Journal of Computers Vol. 33 No. 3, June 2022, pp. 171-185
doi: 10.53106/199115992022063303014

Image Domain Generalization Method based on Solving Domain Discrepancy Phenomenon

Zhi Tan*, Zhao-Fei Teng

School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, 102616, Beijing, China
tanzhi@bucea.edu.cn, Tengzf20162020@163.com

Received 21 September 2021; Revised 13 December 2021; Accepted 13 January 2022

Abstract: In order to solve the problem that the recognition performance is obviously degraded when the model trained by known data distribution transfer to unknown data distribution, domain generalization method based on attention mechanism and adversarial training is proposed. Firstly, a multi-level attention mechanism module is designed to capture the underlying abstract information features of the image; Secondly, increases the loss limit of the generative adversarial network，the virtual enhanced domain which can simulate the target domain of unknown data distribution is generated by adversarial training on the premise of ensuring the consistency of data features and semantics; Finally, through the data mixing algorithm, the source domain and virtual enhanced domain are mixed and input into the model to improve the performance of the classifier. The experiment is carried out on five classic digit recognition and CIFAR-10 series datasets. The experimental results show that the model can learn better decision boundary, generate virtual enhanced domain and significantly improve the accuracy of recognition after model transplantation. Comparing to the previous method, our method improves average accuracy by at least 2.5% and 3% respectively. Experiments on five classic digit recognition and CIFAR-10 series datasets which significantly improves the classification average accuracy after model transfer.

Keywords: attention mechanism, generative adversarial network, domain generalization, image recognition

1 Introduction

In the domain of transfer learning, datasets are usually divided into source domain and target domain. The data and labels of the source domain are easy to obtain, but the data labels of the target domain are expensive or seriously missing. Due to the lack of data, only the datasets with known data distribution can be directly used to train the network model. However, when the training model is transplanted to the datasets with unknown data distribution, the robustness of the model often decreases obviously. The reason for this problem is called domain discrepancy [1]. Therefore, how to solve the problem of domain discrepancy has attracted extensive attention in various fields of deep computer vision domain generalization.

In response to the above problems, Qiao et al. [2] considered the worst-case expression of data distribution near the source domain in the feature space, data enhancement by extending virtual domains. Although the model has achieved excellent performance on the benchmark datasets, there are some defects in the process of expanding the virtual enhanced domain, such as low data semantic consistency, poor simulation target domain distribution, and insufficient image key feature information extraction. Therefore, aiming at the shortcomings of the current model, this paper proposes an improved algorithm model. The experimental results show that the improved model can obtain higher average recognition accuracy, effectively solve the problem of domain discrepancy, and improve the robustness of model transfer. Multi-level parallel attention (MLPA) model is used to solve the shortage of feature extraction. The data mixing algorithm and the loss limit of the generative adversarial network are used to improve the semantic consistency stability of image sample level and feature level, simulate more unknown distributions and improve the performance of the classifier. The final experimental results show that the test accuracy of the improved model in the target domain is significantly higher than that of the original model. The specific work of this paper is as follows:

Therefore, in view of the shortcomings of the current model, we consider the limitation of semantic consistency in the representation of image sample level and feature level at the same time and propose an attention gen-

* Corresponding Author

erative adversarial domain generalization model (AGADG) to simulate more unknown distributions and realize greater domain transportation. The main work of this paper is to make a series of improvements to the single domain generalization method proposed by Qiao et al. The final experimental results show that the test accuracy of the improved model in the target domain is significantly improved compared with the original model. The specific work of this paper is as follows:

(1) A multi-level parallel attention mechanism module (MLPA) is proposed, which can fully capture the detailed features of the image, fuse the multi-level features and establish the relationship between the feature information.

(2) A data mixing algorithm is proposed, which combines the generated virtual enhanced domain data with the source domain data, which can realize the expansion of data distribution and improve the recognition performance of the classifier.

(3) Combining adversarial training and generative adversarial loss is used to enhance the semantic consistency of sample level and feature level, improve the complexity of generating virtual enhanced domain.

(4) Combine the above parts to get the attention generative adversarial domain generalization (AGADG) model.

2 Related Work

In the domain of unsupervised deep learning, in order to improve the portability of neural networks and solve the distribution offset between the source domain and the target domain, there has been a lot of work domain adaptation [2-7] was studied. Among them, Chen et al. [3] proposed an unsupervised domain adaptation method based on a stepwise feature alignment network by allowing a large amount of labeled source domain data and a large amount of unlabeled data target domain for large-scale training. Cai et al. [4] proposed the maximum square loss, which solved the problem of uneven probability distribution caused by unsupervised domain adaptation to minimize entropy through the method of linear growth gradient. Tsai et al. [6] verified that the output of the segmented image can adapt to the scene distribution and semantic information of the source and target domain images and proposed a multi-layer pixel-level semantic segmentation method based on adversarial learning.

At the same time, the idea of adversarial learning [8-13] has also been widely applied to the research work of domain generalization. Vu et al. [8] used the idea of adversarial learning to propose a semantic segmentation domain adaptation method based on entropy loss. Among them, Sinha et al. [9] designed a principled adversarial training algorithm, which first generated some new images that maximize the risk, and the model parameters were optimized for those adversarial images. In order to resist the imperceptible adversarial disturbance, the loss of the new image is absorbed to punish the original and the new difference. Peng et al. [10] proposed that a powerful generalization model of pose estimation can be obtained by combining adversarial learning methods with traditional data enhanced training. Volpi et al. [11] considered the worst-case expression of the data distribution close to the source domain in the feature space. They also proposed an iterative process of domain adaptation against data expansion, which uses samples from the virtual target domain to expand the datasets, has good performance when transplanted to image recognition and semantic segmentation tasks.

At present, meta-learning [12-15] has begun to be studied by more and more scholars, and at the same time, meta-learning is applied to computer vision tasks. The essence of meta-learning is to use a small amount of data in multiple learning tasks to achieve the fastest solution to new tasks and improve the generalization performance of the model. Guo et al. [12] proposed a method of facial recognition based on meta-learning in to learn a facial recognition model that can directly recognize without model update. Finn et al. [13] proposed the MAML method to find an optimal initialization weight more quickly so that the model can adapt to new tasks more quickly. This model has been widely used in small sample learning and reinforcement learning. To evaluate the quality of various damaged images and better adapt to the unknown degree of damage, Zhu et al. [14] proposed a method of non-reference image quality evaluation based on deep meta-learning, which can be obtained by fine-tuning the previously trained model parameters. High-quality evaluation model. Researchers not only pay attention to the application of domain adaptation and domain generalization in image recognition [16-17], but also improve the performance of computer vision tasks such as object detection [18-20] and semantic segmentation [21], solve the problem of domain discrepancy. In this context, Qiao et al. designed a method of meta learning against the enhanced domain. Experimental results show that this method can improve the performance of image recognition and semantic segmentation.

