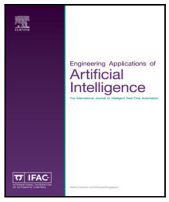
Engineering Applications of Artificial Intelligence 129 (2024) 107602

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Enhanced transfer learning with data augmentation

Jianjun Su [a], Xuejiao Yu [a], Xiru Wang [a], Zhijin Wang [b], Guoqing Chao [a,*]

[a] Harbin Institute of Technology, 2 West Culture Road, Weihai, 264209, Shandong, China
[b] Jimei University, 185 Yinjiang Road, Xiaman, 361021, Fujian, China

ARTICLE INFO

Keywords:
Transfer learning
Domain adaptation
Data augmentation
Unsupervised
Image classification
Convolutional neural network

ABSTRACT

Traditional machine learning methods require the assumption that training and test data are drawn from the same distribution, which proves challenging in real-world applications. Moreover, deep learning models require a substantial amount of labeled data for training in classification tasks and limited samples may lead to overfitting. In many real-world scenarios, there is an insufficient supply of labeled samples within the target domain for learning. Transfer learning offers an effective solution, allowing knowledge from a source domain to be transferred to a target domain. Additionally, data augmentation enhances model generalization by increasing data samples, particularly beneficial when dealing with limited target domain data. In this paper, we synergistically enhance the model's performance on classification tasks by integrating transfer learning techniques with a data augmentation strategy. By conducting numerous experiments across various datasets, we verified the effectiveness of our proposed approach.

1. Introduction

In recent years, significant success have been achieved in image classification tasks through the utilization of deep neural networks, particularly exemplified by convolutional neural networks (Lu and Weng, 2007; Kang et al., 2022). In 2015, the concept of residual networks (He et al., 2016) was introduced to address the issue of network degradation resulting from the increased depth of neural networks, and it performed well in ImageNet classification tasks. While traditional deep neural network has obtained a great success, an underlying assumption of this approach is that the training and test data share an identical probability distribution. In numerous real-world scenarios, this assumption could be challenging due to the complexities associated with gathering new instances that exhibit identical attributes, dimensions, and distributions. Furthermore, given the labor-intensive and time-consuming nature of labeling data, there is an inclination to leverage existing labeled data instead of procuring new labeled datasets. Nevertheless, when the distribution of the training data deviates from that of the test data, the model's performance on the test data tends to deteriorate. To tackle this problem, more and more researchers are focusing on transfer learning (Wang and Deng, 2018). The core objective of transfer learning is to acquire a model from annotated source domain data and subsequently extend its applicability to the target domain, which can be achieved through the reduction of dissimilarities between the source and target domains.

Transfer learning (Chao et al., 2023b) is a popular machine learning paradigm, dedicated to scenarios where disparities in the distributions of both the task and domain arise between the source and target domains. Within the framework of transfer learning, a domain encompasses the feature space and marginal probability distribution, while a task encompasses the label space and the target prediction function. Therefore, diverse scenarios arising from distinct domains or tasks give rise to various configurations for transfer learning. One is domain adaptation (Pan and Yang, 2009) and its unsupervised version unsupervised domain adaptation mainly focuses on covariate shift. Unsupervised domain adaptation concentrates on the situation that conditional probability distributions are the same ($p_S(y|x) = p_T(y|x)$), while marginal probability distributions differ ($p_S(x) \neq p_T(x)$). In this scenario, the primary goal of unsupervised domain adaptation is to reduce the disparity in marginal distributions between the source and target domains.

Data augmentation enhances model performance and addresses the constraint of limited data within the target domain by augmenting the samples within the data space. Furthermore, data augmentation (Shorten and Khoshgoftaar, 2019) has demonstrated its efficacy in enhancing the models' capacity for generalization to new, unknown samples. Given the significant reliance of deep neural networks on extensive datasets for model training to mitigate overfitting and transfer learning can alleviate the heavy data burden in the target domain, this paper tries to synergistically enhance the models' generalization capabilities by integrating transfer learning and data augmentation strategies. The most commonly-used data augmentation methods are

* Corresponding author.
E-mail address: guoqingchao@hit.edu.cn (G. Chao).

https://doi.org/10.1016/j.engappai.2023.107602
Received 26 August 2023; Received in revised form 19 November 2023; Accepted 21 November 2023

color space transform, geometric space transform, kernel filter, random erasure, etc. Recently, several data augmentation algorithms rooted in generative models have emerged. These algorithms employ generative adversarial networks to produce images within the target domain, effectively augmenting the target domain data.

The main contributions of this paper can be summarized as follows:

1. In this paper, we have regarded data augmentation and transfer learning as effective strategies to solve the data scarcity problem of classification tasks in target domain. Subsequently, we have synergistically integrated these approaches to enhance the overall classification performance. To the best of our knowledge, this is the pioneering endeavor to comprehend and harness these strategies to promote classification performance within target domains.
2. Prior transfer learning methodologies primarily concentrate on aligning the distribution between the source and target domains (Farahani et al., 2021), without explicitly considering the performance on forthcoming unknown target domain data. In this article, we initially partition the target domain data into training and test subsets. We conduct data augmentation on train data and align the source domain data to them, and then test the accuracy on test data. Consequently, we are able to quantify the generalization capability of our transfer learning model.
3. We conducted experiments on a total of five datasets, encompassing four public datasets as well as a medical image dataset. By comparing against traditional transfer learning approaches, we have demonstrated the superiority of our proposed method.

The remainder of this paper is structured as follows: Section 2 provides an overview of related research. In Section 3, we present the enhanced transfer learning methodology and delve into the data augmentation strategy. Section 4 showcases the experimental outcomes and subsequently dissects the findings. In Section 5, we provide conclusions and outline potential directions for future research.

2. Related works

In this section, we will introduce four related studies: convolutional neural network, domain adaptation, maximum mean discrepancy (MMD), and data augmentation.

