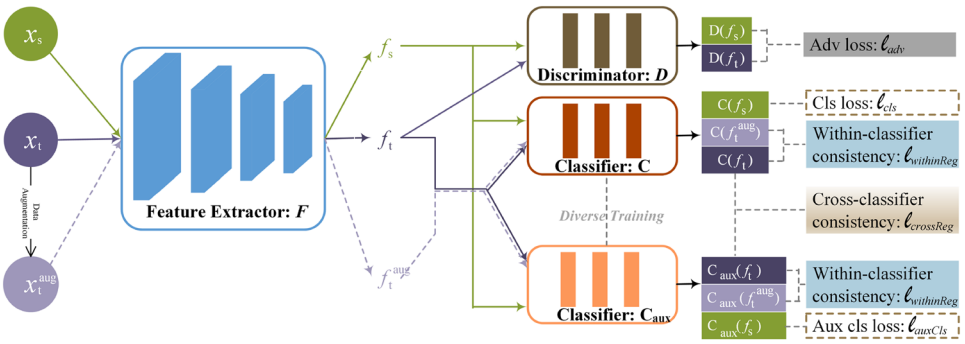

Fig. 2 An unsupervised domain adaptive scheme combining domain adversarial training and dual-consistency regularizations. During training, we optimize the network parameters of F, D, C, and C_{aux} to take into account both global feature alignment and reliable class-level information. In the test phase, the final predictions are averaged with the outputs of the two classifiers C and C_{aux}

learning field. Conventional methods only enforce the consistency of one classifier outputs for the single sample with different perturbations, while Liu et al. [26] further consider the relationship between samples and allows the classifier outputs of the perturbed samples to maintain the original relationships. We exploit consistency training within two classifiers under unsupervised domain adaptation setting in the medical field.

Methods

In this paper, we aim for unsupervised cross-domain adaptation for medical image analysis [1]. There are labeled source domain $\mathcal{D}_s = \{X_s, Y_s\}$ and unlabeled target domain $\mathcal{D}_t = \{X_t\}$, where the two domains $\mathcal{D}_s$ and $\mathcal{D}_t$ follow related but different data distributions, and their label space is the same. The goal is to train a medical image analysis model with $\mathcal{D}_s$ and $\mathcal{D}_t$, which can achieve reliable performance on the target domain. In this section, we first introduce the widely-adopted adversarial training scheme for the global feature alignment. Then, we present the details of the proposed dual-consistency modules for robust alignment for the medical analysis. Finally, the training procedures of our entire adaptation framework will be given.

Adversarial training adaptation

Inspired by the generative adversarial networks (GAN) [12] which can match any data distributions from noise, Ganin et al. propose the Domain-Adversarial Neural Networks (DANN) to align the distributions between the source images and the target images, which include three components: a feature extractor F to learn domain-invariant features, a classifier C to make the predictions and a discriminator D to distinguish the extracted source and the target features. F is trained to fool D by making the extracted features indistinguishable, so that the domain-invariant features are achieved with the minmax game between F and D. F, C and D are neural networks, which are parameterized by θ_f, θ_c and θ_d, respectively. The final prediction model is composed of $F \circ C$, and the objectives of DANN can be explained as follows:

$$\min_{\theta_f,\theta_c} \max_{\theta_d} \ell_{cls}(F, C) + \lambda_d \ell_{adv}(F, D), \tag{1}$$

where

$$\ell_{cls} = \mathbb{E}_{x_s,y_s\sim\mathcal{D}_s}[-y_s \log C(F(x_s))], \tag{2}$$

$$\ell_{adv} = \mathbb{E}_{x_t\sim\mathcal{D}_t}[\log D(F(x_t))] + \mathbb{E}_{x_s\sim\mathcal{D}_s}[\log(1 - D(F(x_s)))], \tag{3}$$

where ℓ_{cls} indicates the classification loss on the labeled source data, ℓ_{adv} indicates the adversarial loss between the global source and the target features, and λ_d is the hyperparameter that balances their effects.

Dual-consistency regularizations

Although the above adversarial training adaptation can alleviate the domain shift, it does not consider the class-level information of the features and fully exploit the knowledge within the unlabeled target data, which will lead to sub-optimal adaptation performance. In addition, in the medical field, falsely aligning the features with different classes results in misdiagnosis, which is unacceptable. To address these issues, we further introduce an auxiliary classifier C_{aux} and propose the dual-consistency regularizations (as shown in Fig. 2) to enhance the robust distribution matching with the class-level information. The dual-consistency regularizations include cross-classifier consistency and within-classifier consistency, which are detailed in following.

Cross-classifier consistency

In order to maintain the discriminative information and avoid negative transfer, we attempt to make the best of information of the classifiers during feature alignment via cross-classifier consistency.

Specifically, conventional adversarial training treats each target instances equally, which may hurt the features' discriminative information. We denote that the unreliable target

instances are not aligned and have relatively large domain gap with the source data. In this case, the feature extractor tries hard to align those unreliable target instances, and will imposes the same effects on the reliable instances, which results in the original aligned features (denoted as reliable instances) falsely mapped after alignment. Therefore, we attempt to detect the unreliable target instances, so that we can put more efforts to align these unreliable target instances while leave the reliable target instances unchanged.

Considering the unreliable target instances which should be far from the support of the source distribution, they are not discriminative with respective to the source classifier so that different source classifiers (i.e., *trained with different seed*) have various behaviors on these target instances. Thus, it is straightforward to further introduce an auxiliary classifier to detect these unreliable target instances during alignment. First, we use the diverse training to increase the diversity between C and C_{aux} for better detection of the unreliable target instances. In this case, unreliable target instances have large discrepancy between two classifiers and reliable target instances have relative small discrepancy. Then, F is trained to minimize their discrepancy via the cross-classifier consistency for the unreliable feature alignment, and the reliable target features are less affected. Thus, the features' discriminative information can be maintained. The corresponding losses function can be formulated as follows:

$$\begin{aligned}\min_{\theta_f,\theta_c,\theta_{c_{aux}}} \max_{\theta_d} \; & \ell_{cls}(F,C)+\ell_{auxCls}(F,C_{aux}) \\ & +\lambda_d\ell_{adv}(F,D)-\lambda_{reg_1}\ell_{crossReg}(C,C_{aux}), \end{aligned} \tag{4}$$

$$\begin{aligned}\min_{\theta_f,\theta_c,\theta_{c_{aux}}} \max_{\theta_d} \; & \ell_{cls}(F,C)+\ell_{auxCls}(F,C_{aux}) \\ & +\lambda_d\ell_{adv}(F,D)+\lambda_{reg_1}\ell_{crossReg}(F), \end{aligned} \tag{5}$$

where

$$\ell_{auxCls} = \mathbb{E}_{x_s,y_s\sim\mathcal{D}_s}[-y_s \log C_{aux}(F(x_s))], \tag{6}$$

$$\ell_{crossReg} = \mathbb{E}_{x_t\sim\mathcal{D}_t}|C(F(x_t)) - C_{aux}(F(x_t))|, \tag{7}$$

where ℓ_{auxCls} indicates the cross-entropy loss with the labeled source data for C_{aux} and $\ell_{cross_{reg}}$ indicates the cross-consistency regularization between two classifiers C and C_{aux} for the same x_t.