3 Algorithm Design

Chen et al. [3] proposed a stepwise feature alignment network model that can solve the unsupervised domain adaptive classification problem by allowing large-scale training of a large number of labeled source domain data and a large number of unlabeled target domain data. Sinha et al. [9] defined the worst case outside the source domain data distribution through the idea of adversarial training. In order to realize the domain generalization outside the source domain data distribution and solve this worst-case problem, it is necessary to train in advance, simulate the distribution of data outside the source domain through the data enhanced method, and generalize the robustness model to the unknown target domain. The specific form of the worst problem in the semantic space is shown in Equation 1:

$$\min_{\theta} \sup_{P_s} \left\{ E_{P_s} \left[L_C(\theta;(X,Y)) \right], D_\theta \left(p_s, p_{adv} \right) \le \rho \right\}. \tag{1}$$

Where P_s and P_{adv} respectively represent the data distribution of the source domain and the unknown domain outside the source domain, θ represents the parameters of the model, L_c represents the objective function of the model, D_θ represents the distance measurement of the two probability distributions P_s and P_{adv}, and ρ represents the distance of the domain offset. At the computational level, ρ is very difficult to determine the deep neural networks. therefore, considering the difference of data distribution, after Volpi et al. [11] reconstruction, this problem can be effectively solved by using the form of Lagrange relaxation. Therefore, the worst problem can be transformed into Equation 2:

$$\min_{\theta} \sup_{P_s} \left\{ E_{P_s} [L_C(\theta;(X,Y))] - \lambda D_\theta \left(P_s, P_{adv} \right) \right\}. \tag{2}$$

Where λ is a hyper-parameter. In order to further solve the worst problem and improve the generalization performance of the model, an improved domain generalization model is constructed by combining multi-level parallel attention mechanism, generative adversarial network and adversarial training. The structure diagram of AGADG model is shown in Fig. 1. The model is composed of feature extraction network, discriminator network and generative adversarial network. In addition, because the image can only extract local features after convolution operation, and the attention mechanism plate can obtain detailed edge features and give different weights to different information, the MLPA model is fused in the feature extraction network and the generative adversarial network, and finally give weight association to different levels of feature information, so as to improve the robustness of the model. The multi-level parallel attention mechanism model will be introduced in Section 3.1, and the weight dependence of different information will be established; In Section 3.2, a data mixing algorithm is designed to realize data divergence, improve the performance of classifier and enhance the robustness of model; In sections 3.3, adversarial data enhanced is introduced respectively. Realize adversarial training by combining task models and generative adversarial networks, more unknown data distributions related to the source domain are generated, and the worst problem is optimized to achieve maximum domain transportation.

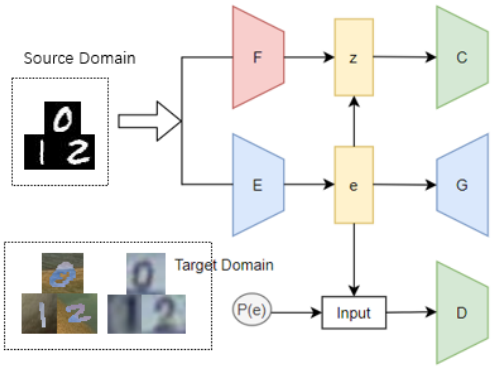


Fig. 1. The overall structure of AGADG model

3.1 Multi-Level Parallel Attention

In order to make full use of the multi-level feature information of image features, obtain the high correlation information between pixels and significant parts, and establish the dependence of global features, MLPA model is designed. MLPA model can dynamically give weights to different levels of features, so as to obtain relatively large weights for boundary key feature information, reduce the weight parameters of irrelevant features, and finally output a more accurate prediction feature map. The specific structure of MLPA model is shown in Fig. 2.

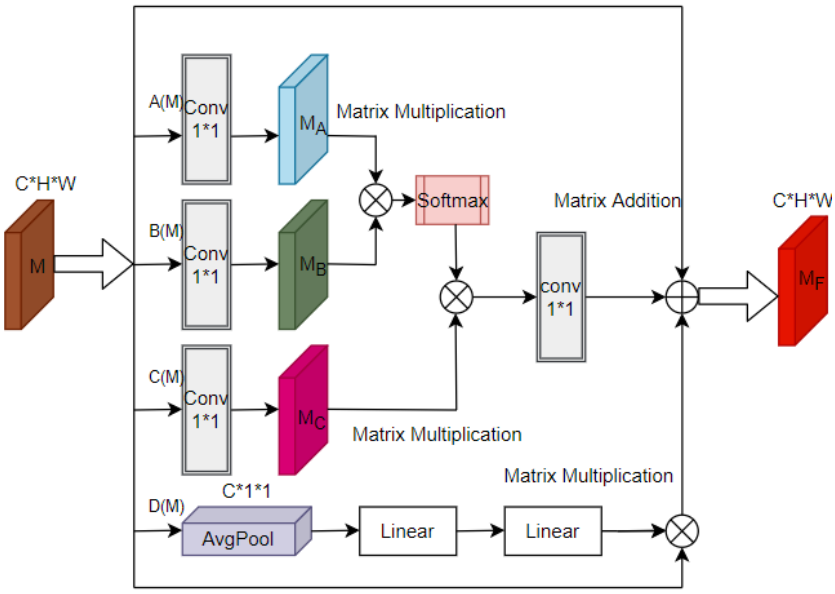


Fig 2. Structure diagram of multi-layer parallel attention model

The MLPA model is composed of multiple branches. The input of the MLPA model is the characteristic diagram output through convolution operation of convolution layer. At the same time, the input feature map is processed in multi branch parallel, and finally the feature vectors of different branches are fused to obtain the output prediction feature map. Suppose the characteristic diagram of the input MLPA model is $M \in R^{C/2*H*W}$, where C, H and W represent the number, height and width of channels of the input characteristic diagram M respectively.