2.1. Convolutional neural network

Convolutional neural network could encode the original data to learn potentially useful features (Rawat and Wang, 2017), and it has obtained remarkable success in image classification. He et al. (2016) proposed ResNet in 2015, with shortcut connection, ResNet could effectively prevent the network degradation problem caused by adding too many layers to neural network. It incorporates residual networks into each successive layer of the convolutional neural network. Subsequently, it defines the structure of the residual block as follows:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x},$$

where $\mathbf{x}$ and $\mathbf{y}$ symbolize the input and output vectors of each block. The function $\mathcal{F}(\mathbf{x}, \{W_i\})$ embodies the residual mapping that is to be learned for each block. There are over 23 million trainable parameters in ResNet50 model, it is composed of five stages, each containing a convolution block and an identity block. Within both the convolution block and the identity block, there are three convolutional layers.

2.2. Maximum Mean Discrepancy (MMD)

Maximum Mean Discrepancy (MMD) (Ben-David et al., 2010, 2006) is commonly used as a loss function in domain adaptation. The main idea of MMD is that if two random variables possess identical moments of any order, their distributions are considered consistent. The adoption of MMD necessitates the selection of a kernel function, which facilitates the transformation of source and target data into a reproducing kernel Hilbert space. Subsequently, the mean discrepancy between the source and target domains is computed following the mapping process. Numerous domain adaptation strategies built upon MMD have been proposed in the literature (Long et al., 2013; Zhu et al., 2019; Ouyang and Key, 2021). Assigning the symbol p to represent the source domain distribution and the symbol q to indicate the target domain distribution, MMD introduces the subsequent disparity metric:

$$d_{\mathcal{H}(p,q)} = \|\mathrm{E}_p[\phi(\mathbf{x}^S)] - \mathrm{E}_q[\phi(\mathbf{x}^T)]\|^2_{\mathcal{H}}.$$

Here, $\mathcal{H}$ indicates the reproducing kernel Hilbert space (RKHS), $\phi(\cdot)$ indicates the kernel function that transforms the initial samples into the RKHS, and the kernel k is given by $k(\mathbf{x}^S, \mathbf{x}^T) = \langle \phi(\mathbf{x}^s), \phi(\mathbf{x}^T) \rangle$, where $\langle \cdot, \cdot \rangle$ represents the inner product of vectors (Gretton et al., 2012). The equivalence of distributions between the source domain and the target domain is achieved when $D_{\mathcal{H}}(p, q) = 0$. In practical implementations, an estimation of MMD is obtained by quantifying the squared distance between the empirical kernel mean embeddings, denoted as follows:

$$\begin{aligned}\hat{d}_{\mathcal{H}}(p,q) &= \| \frac{1}{n^S} \sum_{\mathbf{x}_i \in D_S} \phi(\mathbf{x}_i) - \frac{1}{n^T} \sum_{\mathbf{x}_j \in D_T} \phi(\mathbf{x}_j) \|^2_{\mathcal{H}} \\ &= \sum_{i=1}^{n^S} \sum_{j=1}^{n^S} \frac{1}{{n^S}^2} k(\mathbf{x}_i^S, \mathbf{x}_j^S) + \sum_{i=1}^{n^T} \sum_{j=1}^{n^T} \frac{1}{{n^T}^2} k(\mathbf{x}_i^T, \mathbf{x}_j^T) \\ &\quad - \sum_{i=1}^{n^S} \sum_{j=1}^{n^T} \frac{2}{n^S n^T} k(\mathbf{x}_i^S, \mathbf{x}_j^T).\end{aligned}$$

In this context, $\hat{d}_{\mathcal{H}}(p, q)$ acts as an unbiased estimation of $d_{\mathcal{H}}(p, q)$, with D_s representing the source domain distribution and D_t signifying the target domain distribution.

2.3. Domain adaptation

The objective of transfer learning is to construct a model capable of transferring the acquired knowledge from the source domain to the target domain. A classical survey on transfer learning by Pan and Yang (2009) introduced the concept of unsupervised domain adaptation, which posits that the conditional probability distributions are equivalent across domains ($p_s(y|x) = p_t(y|x)$), while the marginal probability distributions differ ($p_s(x) \neq p_t(x)$). The labeled source domain data is represented as $D_S = \{x_i^S, y_i^S\}_{i=1}^{n^S}$, and the unlabeled target domain data is denoted as $D_T = \{x_j^T\}_{j=1}^{n^T}$. Under the assumption that the source domain and target domain exhibit equivalent conditional distributions, the ratio of the joint probability distribution between these two domains (Patel et al., 2015) can be restated as follows:

$$\frac{p_T(x,y)}{p_S(x,y)} = \frac{p_T(x)p_T(y|x)}{p_S(x)p_S(y|x)} = \frac{p_T(x)}{p_S(x)}.$$

Thus, the divergence between the source domain and the target domain can be estimated by $(p_T(x)/p_S(x))$. Domain adaptation enables the established knowledge transfer from the labeled source domain to the unlabeled target domain by uncovering the domain-invariant structures that bridge disparate domains (Pan and Yang, 2009). In the past several years, numerous models have been introduced to address the challenge posed by the divergence between source domain and target domain distributions. These methods could be categorized into shallow and deep architectures (Farahani et al., 2021), and deep domain adaptation methodologies can also be classified into three categories: discrepancy-based, adversarial-based, and reconstruction-based approaches.

1. Discrepancy-based method. The discrepancy-based domain adaptation normally adopt MMD or the variant of MMD as loss function. Some kernel functions will be used to project the source and target domains into a reproducing kernel Hilbert space, and then the mean discrepancy or its variants are used to measure the gap between the source and target domains.
2. Reconstruction-based method. An encoder and decoder architecture is employed to harmonize the distribution between the source and target domains. Initially, this approach employs a shared encoder to project samples from both the source and target domains into a high-dimensional feature space. Subsequently, it employs the decoder to reconstruct the target domain samples in a way that maximally preserves their original characteristics.
3. Adversarial-based method. Inspired from the idea of generative adversarial networks (Goodfellow et al., 2014), this class of methods encompass two key components: a generator and a discriminator. The generator is used to encode samples from both the source domain and the target domain into a shared feature space and the discriminator is employed to determine the origin of each sample. When the discriminator cannot distinguish the origin of the samples well, it means that the source domain and target domain share an identical distribution after encoding (Ganin and Lempitsky, 2015).

Maximum mean discrepancy (MMD) (Gretton et al., 2012), Correlation Alignment (CORAL), Kullback–Leibler (KL) divergence, and contrastive domain discrepancy are among the most commonly-used distance metric functions in domain adaptation.