As shown in Eq. (4), we will train C and C_{aux} to minimize the source classification loss and to maximize $\ell_{crossReg}$ for the diverse classifiers, which enables the two classifiers to capture the class-level information from different perspectives. Therefore, the two classifiers can be used to detect the unreliable target instances via their disagreement on the extracted target features $F(x_t)$, the reliable target instances which are close to the source domain will have relative similar outputs due to supervised training on source domain, and unreliable target instances will deliver very different outputs.

On the contrary, in Eq. (5), F will be updated to enforce the extracted target features $F(x_t)$ to yield similar outputs with the diverse classifiers, which means the target features will be pushed towards to source domain for feature distribution matching. Compared with traditional adversarial training adaptation, our cross-classifier consistency loss can leverage the class-level information to learn domain-invariant features *adaptively*, since the more unreliable target instances will receive larger $\ell_{crossReg}$, and the reliable target instances will be less affected for robustness.

We will iteratively optimize Eqs. (4) and (5) until convergence. We empirically demonstrate that the proposed cross-consistency regularization can implicitly alleviate the negative transfer caused by the feature deterioration with the help of class-level information.

Within-classifier consistency

Although the above model can deliver improved results on our medical image adaptation experiments, it still highly relies on the robustness of the prediction model ($F \circ C$ or $F \circ C_{aux}$) for the success of unreliable target instances detection. Without extra constraints, during the diverse training, the two classifiers may overfit the source domain and try hard to increase their discrepancy, which can hurt their generalization. Thus, the corresponding feature extractor F may falsely generate target features with wrong classes due to the weak prediction model on the target domain, especially on the early adaptation stage. On the other hand, improving the robustness of the prediction model can also enhance its generalization on the target domain, which brings about better performance.

To address this issue, we propose a within-classifier consistency regularization $\ell_{withinReg}$ to enforce the robustness of the prediction model. We claim that a robust prediction model should be invariant to the perturbation on the input. Specifically, we augment the target data x_t to derive x_t^{aug}. In general, x_t and x_t^{aug} should preserve the same class information, thus, the within-classification consistency enforce the prediction model to deliver the similar outputs. The corresponding loss can be formulated as follows:

$$\begin{aligned}&\min_{\theta_c,\theta_{c_{aux}},\theta_f} \; \ell_{withinReg}(F,C,C_{aug}) \\ &= \mathbb{E}_{x_t\sim\mathcal{D}_t}\left\{\mathrm{KL}[C(F(x_t))||C(F(x_t^{aug}))]\right. \\ &\quad \left.+\,\mathrm{KL}[C_{aux}(F(x_t))||C_{aux}(F(x_t^{aug}))]\right\}, \end{aligned} \tag{8}$$

As shown in Eq. (8), C_{aux}. $\mathrm{KL}[\cdot||\cdot]$ denotes the Kullback–Leibler (KL) divergence, we propose to minimize the KL divergence between the original output and the augmented output within both classifiers C and C_{aux}. Therefore, the whole prediction model will be more robust by avoiding features overfitting and deterioration during features alignment,

where the final adaptation performance can be significantly improved on the medical image applications.

Training procedure

Algorithm 1 Pseudo-code of the proposed model

Input: Labeled source dataset $\{X_s, Y_s\}$, unlabeled target dataset $\{X_t\}$, mini-batch size B, learning rates ζ;
Output: θ_f, θ_c and $\theta_{c_{aux}}$;
1: **for** i = 1 to N **do**
2: **for** each mini-batch **do**
3: Update F, C, and C_{aux} with $\mathcal{L}_{pred}$ for the main prediction model:
$\theta_f \leftarrow \mathrm{SGD}(\nabla_{\theta_f}(\mathcal{L}_{pred}), \theta_f, \zeta)$;
$\theta_c \leftarrow \mathrm{SGD}(\nabla_{\theta_c}(\mathcal{L}_{pred}), \theta_c, \zeta)$;
$\theta_{c_{aux}} \leftarrow \mathrm{SGD}(\nabla_{\theta_{c_{aux}}}(\mathcal{L}_{pred}), \theta_{c_{aux}}, \zeta)$;
4: Update D with $\mathcal{L}_{adv}$ for adversarial alignment:
$\theta_d \leftarrow \mathrm{SGD}(\nabla_{\theta_d}(-\mathcal{L}_{adv}), \theta_d, \zeta)$;
5: Update C and C_{aux} with $\mathcal{L}_{div}$ for diverse classifiers:
$\theta_c, \theta_{c_{aux}} \leftarrow \mathrm{SGD}(\nabla_{\theta_c, \theta_{c_{aux}}}(-\mathcal{L}_{div}), \theta_c, \theta_{c_{aux}}, \zeta)$;
6: **end for**
7: **end for**

The overall objective functions for our medical adaptation model can be summarized as follows:

$$\min_{\theta_f, \theta_c, \theta_{c_{aux}}} \mathcal{L}_{pred} = \ell_{cls} + \ell_{auxCls} + \lambda_d \ell_{adv} + \lambda_{reg_1} \ell_{crossReg} + \lambda_{reg_2} \ell_{withinReg}, \quad (9)$$

$$\max_{\theta_d} \mathcal{L}_{adv} = \ell_{adv}, \quad (10)$$

$$\max_{\theta_c, \theta_{c_{aux}}} \mathcal{L}_{div} = \ell_{crossReg}, \quad (11)$$

where λ_d, λ_{reg_1} and λ_{reg_2} are hyper-parameters to balance impacts of the related losses. We proceed with the training by alternately optimizing the involved network parameters with respect to their corresponding objective function shown in Eqs. (9), (10) and (11), respectively. Our proposed model is trained in the end-to-end manner, and the detailed optimization procedure is summarized in Algorithm 1. Besides, according to our experiments, the proposed dual-consistency regularizations are helpful to stabilize adaptation training. During test, we average the outputs of both classifiers C and C_{aux} for the final prediction.

Experiment

In this section, we extensively evaluate our proposed model on the cross-domain medical analysis adaptation tasks, which include skin diseases classification adapted from smartphone images to dermoscopy images, and the *novel coronavirus disease 2019* (COVID-19) detection adapted from typical pneumonia to COVID-19. First, we introduce the experimental settings, which include the relevant datasets, evaluation metrics and implementation details. Next, we introduce some baselines and recent state-of-the-art methods for the comparison. In addition, several visualization results and ablation studies are presented to verify the effectiveness of our proposed model.