Firstly, in the three branches $A(M)$, $B(M)$ and $C(M)$, the 1 * 1 convolution operation is performed on M, and their channel dimension C is reduced to the original 1/2. The dimensions of the new characteristic graph are $M_A \in R^{C/2*H*W}$, $M_B \in R^{C/2*H*W}$ and $M_C \in R^{C/2*H*W}$ respectively. Secondly, reshape the feature maps M_A and M_B to obtain $M_{A1} \in R^{H*W*C/2}$ and $M_{B1} \in R^{H*W*C/2}$, then transpose M_{A1} and perform matrix multiplication with M_{B1} to obtain the final pixel correlation feature map M_{AB}, where the expression of M_{AB} is shown in Equation 3:

$$M_{AB} = \frac{\exp\left(M_{A1i}^{T} * M_{B1i}\right)}{\sum_{i=1}^{N} \exp\left(M_{A1i}^{T} * M_{B1i}\right)}. \tag{3}$$

Where $i \in [1, N]$, $M_{AB} \in R^{H*W*H*W}$. M_{AB} is subjected to Softmax normalization processing and $M_C \in R^{C/2*H*W}$ is subjected to matrix multiplication operation, and the normalization operation is performed again, and finally the feature map $M_{ABC} \in R^{C/2*H*W}$. The specific calculation of M_{ABC} is shown in Equation 4.

$$M_{ABC} = \frac{\exp\left(M_{Ci}^{T} * M_{ABi}\right)}{\sum_{i=1}^{N} \exp\left(M_{Ci}^{T} * M_{ABi}\right)}. \tag{4}$$

The characteristic diagram M_{ABC} carry out 1 * 1 convolution operation and restore to the original number of channels C. So far, the output characteristic diagram $M_{F1} \in R^{C*H*W}$ of three branches $A(M)$, $B(M)$ and $C(M)$ is obtained. In branch $D(M)$, firstly, the global pooling operation is performed on the input characteristic graph M, and the characteristic graph m is compressed into $M_D \in R^{C*1*1}$. then, through the dimensionality reduction operation of the full connection layer, the number of channels C is reduced to $C/16$, and the ReLU activation function is used for nonlinear processing. In order to ensure that the size of the output feature map is equal to that of the input feature map, it is up sampled through a full connection layer to restore the number of channels of the output feature map to the original size C. Therefore, the dimension of the final feature map of the branch is $M_{F2} \in R^{C*H*W}$.

Through the above operations, the characteristic diagrams of each branch $D(M)$ are weighted and fused. Therefore, the form of the predicted characteristic diagram output by the MLPA model is shown in Equation 5:

$$M_F = \alpha_F M_{F1} + \beta_F M_{F2} + \gamma_F M. \tag{5}$$

Among them, α_F 、 β_F and γ_F are the weight coefficients of features, and the weight size is gradually updated in the continuous learning of the model, so that the MLPA model can be associated with more feature information by establishing the dependence of weight.

3.2 Data Mixing Algorithm

When the data distribution is divergent, it is conducive to expand the data distribution to the field outside the source domain and realize the maximum domain transmission. A data mixing algorithm is designed. Each batch of data input into the network is processed by fusing the source domain and the generated virtual enhanced domain, and finally the consolidated domain data obtained from different domains are fused to realize the divergence of image input. The specific implementation is shown in Algorithm 1.

Algorithm 1. The process of data mixing

Input: Source domain $S \in \{x_i, y_i\}$, $i \in [1, N]$, Virtual enhanced domain S_{advk}, BatchSize bs

Output: Consolidated domain S_{cbk}

Initialize: $S_{cbk} \leftarrow S$, $C_s \leftarrow bs$

for k in K **do:**

$C_s = bs / (k + 1)$

$C_{advk} = bs - C_s$

End for

Return S_{cbk}

The domain exactly the same as the source domain S is constructed as the initial consolidated domain S_{cbk}. Therefore, before generating the virtual enhanced domain S_{advk}, the data X_{cbk} sampled by the consolidated domain S_{cbk} is the data X_s of the source domain S. Suppose that the source domain dataset has several batches,

where bs represents the BatchSize, K represents the number of current virtual enhanced domains, and the initial value is 0. The composition of each batch of S_{cbk} data X_{bs} is shown in Equation 6:

$$X_{bs} = \left\{ X_{s1}, \ldots, X_{sj} \right\}. \tag{6}$$

Among $j \in [1, bs]$, as S_{advk} is generated iteratively, the data of the source domain and the virtual enhanced domain need to be fused to form a consolidated domain S_{cbk}. At this point, the sampling X_{cbk} of each batch of the consolidated domain will come from S and S_{advk} generated by the current iteration. Therefore, after iteratively generating S_{advk}, the composition of each batch of data X_{bs} in S_{cbk} will become as shown in Equation 7:

$$X_{bs} = \left\{ X_{s1}, \ldots, X_{sj}, X_{advk1}, \ldots, X_{advkm} \right\}. \tag{7}$$

Where $m \in [1, bs]$, X_{sj} and X_{advkm} represent each batch data of S_{cbk} samples j images from the S and m images from the current S_{advk}. When iteratively generating the virtual enhanced domain, the data distribution of the source domain will not quickly spread outside the source domain, so it is necessary to input more source domain data to give certain restrictions. With the increase of the number of virtual enhanced domains, the data distribution of the enhanced domain can gradually simulate the data distribution of more unknown domains, and the dependence on the data of the source domain will be reduced. Let k be the number of virtual enhanced domains generated. Therefore, the expressions for selecting the number of data in each batch constituting S_{cbk} from S and S_{advk} are shown in Equations 8 and 9 respectively:

$$C_s = bs / (k+1). \tag{8}$$

$$C_{advk} = bs - C_s. \tag{9}$$

3.3 Adversarial Data Augmentation

The task model T of AGADG model consists of feature extraction network F and classifier $C(\bullet)$. After the class label information is known in the source domain data, the convolution layer combines with the MLPA module to extract the features, and then the supervised learning can be realized directly through $C(\bullet)$. Through category determination, the extracted features are more sensitive to category information, which is conducive to the establishment of decision boundary. The objective function is shown in Equation 10:

$$L_T(x, y) = -E_{(x_s, y_s) \sim (X_s, Y_s)} \sum_{i=1}^{n} \mathrm{y_s} \log C\left(F\left(x_s\right)\right). \tag{10}$$

Where X_s represents the input image of the source domain, Y_s represents the category label of the source domain data, and F represents the feature extraction network model of image recognition. After the MLPA model is incorporated into the convolution calculation, the task model T used in the digital recognition datasets is shown in Fig. 3.

