Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/13/20230725

3D point cloud domain generalization via adversarial training

Yichen Sun

The Khoury College of Computer Science, Northeastern University, USA

sun.yich@northeastern.edu

Abstract. The purpose of the paper is to tackle the classification problem of 3D point cloud data in domain generalization: how to develop a generalized feature representation for an unseen target domain by utilizing sub-field of numerous seen source domain(s). We present a novel methodology based on both adversarial training to learn a generalized feature representations across subdomains in domain adaptation called 3D-AA. We specifically expand adversarial autoencoders by applying the Maximum Mean Discrepancy (MMD) measure to align the distributions across several subdomains, and then matching the aligned distribution to any given prior distribution via adversarial feature learning. In this manner, the learned 3D feature representation is supposed to be universal to the observed source domains due to the MMD regularization and is expected to generalize well on the target domain due to the addition of the prior distribution. We applied an algorithm to train two different 3D point cloud source domains with our framework. The combination of multiple loss functions on 3D point cloud domain generalization task show that our applied algorithm performs better and learn more generalized features for the target domain than the source-only algorithm which only utilized the MMD measurement.

Keywords: 3D point cloud, domain generalization, adversarial training, domain alignment, deep learning.

1. Introduction

In the computer vision 3D applications society, 3D point cloud classification tasks is a vital mechanism to guaran- tee a promising and profound results of widely used 3D recognition applications (i.e., robot vision, and self-driving vehicle). While the issue of some existing methodologies is that the trained model will be influenced negatively due to the domain shift. It means a model and an algorithm trained with the source domains may perform poorly on the target domain. Some researchers design methodologies with unsupervised domain adaptation (UDA) [1-4] to mitigate domain shift's impact on 3D point cloud data. Most UDA techniques are intended for 2D vision applications and concentrate on aligning global image features across several domains [5-7]. While in 3D point cloud data analysis jobs, knowledge of regional and local geometry is essential for attaining better learning results [8]. Zhou et al. initially presents UDA on the task to estimate 3d key points relying on the regularization of the multi-view consistency term [9]. However, this approach cannot be used for more generalizing tasks, such as classification.

Recently, domain generalization is an increasing interested research topic not only in machine learning research also in deep neural network architecture. The purpose of Domain Generalization (DG) is to develop a model that can generalize to an unknown test domain in a difficult situation where one

DOI: 10.54254/2755-2721/13/20230725

or more distinct, but related domains are specified [2]. But most of research papers focus on 2 dimensional pictures and images domain(s). This means the domain generalization resolve most 2D domain problems but leave spaces for 3D point cloud domain(s). Not too many researchers investigate whether the Domain Generalization works or not in 3D point cloud domain(s). We consider a new methodology to design or develop a new domain generalization algorithm. This algorithm performs better transfer learning ability in 3D domain problems without seeing target domains. With our domain generalization method, we can achieve a better accuracy on detection procedure of self-driving vehicle, and other 3D figures or objects.

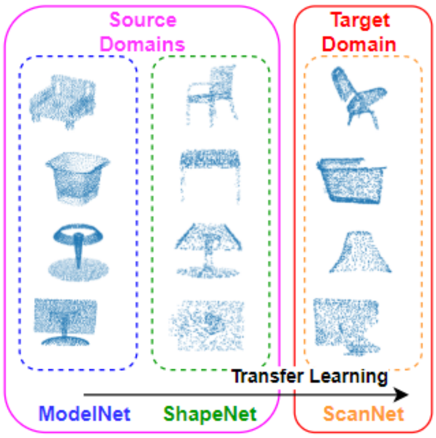

Figure 1. A problem statement for 3D point cloud domain generalization.

Previous works on domain generalization targeted on de- veloping different algorithms and approaches to learn future representations on 2D source domains. For instance, Wang et al. presented several examples from the dataset PACS for domain generalization [10,11]. The training set in the dataset is composed of images from the sketch, cartoon, and art painting domains with domain generalization to develop a generalized model that excels at the unseen target domain of photographs. Qin et al. proposed the PointDAN model and algorithm to learn a multi-scale 3D domain adapation network for point cloud representation based on jointly alignments of the global and local features in multi-level [8]. To learn domain invariant features, Ghifary et al. proposed a multi-task autoencoder [12].