Experiment settings

Skin diseases classification: we selected three skin datasets for our experiments, namely the ISIC dataset,[1] the HAM10000 dataset [40] and the PAD-UFES-20 dataset [32]. The ISIC and HAM10000 datasets are the most wildly-used dermoscopy images for the skin diseases classification tasks. PAD-UFES-20 dataset is a new skin lesion benchmark, which collects the clinical skin lesion images with the smartphone devices. The dataset is involved with 1373 patients, 1641 skin lesions and contains 2298 images for six different diagnoses, and 58.4% of the skin lesions are confirmed by the biopsy. PAD-UFES-20 aims for supporting the skin diseases detection with public clinical images rather than the professional dermoscopy images. In this medical analysis task, we use ISIC and HAM10000 datasets as the source domain, and PAD-UFES-20 dataset is used as the target domain, i.e.,

Dermoscopy→Smartphone. Both domains share six categories of skin diseases, which are actinic keratosis, basal cell carcinoma, malignant melanoma, nevus, squamous cell carcinoma and seborrheic keratosis. Benign lesions include actinic keratosis, nevus and seborrheic keratosis, and malignant lesions includes basal cell carcinoma, malignant melanoma and squamous cell carcinoma. During the experiments, we resize all images to 224 × 224, and the statistics of the dataset are shown in Table 1.

COVID-19 classification: COVID-19 is the new virus which results in related but different symptom to the typical pneumonia. Therefore, we consider to classify COVID-19 samples adapted from the typical pneumonia domain. Following [49], we adopt three public datasets, which includes the COVID-19 chest radiograph dataset,[2] the COVID-19 X-ray dataset [8] and the RSNA pneumonia challenge dataset[3] in our experiments. There are four categories, which are normal, COVID-19, bacterial pneumonia and viral pneumonia, the bacterial pneumonia category and viral pneumonia category are regarded as to the typical pneumonia category. Following [50], some normal instances and all typical

[1] https://www.kaggle.com/datasets/rajivaiml/isic-skin-cancer-dataset.

[2] https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[3] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.

Table 1 Statistics of the skin disease dataset

Set	Domain	Categories						Total
		NEV	ACK	SEK	BCC	MEL	SCC	
Training	HAM10000/ISIC	4702	97	54	4149	4131	144	13,277
	PAD-UFES-20	170	532	164	575	38	123	1602
Test	PAD-UFES-20	43	75	33	132	2	31	316

NEV nevus, *ACK* actinic keratosis, *SEK* seborrheic keratosis, *BCC* basal cell carcinoma, *MEL* malignant melanoma, *SCC* squamous cell carcinoma

Table 2 Statistics of the data set used, where typical pneumonia is the source domain and COVID-19 is the target domain

Set	Domain	Categories			Total
		Normal	Pneumonia	COVID-19	
Training	Pneumonia	5613	2306	0	7919
	COVID-19	2541	0	258	2799
Test	COVID-19	885	0	60	945

pneumonia instances were selected to construct the source domain, and the remaining normal instances and all COVID-19 instances were used as the target domain, i.e., **Typical Pneumonia→COVID-19**. During training, the number of instances in the source domain is 7919 and the number of instances in the target domain is 2799, then the test domain contains 945 instances, and the total data number is 11,663. The detailed statistics of the dataset are shown in Table 2. Compared with skin diseases datasets, COVID-19 detection task is lack of samples and more class-imbalanced, which can further verify the effectiveness of our method during this challenging setting.

Evaluation metrics We adopt Accuracy (%), Macro-Precision (%), Macro-Recall (%), and Macro-F1-measure (%) as the diagnostic metrics in our experiments. Specifically,

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \\
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\
\text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{F1-measure}_{\text{i}} &= \frac{2 * \text{Precision}_{\text{i}} * \text{Recall}_{\text{i}}}{\text{Precision}_{\text{i}} + \text{Recall}_{\text{i}}} \\
\text{Macro-F1-measure} &= \frac{1}{K}\sum_{\text{i}}^{K} \text{F1-measure}_{\text{i}}
\end{aligned} \tag{12}$$

where TP, FP, TN and FN represent numbers of true positive, false positive, true negative and false negative predictions, respectively. K indicates the number of categories

Implementation details: we implement our approach based on PyTorch. For fair comparison, we follow similar settings to the [50]. The ResNet-18 [17] pre-trained on ImageNet [10] is used as the backbone for feature extractor F. The classifier C and the discriminator D are fully connected networks with dense layers. We used an SGD optimizer with a batch size of 16 on a single GPU to train all the neural network modules. The learning rate was set to 0.001. In addition, We do not heavily tune the hyper-parameters due to the proposed robust dual-consistency regularizations, and we select all these hyper-parameters (λ_d, λ_{reg_1}, λ_{reg_2}) from $\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$. For skin diseases adaptation experiments, we found that setting $\lambda_d = 0.01$, $\lambda_{reg_1} = 1$, $\lambda_{reg_2} = 0.01$ obtains the best performance. While for the challenging COVID-19 detection task, setting $\lambda_d = 0.01$, $\lambda_{reg_1} = 0.01$, $\lambda_{reg_2} = 0.1$ obtains the best performance. Implemented source code address: https://github.com/gitMrZheng/LRFA-DA.

Experimental results and analysis

We mainly compare our proposed method with the unsupervised domain adaptation methods, which include the previous typical approaches, i.e., DDC [42], DANN [11], ADDA [41], CDAN [28], MCD [33], JAN [29], and the recent state-of-the-art domain adaptation methods, i.e., AFN [43], MCC [19], BSP [7]. The detailed compared methods are introduced as follows:

- DDC [42] uses max mean discrepancy (MMD) to measure the discrepancy between two domains.
- DANN [11] introduces a discriminator to achieve domain distribution matching via adversarial training.
- ADDA [41] shares a similar idea with DANN, while use two separated feature extractors.

Table 3 Comparison of dermoscopy→smartphone for skin cancer diagnosis

Methods	Accuracy (%)	Macro-F1 (%)	Macro-Precision (%)	Macro-Recall (%)
DDC [42]	65.49	65.46	66.35	66.14
DANN [11]	68.67	68.37	68.77	68.37
ADDA [41]	66.67	66.65	66.81	66.90
CDAN [28]	71.52	71.32	71.56	71.29
MCD [33]	66.77	64.06	67.76	64.97
AFN [43]	66.37	66.35	66.49	66.58
MCC [19]	66.14	65.44	66.56	65.64
JAN [29]	69.62	68.23	71.84	68.83
BSP [7]	68.04	66.53	70.09	67.23
Source-Only	64.72	63.60	63.80	63.77
Our model	**75.47**	**75.60**	**76.23**	**75.94**

The best performances are indicated with **bold**

- CDAN [28] incorporates label information into DANN, which aims for conditional domain adaptation.
- MCD [33] replaces the discriminator with the classifiers to align the source and the target domain by maximizing the classifiers' discrepancy.
- AFN [43] tends to adaptively increase the feature norm to increase the features' transferability.
- MCC [19] minimizes the classifier's confusion among different classes to implicitly achieve domain alignment, which can be used under versatile domain adaptation settings.
- JAN [29] unites distributions and conditional distributions through the process of dimensionality reduction and constructs new feature representations with the goal of simultaneously minimizing the differences between edge distributions and conditional distributions across domains
- BSP [7] proposes Batch Spectral Penalization to increase the optimization term and achieve the double improvement of Transferability and Discriminability.

