A Coherent Cooperative Learning Framework Based on Transfer Learning for Unsupervised Cross-Domain Classification

Xinxin Shan, Ying Wen$^{(\boxtimes)}$, Qingli Li, Yue Lu, and Haibin Cai

Shanghai Key Laboratory of Multidimensional Information Processing,
School of Communication and Electronic Engineering,
East China Normal University, Shanghai, China
`ywen@cs.ecnu.edu.cn`

Abstract. In the practical application of medical image analysis, due to the different data distributions of source domain and target domain and the lack of the labels of target domain, domain adaptation for unsupervised cross-domain classification attracts widespread attention. However, current methods take knowledge transfer model and classification model as two separate training stages, which inadequately considers and utilizes the intrinsic information interaction between modules. In this paper, we propose a coherent cooperative learning framework based on transfer learning for unsupervised cross-domain classification. The proposed framework is constructed by two classifiers trained by transfer learning, which can respectively classify images of source domain and target domain, and a Wasserstein CycleGAN for image translation and data augmentation. In the coherent process, all modules are updated in turn, and the data is transferred between different modules to realize the knowledge transfer and collaborative training. The final prediction is obtained by a voting result of two classifiers. Experimental results on three pneumonia databases demonstrate the effectiveness of our framework with diverse backbones.

Keywords: Unsupervised cross-domain classification · Transfer learning · Collaborative training · Wasserstein CycleGAN

1 Introduction

Many of the basic assumptions in machine learning are based on the fact that the data of source domain (training dataset) and the target domain (testing dataset) is independent and identically distributed. When the data distributions of source domain and target domain are different, domain adaptation for

cross-domain classification becomes an effective solution [9]. Generally, similarity [19], data global structure [16,31] and feature alignment [18] are advisable solutions to cross-domain classification. In addition, many measurement criteria such as maximum mean discrepancy (MMD) [27] and multiple kernel MMD [15] are generally applied in domain adaptation and have inspired many recognition models [6,8,28,35]. Fine-tuning pre-trained models [20,23] and domain adversarial neural network [5,32] are also two efficient techniques for cross-domain classification. Thus, domain adaptation is critical to promote the generalization ability of neural network [7].

When lacking the labels of target domain in cross-domain classification, unsupervised domain adaptation has more profound significance [2]. For instance, CycleGAN [33] and its variants [14,17,29] can translate images from target to source [13,34] to make the distribution approximate for classification [24]. However, there are still some issues that need further consideration. (i) Class-imbalanced datasets [11]: commonly exist in medical datasets, which may result in over-fitting during model training process and degrade the performance. (ii) Lack of labeled medical images [13]: leads to poor use of supervised learning. Fortunately, transfer learning [7,15] and unsupervised learning [2,18,26] are employed to deal with such challenges. (iii) Separate training of different modules: current methods separately train knowledge transfer model and classification model from beginning to end [4]. That is, the training process is divided into two separate stages, which ignores knowledge transfer and information interaction between modules.

To address the above issues, we propose a coherent cooperative learning (CCL) framework based on transfer learning for unsupervised cross-domain classification, which is constructed by a proposed Wasserstein CycleGAN (WCycle-GAN) for image translation and two classifiers for prediction. First, by training the WCycleGAN with the original images from the source domain and the target domain, we obtain a class-balanced dataset used to fine-tune two classifiers that are convolutional neural networks (CNNs) pre-trained on ImageNet. Specially, the classifier of the target domain uses the proposed cooperative mechanism and MMD criterion to achieve unsupervised cross-domain classification. Finally, we input both the probe image and its translated image generated by WCycleGAN into two classifiers, and get the final prediction by a voting strategy.

There are three contributions in the proposed method. (i) The proposed WCycleGAN and two classifiers are iteratively updated in a united process. (ii) The proposed collaborative training makes different modules complement each other. (iii) Knowledge transfer is reflected from three aspects in CCL: image translation by WCycleGAN, transfer learning by fine-tuning the CNN pre-trained on ImageNet to identify pneumonia, and the parameter passing in fine-tuning.