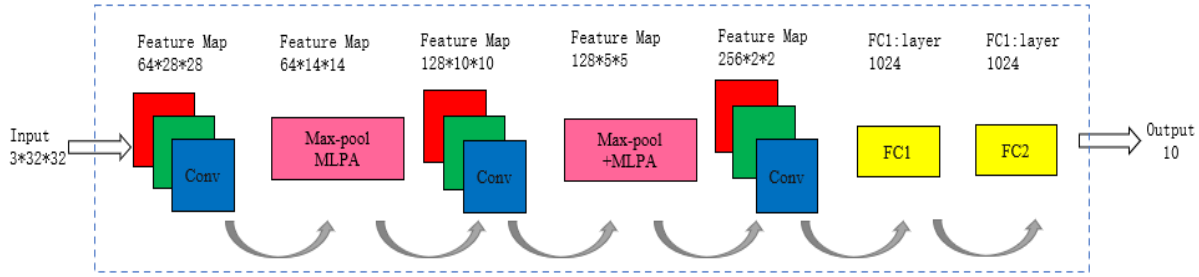


Fig. 3. Structure of feature extraction and image recognition model F

Because the process of domain generalization does not need to access the data of the target domain, and the number of data in the source domain is limited, the method of data enhanced in the source domain must be adopted to enhance the distribution of data. The model combined with the generative adversarial network and adversarial training can reconstruct and learn the image more efficiently, so as to obtain more unknown data distribution outside the source domain and simulate more data distribution different from the source domain. Semantic consistency is achieved at both feature level and sample level to improve the transplantation performance of the model. The generative adversarial network is defined as W, and its parameter is θ_w. W is composed of self-encoder and a discriminator $D(\bullet)$. the self-encoder is composed of encoder $E(e|x)$ and decoder $G(x|e)$, where x represents the input image and e represents the embedded feature. After the encoder convolution operation, the weight dependence is constructed through the MLPA module, and the relationship between features is established from the underlying texture features of the image, so that the network can learn to produce more features of decision boundary, which is very important to the prediction effect of features. X_{advk} is generated in the virtual enhanced domain S_{advk} through iteration. Although the method of relaxing the data distribution of the source domain and enhanced domain designed by Qiao et al. can expand the domain transmission, it has an adverse impact on the semantic consistency of the virtual enhanced domain. Therefore, based on the work of Qiao et al., the restriction conditions are improved by combining the generation of generative adversarial network and adversarial training. At the same time, constraints are carried out at the feature level and sample level to maximize the domain transfer from the source domain to the virtual enhanced domain.

The network model based on convolution operation alone can not establish the dependence between key information, resulting in the generation of fuzzy depth sample map, resulting in the lack of edge detail feature information. Taking the source domain image X_s as the input, the MLPA model is integrated into the convolutional self-encoder for generative adversarial network, so that the image edge detail information can be fully utilized, which can overcome this problem.

Firstly, the generative adversarial network W is pre-trained, and the maximum reconstruction loss of the generated virtual enhanced domain is fixed. In the process of model training, first complete the feature mapping of the source domain data X_s, and then realize the category mapping of the source domain. Then, S_{advk} is generated step by step in an iterative way, the feature mapping of the virtual enhanced domain is carried out, and the domain training working model is combined to judge whether the feature distribution of the generated virtual enhanced domain is outside the source domain distribution, and the parameters of the model are constantly updated, so as to achieve the purpose of maximizing domain transportation and domain generalization.

The newly generated S_{advk} and S are processed by data mixing algorithm. Combined with the feature extraction network containing MLPA and the generative adversarial network for iterative adversarial training, classifier $C(\bullet)$ is easier to expand the distribution of source domain S through data. Generate more virtual enhanced samples through iterative process, and X_{advk} simulates the distribution of target domain. Input the source domain data into the generative adversarial network to obtain the target domain. The generation loss function about the generator is shown in Equation 11:

$$L_G = \min_{(E,G)} \sum \left\| x_s - G\left(E\left(x_s\right)\right)\right\|^2. \tag{11}$$

The classification loss function about the discriminator is shown in Equation 12:

$$L_D = \max_D E_{x_s \sim X_s} \log\left(D\left(E\left(x_s\right)\right)\right) + E_{x_{advk} \sim X_{advk}} \log\left(1 - D\left(G\left(x_{advk}\right)\right)\right). \tag{12}$$

Wassersein distance is used as the distance measure of generative data loss, and whether S_{advk} has the same data distribution as S is judged. At the same time, it maximizes the reconstruction error, ensures that the generative image and the original image can spread to the data distribution outside the source domain at the sample level, and realizes the maximum domain transmission. The data relationship between the calculation input and generation using Wasserstein distance as measurement is shown in Equation 13:

$$L_r = \min_W \left\| x_{advk} - W\left(x_{advk}\right)\right\|^2. \tag{13}$$

Although the above generative loss can realize domain transfer at the sample level, it has an impact on semantic consistency, resulting in misclassification of generative samples. The loss of semantic consistency at the design feature level can effectively solve this problem.

Adversarial training makes the generator and discriminator sensitive to disturbance information. The classifier $C(\bullet)$ after model feature extraction is regarded as a discriminator in the virtual enhanced domain. Through the iterative process, the two discriminators can ensure the consistency of sample semantics. In the training process, the features of the virtual enhanced domain are sent to $C(\bullet)$ for classification and discriminator $D(\bullet)$ to judge whether it is true or false. $C(\bullet)$ determines whether its data distribution conforms to the category of the current data distribution, in other words, whether the virtual enhanced domain has been extended to the unknown distribution outside the source domain. At the same time, the function of $D(\bullet)$ is the same as that of the ordinary generative adversarial network to judge the authenticity of the image. Through iterative adversarial training, adversarial training can effectively ensure semantic consistency. Wasserstein distance is used as the distribution distance between virtual enhanced domain feature z_{advk} and consolidated domain data feature z_{cbk}. The specific form is shown in Equation 14:

$$L_{con} = \frac{1}{2}\left\| \mathbf{z}_{cbk} - \mathbf{z}_{advk}\right\|_2^2. \tag{14}$$

Therefore, combined with the above-mentioned parts, the source domain S is used to train the model, and then the virtual enhanced domain S_{adv} is iteratively generated to gradually optimize the model parameters. The adversarial training iteration process is actually to maximize the generation of data that can simulate the distribution of the target domain, as shown in Equation 15:

$$X_{advk} \in \underset{x \in X}{\operatorname{argmax}} \left\{ L_T\left(\theta; x, y\right) + \alpha L_D(\theta_w; x) + \beta L_r\left(\theta; x\right) - \gamma L_{con}\left(\theta_w; z\right)\right\}. \tag{15}$$

Where α, β and γ represent three penalty term parameters respectively, and the values are 1, 1 and 2000 respectively according to the experiment. The implementation process of AGADG model is shown in detail in Algorithm 2.