Although some encouraging research and findings have been reported, 2D domain generalization methodologies may experience lower performance values with the observed 3D point cloud source domain data. Or, to put it another way, by concentrating on 2D learning a representation by reducing the difference between the observed source domains, the learnt representation may generalize well for all the 2D domains but poorly for the 3D point cloud domains. In this paper, we applied an innovative framework and multiple strategies for 3D point cloud domain generalization that aims to learn a universal representation across 3D domains not only by combining the seen source domains but also by evaluating the performance of 3D point data with our methods. We primarily applied methodology that focuses on 3D point cloud data in domain Generalization with both adversarial training network (3D-AA). Adversarial training procedure generates a set of adversarial samples across domains that are mis-classified by the neural network [13], and then use correct labeled data error to make better predictions. They improve the presence of domain shifts or dataset bias by reducing the difference between the source domains and target domain distributions and therefore improve 3D point cloud domain generalization. Zhu et al. proposed an innovative neural network learning a transfer model by aligning the corresponding subdomain distributions across different domains [14].

Note our work is different from the multi-scale 3D domain adaption network for point cloud representation proposed in [8], which exploits 3D point cloud data in domain adaptation field and

DOI: 10.54254/2755-2721/13/20230725

describe hierarchically scaled characteristics with SA nodes across objects and domains. In our methods, we used an MMD-based regularization strategy to reduce the gap between domains [15], which eliminates the need for data instance correspondence to investigate 3D point cloud domain generalization tasks. This algorithm is called 3D-AA, which achieves better model transferring ability in terms of classification loss, adversarial loss, MMD loss, autoencoder and decoder discrepancy loss. The paper further validates our model with new benchmark that tests and verifies its effectiveness.

In deep learning, 3D point cloud is a popular representation of 3D geometry object, there are some research and findings that study the performance of point cloud data in proposed methodologies. Liu et al. proposed a novel neural network architecture called MeteorNet, which sepcifically learns representation for 3D point cloud data sequences [16]. In MeteorNet, the network addressed points within the same frame, points in different frames, and neighboring points in local structure. The model architecture is related to multi-Scale 3D Domain Adaption Network (PointDAN) [8]. In PointDAN, the network for point cloud data has two alignment methods: A local alignment model LA that aims to learn the discriminative local structures for aligning domains, and a global alignment GA that aims to distinguish align global features across domains with em- ployed adversarial-training strategy. The local alignment LA and global alignment GA are jointly aligning the global and local features in multi-level.

Besides, our work is also related to Generative Adversarial Network (GAN) [17]. There are two different types of net- works in GAN: a generative model (G) that aims to capture the distribution of the training data for data generation and a discriminative model (D) that aims to distinguish between instances taken from G and the original data sampled from the training dataset. In a collaborative, competitive training process, the generative model G and the discriminative model D are used: 1) Train D to differentiate between real instances and phony instances produced by G. 2) Train G to deceive D using the instances it generates. Numerous algorithms and frameworks has been suggested recently built upon the GAN approach. For instance, many domain adaptation methods are based on adversarial training [18-22], such as the gradi- ent reversal layer [23]. AAE was suggested by Makhzani et al. [24] to train the encoder and the decoder using an adversarial learning approach. Li et al. proposed jointly optimize a multi-domain autoencoder regularized by the MMD distance, and discriminator, and a classifier in an adversarial training setting to learn a feature representation [15]. These deep learning techniques for 3D point cloud domain generalization task serve as inspiration for the following experiment. However, most algorithms only address 2D data, and these works only project 2D images and do not study models directly deal with 3D point cloud data.

2. Related works

As mentioned in the previous sections, the purpose pf adversarial training is to learn a more accurate classifier to be used for the target domain by combining data from the partic- ular classes and features of source domain(s). The distinction between them is that for domain adaptation, some unlabeled data and even a few labeled data from the target domain are used to capture attributes of the target domain for model adaptation [15]. While several strategies have been put out for domain adaptation, 3D point cloud domain generalization has received much less attention.

3. Main works

3.1. Overall

A basic understanding behind 3D domain generalization is that the source domains and target domain contain a feature space, which a trained model can learn the difference and characteristics from source domains to generalize well on the target domain. In our framework, we perform two source domains alignment to learn a universal feature representation of samples in training data in encoder and decoder steps. Then, we integrate features learnt by hidden nodes in two coders to compute several losses function and propagate forward to an adversarial loss discriminator and a sub-classification model.