2   Method

The proposed CCL is constructed by a WCycleGAN G and two classifiers C_t, C_s, as shown in Fig. 1. The details will be introduced in the following section.
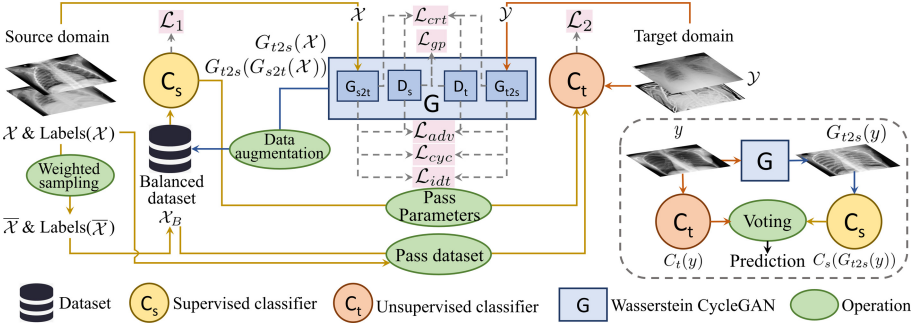
Fig. 1. The training (a) and testing (b) diagrams of the proposed method.

2.1 Wasserstein CycleGAN in the Proposed Method

WCycleGAN is designed for data augmentation and domain translation. We are inspired by Wasserstein GAN [1], Wasserstein GAN with gradient penalty [10] and CycleGAN [33] to develop the proposed WCycleGAN. The inputs of the source domain and the target domain are respectively described as labeled $\mathcal{X} = \{x_i\}_{i=1}^n$ and unlabeled $\mathcal{Y} = \{y_i\}_{i=1}^n$, where n is the number of images in a batch.

To solve the issue of class-imbalanced dataset, we construct an extreme class-balanced dataset $\mathcal{X}_B = \{x_i\}_{i=1}^m$ by G, where m is the total number of images and the number of images per class is the same. As shown in Fig. 1 (a), in order to construct $\mathcal{X}_B$, we first obtain a balanced dataset $\overline{\mathcal{X}}$ by performing weighted sampling on $\mathcal{X}$. Then we supplement $\overline{\mathcal{X}}$ with some intermediate images, which are generated by G during the calculation of losses. In this way, some images may be resampled for $\mathcal{X}_B$. Next, we will elaborate on the acquisition of the intermediate images during training G.

We respectively denote $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ as the data distributions of the source domain and the target domain. There are four components in WCycleGAN: the two discriminators D_s and D_t can distinguish the domain of images; the two generators are able to translate images, i.e. $G_{s2t}: \mathcal{X} \to \mathcal{Y}$ and $G_{t2s}: \mathcal{Y} \to \mathcal{X}$.

The adversarial loss constrains the generators and discriminators:

$$\mathcal{L}_{adv} = \underset{x \sim P_{\mathcal{X}}}{\mathbb{E}}[-D_s(G_{s2t}(x))] + \underset{y \sim P_{\mathcal{Y}}}{\mathbb{E}}[-D_t(G_{t2s}(y))] \tag{1}$$

In order to enforce G_{s2t} and G_{t2s} to be cycle-consistent with each other, we calculate the cycle consistency loss:

$$\mathcal{L}_{cyc} = \underset{x \sim P_{\mathcal{X}}}{\mathbb{E}}[\|G_{t2s}(G_{s2t}(x)) - x\|_1] + \underset{y \sim P_{\mathcal{Y}}}{\mathbb{E}}[\|G_{t2s}(G_{s2t}(y)) - y\|_1] \tag{2}$$