Algorithm 2. The implementation process of AGADG model

Input: Source domain $S \in \{x_i, y_i\}$, $i \in [1, N]$, Number of synthetic domains K

Output: Learned model parameters θ

Initialize: $S_{cbk} \leftarrow S$, $C_s \leftarrow bS$

for *time* in *iteration* **do:**

Generate S_{advk}

Append S_{advk} to S_{adv}

According to the formula (7)-(9) to get S_{cbk}

for i in T_{adv} **do**

Samples X_{cbk} from S_{cbk}

According to the formula (11)(12) to train W

According to the formula (15) to train T

for j in k **do**

Update θ using $\theta \leftarrow \theta - \alpha \nabla_\theta L_T\left(\theta; X_{S \cup S_{adv}}, Y_{S \cup S_{adv}}\right)$

End for

Return θ

4 Experiment and Result Analysis

In order to verify the effectiveness of the model, the test performance of AGADG model is compared with the previous research work on domain generalization. In order to compare the experimental results obviously, two kinds of image recognition datasets are used for experiments. One group is five classical digital datasets MNIST [22], MNIST-M [23], SVHN [24], SYN [23] and USPS [25]. The number of iterations of the experiment is 10^5, the BatchSize is 32, and Adam optimizer is used at the same time, and the learning rate is 0.0001. Another group of experiments used CIFAR-10 [26] and CIFAR-10-C [27] datasets, with 10^5 iterations, BatchSize is 128, SGD optimizer which learning rate is 0.001 and linear decay learning rate is 0.1. First, use the source domain data to pre-train the generative adversarial network. After the pre-training is completed, input the source domain data into the task model and the generative adversarial network to obtain the characteristics of the source domain data and the generated virtual enhanced domain data, which are obtained through the data mixing algorithm combine domains and adversarial training task models and generative models to obtain data with unknown target domain data distribution, and finally obtain a model that can realize cross-domain generalization. The two groups of experimental results show that the portability of AGADG model is greatly improved compared with the previous models, which verifies the feasibility of AGADG model. Section
4.1 introduces the datasets used in the experiment, the evaluation indicators of the experimental results and the display of the experimental results. Section 4.2 analyzes the experimental results and summarizes the performance of the model.

4.1 Experimental Results

In order to eliminate the unforeseen errors between the datasets used in the experiment, first ensure that the number of pixels and channels of the image in the source domain and the test target domain are consistent, avoid the contingency of the experiment, preprocess the MNIST datasets of single channel participating in the training, turn the datasets images into RGB three channel images with 32 * 32 pixels, and use four sets of SVHN, MNIST-M, SYN and USPS as the target domain to test the cross-domain generalization and robustness of the model. As an improved algorithm model, AGADG model has a great correlation with previous work, so it directly uses the result data of the original work to compare with the experimental results. The average recognition accuracy of the model in the four sets of target domain data sets is used as an index to evaluate the cross-domain generalization performance of the model. The experimental results show that the recognition accuracy of the proposed AGADG model on the four digital data sets has been improved, and the average recognition accuracy has increased by

2.5%, compared with the previous work. The specific experimental results of four digit recognition test sets are shown in Table 1.

Table 1. Comparison of experimental results on digit recognition datasets

Method	SVHN	MNIST-M	SYN	USPS	Avg.
d-SNE	0.2622	0.5098	0.3783	0.9316	0.5205
ERM	0.2783	0.5272	0.3695	0.7694	0.4929
GUD	0.3551	0.6041	0.4532	0.7726	0.5462
MADA	0.4255	0.6794	0.4895	0.7853	0.5949
AGADG	0.4457	0.7006	0.5073	0.8290	0.6206

CIFAR-10-C is a new datasets formed in the original CIFAR-10 datasets, which first ensures that the images are RGB three channels and the pixels are 32 * 32, and processes all pictures in different categories and levels without affecting the categories. In order to be consistent with the experimental process of the previous work, CIFAR-10 is also used as the training set training model, twelve different types of image processing in four main forms of Noise, Blur, Weather and Digital are selected as the test set, and the accuracy of each test set is recorded. Finally, the average accuracy of twelve groups of test results is selected as the discrimination index of the experimental results. The specific experimental results of CIFAR-10-C dataset recognition are shown in Table 2.

Table 2. Comparison of experimental results on CIFAR-10-C datasets

	d-SNE	ERM	GUD	MADA	AGADG
Fog	0.6599	0.6592	0.6829	0.6936	0.6930
Snow	0.7546	0.7436	0.7675	0.8059	0.7590
Frost	0.6225	0.6157	0.6994	0.7666	0.7432
Zoom	0.5847	0.5997	0.6295	0.6804	0.6854
Defocus	0.5371	0.5371	0.5641	0.6118	0.6568
Glass	0.5048	0.4944	0.5345	0.6159	0.6214
Speckle	0.4530	0.4131	0.3845	0.6088	0.7645
Shot	0.3993	0.3541	0.3687	0.6058	0.7621
Impulse	0.2795	0.2565	0.2226	0.4518	0.6104
Jpeg	0.7020	0.6990	0.7422	0.7714	0.7849
Pixelate	0.3846	0.4107	0.5334	0.5225	0.5589
Spatter	0.7340	0.7536	0.8027	0.8062	0.8055
Avg.	0.5696	0.5615	0.5826	0.6559	0.6871

