Mitigating Both Covariate and Conditional Shift for Domain Generalization

Jianxin Lin\textsuperscript{1,2}, Yongqiang Tang\textsuperscript{1}, Junping Wang\textsuperscript{1,2}, and Wensheng Zhang\textsuperscript{1,2}

\textsuperscript{1}Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
\textsuperscript{2}School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, China
linjianxin2020@ia.ac.cn, yongqiang.tang@ia.ac.cn, junping.wang@ia.ac.cn, zhangwenshengia@hotmail.com

Abstract: Domain generalization (DG) aims to learn a model on several source domains, hoping that the model can generalize well to unseen target domains. The distribution shift between domains contains the covariate shift and conditional shift, both of which the model must be able to handle for better generalizability. In this paper, a novel DG method is proposed to deal with the distribution shift via Visual Alignment and Uncertainty-guided belief Ensemble (VAUE). Specifically, for the covariate shift, a visual alignment module is designed to align the distribution of image style to a common empirical Gaussian distribution so that the covariate shift can be eliminated in the visual space. For the conditional shift, we adopt an uncertainty-guided belief ensemble strategy based on subjective logic and Dempster-Shafer theory. The conditional distribution given a test sample is estimated by the dynamic combination of that of source domains. Comprehensive experiments are conducted to demonstrate the superior performance of the proposed method on four widely used datasets, i.e., Office-Home, VLCS, TerraIncognita, and PACS.

Keywords: Domain generalization; Domain adaptation; Image style normalization; Uncertainty-based ensemble

1 Introduction

Computer vision has made great progress in recent years with the help of deep learning under the basic assumption that all data are independently and identically distributed. However, in practical applications, images collected by different devices in different environments often follow different distributions. In out of distribution scenarios, existing deep learning models suffer from the distribution shift and fail to generalize well [1].

To tackle the distribution shift problem, great efforts have been made in Domain Adaptation (DA), which generally aims to transfer knowledge from a labeled source domain to an unlabeled target domain so that the learned model can perform well on the target domain [1]. However, DA requires the target domain to be accessible which is hard to meet when the target domain changes dynamically. Furthermore, DA needs to retrain the model when applying it to another target domain, and it is time-consuming. In recent years, Domain Generalization (DG) has attracted much attention, which tries to enable the model to generalize to unseen domains utilizing multiple source domains. Without access to target domains, DG improves the generalizability of models in out of distribution scenarios and has broad application prospects.

Among existing works, domain-invariant representation learning is a classic approach to DG. Let $X$ denote the input variable, i.e., an image, and $Y$ denote the output variable, i.e., a predicted label. As analyzed in previous works, traditional models often suffer from covariate shift [2], i.e., $P_i(X) \neq P_j(X)$, and conditional shift [3], i.e., $P_i(Y|X) \neq P_j(Y|X)$, in out of distribution scenarios. Some works [4], [5] try to learn a representation space $Z = F(X)$ where the marginal distribution $P(Z)$ keeps the same across source domains so that the covariate shift can be eliminated assuming that $P(Y|X)$ keeps stable. Another line of research [6] tries to align the class-conditioned distribution $P(Z|Y)$ in the representation space for a fine-grained distribution matching assuming that $P(Y)$ keeps stable. These works often hold impractical assumptions and fail to eliminate both covariate and conditional shift. Additionally, though distributions of source domains are aligned, there is no guarantee that the distributions of unseen domains would be also aligned with that of source domains.

In this paper, we propose a new approach to eliminate both the covariate shift and conditional shift. For the covariate shift, we propose to align $P(X)$ in visual space. In most scenarios, the domain shift is mainly caused by the image style which can be represented as feature statistics [7]. We attempt to model the common real-world style distribution, i.e., the real-world distribution of feature statistics, which has yielded domain-specific style distributions with different selection biases. After that, we normalize all the domain-specific style distribution to the common style distribution so that the covariate shift is eliminated via visual alignment. For the conditional shift, instead of aligning $P(Y|X)$...
across source domains, we design a nonlinear ensemble scheme based on uncertainty modeling to dynamically approximate \( P(Y^n | X = x^i) \) given a test sample \( x^i \). The subject logic and Dempster-Shafer theory of evidence are first introduced to solving DG in this paper. Comprehensive experiments have conducted on four widely used datasets to demonstrate the effectiveness of VAUE.