Experiment results on dermoscopy→smartphone

The performance of our model on the skin disease cross-domain adaptation task are presented in Table 3, together with the results of the recent domain adaptation methods based on the same protocol. Source-Only is only 64.72% Accuracy, with Macro-F1-measure, Macro-Precision and Macro-Recall all coming close at 63.60%, 63.80% and 63.77% respectively. From the experiments, it can be found that the performance is significantly improved after introducing the domain adversarial approach on the recognition and diagnosis of skin lesion images, where the accuracy of DANN is 68.67%, while the conditional adversarial domain adaptation (CDAN) reaches 71.52%, then it is reasonable to believe that the adversarial domain adaptation approach is effective on medical images of unlabeled target domains.

In addition, MCD achieves limited improvements compared with the baseline. We speculate that the less discriminative classifiers without any constraints may lead to false detection of the unreliable target samples. The DDC method and the MCC method are not able to better handle medical tasks with large cross-domain variations in dermatoscope-to-smartphone adaptation. In contrary, our proposed dual-consistency regularization model, which involves class-level information with cross-classifier consistency, and robust training with the within-classifier consistency during alignment. Therefore, unlike previous methods, the performance in terms of all metrics (Macro-F1-measure, Macro-Precision, Macro-Recall, and Accuracy) are significantly improved compared with the Source-Only baseline. The results demonstrate the superiority of the proposed model on the medical domain adaptation task.

Experiment results on typical pneumonia→COVID-19

Table 4 compares the performance of our method, recent domain adaptation methods. First, for novel coronavirus COVID-19 classification, the Source-Only model has a lower Macro-F1 and Macro-Precision of 62.86% and 60.35%, respectively. Then the JAN and BSP models have better performance compared to other unsupervised methods, and they achieved 95.98% accuracy. Most previous domain adaptation methods can improve the performance of Source-Only model in terms of Macro-F1-measure, and their both Macro-Precision and Macro-Recall are increased. However, the Macro-F1 performance of ADDA [41] becomes worse to some extent, which indicates separating feature extractor may be harmful in the complex medical field. On typical pneumonia to COVID-19 adaptation, DANN, AFN and MCC

Table 4 Comparison of typical pneumonia→COVID-19 for new virus diagnosis

Methods	Accuracy (%)	Macro-F1 (%)	Macro-Precision (%)	Macro-Recall (%)
DDC [42]	94.18	64.53	78.41	60.38
DANN [11]	93.97	75.59	74.68	76.58
ADDA [41]	86.24	60.86	58.88	67.02
CDAN [28]	94.50	77.56	76.76	78.42
MCD [33]	89.05	72.44	67.53	87.42
AFN [43]	93.02	69.69	70.35	69.08
MCC [19]	92.70	65.82	67.93	64.25
JAN [29]	95.98	82.83	83.36	82.32
BSP [7]	95.98	78.73	89.42	72.99
Source-Only	85.40	62.86	60.35	72.23
Our model	**96.93**	**86.52**	**85.71**	**87.37**

The best performances are indicated with **bold**

Table 5 Ablation study of the proposed method for investigating the effects of the dual-consistency regularizations on the skin diseases classification task

$\ell_{cls}, \ell_{auxCls}$	ℓ_{adv}	$\ell_{crossReg}$	$\ell_{withinReg}$	Accuracy (%)	Macro-F1 (%)	Macro-Precision (%)	Macro-Recall (%)
✓				67.25	66.43	68.98	67.04
✓	✓			68.51	69.27	69.81	69.59
✓		✓		70.73	70.16	70.90	70.19
✓			✓	68.51	68.91	68.92	68.90
✓	✓	✓		72.47	71.44	72.22	71.46
✓	✓	✓	✓	75.47	75.60	76.23	75.94

methods failed to better focus on the class information of the target domain samples, and were more ineffective in diagnosing COVID-19 compared to other methods. However, our proposed model further utilizes the within-consistency regularization to enhance the robustness of both classifiers, thus, obtains much advanced performance (96.93% accuracy and 86.52% Macro-F1), where the Marco F1 score outperforms the second-best method by around 4 percentage points.

Ablation study

To further evaluate the effectiveness of our method, we conducted an ablation study, as shown in Table 5. The feature adversarial adaptation module, all the cross-classifier consistency module, and the within-classifier consistency module made important contributions. It is clear from the ablation studies, classification loss alone does not yield good results, indicating the existence of task differences and feature distribution differences. Adding within-classifier consistency, or the cross-classifier consistency criterion or feature adversarial adaptation alone, only yielded small improvements as show in Table 5. While, combining both cross-classifier consistency and feature adversarial adaptation can further improve the diagnostic performance, which indicates using class-level information during alignment is helpful. Finally, adding the within-classifier consistency can significantly improve the results, since the class-level information is more accurate due to the robust classifiers.

Model analysis

To demonstrate more intuitively that our proposed method is superior, we visualize the t-SNE embedding of feature changes before and after to understand the domain adaptation from the perspective of domains (Fig. 3) and categories (Fig. 4), respectively. Specifically, it includes the results from Source-Only, DANN, MCD and our model. The original feature distribution shows that the two domains are separated (Fig. 3a). DANN and MCD (Fig. 3b, c) show that the feature distributions of the source and target domains are more consistent after domain adaptation. Our model makes the inter-domain distance smaller (Fig. 3d), which indicates that two domains are aligned together. Some of the benign and malignant skin diseases were difficult to tell apart for diagnosis (Fig. 4a). While DANN and MCD align the distribution of features so that the performance are improved, they produced fewer abnormal classes in clean categories (Fig. 4b, c). Our model can further distinguish the degree of skin damage by

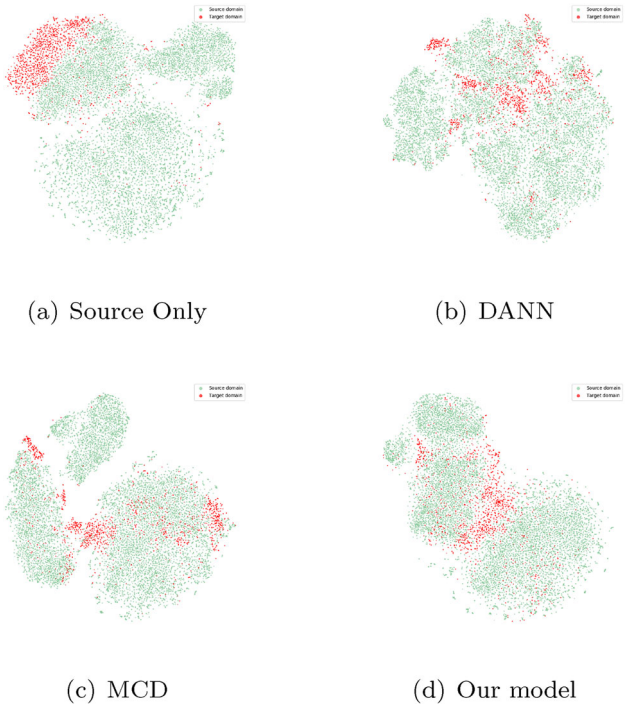

Fig. 3 The visualization of the domain of t-SNE (**a**) without adaptation, after adaptation in DANN (**b**), MCD (**c**) and our model (**d**). Green indicates the source domain and red indicates the target domain