4.2 Result Analysis

During the model training, in order to ensure that the trained model is more convincing in the experimental comparison, five experiments are carried out on the number recognition experiment, and the accuracy of different test sets is recorded in the five experiments. At the same time, the results of each test set are displayed. The specific experimental results of four groups of digital datasets for five times are shown in Fig. 4. The trained AGADG model is transplanted to different test sets. The comparison between the model and the previous model on the four sets of digit recognition datasets is shown in Fig. 5(a). The evaluation of the average recognition accuracy of these models in the digit recognition experiment is shown in Fig. 5(b). It can be clearly seen that compared with the previous models, the AGAGG model has a significant improvement in the cross-domain recognition generalization performance of the digit recognition experiment. The recognition accuracy can reach 62.06%, which is at least 2.5% higher than previous models

As shown in Fig. 4, five experiments show in terms of model stability, the result curve of five tests of AGADG model on SVHN datasets fluctuates, while the prediction results on MNIST-M, SYN and USPS datasets basically tend to be stable. The stability of the model for SVHN dataset needs to be improved. In terms of accuracy distribution, the model performs best on the USPS datasets, which can reach 82%, which is the most improved compared with the other three groups of test sets. This is because the data distribution of USPS is similar to MNIST, the difference between USPS and MNIST domain is small, and it is easier to be simulated.

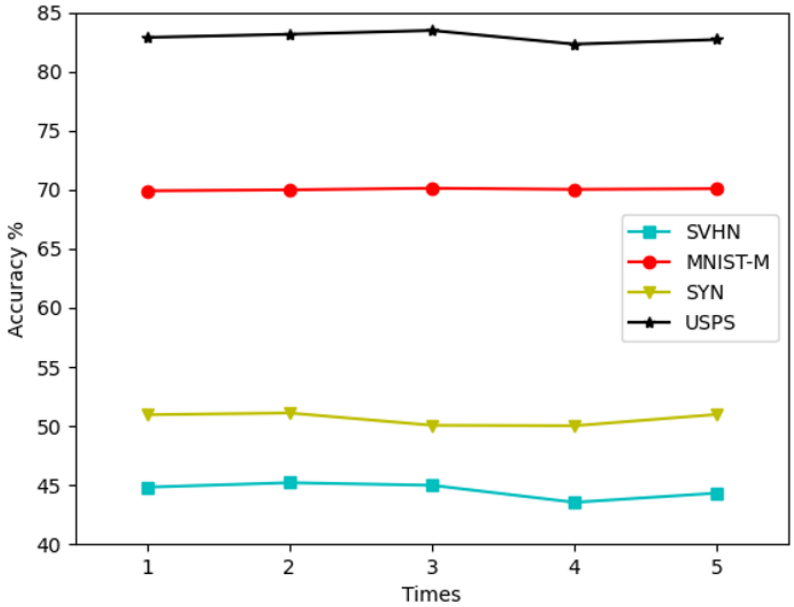
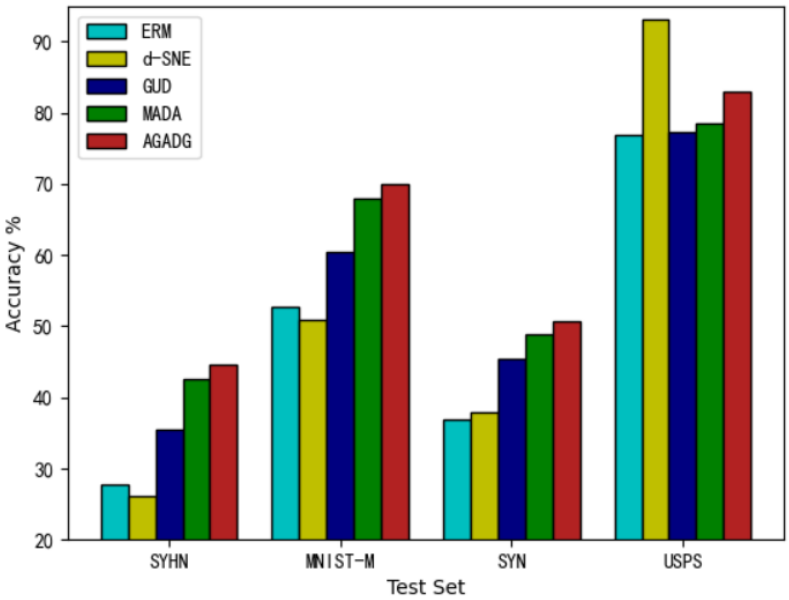
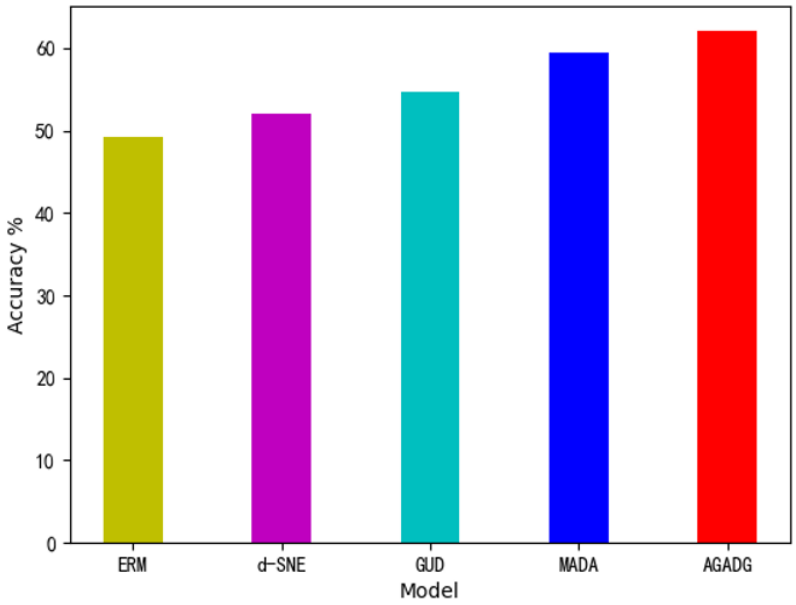


Fig 4. Results of multiple experiments on the digital recognition datasets

As shown in Fig. 5, comparing the test results of four groups of digital datasets through AGADG model with the prediction results of previous d-SNE [28], ERM [29], GUD [13] and MADA [2], it can be seen that the accuracy of ADAGD model is higher than the experimental results of previous models. At the same time, the test set SVHN, the accuracy improvement ratio of MNIST-M and SYN is significantly higher than that of USPS datasets. Compared with MADA, although AGADG model improves the accuracy of USPS datasets, it is still lower than d- SNE. This is because d- SNE only makes a lot of improvement on the recognition accuracy of USPS datasets. d- SNE performs poorly in the transplantation performance of the other three groups of datasets, and the accuracy of AGADG model is much higher than that of d- SNE model in these three groups of data results.