To further ensure the generation capacity of the generator, we also introduce the identity mapping loss [26]:

$$\mathcal{L}_{idt} = \underset{x \sim P_{\mathcal{X}}}{\mathbb{E}}[\|G_{t2s}(x) - x\|_1] + \underset{y \sim P_{\mathcal{Y}}}{\mathbb{E}}[\|G_{s2t}(y) - y\|_1] \tag{3}$$

It is noted that G_{t2s} and G_{s2t} are generators shown in Fig. 1 (a), from which we can get $G_{t2s}(G_{t2s}(x))$ of Eq. (2) and $G_{t2s}(x)$ of Eq. (3) as the intermediate images for data augmentation. The overall loss of generators is:

$$\mathcal{L}_{gen} = \mathcal{L}_{adv} + \mathcal{L}_{cyc} + \mathcal{L}_{idt} \tag{4}$$

Then, we take D_s for example to illustrate the optimal objective of discriminators, and its critic loss is:

$$\mathcal{L}_{crt}(D_s, G_{t2s}, \mathcal{X}, \mathcal{Y}) = \underset{y \sim P_{\mathcal{Y}}}{\mathbb{E}} [D_s(G_{t2s}(y))] - \underset{x \sim P_{\mathcal{X}}}{\mathbb{E}} [D_s(x)] \tag{5}$$

On this basis, the gradient penalty term is added to constrain the gradient norm of the outputs of discriminator:

$$\mathcal{L}_{gp}(D_s, \mathcal{X}) = \underset{\hat{x} \sim P_{\hat{\mathcal{X}}}}{\mathbb{E}} [\|(\nabla_{\hat{x}} D_s(\hat{x})\|_2 - 1)^2] \tag{6}$$

where $\hat{x} = \epsilon x + (1 - \epsilon)G_{s2t}(x)$ represents sampling from source data x and generating data $G_{s2t}(x)$, and $\epsilon \in U([0, 1])$ is a random number. Therefore, the overall loss of D_s is:

$$\mathcal{L}_{dis}(D_s, G_{t2s}, \mathcal{X}, \mathcal{Y}) = \lambda_1 \mathcal{L}_{crt}(D_s, G_{t2s}, \mathcal{X}, \mathcal{Y}) + \lambda_2 \mathcal{L}_{gp}(D_s, \mathcal{X}) \tag{7}$$

where λ_1 and λ_2 are two hyper-parameters. Similarly, $\mathcal{L}_{dis}(D_t, G_{s2t}, \mathcal{Y}, \mathcal{X})$ is the loss for D_t. Finally, the optimization task of WCycleGAN is formulated by:

$$\mathcal{L}_{total} = \mathcal{L}_{gen} + \mathcal{L}_{dis}(D_s, G_{t2s}, \mathcal{X}, \mathcal{Y}) + \mathcal{L}_{dis}(D_t, G_{s2t}, \mathcal{Y}, \mathcal{X}) \tag{8}$$

The WCycleGAN G is an important part of CCL: $\mathcal{X}_B$ is formed with the help of G, and G is able to translate images to a given domain.

2.2   Cooperative Training of Classifiers Based on Transfer Learning

The classifiers C_s and C_t in a collaborative way are trained with a batch during each iterative update, and they are respectively designed for the classification tasks of the source domain and the target domain.

Supervised Classifier C_s is obtained by fine-tuning a CNN that pre-trained on ImageNet. It can be used to classify images of the source domain and conditionally assist the training of C_t.