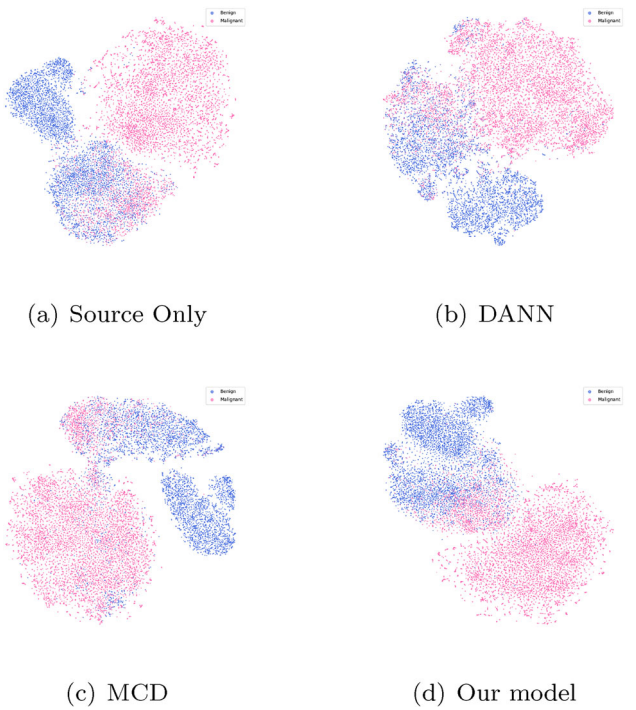

Fig. 4 The visualization of the categories of t-SNE (**a**) without adaptation, after adaptation of DANN (**b**), MCD (**c**) and our model (**d**). Blue color indicates benign skin disease and pink color indicates malignant skin disease

within-classifier consistency regularization (Fig. 4d), where benign and malignant samples are more separable.

We also interpret the results of the analytical study more significantly by visualizing the confusion matrix, as shown in Fig. 5. Each row of the matrix represents the real class, while each column represents the predicted class. In the experimental sample, 151 cases of benign skin lesions and 165 cases of malignant skin lesions with our approach. Deviation in the number of instances between the prediction results of different methods for skin cancer and the true value. Our method has lower instances of class I error (false positives) and class II error (false negatives), then it is obvious that the predictive diagnostic results of our method for skin cancer are balanced and superior.

In order to understand the fluctuation of the training loss curves of each model, Fig. 6a–f shows the comparison of the training loss curves of DDC, AFN, DANN, MCC, CDAN and our model, respectively, and it can be clearly seen that our model converges faster and has less fluctuation of the loss curves compared with the other models.

ROC curves were plotted to better understand the diagnostic performance of the model by its ability to diagnose COVID-19 and skin diseases as a function of the discrimination threshold, as shown in Fig. 7. The ROC visualization shows the superior performance of the proposed model with higher true positive rate and lower false positive rate in the threshold setting compared to the baseline approach. The area under the ROC curve of our model for COVID-19 and skin diseases is 0.890, indicating that positive cases are well differentiated from negative controls. The ROC curves particularly illustrate the benefits of our feature alignment approach in improving sensitivity without sacrificing specificity. By taking class-level information into account during domain adaptation, the model is able to better distinguish diagnostic categories.

As for the COVID-19 classification task, we used the gradient-weighted class activation mapping Grad-CAM method to visualize the location of the chest radiographic region of interest for the diagnosis of neocoronary pneumonia. The left image of Fig. 8 shows the original input image of the COVID-19 patient, while the middle image shows the Grad-CAM obtained on the Source-Only method, and the right image shows the adaptive results of our model. It can be observed that the Grad-CAM visualization of the pathological region obtained from the proposed model lies on the infection position, which is adapted and interpretable.

Parameter analysis

For the skin disease and COVID-19 detection experiments, hyperparameter values $\{\lambda_d, \lambda_{reg_1}, \lambda_{reg_2}\}$ were selected by tuning over the following ranges from $\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$. We further evaluated parameter sensitivity by assessing the

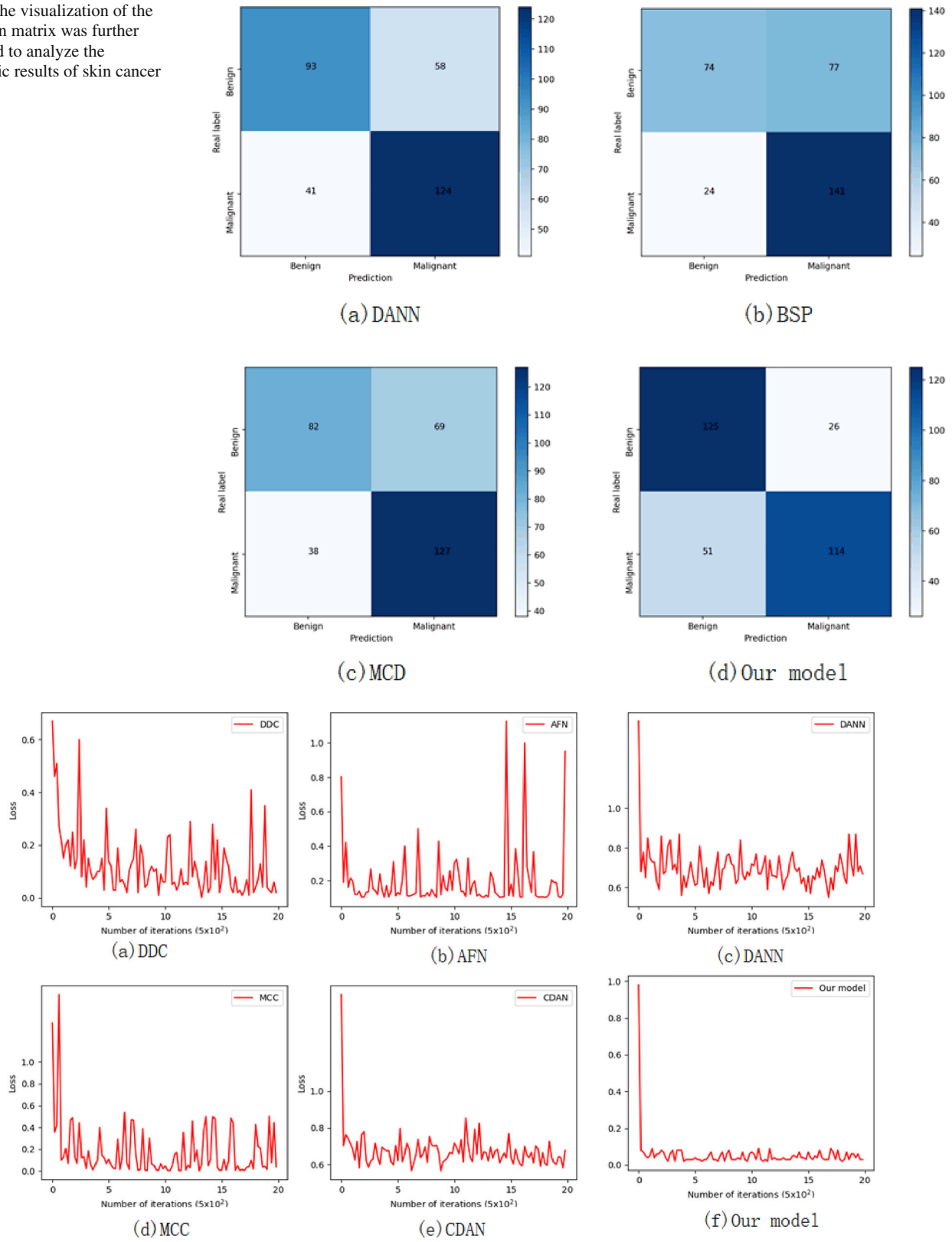

Fig. 5 The visualization of the confusion matrix was further evaluated to analyze the diagnostic results of skin cancer