(a) Comparison of accurary of different models on four digit datasets



(b) Comparison of average accurary of different models on digit recognition experiment

Fig. 5. Comparison of average accuracy of different models

When training the model on CIFAR-10 datasets, the feature extraction part uses a 16 layer wide residual network [30] with a width of 4, add the MLPA mechanism model to the first layer of each residual block. Selects 12 different processing categories in CIFAR-10-C as the test set for testing, lists 4 types of data samples as shown in Fig. 6.

The comparison results between the test results of each test set and the previous model are shown in Fig. 7(a). As shown in Fig. 7(a) histogram, the experimental results of AGADG model in Speckle, Shot and Impulse test sets are more than 15% higher than those of MADA model, and the recognition accuracy of Defocus, Jpeg and Pixelate test sets is also significantly improved. However, the performance on fog, snow and frost is poor, which is slightly lower than the test results of MADA model. It can be clearly seen from the histogram in Fig. 7(b). the average recognition accuracy of this model in the CIFAR-10 recognition experiment is at least 3% higher than that of previous models. It can be seen that AGADG model is more robust than previous models for cross-domain generalization.

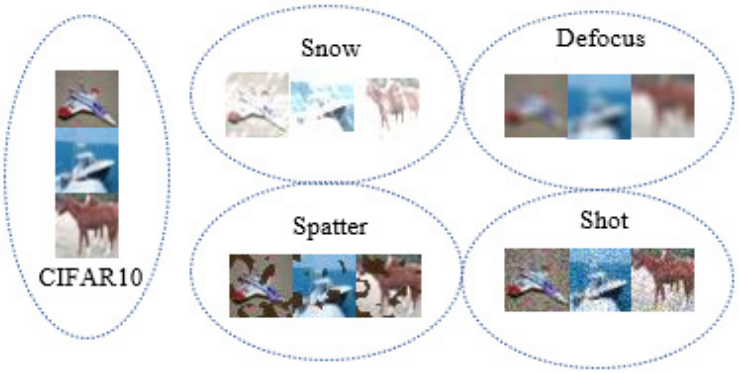
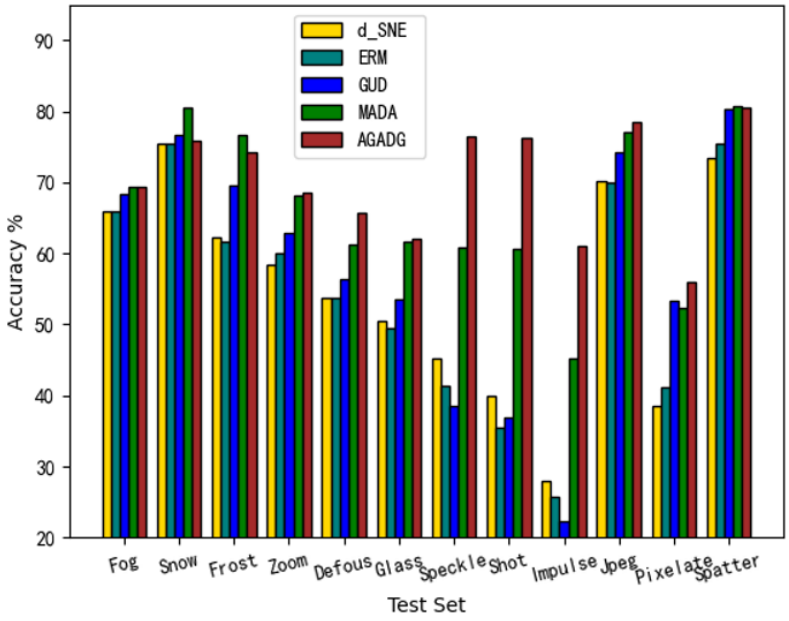
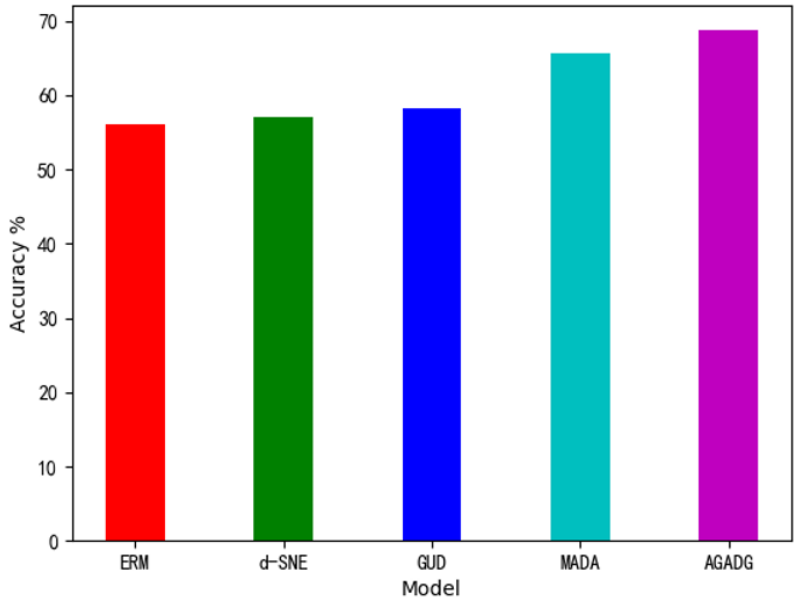


Fig. 6. On the left is the original image of CIFAR10, on the right are four corrosion types of CIFAR-10-C samples



(a) Comparison of accurary of different models on CIFAR-10-C datasets



(b) Comparison of average accurary of different models on CIFAR-10 recognition experiment

Fig. 7. Comparison of accuracy of different models

Conclusion

In this paper, a method of attention generative adversarial domain generalization is proposed. This method uses the multi-level attention mechanism model to fully extract the detailed features of the image, and then increases the loss limit of the generative adversarial network on the basis of ensuring the semantic consistency of the sample level and the feature level to generate the virtual enhanced domain, and finally combines the data mixing algorithm to obtain the merge The domain is used for adversarial training. Use the source domain to train the model and transfer the model to multiple target domains. The two sets of experimental results show that the AGADG model can perform higher average recognition accuracy than previous models, and effectively solve the problem of domain shift.