2 Related Work

2.1 Style Randomization

In many scenarios, domain shift existing in image datasets are mainly caused by changes in image style. For example, a model trained on images of paintings often fails to perform well on images of photo. Intuitively, many works attempt to make the models robust to difference of image style across domains by randomizing the style of training images so that the correlation between image style and label can be eliminated to some extent. Previous work [7] in image style transfer changes the style of an image to that of another image by exchanging the statistics of intermediate features of convolution networks. The feature statistics, namely the mean and variance of each channel, have encoded the style information of an image to some extent.

Inspired by this, some works [8], [9] try to randomize the style by random interpolation of feature statistics. However, vanilla randomization of feature statistics is too tricky and hard to be related to the elimination of covariate shift. In this paper, we try to normalize the distribution of style by visual alignment modules, which normalize the distribution of feature statistics to a common Gaussian distribution parameterized by statistics of a batch of samples instead of plain randomization.

2.2 Evidential deep learning

Evidential Deep Learning is proposed to quantify classification uncertainty using the theory of subjective logic [10]. Evidential deep learning constructs evidence collector with deep neural network, whose output is unnormalized and nonnegative. Unlike the usual classifier which provides a deterministic posterior distribution, evidential deep learning provides a Dirichlet distribution defined on a simplex, which is parameterized with the collected evidence. Each posterior distribution sampled from the simplex is attached with a certain probability density, which models the uncertainty of the prediction.

Han et al. [11] apply evidential deep learning to model the uncertainty of prediction from each view, and then the uncertainty-based ensemble using Dempster-Shafer theory is designed to produce robust and trustworthy predictions. From the perspective of posterior distribution modeling, we adopt evidential deep learning and the Dempster-Shafer theory to model the posterior distribution given a test sample by dynamically integrating the posterior distributions modeled by source domain-specific classifiers to eliminate the distribution shift between training data and test data.

3 Proposed Method

In this section, we first introduce the notations used in this paper. Then the proposed VAUE is detailed.

3.1 Notations

Let the input sample and output label spaces be denoted as \( X \) and \( Y \) respectively. A domain is a set of data sampled from a joint distribution, which can be denoted as \( D = \{ (x_i, y_i) \}_{i=1}^N \sim P(X, Y) \), where \( x \in X \subset \mathbb{R}^d, y \in Y \subset \mathbb{R} \), and \( P(X, Y) \) denotes the joint distribution of the sample and label. \( X, Y \) are the corresponding random variables. \( C = |Y| \) denotes the number of classes. Given \( N \) source domains \( \{ D_i \}_{i=1}^N \) which follow different distributions, DG aims to learn a model which can generalize well on unseen target domains with unknown distribution shifts. In this paper, vectors are shown in bold, and the subscript indicates the corresponding dimension of the vector, e.g., \( v \) and \( v_i \).

3.2 Covariate Shift

**Assumption 1 (Independence Assumption).** Let \( f_{sem}(X) \) and \( f_{sty}(X) \) be the semantic and style component which are extracted from \( X \), which are independent with each other so that \( P(fsem(X), fstyx(X)) = P(fsem(X)) \times P(fstyx(X)) \). \( P(fsem(X)) \) keeps stable, while \( P(fstyx(X)) \) changes on various domains.

Under Assumption 1, we can align \( \{ P^i(X) \}_{i=1}^N \) of source domains by normalizing \( \{ P^i(fstyx(X)) \}_{i=1}^N \) to the same target distribution. According to previous work [7], we know that feature statistics of the features in intermediate layers of deep networks, i.e., means and standard deviations of each individual feature channels, have encoded the style information of images. Given an image, the intermediate feature after a certain layer of a network is denoted as \( z \in \mathbb{R}^{KH} \). The vector of feature statistics \( \mu \in \mathbb{R}^K \) and \( \sigma \in \mathbb{R}^K \) can be calculated as follows:

\[
\mu_k(z) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} z_{khw} \\
\sigma_k(z) = \sqrt{\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (z_{khw} - \mu_k(z))^2}
\]