2.4. Data augmentation

Since deep neural network relies on a large amount of input data for model training, increasing the quantity of the data set by data augmentation can effectively improve the generalization ability of the model (Van Dyk and Meng, 2001). Data augmentation has several advantages over other training methods in deep learning. Firstly, it helps to prevent overfitting by artificially increasing the size of the training dataset, which can improve the generalization performance of the model. Secondly, it can be used to improve the performance of deep learning models in domains with limited data. Thirdly, it can reduce the need for manual data labeling, which can be time-consuming and expensive. Finally, data augmentation can improve the robustness of the model to variations in the input data, such as changes in lighting, orientation, and scale. Nowadays, data augmentation has been widely used in deep neural network (Shorten and Khoshgoftaar, 2019; Chao et al., 2023a). Herein we introduce some data augmentation techniques as follows:

1. Flipping is the simplest data augmentation strategy, it is generally divided into horizontal and vertical flipping, but vertical flipping occasionally leads to label changes, so the more common choice is horizontal flipping.
2. Color space augmentation (Tellez et al., 2019) involves modifying the brightness, contrast, and saturation of an image through straightforward matrix operations, thereby either augmenting or reducing these attributes.
3. Cropping changes the size of the input image by a pre-configured threshold. Depending on the selected cropping threshold, this transformation might not necessarily preserve the labels of the data.
4. Rotation within the range of 1 to 20 degrees or −1 to −20 degrees have demonstrated helpful in tasks involving digit recognition. However, as the degree of rotation increases, the transformation cannot preserve the original labels of the data.
5. Translation, which involves shifting images horizontally or vertically (left, right, up, or down), serves as a valuable transformation to mitigate potential positional bias within the data.
6. Noise injection could help deep neural network learn more robust features. This process involves the introduction of a matrix comprising random values drawn from a Gaussian distribution.

3. Method

Within this section, we provide the proposed enhanced transfer learning model. We leverage data augmentation to augment the volume of the target domain data, followed by the application of transfer learning models to mitigate the divergence between the source domain and the target domain. The general process of our enhanced transfer learning with data augmentation could be depicted as Algorithm 1.

Algorithm 1 Enhanced transfer learning with data augmentation

Input: Source domain, target domain, strategy of data augmentation, backbone network, distance metric function, maximum iteration number.
Output: Classification model of target domain.
Initialize parameters of backbone network.
Implementing Data Augmentation Algorithms to Expand the Scale of Target Domain Data.
repeat
    Using the backbone network as an encoder to map source and target domain data to a high-dimensional feature space.
    Computing distributions differences between domains using a distance metric function in high-dimensional feature space.
    Backpropagation calculates the parameter gradient and updates the parameters.
until Reach maximum number of iterations

3.1. Data augmentation

For these four public data sets, we implement data augmentation on train data D_{ttr}. After comparing several data augmentation algorithms' combination. We finally choose the strategies as follows:

1. Horizontal axis flipping.
2. Noise injection. Add Gaussian noise into images. The Gaussian distribution is set with kernel size 5 and standard deviation(σ) as 3.
3. Random rotation. Choose a random degree from −5° to 5°.
4. Center cropping. Set the cropping size as raw image size $*$ 0.9.
5. Color space augmentation. Set brightness, contrast and saturation degree from 0.8 to 2.

Fig. 1 depicts the original image, while Fig. 2 illustrates the images obtained after conducting specific data augmentation operations. Aug 1 implements horizontal axis flipping; Aug 2 implements random rotation; Aug 3 implements color space augmentation; Aug 4 implements noise injection. The notation "Aug 2 + Aug 1" indicates the sequential application of Aug 2 after Aug 1, similarly, "Aug 3 + Aug 1" denotes the sequential application of Aug 3 following Aug 1.

3.2. Enhanced transfer learning

Given the source domain data $D_S = \{(\mathbf{x}_i^S, \mathbf{y}_i^S)\}_{i=1}^{n^S}$, and the target domain training data $\hat{D}_{ttr}$. Resorting to Resnet50 (He et al., 2016) as the foundational backbone, we extract transferable features from the raw image data space. The loss function of the convolutional neural network is represented as follows:

$$\mathcal{L}_{NN} = \frac{1}{n^S} \sum_{i=1}^{n^S} J, \tag{1}$$

Fig. 1. An image sample of image-CLEF data set.

where $J(\cdot,\cdot)$ indicates the cross-entropy loss. The classification model for the source domain is acquired through the minimization of the expected risk linked with the labeled source domain data:

$$\begin{aligned} R_S(h) &= \mathbb{E}_{(x,y)\sim p_S(x,y)}[\mathcal{L}(f(x),y)] \\ &= \sum_{y\in\mathcal{Y}} \int_{\mathcal{X}} \mathcal{L}(f(x),y) P_S(x,y) dx, \end{aligned} \tag{2}$$

where $P_S(x,y)$ represents the joint probability distribution of source domain, and $\mathcal{L}(f(x),y)$ represents the classification loss. However, for the transfer learning tasks, our target is to increase the accuracy on target domain, thus, we could rewrite the above equation as follows:

$$\begin{aligned} R_T(h) &= \mathbb{E}_{(x,y)\sim p_T(x,y)}[\mathcal{L}(f(x),y)] \\ &= \sum_{y\in\mathcal{Y}} \int_{\mathcal{X}} \mathcal{L}(f(x),y) P_T(x,y) dx \\ &= \sum_{y\in\mathcal{Y}} \int_{\mathcal{X}} \mathcal{L}(f(x),y) \frac{P_T(x,y)}{P_S(x,y)} P_S(x,y) dx \\ &= \mathbb{E}_{(x,y)\sim p_S(x,y)} \left[\frac{P_T(x)\cancel{P_T(y|x)}}{P_S(x)\cancel{P_S(y|x)}} \mathcal{L}(f(x),y) \right], \end{aligned} \tag{3}$$

where $P_S(x,y)$ and $P_T(x,y)$ represent the joint probability distributions of source domain and target domain, respectively. In order to minimize the expected risk $R_T(h)$, it is imperative to simultaneously minimize the loss function $\mathcal{L}(h(x),y)$ and synchronize the marginal distributions of the source domain and the target domain.