In the current iteration, after constructing $\mathcal{X}_B$ by G and feeding $\mathcal{X}_B$ into C_s, we can get the feature maps $\mathcal{F} = \{f_i\}_{i=1}^m \in \mathbb{R}^{m \times (D \times K)}$ before the final softmax layer, where D is the dimension of images and K is the number of classes. Each $x_i \in \mathcal{X}_B$ in the source domain is labeled as $Label(x_i)$, so $f_{i,Label(x_i)}$ is a D-dimensional feature map of x_i. The objective function of C_s is a combination of the cross entropy loss $\mathcal{L}_S$ and the center loss $\mathcal{L}_C$ [30] balanced by a hyper-parameter λ_3:

$$\mathcal{L}_1 = \mathcal{L}_S + \lambda_3 \mathcal{L}_C = -\sum_{i=1}^{m} log\left(\frac{e^{f_{i,Label(x_i)}}}{\sum_{k=1}^{K} e^{f_{i,k}}}\right) + \frac{\lambda_3}{2}\sum_{i=1}^{m}\left\|f_{i,Label(x_i)} - c_{l_i}\right\|_2^2 \quad (9)$$

where c_{l_i} is the feature center of the l_i class, and l_i is the label of the i-th images.

Unsupervised Classifier C_t is also obtained by fine-tuning the CNN that pre-trained on ImageNet, which is an unsupervised classifier and can classify images of the target domain. During each iterative training of C_t, we exploit labeled $\mathcal{X}$, $\mathcal{X}_B$ and unlabeled $\mathcal{Y}$.

The unsupervised classifier C_t needs to receive the parameters and images from the labeled source domain, so as to realize unsupervised domain adaptation well. During the iterative training process, as shown in Fig. 1 (a), we design a cooperative mechanism to control C_s to pass parameters to C_t under two conditions: (i) the training/verification accuracy of C_s does not reach the threshold τ; (ii) the predictions of C_t per batch all belong to a certain class (over-fitting).

When the above conditions are not satisfied in the current iteration, C_t will train by itself to update, in which MMD [27] criterion is utilized to minimize the distribution distance between $\mathcal{X}$ and $\mathcal{Y}$. The squared MMD distance is:

$$d(\mathcal{X},\ \mathcal{Y})^2 = \left\|\frac{1}{n}\left(\sum_{x_i \in \mathcal{X}} \phi(x_i) - \sum_{y_i \in \mathcal{Y}} \phi(y_i)\right)\right\|_H^2 \quad (10)$$

where $\phi(\cdot)$ is the mapping function corresponding to the Gaussian kernel, and the subscript H means the distance is measured by using $\phi(\cdot)$ to map the data into reproducing kernel Hilbert space.

By combining Eq. (10) with Eq. (9) using a hyper-parameter λ_4, the optimized object of C_t is:

$$\mathcal{L}_2 = \mathcal{L}_1 + \lambda_4 d(\mathcal{X},\ \mathcal{Y})^2 \quad (11)$$

2.3   Prediction

After the above iterative training, C_s and C_t are able to classify images of the corresponding domain, and G can translate the original images to a given domain. The output of each classifier is the predicted probability for each class. For obtaining more reliable prediction, we utilize a voting strategy of Fig. 1 (b) to fuse predictions.

For a probe y from the target domain, we can obtain $y' = G_{t2s}(y)$ translated by G, whose distribution is as consistent as possible with $\mathcal{X}$. Then we use both y and y' for prediction. The label of y is expressed as $Label(y) = l_k$, where l_k represents the label of the k-th class. The indicator k is worked out by:

$$k = \begin{cases} mode(\arg\max_k \left\{C_s^k(y'), C_t^k(y)\right\}), & \text{if it is unique} \\ \arg\max_k (C_s^k(y') + C_t^k(y)) & \text{else} \end{cases} \quad (12)$$

where the '$mode$' of Eq. (12) chooses a value that appears the most frequently.

3   Experiments

3.1   Databases and Settings

Databases. We use three databases in the experiments, and their information is shown in Table 1. The Chest X-Ray[1] is divided into training dataset and testing dataset [12]. Single lesion[2] and Multiple lesions[3] are the training datasets of two open lesion recognition competitions. We name them according to the number of lesions: each image consists of at most/least one lesion.