To ensure for the learnt feature space to possess the stated characteristic, our method is built on AAE [24] [15], which is a popular suggested probabilistic autoencoder model recently, to the multi-domain setting for learning cross-domain invariant features using MMD in an adversarial learning setting. We intend to learn a feature space in 3D point cloud dataset within two source domains by minimizing the distribution differences on them depend on MMD loss function. Thereby, we term our applied methodology by 3D point cloud adversarial autoencodeer (3D-AA). We extend 3D-AA to the supervised learning context by integrating a classification layer to include label information within training, making the learnt feature space discriminated to labels. Also, the target domain passes through the same procedure as two sources domain but without feature alignment process. In this manner, the final classifier and feature space are learnt concurrently. In the parts that follow, we go into further depth about our applied model.

Notation: 3D point cloud data includes three domains in total: ModelNet, ShapeNet, ScanNet. We denote by MD = [x1, x2, x3, ..., xi], SH = [y1, y2, y3, ..., yi], and SC = [z1, z2, z3, ..., zi] as i represents the total number of inputs in domains. We can pick two of three as the inputs for source domains SD $\in$ {1, 2} and the one left as the inputs for target domain TD $\in$ {1}.

3.2. Source-domain MMD alignment

In our framework, we use PointNet [25] which is a multi-level model that aims to represent the 3-dimensional coordinates (x, y, z) to learn an evaluation function f : X $\rightarrow$ R that projects raw inputs into a shared information for cross- domain samples. The PointNet model is highly effective and efficient to directly consume point cloud data.

3.3. MMD alignment adversarial training

In this section, we describe how 3D-AA extends AAE [15] for 3D domain generation. The architecture of 3D-AA is presented in Figure 2. In 3D-AA, we have an encoder E(.) to evaluate inputs to hidden nodes and a decoder D(.) to recover information from hidden nodes. All the domains, including the target domain in the prediction phrase, share the pair of encoder and decoder. Over all of the observed source domains, the autoencoder's discrepancy loss is calculated as

$$\ell_{ED} = \sum_{i=1}^{N} \|D(E(.)) - E(.)\|$$

The feature extraction G(.) will be the input for autoencoder in the training process. Then, we pass the extracted feature representations of our source domains samples into encoder phrase E(G(.)) Following GAN [17], MMDAAT can be written as the following optimization problem,

$$\ell ED + RMMD + \ell ce(S1, S2),$$

where $L_{ce} = -(y \log(p) + (1 - y) \log (1 - p))$ and Dis(.) is the discriminator to separate the true nodes samples on the generated nodes by E(.).

3.4. Adversarial alignment

Adversarial alignment enables competing models to collaborate on theory development for cumulative study. In the framework, we employ adversarial loss function to discover features that distribute several adversarial domains as well as the unseen domain. Thereby, we apply the idea from Mao et al. to substitute the log probabilistic term by the least- squared term formulated as follows [26],

$$\ell_{adv} = Adv_{\,x\sim p(x)}\,[d(x)^2] + Adv_{\,y\sim p(y)}\,[1 - D(E(y))^2]$$

As mentioned in Li et al. [15], the minimax objective can be achieved by minimizing x^2 divergence. 3D-AA minimizes the restricting error between the autoencoder and autodecoder.

3.5. Classifier module

To integrate label representations into hidden code learning in 3D-AA, we simply add a sub-classification layer on top of the hidden layer. According to MMD and Adversarial alignment procedure, we have reduced the internal statis- tical variation and used adversarial learning to make the features consistent, and then we can use this feature to do the classification on source domains and target domain. As Zhu et al. mentioned that the transfer learning methods mostly combine learning algorithm with adaption layer and a universal domain adaptation loss [14]. The official representation may be,

$$min\frac{1}{n}\ L_{ce} + \ell_{ED}$$

where L_{ce} is the cross-entropy loss function refers to classification loss and ℓ_{ED} is domain adaptation loss (differentiate autoencoder and autodecoder).

3.6. Overall objectives

To learn cross representations among all of three align- ments, we propose a method to calculate a total of three loss function alignments which can be formulated as follows,

$$Ltotal = RMMD + \ell adv + \ell ED,$$

and this can be optimized directly by back-propagation and gradient descent algorithms. This method focuses on make distribution alignments of the source domains by universally learning the multiple 3D point cloud data MMD loss, adversarial loss, and classification loss.

Table 1. Splits of PointDA-10 dataset.