The selection of distance metric functions between source domain data and target domain data varies based on the specific transfer learning models employed. By combining the classification loss with the distance metric function between source domain data and target domain data, the total loss function can be expressed as follows:

$$\min_{\theta} \frac{1}{n^S} \sum_{i=1}^{n^S} J\left(f(\mathbf{x}_i^S), y_i^S\right) + \lambda \sum_{\ell=fc_1}^{fc_n} d_k\left(D_S^{\ell}, D_T^{\ell}\right), \tag{4}$$

where the parameter λ serves as a trade-off coefficient. $\lambda = 0$ means only considering the cross-entropy loss of classification task and $\lambda \to \infty$ means only considering the discrepancy between source and target domain.

4. Experiments

In this section, we conduct a comparative analysis between our enhanced transfer learning approach and conventional transfer learning models across four publicly available datasets, as well as a dataset comprising pneumonia X-ray images. The transfer learning models include: DAN (Long et al., 2015), DAAN (Yu et al., 2019), DANN (Ganin and Lempitsky, 2015), BNM (Cui et al., 2020) DeepCORAL (Sun and Saenko, 2016) and DSAN (Zhu et al., 2020). The four public data sets are Office-31, Office-Caltech10, Image-CLEF and Office-Home.

Table 1
The statistics of four datasets.

Dataset	#Domain	#Domain category	#Class	#Sample
Office31	3	A, D, W	31	2817
Office-Caltech10	4	C, A, D, W	31	2533
Office-Home	4	A, C, P, R	65	15 500
Image-Clef	3	C, I, P	12	1800

Table 2
Introduction of pneumonia X-ray dataset.

Domain	Domain category	#Positive	#Negative	#Total
Source domain	Normal pneumonia	3418	1020	4438
Target training domain	COVID-19	460	246	706
Target test domain	COVID-19	116	113	229

4.1. Datasets

We adopt four general datasets and one medical dataset to perform our algorithm and make comparison with other algorithms. An overview of the datasets is provided below (see Table 1): **Office-31** (Saenko et al., 2010) serves as a benchmark dataset for domain adaptation, comprising 4110 images categorized across 31 classes. Office-31 consists of three distinct domains: **A** short for Amazon (comprising images from amazon.com), **W** short for Webcam (including images captured by webcams), and **D** short for DSLR (containing images acquired using digital SLR cameras). Given the presence of imbalanced images within each domain, to ensure unbiased assessment, we evaluated six distinct combinations: (**A→D, A→W, D→A, D→W, W→A, W→D**). It is noteworthy that in the notation $X \to Y$, X represents the source domain, and Y represents the target domain.

Office-Caltech-10 (Gong et al., 2012) is a standard dataset utilized for domain adaptation study. It comprises the 10 common categories present in both Office-31 and Caltech-256 (**C**). Office-Caltech-10 encompasses four domains: Caltech, Amazon, Webcam, and DSLR. We conducted evaluations on six distinct combinations: (**C → A, C → W, C → D, A → C, W → C, D → C**).

Office-Home (Venkateswara et al., 2017) serves as a standard dataset for domain adaptation purposes. The dataset includes an average of about 70 images per class, while certain classes may have up to 99 images. Office-home contains four domains: Art (which contains images consisted of art pictures in the form of sketches, paintings, and decorations), Clipart, Product (which contains images consisted of background-free object), Real-World-images (which contains images acquired by camera). We evaluated twelve different combinations (**A → C, A → P, A → R, C → A, C → P, C → R, P → A, P → C, P → R, R → A, R → C, R → P**).

Image-CLEF-DA, a benchmark dataset for domain adaptation, is structured around 12 shared categories across three distinct domains. Each category contains 50 images and each domain contains 600 images. The dataset comprises three distinct domains: Caltech-256 (**C**), ImageNet ILSVRC 2012 (**I**), and Pascal VOC 2012 (**P**). We conducted experiments across six distinct combinations: (**C → I, C → P, I → C, I → P, P → C, P → I**) (see Fig. 3).

We also focus on medical image field, and implement our enhanced transfer learning on a pneumonia X-ray dataset. Detailed information about this dataset can be found in Table 2.

To simulate the real-world scenarios, We use all the source domain data as training source domain data, and split target domain data into 30% as training data and 70% as test data. For the normal transfer learning models, we utilize the training source domain data alongside the training target domain data to compose the training set, while the

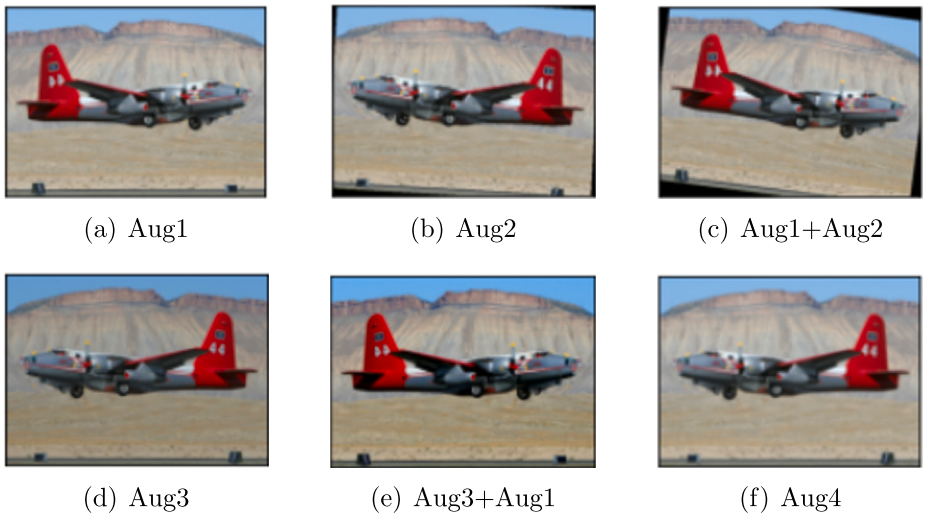


Fig. 2. Examples obtained after specific data augmentation operation.