Table 1. The number of images and examples of (a) Chest X-Ray database; (b) Single lesion database; (c) Multiple lesions database. Their data distributions vary due to different imaging equipment/standards.

Name	Normal		Pneumonia		Sum	Class-imbalanced
Chest X-Ray (for training)	1349		3883		5232	✓
Chest X-Ray (for testing)	234		390		624	✓
Single lesion	20025		5659		25684	✓
Multiple lesions	15504		4509		20013	✓

Comparative Methods. Kermany *et al.* [12], Ayan *et al.* [3] and Gu *et al.* [7] (a two-step progressive transfer learning technique) are all the transfer learning methods for medical image classification. Additionally, we introduce the domain-adversarial training of neural networks (DANN) [5] and [7]-GAN (an adversarial learning technique with CycleGAN) for cross-domain classification.

Settings. All the experiments are carried out on Intel® Xeon® Gold 6230, and CCL is implemented with Pytorch 1.6.0. We set $\lambda_1 = 0.5$ [33] and $\lambda_2 = 10.0$ [10] in Eq. (7), $\lambda_3 = 1.0$ [30] in Eq. (9) and $\lambda_4 = 2.0$ [27] in Eq. (11). During training, the initial learning rate for SGD optimizer is 0.001, the weight decay is 0.0005 and the threshold τ of cooperative mechanism is 0.8 (see the Supplement Sect. 1 for the related experiments). The number per batch $n = 32$, and one batch constitutes n mini-batches including 1 images in training WCycleGAN. We can load a simple pre-trained WCycleGAN to speed up the convergence, and the training can be completed within 10 epochs with the optimal model stored

---

[1] https://data.mendeley.com/datasets/rscbjbr9sj/2/files/.
[2] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.
[3] https://god.yanxishe.com/18.

after every 50 batches. The visualization of the training process for verifying the performance of CCL can be seen from the Supplement Sect. 2.

Evaluation Metrics. The evaluation metrics are accuracy, precision, recall and F1 score. We take the normal diagnosis as the positive in this paper.

3.2   Performance of the Supervised Classifier

Table 2. Results of supervised classification on Chest X-Ray.

Method	Accuracy	Precision	Recall	F1 score	Backbone
Kermany *et al.* [12]	83.97%	65.81%	88.51%	0.75	Inception v3
Ayan *et al.* [3]	82.00%	73.08%	85.00%	0.75	Xception
Ayan *et al.* [3]	87.00%	78.63%	85.19%	0.82	Vgg16
Gu *et al.* [7]	88.00%	47.86%	91.80%	0.63	ResNet 152
CCL	89.42%	79.91%	90.78%	0.85	MobileNet v2
CCL	88.94%	76.92%	92.31%	0.84	Inception v3
CCL	**91.35%**	**80.34%**	**95.92%**	**0.87**	Vgg16

We first verify the performance of the supervised classifier C_s on Chest X-Ray. Consider the experimental results with diverse backbones, we choose MobileNetV2 (fine-tune the last 6 layers) [21], InceptionV3 (fine-tune the classification layer) [25] and Vgg16 (fine-tune the last 5 layers) [22] as backbones. From Table 2, C_s based on diverse backbones is always superior compared with other methods, which indicates that C_s is capable of classification in the same domain very well.

3.3   Evaluation of Unsupervised Cross-Domain Classification

We follow the common leave-one-domain-out strategy as [18], and use the three databases in pairs to test the performance of CCL on cross-domain classification. We get the average result of 3 runs. To ensure the comparison and repeatability of experimental results, we use the existed divided datasets for training/testing/validation, instead of using cross-validation that will change the composition of datasets. All samples of the source domain are used for training, and Chest X-Ray (for training) and Chest X-Ray (for testing) in the target domain are respectively used for validation and testing to ascertain the iterations when convergence. Then we can apply the iterations as stop criteria to test other datasets in the target domain.