As analyzed in Assumption 1, we can approximately represent \( f_{sty}(X) \) with \( [\mu, \sigma] \). Hence, we can normalize the distribution of \( [\mu, \sigma] \) of images to align \( P(X) \) in visual space. Assuming that all domain-specific feature statistics are sampled from the real-world style distribution with various selection biases, given a batch of images, we aggregate samples from all domains to approximate the real-world style distribution with an empirical distribution of feature statistics \( P^{emp}(f_{sty}(X)) \), which is formalized as a Gaussian distribution, and parameters are calculated as follows:
Proceedings of CCIS2022

\[ m_\mu = \frac{1}{B} \sum_{b=1}^{B} \mu(z^b) \]
\[ m_\sigma = \frac{1}{N} \sum_{b=1}^{B} \sigma(z^b) \]
\[ \Sigma_\mu = \frac{1}{B-1} \sum_{b=1}^{B} (\mu(z^b) - m_\mu)(\mu(z^b) - m_\mu)^T \]
\[ \Sigma_\sigma = \frac{1}{B-1} \sum_{b=1}^{B} (\sigma(z^b) - m_\sigma)(\sigma(z^b) - m_\sigma)^T \]

Then we can get that:

\[ p_{emp}(\mu, \sigma) = g(\mu | m_\mu, \Sigma_\mu) \times g(\sigma | m_\sigma, \Sigma_\sigma) \]  

(4)

We normalize \( \{p_i(f_{sty}(X))\}_{i=1}^{N} \) to \( p_{emp}(f_{sty}(X)) \) for the marginal distribution alignment. For each image, we replace the original feature statistics as the new values sampled from the empirical distribution:

\[ \mu_s, \sigma_s \sim p_{emp}(f_{sty}(X)) \]
\[ z = (z - \mu(z))/\sigma(z) + \mu_s \]

(5)

Where / and \( \odot \) denote channel-wise division and multiplication respectively. By normalizing the distribution of feature statistics to the empirical distribution \( p_{emp}(f_{sty}(X)) \), which is a reasonable substitute for the real-world style distribution, the visual style distributions \( P(f_{sty}(X)) \) of source domains are aligned. The operation of visual alignment is a plug-and-play module, which can be flexibly inserted into the neural networks at different positions like Batch Normalization layers. Each visual alignment module is randomly enabled according to a probability factor \( \tau \) to control the ratio of samples to be transformed.

3.3 Conditional Shift

To overcome the conditional shift, we design an uncertainty-guided belief ensemble strategy to dynamically approximate \( P(Y|X) \) given a test sample based on the Subjective Logic (SL) and Dempster-Shafer theory of evidence. For more details, we refer readers to published works [10], [11], [12], [13].

3.3.1 Subjective Logic

We construct a linear domain-specific classifier for each source domain following a shared feature extractor, which is expected to fit the domain-specific posterior distribution. Instead of the softmax operation, we choose \( \exp(\cdot) \) to make outputs of linear classifiers nonnegative. The unnormalized non-negative outputs of each domain-specific classifier, \( \{e_i^1\}_{i=1}^{N}, e_i^1 \in \mathbb{R}^c \), can be regarded as the collected evidences in favor of a sample to be classified into a certain category.

Following SL, we define the belief masses, \( \{b_i^1\}_{i=1}^{N}, b_i^1 \in \mathbb{R}^c \), and uncertainty masses, \( \{u_i\}_{i=1}^{N}, u_i \in \mathbb{R} \), for \( i \)-th domain-specific classifier as follows:

\[ b_i^1 = \frac{e_i^1}{S_i}, u_i^1 = \frac{C_i}{S_i}, \text{ and } u_i + \sum_{c=1}^{C} b_i^c = 1 \]

(6)

where \( S_i = \sum_{c=1}^{C} (e_i^c + 1) \). Under this definition, the uncertainty mass is inversely proportional to the total evidence.