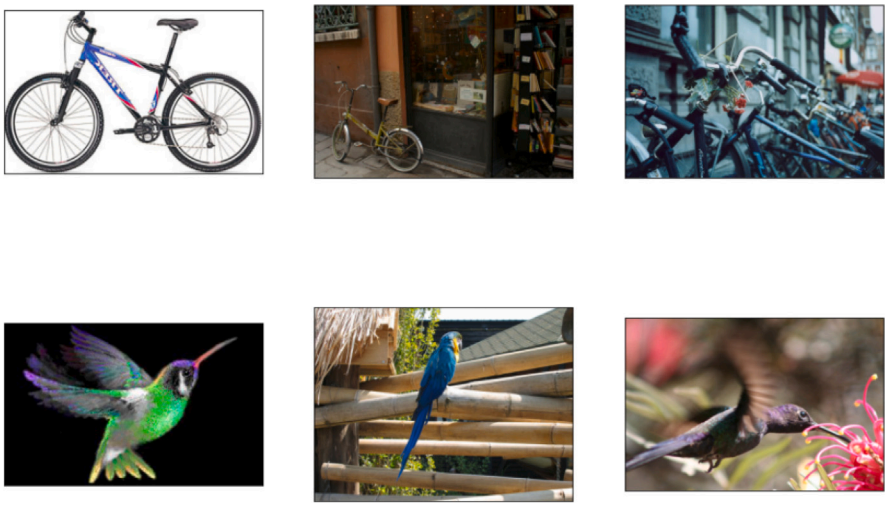


Fig. 3. Examples of image-CLEF data set.

Table 3
The parameter settings.

Parameter	Base LR	Weight decay	LR gamma	LR decay	Momentum	n epoch	Random seed
Value	0.003	0.0005	0.0003	0.75	0.9	20	1

Table 4
The accuracy (%) on Office-31 dataset through unsupervised domain adaptation (ResNet50), the version "+DA" indicates the enhanced transfer learning method with data augmentation operations.

Method	A→W	A→D	D→A	D→W	W→A	W→D	Avg
Resnet (He et al., 2016)	68.4	68.9	62.5	96.7	60.7	99.3	76.0
DAN (Long et al., 2015)	83.7	78.9	**67.5**	98.0	67.4	99.5	82.5
DAN+DA	**85.0**	**83.7**	66.7	**98.7**	**67.9**	**100.0**	**83.7**
DAAN (Yu et al., 2019)	71.9	76.6	64.5	**98.0**	64.3	99.6	79.1
DAAN+DA	**74.4**	**77.7**	**64.6**	97.8	**64.3**	**100.0**	**79.8**
DANN (Ganin and Lempitsky, 2015)	82.3	83.0	67.1	98.0	**68.9**	**100.0**	83.2
DANN+DA	**87.1**	**83.4**	**67.5**	**98.4**	68.8	99.7	**84.2**
BNM (Cui et al., 2020)	90.0	87.2	71.2	**99.2**	71.5	99.6	86.4
BNM+DA	**90.1**	**88.3**	**72.0**	99.0	**71.5**	**99.6**	**86.7**
DeeepCoral (Sun and Saenko, 2016)	79.3	**81.9**	**65.4**	**98.0**	**66.1**	99.6	**81.7**
DeeepCoral+DA	**79.5**	81.3	64.8	97.8	65.9	**99.7**	81.5
DSAN (Zhu et al., 2020)	87.4	85.5	74.3	**98.8**	**70.9**	100.0	86.1
DSAN+DA	**89.9**	**87.0**	**74.9**	98.1	70.5	**100.0**	**86.7**

Table 5
The accuracy (%) on the Office_caltech_10 dataset through unsupervised domain adaptation (ResNet50), the version "+DA" indicates the enhanced transfer learning method with data augmentation operations.

Method	C→A	C→D	C→W	A→C	D→C	W→C	Avg
DAN	95.1	91.9	96.2	92.9	91.5	93.0	93.4
DAN+DA	**96.3**	**94.2**	**96.8**	**93.6**	**91.5**	**93.7**	**94.3**
DAAN	94.8	93.1	94.1	92.0	90.1	91.3	92.5
DAAN+DA	**96.1**	**94.0**	**97.8**	**92.3**	**90.7**	**91.3**	**93.7**
DANN	94.5	89.6	95.6	94.2	92.1	93.1	93.1
DANN+DA	**96.2**	**90.8**	**96.8**	**94.4**	**92.6**	**93.8**	**94.1**
BNM	95.3	94.2	**98.4**	95.0	94.2	94.7	95.3
BNM+DA	**96.9**	**96.3**	98.3	**95.3**	**95.3**	**95.4**	**96.3**
DeeepCoral	94.8	91.4	**97.3**	91.9	91.2	92.1	93.1
DeeepCoral+DA	**95.1**	**95.0**	97.1	**92.4**	**91.4**	**92.6**	**94.0**
DSAN	95.6	90.8	96.8	94.7	95.5	94.9	94.7
DSAN+DA	**96.5**	**93.1**	**97.8**	**96.5**	**96.0**	**95.6**	**95.9**

Table 6
The accuracy(%) on the Image-ClEF dataset through unsupervised domain adaptation (ResNet50), the version "+DA" indicates the enhanced transfer learning method with data augmentation operations.

Method	c→i	c→p	i→c	i→p	p→c	p→i	Avg
DAN	90.7	78.6	94.3	98.9	91.4	91.2	90.8
DAN+DA	**90.9**	**79.6**	**94.4**	**98.9**	**91.9**	**91.5**	**91.2**
DAAN	86.6	**75.7**	91.6	78.4	89.5	90.5	85.3
DAAN+DA	**89.3**	75.6	**92.7**	**79.3**	**90.1**	**91.2**	**86.4**
DANN	90.2	78.1	94.6	78.4	91.6	90.7	87.2
DANN+DA	**91.3**	**79.9**	**95.1**	**79.1**	**92.2**	**91.7**	**88.3**
BNM	91.7	78.9	**95.5**	78.1	93.8	92.7	88.4
BNM+DA	**92.5**	**78.9**	95.1	**78.7**	**94.9**	**92.7**	**88.8**
DeeepCoral	88.0	77.6	93.1	**78.6**	90.2	90.0	86.2
DeeepCoral+DA	**88.8**	**77.7**	**93.4**	78.4	**90.4**	**90.3**	**86.5**
DSAN	90.8	79.9	95.3	78.1	92.9	91.9	88.1
DSAN+DA	**91.9**	**80.1**	**95.5**	**79.6**	**94.1**	**93.7**	**89.2**

test target domain data is used as the test set. For the enhanced transfer learning models, initially, we apply data augmentation operations to the training target domain data, thereby generating augmented training samples. Subsequently, we compile the training set by combining the source domain data and the augmented training target domain data. This training set is then used for model training and subsequently evaluated for performance on the test target data.