Future work needs further research: exploring more effective limiting methods for expanding the source do-

main distribution, optimizing the stability of the model, in order to improve the generalization effect of the model.

References

[1] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus, Intriguing properties of neural networks, in: Proc. 2014 International Conference on Learning Representations, 2014.
[2] F. Qiao, L. Zhao, X. Peng., Learning to Learn Single Domain Generalization, in: Proc. 2020 IEEE Conference on Computer Vision and Pattern Recognition, 2020.
[3] C.-Q. Chen, W.-P. Xie, W.-B. Huang, Y. Rong, X.-H. Ding, Y. Huang, T.-Y Xu, J.-Z. Huang, Progressive feature alignment for unsupervised domain adaptation, in: Proc. 2019 IEEE International Conference on Computer Vision and Pattern Recognition, 2019.
[4] M.-H. Chen, H.-Y. Xue, D. Cai, Domain adaptation for semantic segmentation with maximum squares loss, in: Proc. 2019 IEEE International Conference on Computer Vision, 2019.
[5] K. Park, S. Woo, D. Kim, D. Cho, I. Kweon, Preserving semantic and temporal consistency for unpaired video-to-video translation, In: Proc. 2019 ACM International Conference on Multimedia, 2019.
[6] Y. Tsai, W. Hung, S. Schulter, K. Sohn, M. Yang, M. Chandraker, Learning to adapt structured output space for semantic segmentation, in: Proc. 2018 IEEE International Conference on Computer Vision and Pattern Recognition, 2018.
[7] P. Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Mathematical Programming 171(1)(2018) 115-166.
[8] T. Vu, H. Jain, M. Bucher, M. Cord, P. Pérez, Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proc. 2019 IEEE International Conference on Computer Vision and Pattern Recognition, 2019.
[9] A. Sinha, H. Namkoong, J. Duchi, Certifying some distributional robustness with principled adversarial training, in: Proc. 2018 International Conference on Learning Representations, 2018.
[10]X. Peng, Z.-Q. Tang, F. Yang, R. Feris, D. Metaxas, Jointly Optimize Data Augmentation and Network Training: Adversarial Data Augmentation in Human Pose Estimation, in: Proc. 2018 IEEE International Conference on Computer Vision and Pattern Recognition, 2018.
[11]R. Volpi, H. Namkoong, O. Sener, J.C. Duchi, V. Murino, S. Savarese, Generalizing to unknown domains via adversarial data augmentation, in: Proc. 2018 Conference and Workshop on Neural Information Processing Systems, 2018.
[12]J.-Z. Guo, X.-Y. Zhu, Z. Lei, S.Z. Li, Decomposed Meta Batch Normalization for Fast Domain Adaptation in Face Recognition, in: Proc. 2020 IEEE International Conference on Computer Vision and Pattern Recognition, 2020.
[13]C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proc. 2017 International Conference on Machine Learning, 2017.
[14]H.-C. Zhu, L. Li, J.-J. Wu, W.-S. Dong, G.-M. Shi, MetaIQA: Deep Meta-learning for No-Reference Image Quality Assessment, in: Proc. 2020 IEEE International Conference on Computer Vision and Pattern Recognition, 2020.
[15]J.-Y. Wang, X.-T. Zhu, S.-G. Gong, W. Li, Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-Identification, in: Proc. 2018 IEEE International Conference on Computer Vision and Pattern Recognition, 2018.
[16]A. Madani, M. Moradi, A. Karargyris, T. Mahmood, Semi-Supervised Learning with Generative Adversarial Networks for Chest X-Ray Classification with Ability of Data Domain Adaptation, in: Proc. 2018 IEEE 15th International Symposium on Biomedical Imaging, 2018.
[17]M. Choi, J. Choi, S. Baik, T.H. Kim, K.M. Lee, Scene-Adaptive Video Frame Interpolation via Meta-Learning, in: Proc. 2020 IEEE International Conference on Computer Vision and Pattern Recognition, 2020.
[18]T. Wang, X. Zhang, L. Yuan, J. Feng, Few-shot Adaptive Faster R-CNN, in: Proc. 2019 IEEE International Conference on Computer Vision and Pattern Recognition, 2019.
[19]N. Inoue, R. Furuta, T. Yamasaki, K. Aizawa, Cross-Domain Weakly-Supervised Object Detection Through Progressive Domain Adaptation, in: Proc. 2018 IEEE International Conference on Computer Vision and Pattern Recognition, 2018.
[20]X.-D. Wang, Z.-W. Cai, D.-S. Gao, N. Vasconcelos, Towards Universal Object Detection by Domain Attention, in: Proc. 2019 IEEE International Conference on Computer Vision and Pattern Recognition, 2019.
[21]G.-L. Kang, Y.-C. Wei, Y. Yang, Y.-T. Zhuang, Alexander G. Hauptmann, Pixel-Level Cycle Association: A New Perspective for Domain Adaptive Semantic Segmentation, in: Proc. 2020 Advances in Neural Information Processing Systems, 2020.
[22]Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, in: Proc. 1998 IEEE, 86(11)(1998) 2278-2324.
[23]Y. Ganin, V. Lempitsky, Unsupervised Domain Adaptation by Backpropagation, in: Proc. 2015 International Conference on Machine Learning, 2015.
[24]Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A. Ng, Reading digits in natural images with unsupervised feature learning, in: Proc. 2011 Conference and Workshop on Neural Information Processing Systems, 2011.
[25]J.S. Denker, W.R. Gardner, H.P. Graf, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, H.S. Baird, I. Guyon, Neural network recognizer for hand-written zip code digits, in: Proc. 1989 Conference and Workshop on Neural Information Processing Systems, 1989.

[26]A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images, Master's thesis, Department of Computer Science, University of Toronto, 2009.
[27]H. Dan, D. Thomas, Benchmarking neural network robustness to common corruptions and perturbations, in: Proc. 2019 International Conference on Learning Representations, 2019.
[28]X. Xu, X. Zhou, R. Venkatesan, G. Swaminathan, O. Majumder, D-sne: Domain adaptation using stochastic neighborhood embedding, in: Proc. 2019 IEEE International Conference on Computer Vision and Pattern Recognition, 2019.
[29]V. Koltchinskii, Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008, volume 2033.Springer Science & Business Media, 2011.
[30]S. Zagoruyko, N. Komodakis, Wide residual networks, in: Proc. 2016 BMVC, 2016.