The output of a classical deep learning classifier is a probability assignment over all classes. SL parameterizes a Dirichlet distribution with the evidence, which is a probability density function (pdf) for all possible probability assignments over all classes. In other words, Dirichlet distribution is a pdf defined on a \( C \)-dimensional unit simplex, \( S_c = \{ p | \sum_{c=1}^{C} p_c = 1 \text{ and } \forall c, p_c \geq 0 \} \), where every point is a \( C \)-dimensional probability assignments. Specifically, the parameters of Dirichlet distribution for \( i \)-th domain-specific classifier is defined as \( a_i^1 = e_i^1 + 1 \), and then the \( i \)-th Dirichlet distribution can be denoted as:

\[ Dir(p|\alpha) = \begin{cases} \frac{1}{\Gamma(a)} \prod_{c=1}^{C} p_c^{a_c-1} & \text{for } p \in S_c, \\ 0 & \text{otherwise} \end{cases} \]

where \( B(\cdot) \) is a \( C \)-dimensional multinomial beta function. Under above definitions, given a sample, all domain-specific classifiers will output their evidences collected from the sample, \( \{e_i^c\}_{i=1}^{N} \). And then belief masses \( \{b_i^c\}_{i=1}^{N} \) for each category and uncertainty masses \( \{u_i\}_{i=1}^{N} \) can be derived. What is more, Dirichlet distributions of domain-specific classifiers \( \{Dir(p|\alpha)\}_{i=1}^{N} \) will be formalized with derived evidences.

3.3.2 Reduced Dempster’s Combinational Rule

We adopt a reduced Dempster’s combinational rule [11] to nonlinearly combine the predictions of all domain-specific classifiers.

Definition 1 (Reduced Dempster’s Combinational Rule). Given two sets of masses \( M^1 = \{b^1, u^1\} \) and \( M^2 = \{b^2, u^2\} \), the combination \( M = \{b, u\} \) can be calculated as follows:

\[ b_c = \frac{1}{1-F}(b_1^1 b_2^2 + b_1^2 b_2^1 + b_2^1 u_2^1 + b_2^2 u_1^1), u = \frac{1}{1-F} u_1^1 u_2^2, \]

(8)

where \( F = \sum_{i=1}^{2} b_i^1 b_i^2 \) reflecting the conflict between two mass sets. The combination can be denoted as \( M = M^1 \ominus M^2 \).

All mass sets given by domain-specific classifiers can be combined as \( M = M^1 \ominus M^2 \ominus \cdots \ominus M^N \). By doing so, we can formalize the overall Dirichlet distribution based on the combined belief masses and uncertainty mass \( b, u \). Specifically, parameters of the combined Dirichlet distribution can be derived as follows:

\[ s = \frac{c}{u}, e = b \times s \text{ and } \alpha = e + 1. \]

3.3.3 Single-Domain and Cross-Domain Training

After illustrating the definitions of evidence, belief, uncertainty and the combination rule, we now specifically
show the detailed training process. Given an input sample, the $i$-th domain-specific classifiers will output a mass set $\mathcal{M}^i$ and a Dirichlet distribution $\text{Dir}(\mathbf{p}|\alpha^i)$. Let $\mathbf{x}$ denote an input sample, and $\mathbf{y}$ denote the corresponding one-hot label. To train this domain-specific classifier with the input sample $\mathbf{x}$, we adopt following loss functions [10]:

$$L_{\text{ec}}(\mathbf{x}, \mathbf{y}, \alpha^i) = \mathbb{E}_{\text{Dir}(\mathbf{p}|\alpha^i)} \left[ \sum_{c=1}^{C} y_c \log \left( p_c \right) \right] + \lambda \text{KL}[\text{Dir}(\mathbf{p} | \alpha^i) \parallel \text{Dir}(\mathbf{p} | \mathbf{1})]$$

(10)

where $\alpha^i = \mathbf{y} + (1 - \mathbf{y}) \odot \alpha^i$, $\mathbf{1}$ is a vector with all elements equal to 1, and $\psi(\cdot)$ is the digamma function. In this paper, $\lambda$ is set to 0.01. The first term is an expectation of cross entropy computed over the Dirichlet distribution. And the second term is proposed to enforce the evidence for incorrect labels to shrink to 0 [10].