Fig. 6 Understanding the dynamics of the model diagnosis COVID-19 disease training process and comparing the training loss curves of visualized models

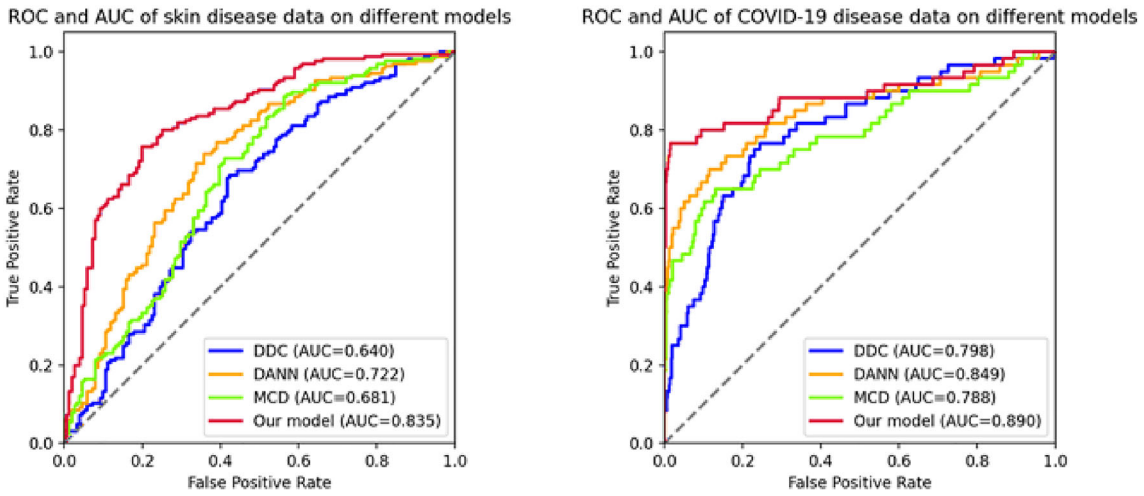

Fig. 7 The left and right panels show the ROC curves for the diagnosis of skin diseases on COVID-19 on different models, respectively

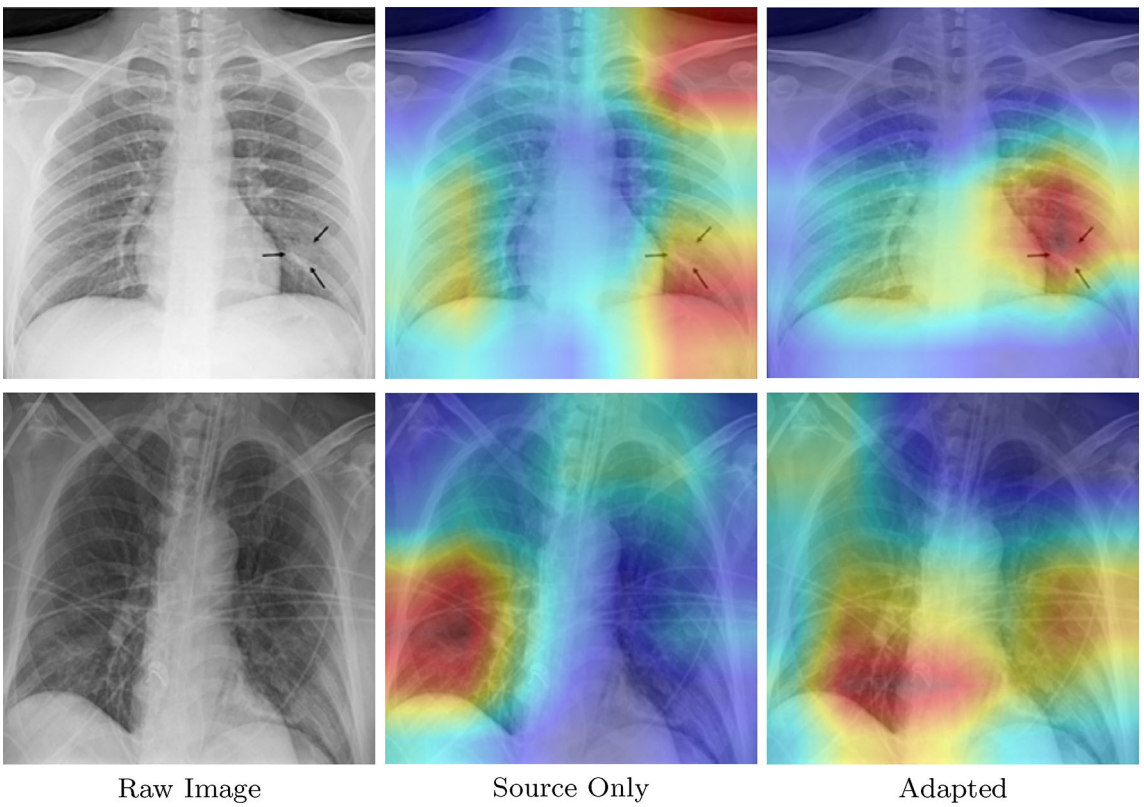

Fig. 8 Visualization results of the Grad-CAM method on COVID-19 cases, with the shaded part of the image showing the lung region and the heat map showing the activation weights of the model

COVID-19 detection performance while varying one parameter at a time. The additional results in Table 6 demonstrate the robustness of the proposed method to hyperparameter choices within the tuned ranges. Performance remains stable across $\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and with optimal COVID-19 diagnosis achieved setting $\lambda_d = 0.01$, $\lambda_{reg_1} = 0.01$,$\lambda_{reg_2} = 0.1$. This analysis highlights the insensitivity of our approach to minor parameter variations, as long as values are chosen within reasonable bounds through tuning. The method is able to maintain effectiveness without requiring precise, narrow hyperparameter specifications.