From the results in Table 3, CCL has the best accuracy and F1 score for all tasks. For different classification task, diverse backbones have their own advantages: (i) when training with the large dataset and the limited device memory, the lightweight network such as MobileNetV2 is a priority; (ii) complex CNNs have better ability to prevent over-fitting to some extent.

Table 3. Results of unsupervised cross-domain classification

IDX	Method	Acc	Precision	Recall	F1 score	IDX	Acc	Precision	Recall	F1 score	Backbone
Training on Chest X-Ray. Testing on (a) Single lesion; (b) Multiple lesions.											
(a)	DANN [5]	0.57	0.51	**0.89**	0.65	(b)	0.54	0.44	**0.92**	0.60	–
	[12]	0.24	0.95	0.03	0.05		0.25	0.95	0.03	0.05	InceptionV3
	[3]	0.60	0.55	0.89	0.68		0.53	0.87	0.46	0.60	Vgg16
	[7]	0.77	**0.99**	0.77	0.87		0.71	0.91	0.76	0.83	ResNet152
	[7]-GAN	0.33	0.18	0.83	0.30		0.23	0.17	0.66	0.00	ResNet152
	CCL	0.73	0.81	0.83	0.82		0.62	0.60	0.87	0.71	As DANN
	CCL	**0.78**	**0.99**	0.78	**0.88**		0.74	0.89	0.80	0.84	MobileNetV2
	CCL	0.50	0.56	0.73	0.63		**0.77**	**0.99**	0.77	**0.87**	InceptionV3
	CCL	0.76	0.89	0.82	0.85		0.76	0.92	0.80	0.86	Vgg16
Training on Single lesion. Testing on (a) Chest X-Ray; (b) Multiple lesions.											
(a)	DANN [5]	0.77	0.66	0.73	0.70	(b)	0.72	0.73	0.89	0.80	–
	[12]	0.74	0.62	0.76	0.68		0.73	0.93	0.71	0.80	InceptionV3
	[3]	0.77	0.73	0.70	0.71		0.77	0.76	**0.93**	0.84	Vgg16
	[7]	0.65	0.66	0.53	0.59		0.23	0.02	0.56	0.04	ResNet152
	[7]-GAN	0.38	**0.97**	0.38	0.54		0.73	0.93	0.77	0.84	ResNet152
	CCL	0.78	0.45	0.98	0.62		0.72	0.70	0.92	0.80	As DANN
	CCL	0.71	0.36	**0.77**	0.49		**0.79**	0.89	0.85	0.87	MobileNetV2
	CCL	0.67	0.43	0.60	0.50		0.78	**0.99**	0.79	**0.88**	InceptionV3
	CCL	**0.79**	0.81	0.69	**0.75**		0.72	0.70	0.91	0.79	Vgg16
Training on Multiple lesions. Testing on (a) Chest X-Ray; (b) Single lesion.											
(a)	DANN [5]	0.75	0.81	0.64	0.72	(b)	0.77	0.76	0.92	0.84	–
	[12]	0.69	0.55	**0.91**	0.69		0.77	0.94	0.75	0.83	InceptionV3
	[3]	0.73	0.76	0.62	0.68		0.77	0.76	**0.93**	0.84	Vgg16
	[7]	0.59	0.42	0.45	0.43		0.55	0.66	0.73	0.69	ResNet152
	[7]-GAN	0.38	**0.98**	0.38	0.55		0.58	0.65	0.77	0.70	ResNet152
	CCL	0.77	0.61	0.76	0.68		0.77	0.83	0.87	0.85	As DANN
	CCL	0.71	0.26	0.79	0.39		**0.79**	0.97	0.80	**0.88**	MobileNetV2
	CCL	0.68	0.27	0.71	0.39		**0.79**	**0.98**	0.80	**0.88**	InceptionV3
	CCL	**0.80**	0.72	0.74	**0.73**		0.77	0.80	0.89	0.84	Vgg16

3.4 Visualization and Ablation Experiments

Figure 2 is the visualization of translating original images. Figure 2 (b) and (c) are respectively translated to another domain by CycleGAN and WCycleGAN, from which it is clear that WCycleGAN has the better generation capacity (e.g. the edges of ribs are as clear as the images of the source domain) than CycleGAN. Figure 2 (d) and (e) are respectively the intermediates in Eq. (2) and Eq. (3) and also used for data augmentation, which are similar to Fig. 2 (a) but not identical.