To model the uncertainty well, we design a single-domain training part and a cross-domain training part. For the single-domain part, the data of $i$-th domain are only fed into $i$-th domain-specific classifier and compute the loss $L_{\text{ec}}$. For the cross-domain part, the data of $i$-th domain are fed into all domain-specific classifiers except the $i$-th domain. After that, $N - 1$ mass sets are combined. And the combined Dirichlet distribution is derived, parameters of which are denoted as $\alpha^i$. Hence the loss function can be designed as follows:

$$L_D = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{|\mathcal{D}_i|} \sum_{k=1}^{\left|\mathcal{D}_i\right|} \left[ L_{\text{ec}}(\mathbf{x}^k, \mathbf{y}^k, \alpha^i) \right] + L_{\text{ec}}(\mathbf{x}^i, \mathbf{y}^i, \alpha^i)$$

(11)

Furthermore, we enforce the correlation across different dimensions of the extracted feature before classifiers to shrink to 0. We found that this design can prevent the numerical computation problem during training. Specifically, for a batch of features $\mathbf{z} \in \mathbb{R}^{B \times d}$, the mean of features $\mathbf{m} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{z}_b$, then the decorrelation loss can be designed as:

$$L_{\text{decor}} = \left| \frac{1}{B - 1} (\mathbf{z} - \mathbf{m})^T (\mathbf{z} - \mathbf{m}) \odot (\mathbf{1} - \mathbf{I}) \right|_1$$

(12)

where $\mathbf{I}$ is an identity matrix. The finally loss function can be designed as follows:

$$L = L_D + L_{\text{decor}}$$

(13)

### 3.3.4 Testing

At test time, a test image is fed into the feature extractor firstly, and then the extracted feature is fed into all domain-specific classifiers. After that, all mass sets produced by classifiers are combined based on the reduced Dempster's combinational rule. The class which has the highest combined belief mass is the final prediction. The combined uncertainty mass shows the overall confidence of the prediction. Given a test sample, the really working labeling function is the nonlinear combination of that of source domains, which automatically adjusts the weight of each labeling function of source domains according to the uncertainty of the sample at test time. By this way, the conditional shift between train data and test data is eliminated.

### 4 Experimental Results

#### 4.1 Experimental Setup

##### 4.1.1 Datasets

For demonstrating the effectiveness of the proposed method VAUE, we evaluate it on four widely used DG datasets, namely Office-Home (4 domains, 65 classes, and 75,588 images), VLCS 4 domains, 5 classes, and 10,729 images), TerraIncognita (4 domains, 10 classes, and 24,788 images), PACS (4 domains, 7 classes, and 9,991 images). For all experiments, one domain is selected as the unseen test domain, and the others are treated as training domains.

![Figure 2 Samples from Office-Home, VLCS, TerraIncognita, and PACS](image)

#### 4.1.2 Implementation Details

For all experiments, the networks, which are pre-trained on ImageNet, are trained by AdamW with batch size of 64 for each domain and weight decay of 5e-4. The batch normalization is frozen during the training. The exponential moving average of model parameters with momentum of 0.999 are conducted to make the training processes more stable. We adopt the standard data augmentations following [14]. The visual alignment modules are inserted after 1,2,3-th ConvBlock of ResNet. All results are reported based on the average top-1 classification accuracy over three repetitive runs.

For Office-Home, VLCS and TerraIncognita datasets, following [14], we randomly split each training domain...
4.2 Performance Comparison

4.2.1 Evaluation on Office-Home Dataset

As shown in Table I, we evaluate the proposed VAUE on Office-Home dataset with ResNet-50 as the backbone. We can see that our method achieves the best performance on all four domains and significantly surpasses the second-best competitor by a large margin of 2.1%. As shown in Figure 2 (a), the main domain shift in Office-Home is reflected in the difference of image style across domains, so it is expectable that VAUE performs well due to the visual normalization modules.

4.2.2 Evaluation on VLCS Dataset

The results of the experiments which evaluate the performance of VAUE on VLCS with ResNet-50 as the backbone are summarized in Table II. As shown in Figure 2, the domain shift in VLCS dataset is mainly caused by changes in environments or viewpoints. Intuitively, the ensemble of decisions from different viewpoints would be expectable to perform well. As shown in Table II, we can see that VAUE performs best on most scenarios, and achieves the best accuracy.