Table 6 Parameter analysis on COVID-19

Parameters	Value	Accuracy	Macro-F1	Macro-Precision	Macro-Recall
λ_d	10^{-3}	96.67	85.65	**87.17**	84.27
	10^{-2}	**96.96**	**86.52**	85.71	**87.37**
	10^{-1}	96.03	84.31	82.06	86.98
	1	94.71	78.09	77.67	78.53
λ_{reg1}	10^{-3}	96.19	83.87	86.32	81.77
	10^{-2}	**96.96**	86.52	85.71	87.37
	10^{-1}	96.46	**86.72**	85.42	**88.15**
	1	96.93	85.88	**89.70**	82.82
λ_{reg2}	10^{-3}	93.97	78.98	74.99	85.13
	10^{-2}	96.46	85.10	84.58	85.65
	10^{-1}	**96.96**	**86.52**	**85.71**	**87.37**
	1	96.30	82.87	84.27	81.60

The best results are highlighted in bold

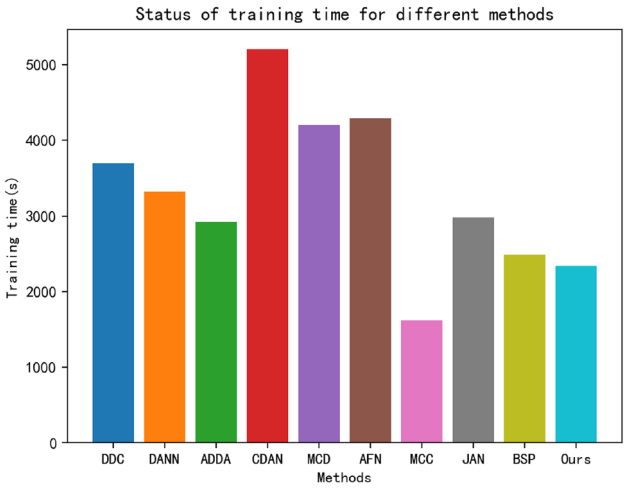

Fig. 9 Comparison of training time of different methods in the diagnostic process of skin diseases

Complexity analysis

To evaluate computational performance, we compared the computation time of the proposed model against state-of-the-art methods, as shown in Fig. 9. Experiments were conducted using a desktop with an Intel Core i9-10900X CPU, NVIDIA GeForce RTX 2080Super GPU, and 64 GB RAM. As illustrated, the proposed model achieves comparable computational efficiency to other benchmark models. The main computational burden of our method involves training the feature extraction network and adversarial discriminator(s).

We also provided FLOPS, model size, trainable parameters and FPS values to analyze the model complexity, which is shown in Table 7. The first model (DDC) does not include any additional modules, e.g., discriminator and classifier, which has relatively less complexity but is not superior in the medical adaptation tasks. Meanwhile, our method has very similar model complexity to the other methods with only one additional classifier, and the corresponding performance is significantly improved with the proposed dual-consistency constraints.

Table 7 The value of model complexity in our model versus competing approaches

Methods	FLOPS (GFlops)	Model size (MB)	Trainable parameters	FPS
DDC [42]	1.82	37.50	11,178,562	154.60
DANN [11]	1.82	45.18	12,888,385	119.68
ADDA [41]	1.82	45.18	12,758,597	135.48
CDAN [28]	1.82	45.18	12,757,571	128.57
MCD [33]	1.82	56.77	11,178,564	121.63
AFN [43]	1.82	46.65	12,210,516	126.27
MCC [19]	1.82	45.18	11,308,866	129.28
JAN [29]	1.82	45.18	11,177,538	127.79
BSP [7]	1.82	45.18	12,758,081	119.87
Our model	1.83	56.77	15,788,357	122.73

Conclusion

In this paper, we propose a novel unsupervised domain adaptation method for cross-domain medical image analysis. Apart from conventional domain adversarial training which may lead to sub-optimal performance in the medical field, we further integrate dual-consistency regularizations for robust features alignment. With a diverse auxiliary classifier, we propose a cross-classifier consistency to focus on aligning unreliable target features with the class-level information. For obtaining robust class-level information and avoid source domain overfitting, we propose a within-classifier consistency regularization to enhance the robustness of both classifiers during training. Extensive experiments on several medical adaptation tasks verify the effectiveness of our method, which is beneficial to various medical image applications. We conduct extensive experiments on several medical adaptation tasks, which demonstrate the effectiveness of our method and will be beneficial to various real-world medical image analysis applications.

Acknowledgements This work was supported in part by the Natural Science Foundation of Guangdong Province (Project No.2022A1515010434), in part by National Natural Science Foundation of China (Project No. 62306052 and No. 62106136), in part by Shantou University under Project NTF20007.

Data availability The data used in this study are available on Kaggle, and detailed information is provided in the paper.