Dataset	Bathtub	Bed	Bookshelf	Cabinet
M Train / Test	106 / 50	515 / 100	572 / 100	200 / 86
S1 Train / Test	599 / 85	167 / 23	310 / 50	1,076 / 126
S2 Train / Test	98 / 26	329 / 85	464 / 146	650 / 149

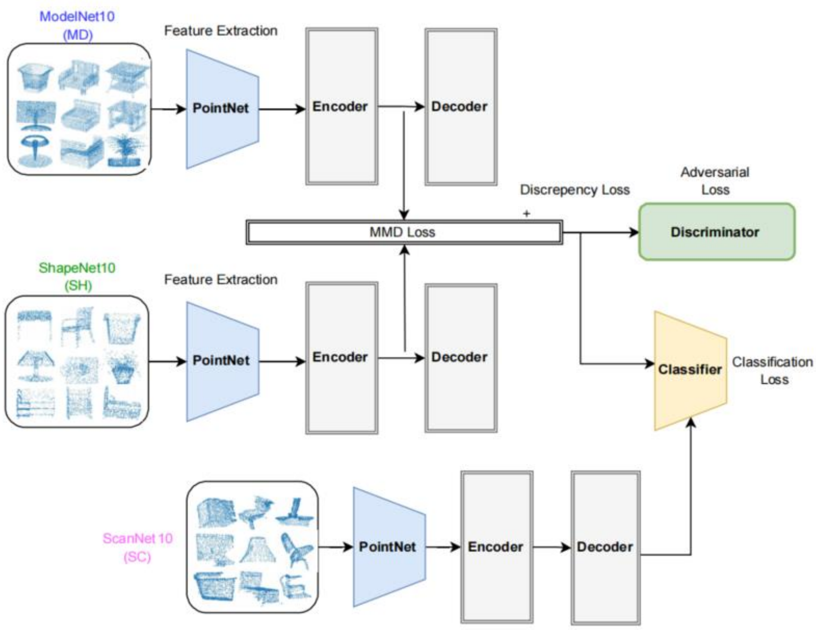

Figure 2. An overview of our designed framework (3D-) for 3D point cloud domain generalization.

We choose two source domains from 3D point cloud dataset and extract their feature based on PointNet model. Next, we perform domain generalization by incorporating feature representations based on Maximum Mean Discrepancy (MMD loss) and discrepancy loss with an adversarial discriminator network. Then, we further integrate a classification sub-domain model to evaluate the specific domain information from source samples and target samples.

4. Experiments

In this section, we conduct experiments on 3D point cloud data problems to evaluate the performance of the framework and methodologies for 3D point cloud for domain generalization. We use the popular benchmark dataset 3D Point Cloud [8] PointDA-10 Dataset with rotation for 3D point cloud classification.

4.1. Experiments setup

Since there is not a 3D point cloud benchmark created for domain generalization, we suggest three datasets with distinct characteristics, namely ModelNet-10, ShapeNet-10, and ScanNet-10, for the assessment of 3D point cloud domain generalization via our method. We take the examples for 10 common classes from ModelNet40 [27], ShapeNet [28], and ScanNet [29], respectively, to construct them. Table 1 and Figure 3 display the statistics and visualization for some samples in PointDA-10 Dataset. We categorize 3 sorts of generalization situations, which are **MD, SH→SC, MD, SC→SH, and SH, SC→MD** with different combination of loss function.

ModelNet-10 (MD): ModelNet40 is a collection of 40 categories of clean 3D CAD models. Because these two objects almost exactly have the same structure, we consider the "nightstand" class in ModelNet-40 to be the "cabinet" class in ModelNet-10 in order to extract overlapped classes [8]. We sample spots on the surface as [30] after receiving the CAD model to completely cover the object.

ShapeNet-10 (SH): 55 categories of 3D CAD models from online libraries are included in ShapeNet. In comparison to ModelNet, ShapeNet has more samples, and its objects' structural variance is higher [8]. To gather the ShapeNet points on the surface, we use uniform sampling to reduce some marginal points.

ScanNet-10 (SC): Real-world indoor scenes that have been scanned and rebuilt can be found on ScanNet. In annotated bounding boxes, we separate 10 classes' instances for categorization [8]. ScanNet is the most difficult yet practical real-world indoor domain.

Evaluation: Given the labeled samples in two source domain and unlabeled samples from target domain for training, all the models with different accuracy method would be computed on the test set of target domain.

Implementation Detail: We choose the idea of PointNet [25] as the foundation of our MMD combinations generalization method, and MMD-AAE [15] as backbone for implementing our proposed method. This approach based on Adam optimizer [23] and trained the network with NVIDIA GPU 2070. The learning rate is chosen as 0.0001. All models have been trained for 20 epochs of batch size 32. Both hyper- parameters are adaptable and flexible to adjust for further ablation study.