4.2.3 Evaluation on TeraIncognita Dataset

TeraIncognita dataset includes photos of wild animals taken at different locations. The domain shift is less related to image style, so the performance improvement is not as obvious as that on other datasets. However, VAUE still achieves the best performance on L100, L43, L46 domains and achieves the second-best average accuracy, as shown in Table III.

4.2.4 Evaluation on PACS Dataset

We also evaluate VAUE on PACS dataset with ResNet-18 and ResNet-50 as backbone respectively to compare our method with a greater variety of methods, as shown in Table IV. We can see that VAUE achieves the best average accuracy with both ResNet-18 and ResNet-50. We notice that there is an obvious performance drop on Photo domain compared to DeepAll, which minimizing the empirical risk by aggregating samples from all source domains. This is mainly due to the ImageNet pretraining [29]. The images of Photo domain are highly similar to images of ImageNet. And the
Table IV Performance comparison on PACS dataset with ResNet-18 as backbone

| Domain    | Art    | Cartoon | Photo | Sketch | Avg. |
|-----------|--------|---------|-------|--------|------|
| DeepAll [30] | 78.93  | 75.02   | 96.60 | 70.48  | 80.25 |
| DSON [31]  | 84.67  | 77.65   | 95.87 | 82.23  | 85.11 |
| DMG [32]   | 76.90  | 80.38   | 93.35 | 75.21  | 81.46 |
| EISNet [33] | 81.89  | 76.44   | 95.93 | 74.33  | 82.15 |
| DGER [30]  | 80.70  | 76.40   | 96.65 | 71.77  | 81.38 |
| RSC [26]   | 83.43  | 80.31   | 99.99 | 80.85  | 85.15 |
| RSC* [26]  | 78.90  | 76.88   | 94.10 | 76.81  | 81.67 |
| MDGHybrid [34] | 81.71 | 81.61   | 96.67 | 81.05  | 85.53 |
| MGFA [35]  | 81.70  | 77.61   | 95.40 | 76.02  | 82.68 |
| FACT [36]  | 85.37  | 78.38   | 95.15 | 79.15  | 84.51 |
| pAaIN [37] | 81.74  | 76.91   | 96.29 | 75.13  | 82.51 |
| EFDmix [38] | 83.90  | 79.40   | 96.80 | 75.00  | 83.90 |
| DSFG [39]  | 83.89  | 76.45   | 95.09 | 78.26  | 83.42 |
| ITL-NET [28] | 83.90 | 78.90   | 94.80 | 80.10  | 84.40 |
| VAUE (ours) | **85.82** | **79.37** | **94.99** | **84.20** | **86.10** |

RSC* denotes the reproduced results from pAaIN [37].

Table V Performance comparison on PACS dataset with ResNet-50 as backbone

| Domain    | Art    | Cartoon | Photo | Sketch | Avg. |
|-----------|--------|---------|-------|--------|------|
| DeepAll [30] | 86.18  | 76.79   | 98.14 | 74.66  | 83.94 |
| DSON [31]  | 87.04  | 80.62   | 95.99 | 82.90  | 86.64 |
| DMG [32]   | 82.57  | 78.11   | 94.49 | 78.32  | 83.37 |
| EISNet [33] | 86.64  | 81.53   | 97.11 | 78.07  | 85.84 |
| DGER [30]  | 87.51  | 79.31   | 98.25 | 76.30  | 85.34 |
| RSC [26]   | 87.89  | 82.16   | 97.92 | 83.35  | 87.83 |
| RSC* [26]  | 81.38  | 80.14   | 93.72 | 82.31  | 84.38 |
| MDGHybrid [34] | 86.74 | 82.32   | 98.36 | 82.66  | 87.52 |
| MGFA [35]  | 86.40  | 79.45   | 97.86 | 78.72  | 85.60 |
| FACT [36]  | 89.63  | 81.77   | 96.75 | 84.46  | 88.15 |
| pAaIN [37] | 85.82  | 81.06   | 97.17 | 77.37  | 85.36 |
| EFDmix [38] | **90.60** | 82.50   | 98.10 | 76.40  | 86.90 |
| DSFG [39]  | 87.30  | 80.93   | 96.59 | 83.43  | 87.06 |
| ITL-NET [28] | 87.10 | 83.30   | 96.10 | 79.30  | 86.40 |
| VAUE (ours) | 88.30  | **84.11** | **95.95** | **86.13** | **88.62** |