References

1. Ahn E, Kumar A, Fulham MJ, Feng D, Kim J (2020) Unsupervised domain adaptation to classify medical images using zero-bias convolutional auto-encoders and context-based feature augmentation. IEEE Trans Med Imaging 39:2385–2394. https://doi.org/10.1109/TMI.2020.2971258
2. Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW (2010) A theory of learning from different domains. Mach Learn 79:151–175
3. Berthelot D, Carlini N, Cubuk ED, Kurakin A, Sohn K, Zhang H, Raffel C (2020) Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In: 8th International conference on learning representations, ICLR, OpenReview.net. https://openreview.net/forum?id=HklkeR4KPB
4. Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D (2017) Unsupervised pixel-level domain adaptation with generative adversarial networks. In: IEEE conference on computer vision and pattern recognition. pp 95–104
5. Bousmalis K, Trigeorgis G, Silberman N, Krishnan D, Erhan D (2016) Domain separation networks. In: Advances in neural information processing systems. pp 343–351
6. Chen S, Wu H, Liu C (2021) Domain invariant and agnostic adaptation. Knowl Based Syst 227:107192
7. Chen X, Wang S, Long M, Wang J (2019) Transferability vs. discriminability: batch spectral penalization for adversarial domain adaptation. In: International conference on machine learning, PMLR. pp 1081–1090
8. Cohen JP, Morrison P, Dao L (2020) COVID-19 image data collection. CoRR abs/2003.11597. arXiv:2003.11597
9. Dai D, Dong C, Xu S, Yan Q, Li Z, Zhang C, Luo N (2022) Ms RED: a novel multi-scale residual encoding and decoding network for skin lesion segmentation. Med Image Anal 75:102293. https://doi.org/10.1016/j.media.2021.102293
10. Deng J, Dong W, Socher R, Li L, Li K, Li F (2009) Imagenet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition. pp 248–255
11. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky VS (2016) Domain-adversarial training of neural networks. J Mach Learn Res 17:59:1-59:35
12. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville AC, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems. pp 2672–2680
13. Gretton A, Sriperumbudur BK, Sejdinovic D, Strathmann H, Balakrishnan S, Pontil M, Fukumizu K (2012) Optimal kernel choice for large-scale two-sample tests. In: Advances in neural information processing systems. pp 1214–1222
14. Gu Y, Ge Z, Bonnington CP, Zhou J (2020) Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. IEEE J Biomed Health Inform 24:1379–1393. https://doi.org/10.1109/JBHI.2019.2942429
15. Guan H, Liu M (2022) Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng 69:1173–1185. https://doi.org/10.1109/TBME.2021.3117407
16. Han C, Lei Y, Xie Y, Zhou D, Gong M (2021) Learning smooth representations with generalized softmax for unsupervised domain adaptation. Inf Sci 544:415–426. https://doi.org/10.1016/j.ins.2020.08.075
17. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition. pp 770–778
18. Hryniewska W, Bombinski P, Szatkowski P, Tomaszewska P, Przelaskowski A, Biecek P (2021) Checklist for responsible deep learning modeling of medical images based on COVID-19 detection studies. Pattern Recognit 118:108035. https://doi.org/10.1016/j.patcog.2021.108035
19. Jin Y, Wang X, Long M, Wang J (2020) Minimum class confusion for versatile domain adaptation. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) Computer vision—ECCV 2020—16th European conference. Springer, Berlin, pp 464–480. https://doi.org/10.1007/978-3-030-58589-1_28
20. Kumar A, Kim J, Lyndon D, Fulham MJ, Feng DD (2017) An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE J Biomed Health Inform 21:31–40. https://doi.org/10.1109/JBHI.2016.2635663
21. Li J, Lü S, Li Z (2022) Unsupervised domain adaptation via softmax-based prototype construction and adaptation. Inf Sci 609:257–275. https://doi.org/10.1016/j.ins.2022.07.068
22. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M (2014) Medical image classification with convolutional neural network. In: 13th international conference on control automation robotics & vision, ICARCV 2014, Singapore, December 10–12, 2014, IEEE. pp 844–848. https://doi.org/10.1109/ICARCV.2014.7064414
23. Li X, Li C, Rahaman MM, Sun H, Li X, Wu J, Yao Y, Grzegorzek M (2022) A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artif Intell Rev 55:4809–4878
24. Liu C, Wu S, Cao W, Shen W, Jiang D, Yu Z, Wong HS (2020) Joint subspace and discriminative learning for self-paced domain adaptation. Knowl Based Syst 205, 106285
25. Liu H, Long M, Wang J, Jordan MI (2019) Transferable adversarial training: a general approach to adapting deep classifiers. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th international conference on machine learning, ICML, PMLR. pp 4013–4022. http://proceedings.mlr.press/v97/liu19b.html
26. Liu Q, Yu L, Luo L, Dou Q, Heng P (2020) Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Trans Med Imaging 39:3429–3440. https://doi.org/10.1109/TMI.2020.2995518
27. Long M, Cao Y, Cao Z, Wang J, Jordan MI (2019) Transferable representation learning with deep adaptation networks. IEEE Trans Pattern Anal Mach Intell 41:3071–3085
28. Long M, Cao Z, Wang J, Jordan MI (2018) Conditional adversarial domain adaptation. In: Advances in neural information processing systems. pp 1647–1657
29. Long M, Zhu H, Wang J, Jordan MI (2017) Deep transfer learning with joint adaptation networks. In: International conference on machine learning. pp 2208–2217
30. Luo Z, Zou Y, Hoffman J, Li F (2017) Label efficient learning of transferable representations acrosss domains and tasks. In: Advances in neural information processing systems. pp 164–176
31. Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ (2020) Deep-covid: predicting COVID-19 from chest x-ray images using deep

transfer learning. Med Image Anal 65:101794. https://doi.org/10.1016/j.media.2020.101794
32. Pacheco AG, Lima GR, Salomão AS, Krohling B, Biral IP, de Angelo GG, Alves FC Jr, Esgario JG, Simora AC, Castro PB et al (2020) PAD-UFES-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones. Data Brief 32:106221
33. Saito K, Watanabe K, Ushiku Y, Harada T (2018) Maximum classifier discrepancy for unsupervised domain adaptation. In: IEEE conference on computer vision and pattern recognition. pp 3723–3732
34. Signoroni A, Savardi M, Benini S, Adami N, Leonardi R, Gibellini P, Vaccher F, Ravanelli M, Borghesi A, Maroldi R, Farina D (2021) Bs-net: learning COVID-19 pneumonia severity on a large chest x-ray dataset. Med Image Anal 71:102046. https://doi.org/10.1016/j.media.2021.102046
35. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR arXiv:abs/1409.1556
36. Sohn K, Berthelot D, Carlini N, Zhang Z, Zhang H, Raffel C, Cubuk ED, Kurakin A, Li C (2020) Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: Advances in neural information processing systems
37. Sun B, Saenko K (2016) Deep coral: correlation alignment for deep domain adaptation. In: Computer vision—ECCV workshops. pp 443–450
38. Tabik S, Gómez-Ríos A, Martín-Rodríguez JL, Sevillano-García I, Rey-Area M, Charte D, Guirado E, Suárez J, Luengo J, Valero-González MA, García-Villanova P, Olmedo-Sánchez E, Herrera F (2020) COVIDGR dataset and covid-sdnet methodology for predicting COVID-19 based on chest x-ray images. IEEE J Biomed Health Inform 24:3595–3605. https://doi.org/10.1109/JBHI.2020.3037127
39. Tian Q, Zhou J, Chu Y (2022) Joint bi-adversarial learning for unsupervised domain adaptation. Knowl Based Syst 248:108903. https://doi.org/10.1016/j.knosys.2022.108903
40. Tschandl P, Rosendahl C, Kittler H (2018) The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5:1–9
41. Tzeng E, Hoffman J, Saenko K, Darrell T (2017) Adversarial discriminative domain adaptation. In: IEEE conference on computer vision and pattern recognition. pp 2962–2971
42. Tzeng E, Hoffman J, Zhang N, Saenko K, Darrell T (2014) Deep domain confusion: maximizing for domain invariance. CoRR arXiv:abs/1412.3474
43. Xu R, Li G, Yang J, Lin L (2019) Larger norm more transferable: an adaptive feature norm approach for unsupervised domain adaptation. In: 2019 IEEE/CVF international conference on computer vision, ICCV, IEEE. pp 1426–1435. https://doi.org/10.1109/ICCV.2019.00151
44. Yang L, Zhong P (2020) Discriminative and informative joint distribution adaptation for unsupervised domain adaptation. Knowl Based Syst 207:106394. https://doi.org/10.1016/j.knosys.2020.106394
45. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Advances in neural information processing systems, pp 3320–3328
46. Yu L, Chen H, Dou Q, Qin J, Heng P (2017) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 36:994–1004. https://doi.org/10.1109/TMI.2016.2642839
47. Zhang C, Zhao Q (2021) Deep discriminative domain adaptation. Inf Sci 575:599–610. https://doi.org/10.1016/j.ins.2021.07.073
48. Zhang Y (2021) A survey of unsupervised domain adaptation for visual recognition. CoRR arXiv:abs/2112.06745
49. Zhang Y, Niu S, Qiu Z, Wei Y, Zhao P, Yao J, Huang J, Wu Q, Tan M (2020) COVID-DA: deep domain adaptation from typical pneumonia to COVID-19. CoRR abs/2005.01577. arXiv:2005.01577
50. Zhang Y, Wei Y, Wu Q, Zhao P, Niu S, Huang J, Tan M (2020) Collaborative unsupervised domain adaptation for medical image diagnosis. IEEE Trans Image Process 29:7834–7844. https://doi.org/10.1109/TIP.2020.3006377
51. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273