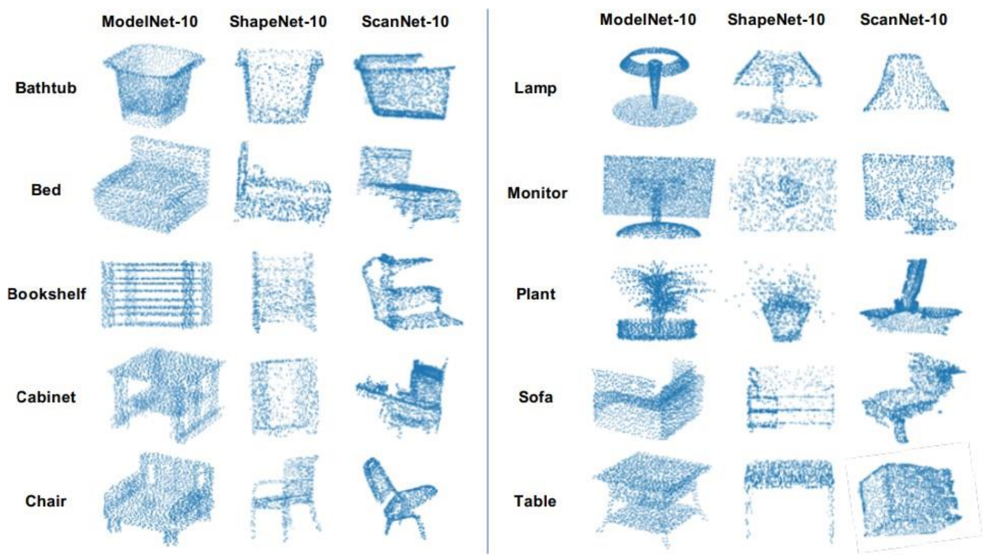

Figure 3. Samples of PointDA-10 dataset. Source: adapted from [8].

DOI: 10.54254/2755-2721/13/20230725

4.2. Baseline method

We compare our proposed combination MMD Generalization with the following baseline methods for domain generalization in terms of 3D point cloud data accuracy.

•Source-only: We adopt a normal cross entropy loss to train a classifier by directly using two sources domain.

•**MMD+dis:** We adopt a combination of MMD loss and discrepancy loss between encoder and decoder to learn the hidden characteristics of point cloud samples and train a source-only.

•adv+dis: We apply adversarial loss and discrepancy loss between encoder and decoder to determine

•MMD+adv+dis: We consider using full methods to predict the accuracy of our models to compare with previous algorithms that consist of two methods.

4.3. Overall performance

The classification results and comparison of models on PointDA-10 dataset are represented in Table 2. The proposed methods slightly outperform on the source-only method on some generalization scenarios. And we apply the two source domains and leave one for target domain strategy to establish 3D domain generalization tasks. We conduct the experiment for 20 epochs and analyze the classification accuracy.

4.4. Ablation study

To analyze the effects of the function alignments, we introduce the ablation study which composed of four pars: source-only, MMD+discrepancy, adversarial+discrepancy, and MMD+adversarial+discrepancy. In Table II, the source-only does pretty good in measuring ModelNet and ScanNet as source domains, however the other two measurements do not have a good performance. For function alignment methodologies, ModelNet and ScanNet source domains do not have good performance, and rest of the other two methods performance slightly better than the source-only. For our strategy, which includes MMD, discrepancy, and adversarial, it has much higher accuracy in predicting all the choices of the source domains. The reason behind the higher performance score is that all three procedures learnt the hidden feature information to prevent lower accuracy.

Table 2. Performance on 3D point cloud data.

Method	md, sh ->sc	md, sc ->sh	sh, sc ->md
source-only	0.1358	0.6453	0.2007
MMD+dis	0.1523	0.5735	0.4159
adv+dis	0.1426	0.5978	0.1227
Ours (MMD+dis+adv)	0.1693	0.6632	0.3954

5. Conclusion

In this paper, we conduct 3D Point Cloud Adversarial Autoencoder (3D-AA) for 3D point cloud domain generalization. The main goal is to learn a multi-domain autoencoder normalized by MMD distance, a discriminator, and a classifier are collaboratively optimized in an adversarial training process to learn a feature representation with 3D point cloud domains. To evaluate the proposed methodology, we build an ablation study experiment to compare the score of each method. In the experiments, we illustrate the combination of our approach over multiple choices of domain generalization methods.