We take Vgg16 as the backbone to do the ablation experiments of method#1 wo passing balanced dataset, method#2 wo passing parameters and method#3 wo generating balanced dataset, which are the operations of Fig. 1. From Fig. 3, it is obvious that method#3 causes severe over-fitting, and method#1 and #2 are also greatly affected. Hence, these operations in CCL are all essential and beneficial.
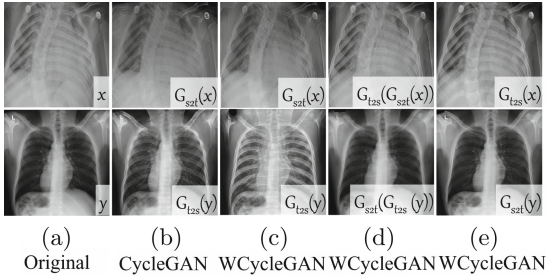
Fig. 2. The visual image translation between Chest X-Ray (row1) and Multiple lesions (row2).
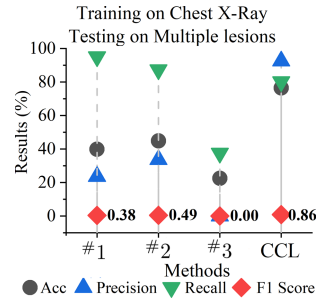
Fig. 3. The results of ablation experiments.

4   Discussion and Conclusion

In this paper, we present an effective framework named CCL based on transfer learning for unsupervised cross-domain classification. The class-balanced dataset of CCL contributes to avoiding over-fitting. Besides, the proposed method can overcome the problem of insufficient labels in medical data by combining transfer learning and unsupervised learning. During the training and testing process, WCycleGAN and two classifiers complement each other by cooperative learning, whose backbones can be flexibly modified to obtain competitive results. Experiments on three pneumonia databases indicate that the propose method achieves promising performance in unsupervised cross-domain classification.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GaN. In: ICML, pp. 1–18 (2017)
2. Armanious, K., Jiang, C., Abdulatif, S., Küstner, T., Gatidis, S., Yang, B.: Unsupervised medical image translation using Cycle-MeDGAN. In: EUSIPCO, pp. 1–5 (2019)
3. Ayan, E., Ünver, H.M.: Diagnosis of pneumonia from chest X-ray images using deep learning. In: EBBT, pp. 7–11 (2019)
4. Chiou, E., Giganti, F., Punwani, S., Kokkinos, I., Panagiotaki, E.: Harnessing uncertainty in domain adaptation for MRI prostate lesion segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 510–520. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_50
5. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**, 189–209 (2016)

6. Ghifary, M., Kleijn, W.B., Zhang, M.: Domain adaptive neural networks for object recognition. In: Pham, D.-N., Park, S.-B. (eds.) PRICAI 2014. LNCS (LNAI), vol. 8862, pp. 898–904. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13560-1_76

7. Gu, Y., Ge, Z., Bonnington, C.P., Zhou, J.: Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. IEEE J. Biomed. Health Inform. **24**(5), 1379–1393 (2020)

8. Han, X., Wang, J., Zhou, W., Chang, C., Ying, S., Shi, J.: Deep doubly supervised transfer network for diagnosis of breast cancer with imbalanced ultrasound imaging modalities. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 141–149. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_14