Table VI Ablation Study on PACS dataset

| Domain    | Art    | Cartoon | Photo | Sketch | Avg. |
|-----------|--------|---------|-------|--------|------|
| VAUE      | **85.82** | **79.37** | **94.99** | **84.20** | **86.10** |
| - w/o VA  | 83.04  | 77.35   | **95.09** | 79.85  | 83.83 |
| - w/o EC  | 85.72  | 78.77   | 94.51  | 82.88  | 85.48 |
| - w/o CD  | 85.27  | 78.03   | 94.89  | 83.30  | 85.37 |
| - w/o UE  | 85.63  | 78.36   | 94.13  | 83.41  | 85.38 |

5 Conclusion

In this paper, we design the visual alignment module for dealing with the covariate shift by aligning the distribution of image style to a common Gaussian distribution. Another uncertainty-guided ensemble strategy is proposed to deal with the conditional shift between training domains and test samples by dynamically adjustment. Experiment results show the excellent performance of the proposed VAUE.

Acknowledgements

The authors are thankful for the financial support by the National Key Research and Development Program of China (No. 2018AAA0102000) and the National Natural Science Foundation of China (62106266, U1936206, 61906191). Yongqiang Tang and Wensheng Zhang are the corresponding authors.

References

[1] M. Long, H. Zhu, J. Wang, S. Wang, and M. I. Jordan, “Un-supervised Domain Adaptation with Residual transfer networks,” in Proc. Advances Neural Inf. Process. Syst., 2016.

[2] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodriguez, N. V. Chawla, F. Herrera, “A unifying view on dataset shift in classification,” in Pattern Recognit., vol. 45, no. 1, pp. 521-530, 2012.

[3] Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang and D. Tao, “Deep domain generalization via conditional invariant adversarial networks,” in Proc. Eur. Conf. Comput. Vis., pp. 624-639, 2018.

[4] K. Muandet, D. Balduzzi, and B. Scholkopf, “Domain generalization via invariant feature representation,” in
Proceedings of CCIS2022

I. Albuquerque, J. Monteiro, M. Darvishi, T. H. Falk, and I. Mitliagkas, “Generalizing to unseen domains via distribution matching,” *arXiv preprint arXiv:1911.00804v6*, 2021.

Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao, “Deep domain generalization via conditional invariant adversarial networks,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 624-639, 2018.

X. Huang, S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1501-1510, 2017.

K. Zhou, Y. Yang, Y. Qiao, and X. Tao, “Domain generalization with mixstyle,” in *Proc. Int. Conf. Learn. Representat.*, 2021.

H. Nam, H. Lee, J. Park, W. Yoon, and D. Yoo, “Reducing domain gap by reducing style bias,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 8690-8699, 2021.

M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” in *Proc. Advances Neural Inf. Process. Syst.*, 2018.

Z. Han, C. Zhang, H. Fu and J. T. Zhou, “Trusted Multi-View Classification with Dynamic Evidential Fusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, doi: 10.1109/TPAMI.2022.3171983.

A. Josang, “Subjective Logic: A formalism for reasoning under uncertainty,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Springer, 2018.

K. Sentz, S. Ferson, “Combination of evidence in Dempster-Shafer theory,” in *Citeeseer*, 2002.

I. Gulrajani, D. Lopez-Paz, “In search of lost domain generalization,” in *Proc. Int. Conf. Learn. Representat.*, 2021.

D. Li, Y. Yang, Y. Song, and T. M. Hospedales, “Deeper, broader and artier domain generalization,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 5542-5550, 2017.

M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, “Invariant risk minimization,” *arXiv preprint arXiv:1907.02893*, 2019.