References

[1] H. Zou, Y. Zhou, J. Yang, H. Liu, H. P. Das, and C. J. Spanos, “Consensus adversarial domain adaptation,” in The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019.

[2] J. Yang, H. Zou, S. Cao, Z. Chen, and L. Xie, “Mobileda: Toward edge domain adaptation,” IEEE Internet of Things Journal, vol. 7, no. 8, pp. 6909–6918, 2020.

[3] J. Yang, J. Yang, S. Wang, S. Cao, H. Zou, and L. Xie, “Advancing imbalanced domain adaptation: Cluster-level discrepancy minimization with a comprehensive benchmark,” IEEE Transactions on Cybernetics, 2021.

[4] J. Yang, H. Qian, H. Zou, and L. Xie, “Learning decomposed hier- archical feature for better transferability of deep models,” Information Sciences, vol. 580, pp. 385–397, 2021.
[5] Y. Xu, J. Yang, H. Cao, Z. Chen, Q. Li, and K. Mao, “Partial video domain adaptation with partial adversarial temporal attentive network,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9332–9341.
[6] Y. Xu, H. Cao, K. Mao, Z. Chen, L. Xie, and J. Yang, “Aligning correlation information for domain adaptation in action recognition,” IEEE Transactions on Neural Networks and Learning Systems, 2022.
[7] Y. Xu, J. Yang, H. Cao, K. Wu, M. Wu, and Z. Chen, “Source-free video domain adaptation by learning temporal consistency for action recognition,” in European Conference on Computer Vision. Springer, 2022, pp. 147–164.
[8] C. Qin, H. You, L. Wang, C.-C. J. Kuo, and Y. Fu, “Pointdan: A multi-scale 3d domain adaption network for point cloud representation,” Advances in Neural Information Processing Systems, vol. 32, 2019.
[9] X. Zhou, A. Karpur, C. Gan, L. Luo, and Q. Huang, “Unsupervised domain adaptation for 3d keypoint estimation via view consistency,” in Proceedings of the European conference on computer vision (ECCV), pp. 137–153.
[10] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen,
[11] W. Zeng, and P. Yu, “Generalizing to unseen domains: A survey on domain generalization,” IEEE Transactions on Knowledge and Data Engineering, 2022.
[12] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, “Deeper, broader and artier domain generalization,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 5542–5550.
[13] M. Ghifary, W. B. Kleijn, M. Zhang, and D. Balduzzi, “Domain generalization for object recognition with multi-task autoencoders,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2551–2559.
[14] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discrim- inative domain adaptation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7167–7176.
[15] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and
[16] Q. He, “Deep subdomain adaptation network for image classification,” IEEE transactions on neural networks and learning systems, vol. 32, no. 4, pp. 1713–1722, 2020.
[17] H. Li, S. J. Pan, S. Wang, and A. C. Kot, “Domain generalization with adversarial feature learning,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5400–5409.
[18] X. Liu, M. Yan, and J. Bohg, “Meteornet: Deep learning on dynamic 3d point cloud sequences,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9246–9255.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley,
[20] S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,”
[21] Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.
[22] J. Yang, H. Chen, Y. Xu, Z. Shi, R. Luo, L. Xie, and R. Su, “Domain adaptation for degraded remote scene classification,” in 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2020, pp. 111–117.
[23] J. Yang, H. Zou, Y. Zhou, Z. Zeng, and L. Xie, “Mind the dis- criminability: Asymmetric adversarial domain adaptation,” in European Conference on Computer Vision, 2020.
[24] J. Yang, H. Zou, Y. Zhou, and L. Xie, “Towards stable and comprehen- sive domain alignment: Max-margin domain-adversarial training,” arXiv preprint arXiv:2003.13249, 2020.——, “Robust adversarial discriminative domain adaptation for real- world cross-domain visual recognition,” Neurocomputing, vol. 433, pp. 28–36, 2021.

DOI: 10.54254/2755-2721/13/20230725

[25] J. Yang, X. Peng, K. Wang, Z. Zhu, J. Feng, L. Xie, and Y. You, “Divide to adapt: Mitigating confirmation bias for domain adaptation of black- box predictors,” arXiv preprint arXiv:2205.14467, 2022.
[26] Y. Ganin and V. Lempitsky, “Unsupervised domain adaptation by back- propagation,” in International conference on machine learning. PMLR, 2015, pp. 1180–1189.
[27] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, “Adver- sarial autoencoders,” arXiv preprint arXiv:1511.05644, 2015.
[28] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.
[29] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2794–2802.