9. He, Y., Carass, A., Zuo, L., Dewey, B.E., Prince, J.L.: Self domain adapted network. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 437–446. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_43

10. Ishaan G., Faruk A., Martin A., Vincent D., Aaron, C.: Improved training of Wasserstein GANs. In: NeurIPS, pp. 1–11 (2017)

11. Jamal, M.A., Brown, M., Yang, M.H., Wang, L., Gong, B.: Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective. In: CVPR, pp. 7610–7619 (2020)

12. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122.e9–1131.e9 (2018)

13. Liu, J., Li, J., Liu, T., Tam, J.: Graded image generation using stratified Cycle-GAN. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 760–769. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_73

14. Liu, J., et al.: Illumination-invariant flotation froth color measuring via Wasserstein distance-based CycleGAN with structure-preserving constraint. IEEE Trans. Cybern. **51**, 839–852 (2020)

15. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. In: ICML, pp. 97–105 (2015)

16. Ma, Y., Xu, Y., Liu, F.: Multi-perspective dynamic features for cross-database face presentation attack detection. IEEE Access **8**, 26505–26516 (2020)

17. McDermott, M.B., et al.: Semi-supervised biomedical translation with cycle Wasserstein regression GaNs. In: AAAI, pp. 2363–2370 (2018)

18. Meng, Q., Rueckert, D., Kainz, B.: Unsupervised cross-domain image classification by distance metric guided feature alignment. In: Hu, Y., et al. (eds.) ASMUS/PIPPI -2020. LNCS, vol. 12437, pp. 146–157. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60334-2_15

19. Ali, N., Neagu, D., Trundle, P.: Classification of heterogeneous data based on data type impact on similarity. In: Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M. (eds.) UKCI 2018. AISC, vol. 840, pp. 252–263. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97982-3_21

20. Othman, E., Werner, P., Saxen, F., Al-Hamadi, A., Walter, S.: Cross-database evaluation of pain recognition from facial video. In: International Symposium on Image and Signal Processing and Analysis, pp. 181–186 (2019)

21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: inverted residuals and linear bottlenecks. In: CVPR, pp. 4510–4520 (2018)

22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR, pp. 1–14 (2015)

23. Su, Y., Fu, Y., Tian, Q., Gao, X.: Cross-database age estimation based on transfer learning. In: ICASSP, pp. 1270–1273 (2010)

24. Sun, Y., Yang, G., DIng, D., Cheng, G., Xu, J., Li, X.: A GAN-based domain adaptation method for glaucoma diagnosis. In: IJCNN, pp. 1–8 (2020)
25. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR, pp. 2818–2826 (2016)
26. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. In: ICLR, pp. 1–14 (2017)
27. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: maximizing for domain invariance. arXiv preprint arXiv:1412.3474, pp. 1–9 (2014)
28. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: CVPR, pp. 5385–5394 (2017)
29. Wang, C., Zhang, F., Yu, Y., Wang, Y.: BR-GAN: bilateral residual generating adversarial network for mammogram classification. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 657–666. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_63
30. Wen, Y., Zhang, K., Li, Z., Qiao, Yu.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
31. Xia, K., Ni, T., Yin, H., Chen, B.: Cross-domain classification model with knowledge utilization maximization for recognition of epileptic EEG signals. IEEE/ACM Trans. Comput. Biol. Bioinf. **18**, 53–61 (2020)
32. Yang, F.E., Chang, J.C., Tsai, C.C., Wang, Y.C.F.: A Multi-domain and multimodal representation disentangler for cross-domain image manipulation and classification. IEEE Trans. Image Process. **29**, 2795–2807 (2020)
33. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV, pp. 2242–2251 (2017)
34. Zhu, Y., et al.: Cross-domain medical image translation by shared latent Gaussian mixture model. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 379–389. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_37
35. Zhu, Y., et al.: Multi-representation adaptation network for cross-domain image classification. Neural Netw. **119**, 214–221 (2019)