4.2. Compared methods

To verify the effectiveness of the enhanced transfer learning approach, we conducted comparisons with the subsequent conventional transfer learning models.

Deep Adaptation Network (DAN) (Long et al., 2015) is a discrepancy-based method, which constructs a total kernel by using multiple kernel functions, and automatically select the optimal kernel function. By incorporating an MK-MMD-based multi-layer adaptation regularizer into the fully connected layers, this approach has the capacity to acquire transferable features that effectively mitigate the cross-domain disparities.

DANN (Ganin and Lempitsky, 2015) constitutes an adversarial-based technique. The fundamental architecture of DANN comprises three key components: a feature extractor, a classifier, and a discriminator. Remarkably, the feature extractor and classifier essentially form a conventional classification model. DANN is inspired by Generative Adversarial Network (Goodfellow et al., 2014). Instead of fake samples, fake features, which are sufficient to make the target and source domains indistinguishable, are generated.

Dynamic Adversarial Adaptation Network (DAAN) (Yu et al., 2019) is an adversarial-based method, which could dynamically adjust the boundary and conditional distribution relationship. The basic network is the same as DANN network, besides, DAAN introduces conditional domain discriminant block and dynamic adjustment factor into the network.

Batch Nuclear-Norm Maximization (BNM) (Cui et al., 2020) could effectively avoid prediction degradation in the situation where models are learned with insufficient labels. BNM is a method to maximize the nuclear-norm of the batch output matrix. This simultaneous optimization strategy serves to enhance both the discriminability and diversity of the predicted outcomes.

DeepCORAL (Sun and Saenko, 2016) utilizes the second-order covariance of features from both the source and target domains to quantify the distribution discrepancy. By combining the loss function and the CORAL distance within the classification model, a domain adaptive model based on the CORAL distance can be constructed.

Deep Subdomain Adaptation Network (DSAN) (Zhu et al., 2020) overcomes the constraint of aligning global distributions by leveraging the inter-subdomain relationships across distinct domains. It employs the network outputs as pseudo-labels for the target domain data.

4.3. Experimental settings

Baseline Methods: For Office-31 dataset, we compared enhanced transfer learning models with standard deep neural network model and normal transfer learning models: deep convolutional neural network (He et al., 2016), DAN (Long et al., 2015), DAAN (Yu et al., 2019), DANN (Ganin and Lempitsky, 2015), DeepCORAL (Sun and Saenko, 2016), BNM (Cui et al., 2020), DSAN (Zhu et al., 2020). For other three datasets, all of the compared methods except standard deep neural network are used. The results of standard deep neural network are collected from Zhu et al. (2020), and other results are all obtained from our experiments.

Implementation Details: On all the tasks, a mini-batch stochastic gradient descent (SGD) approach is employed, and the batch size is set to 32. For the pattern recognition tasks, we choose to employ the Residual network (He et al., 2016). Following the methodology of CDAN (Long et al., 2018), we incorporated a bottleneck layer denoted as fcb with 256 units after the final average pooling layer to facilitate transfer representation learning. The outputs of the bottleneck layer fcb serves as the inputs for the discrepancy function between the source domain data and the target domain. For the parameter settings, see Table 3.

4.4. Results on public datasets

We conducted our enhanced transfer learning approach on four datasets, and denote all the transfer learning tasks as source domain → target domain.

The outcomes of unsupervised domain adaptation on *Office-31* dataset are presented in Table 4. It can be seen that enhanced transfer learning models achieve better performance except BNM model with DA. In **A** → **W**, **A** → **D**, **D** → **A** transfer learning tasks, the performance improvement is obvious. It might be because that in these tasks, the source domain exhibits a higher degree of similarity with the target domains. Enhanced transfer learning methods harness this similarity through data augmentation and domain adaptation, leading to improved performance. Comparing enhanced transfer learning and traditional convolutional neural network, the promotion is significant. The enhanced transfer learning performs comparable with the traditional transfer learning methods on **D** → **W** and **W** → **A** cases. There might be several reasons for the observed results. In these tasks, the data distribution disparities between source and target domains may be smaller, allowing traditional transfer learning methods to effectively capture existing similarities. Another possible explanation is that excessive data augmentation in some cases can lead to overfitting, hindering the model's ability to generalize to the target domain.

The outcomes of unsupervised domain adaptation for the *Office-Caltech10* dataset are displayed in Table 5. It is obvious that the average

Table 7
The accuracy (%) on the Office-home dataset through unsupervised domain adaptation (ResNet50), the version "+DA" indicates the enhanced transfer learning method with data augmentation operations.

Method	A→C	C→A	A→P	P→A	A→R	R→A	C→P	P→C	C→R	R→C	P→R	R→P	Avg
DAN	51.0	57.5	67.9	55.9	74.9	**65.6**	65.8	47.5	67.7	54.4	75.2	79.6	63.5
DAN+DA	**51.4**	**58.3**	**68.5**	**57.0**	**75.8**	68.4	**65.7**	**48.1**	**67.7**	**54.5**	**75.5**	**80.1**	**64.5**
DAAN	48.2	52.9	64.4	54.2	**73.8**	66.2	61.6	43.2	64.3	52.0	72.9	78.3	60.9
DAAN+DA	**48.6**	**55.0**	**65.7**	**54.5**	73.3	**66.2**	**62.0**	**43.3**	**65.3**	**52.0**	**73.4**	**79.7**	**61.6**
DANN	49.6	54.3	65.0	55.1	73.4	**67.4**	63.2	48.7	65.1	55.1	74.1	79.0	62.5
DANN+DA	**50.3**	**55.0**	**66.3**	**55.7**	**75.1**	66.8	**64.2**	**49.8**	**66.2**	**56.3**	**74.6**	**80.5**	**64.3**
BNM	52.1	60.0	69.5	58.3	75.5	67.5	**70.1**	49.0	71.3	54.4	77.7	80.3	65.4
BNM+DA	**52.3**	**61.7**	**72.1**	**60.4**	**77.0**	**70.1**	67.9	**51.2**	**71.3**	**55.6**	**78.7**	**80.6**	**66.6**
DeeepCoral	51.0	56.7	66.3	55.5	74.5	68.2	64.0	47.2	67.1	54.4	**75.0**	**79.6**	63.2
DeeepCoral+DA	**51.2**	**57.3**	**66.3**	**56.8**	**74.5**	**68.4**	**64.8**	**50.1**	**67.4**	**54.7**	74.7	78.7	**67.0**
DSAN	52.6	60.1	**70.0**	61.8	74.5	71.7	**67.5**	54.7	68.0	59.1	77.4	82.3	66.6
DSAN+DA	**53.2**	**60.4**	69.7	**62.8**	**74.7**	**72.2**	67.2	**55.0**	**68.4**	**59.5**	**78.1**	**82.3**	**67.0**