S. Sagawa, P. W. Koh, T. B. Hashimoto, and P. Liang, “Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization,” in *Proc. Int. Conf. Learn. Representat.*, 2017.

H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “Mixup: Beyond empirical risk minimization,” in *Proc. Int. Conf. Learn. Representat.*, 2018.

D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, “Learning to generalize: Meta-learning for domain generalization,” in *Proc. AAAI Conf. Artif. Intell.*, 2018.

X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, “Moment matching for multi-source domain adaptation,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1406-1415, 2019.

M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” in *Proc. Int. Conf. Mach. Learn.*, pp. 97-105, 2015.

Y. Ganin, V. Lempitsky, “Unsupervised domain adaptation by backpropagation,” in *Proc. Int. Conf. Mach. Learn.*, pp. 1180–1189, 2015.

G. Blanchard, A. A. Deshmukh, Ü. Dogan, G. Lee, and C. Scott, “Domain generalization by marginal transfer learning,” *J. Mach. Learn. Res.*, vol. 22, pp. 1-55, Jan. 2021.

M. Zhang, H. Marklund, N. Dhawan, A. Gupta, S. Levine, and C. Finn, “Adaptive risk minimization: learning to adapt to domain shift,” in *Proc. Advances Neural Inf. Process. Syst.*, pp. 25664-23678, 2021.

D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. L. Priel, and A. Courville, “Out-of-distribution generalization via risk extrapolation (rex),” in *Proc. Int. Conf. Mach. Learn.*, pp. 5815-5826, 2021.

Z. Huang, H. Wang, E. P. Xing and D. Huang, “Self-challenging improves cross-domain generalization,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 124-140, 2020.

A. Rame, C. Danette, and M. Cord, “Fishr: Invariant gradient variances for out-of-distribution generalization,” in *Proc. Int. Conf. Mach. Learn.*, pp. 18347-18377, 2022.

B. Gao, H. Gouk, Y. Yang, and T. Hospedales, “Loss function learning for domain generalization by implicit gradient,” in *Proc. Int. Conf. Mach. Learn.*, pp. 7002-7016, 2022.

Z. Xu, D. Liu, J. Yang, C. Raffel, and M. Niethammer, “Robust and generalizable visual representation learning via random convolutions,” in *Proc. Int. Conf. Learn. Representat.*, 2020.

S. Zhao, VM. Gong, T. Liu, H. Fu, and D. Tao, “Domain generalization via entropy regularization,” in *Proc. Advances Neural Inf. Process. Syst.*, pp. 16096-16107, 2020.

S. Seo, Y. Suh, D. Kim, G. Kim, J. Han, and B. Han, “Learning to optimize domain specific normalization for domain generalization,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 68-83, 2020.

P. Chattopadhyay, Y. Balaji, and J. Hoffmann, “Learning to balance specificity and invariance for in and out of domain generalization,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 301-318, 2020.

S. Wang, L. Yu, C. Li, C.-W. Fu, and P.-A. Heng, “Learning from extrinsic and intrinsic supervisions for domain generalization,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 159-176, 2020.

D. Mahajan, S. Tople, and A. Sharma, “Domain generalization using causal matching,” in *Proc. Int. Conf. Mach. Learn.*, pp. 7313-7324, 2021.

M. H. Khan, T. Zaidi, S. Khan, and F. S. Khan, “Mode-guided feature augmentation for domain generalization,” in *Proc. Brit. Mach. Vis. Conf.*, 2021.

Q. Xu, R. Zhang, Y. Zhang, Y. Wang, and Q. Tian, “A fourier-based framework for domain generalization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 14383-14392, 2021.

O. Nuriel, S. Benaim, and L. Wolf, “Permuted adain: Reducing the bias towards global statistics in image classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 9482-9491, 2021.

Y. Zhang, M. Li, R. Li, K. Jia, and L. Zhang, “Exact feature distribution matching for arbitrary style transfer and domain generalization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 8035-8045, 2022.

S. Yuan, Y. Li, D. Wang, K. Bai, L. Carin, and D. Carlson, “Learning to weight filter groups for robust classification,” in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pp. 3041-3050, 2022.