accuracy obtained by the enhanced transfer learning models surpass those of the normal transfer learning models. All enhanced transfer learning models achieve better accuracy than normal transfer learning model, especially on **C → A**, **C → D** and **C → W** tasks.

The outcomes of unsupervised domain adaptation for the *ImageCLEF* dataset are presented in Table 6. We can find that all the average accuracy in enhanced transfer learning models are higher than normal transfer learning models.

The outcomes of unsupervised domain adaptation for the *Office-Home* dataset are displayed in Table 7. It is evident that the enhanced transfer learning models consistently exhibit superior performance. Notably, DeepCORAL combined with data augmentation leads to substantial improvements in average accuracy, with gains exceeding 3%.

4.5. Results on medical dataset

We combine transfer learning models with several different data augmentation strategies. The source domain for this study is the normal pneumonia dataset, while the target domain consists of the COVID-19 dataset. In the method section, we introduced several data augmentation strategies. In current section, we proceed to compare three distinct data augmentation strategies. The details of our data augmentations are listed as follows:

DA1 is based on geometric space transformation, which mainly includes three aspects: random rotation, center cropping and horizontal flipping. Herein the rotation angle range is set from −15° to +15°, the cropping size is set to 0.9 times the original size, and the horizontal flipping is enforced with a probability 0.5.

DA2 adopts color space transformation, which mainly includes three aspects: brightness, contrast, and saturation. For these three parameters, the value range is set to 0.5 to 2. For any original image, a real value is randomly selected from this range as the color space transformation parameter.

DA3 uses both color space transformation and geometric space transformation. It combines DA1 and DA2. However, the random rotation angle range in geometric space transformation is set from −5° to 5°. In addition, an additional noise injection operation is added.

Table 8 presents the comparison of different data augmentation strategies. We come to the following conclusions: the enhanced transfer learning could effectively improve the model's performance on pneumonia X-ray data set.

The choice of data augmentation strategy will greatly affect the final performance of the models. Enhanced transfer learning with DA1 performs exceptionally well in PPV, likely due to two primary factors. DA1 emphasizes geometric space transformations, such as random rotation, center cropping, and horizontal flipping, improving the model's understanding of image spatial characteristics. In medical imaging, where pathology orientation can vary, random rotation enables the model to analyze pathology from different angles, enhancing its ability to accurately detect and predict positive cases. Furthermore, DA1's practice of center cropping images at 90% of their original size effectively reduces noise and unnecessary information along the image edges. This noise reduction sharpens the model's focus on relevant regions, a critical factor for accurate predictions, particularly in medical imaging. By conducting experiments on different data augmentation strategies, we proved that different data augmentation strategies improve the performance of our enhanced transfer learning models. Moreover, it is a feasible data augmentation strategy to inject noise to image data. By injecting noise, the generalization ability of the models have been improved. Although the proposed method demonstrated effectiveness and superiority over compared methods, more exploration can be conducted, such as statistical analysis (Wang and He, 2016; Demšar, 2006) can be added to further enhance the credibility of the results.

Table 8
The accuracy (%) on the pneumonia X-ray dataset (ResNet50), the version "+DA" indicates the enhanced transfer learning method with data augmentation operations.

Method	ACC	PPV	TPR	TNR	F1−score
DAN	88.29 ± 0.52	84.20	94.66	81.76	89.12
DAN+DA1	90.56 ± 1.48	87.46	**95.00**	86.01	91.07
DAN+DA2	90.39 ± 0.87	87.30	94.82	85.84	90.91
DAN+DA3	**91.79 ± 1.83**	**89.45**	**95.00**	**88.50**	**92.14**
DANN	87.77 ± 0.87	83.33	94.82	80.53	88.71
DANN+DA1	89.96 ± 0.87	**87.68**	93.28	**86.55**	90.39
DANN+DA2	89.43 ± 0.78	86.48	93.79	84.96	90.00
DANN+DA3	**90.12**±1.00	86.52	**95.34**	84.78	**90.71**
DAAN	89.17 ± 0.52	88.78	90.00	88.31	89.38
DAAN+DA1	**90.66 ± 0.69**	**91.13**	90.34	**90.97**	**90.73**
DAAN+DA2	90.04 ± 0.52	89.63	**90.86**	89.20	90.24
DAAN+DA3	89.87 ± 1.20	90.85	88.97	90.80	89.90
DeepCoral	88.29 ± 0.52	84.52	**94.13**	82.30	89.07
DeepCoral+DA1	90.31 ± 0.83	**89.28**	91.89	**88.67**	90.57
DeepCoral+DA2	**90.39 ± 0.74**	88.78	92.76	87.96	**90.73**
DeepCoral+DA3	90.13 ± 1.48	88.09	93.10	87.08	90.53
DSAN	91.53 ± 0.70	94.14	88.79	94.33	91.39
DSAN+DA1	**92.49 ± 0.34**	**97.68**	87.24	**97.88**	92.17
DSAN+DA2	92.31 ± 0.61	96.59	87.93	96.81	92.06
DSAN+DA3	**92.49 ± 0.34**	95.57	**89.31**	95.75	**92.34**

5. Conclusion

This paper treat both transfer learning and data augmentation as the solutions to limited target domain data problems, and then designed the enhanced transfer learning methods with data augmentation to improve the image classification performance. The experimental results obtained on several real-world datasets and a medical image dataset verified the effectiveness of the proposed method.

Although deep learning has achieved a great success (Liu et al., 2015), the training of deep learning models requires a large amount of high-quality data and powerful computational resources. The collection and annotation of data need to invest a lot of human and resource costs (Wang and He, 2016), the enhanced transfer learning with data augmentation approach proposed in this paper, on the one hand, expands the target domain dataset by constructing an effective data augmentation strategy so as to improve the generalizability of the model; on the other hand, by transferring the knowledge of similar domains to the target domain of interest, it can effectively solve the scarce data problem, which save great time and resource costs. Thus the proposed method is helpful to the digital economy.

It is noted that some ongoing challenges need further study, particularly the adaptive data augmentation strategy. Current methods rely on fixed rules and transformations, but the best approach can vary across different contexts. Further research is necessary to automate the selection and fine-tuning of augmentation strategies for optimal performance. Moreover, for future work, few-shot learning (Wang et al., 2020; Sun et al., 2019), semi-supervised learning (Chao and Sun, 2012; Sun and Shawe-Taylor, 2010; Chao and Sun, 2018), and active learning (Beluch et al., 2018) are all potential solutions to the limited data problem, and how to ensemble some of them to further promote the performance is worth further research. While this work focuses on image classification task, more tasks should be explored to get a deeper insight into the limited target domain data problem.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China 62276079 and the Research and Innovation Fund of Harbin Institute of Technology IDGAZMZ00210325 and Key Research, and Development Plan of Shandong Province 2021SFGC0104.

References

Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M., 2018. The power of ensembles for active learning in image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9368–9377.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W., 2010. A theory of learning from different domains. Mach. Learn. 79 (1), 151–175.

Ben-David, S., Blitzer, J., Crammer, K., Pereira, F., 2006. Analysis of representations for domain adaptation. Adv. Neural Inf. Process. Syst. 19.

Chao, G., Liu, J., Wang, M., Chu, D., 2023a. Data augmentation for sentiment classification with semantic preservation and diversity. Knowl.-Based Syst. 111038.

Chao, G., Sun, S., 2012. Semi-supervised multitask learning via self-training and maximum entropy discrimination. In: International Conference on Neural Information Processing. Springer, pp. 340–347.

Chao, G., Sun, S., 2018. Semi-supervised multi-view maximum entropy discrimination with expectation Laplacian regularization. Inf. Fusion 45, 296–306.

Chao, G., Zhu, X., Ding, W., Bi, J., Sun, S., 2023b. Special issue on transfer learning. Neural Processing Letters 1–4.

Cui, S., Wang, S., Zhuo, J., Li, L., Huang, Q., Tian, Q., 2020. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3941–3950.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine learning research 7, 1–30.

Farahani, A., Voghoei, S., Rasheed, K., Arabnia, H.R., 2021. A brief review of domain adaptation. Adv. Data Sci. Inf. Eng. 877–894.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning. PMLR, pp. 1180–1189.

Gong, B., Shi, Y., Sha, F., Grauman, K., 2012. Geodesic flow kernel for unsupervised domain adaptation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2066–2073.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A., 2012. A kernel two-sample test. J. Mach. Learn. Res. 13 (1), 723–773.

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

Kang, Y., Chao, G., Hu, X., Tu, Z., Chu, D., 2022. Deep learning for fine-grained image recognition: a comprehensive study. In: Proceedings of the 2022 4th Asia Pacific Information Technology Conference. pp. 31–39.

Liu, J.N., Hu, Y., He, Y., Chan, P.W., Lai, L., 2015. Deep neural network modeling for big data weather forecasting. Information granularity, big data, and computational intelligence 389–408.

Long, M., Cao, Y., Wang, J., Jordan, M., 2015. Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning. PMLR, pp. 97–105.

Long, M., Cao, Z., Wang, J., Jordan, M.I., 2018. Conditional adversarial domain adaptation. Adv. Neural Inf. Process. Syst. 31.

Long, M., Wang, J., Ding, G., Sun, J., Yu, P.S., 2013. Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2200–2207.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28 (5), 823–870.

Ouyang, L., Key, A., 2021. Maximum mean discrepancy for generalization in the presence of distribution and missingness shift. arXiv preprint arXiv:2111.10344.

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

Patel, V.M., Gopalan, R., Li, R., Chellappa, R., 2015. Visual domain adaptation: A survey of recent advances. IEEE Signal Process. Mag. 32 (3), 53–69.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 29 (9), 2352–2449.

Saenko, K., Kulis, B., Fritz, M., Darrell, T., 2010. Adapting visual category models to new domains. In: European Conference on Computer Vision. Springer, pp. 213–226.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.

Sun, Q., Liu, Y., Chua, T.-S., Schiele, B., 2019. Meta-transfer learning for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 403–412.

Sun, B., Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation. In: European Conference on Computer Vision. Springer, pp. 443–450.

Sun, S., Shawe-Taylor, J., 2010. Sparse semi-supervised learning using conjugate functions. J. Mach. Learn. Res. 11, 2423–2455.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., Van Der Laak, J., 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. 58, 101544.

Van Dyk, D.A., Meng, X.-L., 2001. The art of data augmentation. J. Comput. Graph. Statist. 10 (1), 1–50.

Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S., 2017. Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5018–5027.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. Neurocomputing 312, 135–153.

Wang, X., He, Y., 2016. Learning from uncertainty for big data: future analytical challenges and strategies. IEEE Systems, Man, and Cybernetics Magazine 2 (2), 26–31.

Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M., 2020. Generalizing from a few examples: A survey on few-shot learning. ACM Comput. Surv. (csur) 53 (3), 1–34.

Yu, C., Wang, J., Chen, Y., Huang, M., 2019. Transfer learning with dynamic adversarial adaptation network. In: 2019 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 778–786.

Zhu, Y., Zhuang, F., Wang, D., 2019. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 5989–5996.

Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., He, Q., 2020. Deep subdomain adaptation network for image classification. IEEE Trans. Neural Netw. Learn. Syst. 32 (4), 1713–1722.