Class-wise Feature Distribution Matching Regularization for Domain Generalization

Seungmin Lee
Seoul National University
profile2697@gmail.com

Abstract

Domain generalization (DG) aims to learn a model that generalizes well to an unseen domain (target domain), which has a different distribution than known domains (source domains). Many of the previous works try to learn domain-invariant features. These methods adopt a loss that tries to match the whole distributions of the source domains. However, these works are sub-optimal because they rarely utilize task-specific information, such as class labels. Concerning the information, we propose a simple but effective regularizing method called Class-wise Feature Distribution Matching (FDM). The proposed method attempts to induce the network to produce similar features when the labels are the same. By doing this, a model is expected to learn more task-specific and robust features than the previous works.

1. Introduction

Deep learning has been remarkably successful in many areas [10, 14, 26, 19]. However, many studies find out deep learning methods are hard to generalize when they encounter an unseen domain, which has a different data distribution than domains used for training [16, 15, 16, 29, 7, 3, 20]. This problem called *domain shift* [30]. For alleviating the domain shift problem, many studies have been carried out on different assumptions. Domain Adaptation (DA) assumes there are two domains. The first is a fully-labeled source domain, and the other is a sparsely labeled or totally unlabeled target domain. Otherwise, domain generalization (DG) assumes there are some fully-labeled source domains, but the target domain is totally unavailable. DG is a challenge but important research area because generalization to other domains is crucial to make a safe artificial intelligence. Moreover, it is helpful to understand how deep neural networks see the world.

Existing DG studies can be classified into several categories depending on their strategies. Some methods proposed novel model architectures that are robust to domain shift [13, 15]. Others suggested learning algorithms aim to induce a model to fit in a more robust minimum [17, 16, 1]. The others adopted losses to learn domain-invariant features by matching feature distributions of the source domains [9, 22, 18]. Although these domain-invariant feature learning methods work well, the methods are sub-optimal because they do not utilize task-specific information such as class labels explicitly.

Our approach also aims to learn domain-invariant features, but the proposed method explicitly adopts the task-specific information. More specifically, we add a simple consistency loss that induces the model to produce similar features when the labels are the same. By using this constraint, the model is expected to learn features that are adequately semantic and constant across domains.

To demonstrate the proposed method works, we conducted an experiment on PACS [15] dataset using AlexNet [14]. In the test, the proposed regularization method shows improvement compared to a baseline method. Additionally, the proposed method gives comparable or better performance than the majority of the previous methods. We are planning to test our regularization method on other DG benchmarks, such as VLCS [6] or using ResNet [12].

2. Related Works

2.1. Multi-Domain Learning and Multi-Source Domain Adaptation

The primary purpose of Multi-Domain Learning (*MDL*) is to learn a single model that can compactly represent all domains with a smaller number of parameters [31, 24, 2, 25]. For this purpose, Bilen *et al.* [2] adopts domain-specific parameters of instance normalization and batch normalization while using shared parameters in other layers. Rebuffi *et al.* [24] transforms the standard residual network architecture to share a significant amount of parameters between different domains.

Multi-Source Domain Adaptation (*MSDA*) also uses a set of domains, but it additionally utilizes the images of

an unlabeled target domain. The main focus of MSDA is to train a model that works well on the target domain without labels of it. Even though many studies have been conducted on single-source domain adaptation, there are a limited number of researches on MSDA [32, 5, 11, 23]. Chang *et al.* [5] proposes a domain-specific batch normalization with shared weights parameters and extends their method to MSDA. Peng *et al.* [23] suggests reducing moment distances between different domains. The moment distance measures the difference between feature distributions of two domains without concerning the task at hand.

MDL and MSDA are closely related to DG since DG also utilizes many domains as training data. However, DG is different from MDL in that the primary focus of DG is to learn semantic and domain-invariant features, not to learn compact representations. Furthermore, DG is more challenging than MSDA because the target domain is totally unavailable in DG.

2.2. Domain Generalization

Even though existing DG methods basically aim to learn domain-invariant features, the methods can be classified into several groups based on their approaches. The first group proposes a novel architecture [13, 15]. The methods separate domain-specific parameters and domain-agnostic parameters. After that, they only extract and utilize the domain-agnostic parameters for the unseen domain. The second group of methods suggests optimization algorithms that adopt episodic learning or self-supervised learning [17, 16, 1, 4]. For example, MLDG [16] constructs an episode by splitting the source domains into training domains and test domains in each iteration. The final group of methods uses losses that aims to learn domain-invariant features [9, 22, 18]. These methods often adopt maximum mean distribution (*MMD*) constraints. However, MMD simply tries to match the feature distributions of all available source domains without concerning the task at hand. Therefore, methods that adopted MMD can be sub-optimal [28, 27]. Otherwise, we propose a regularizing method using a simple consistency loss that explicitly utilizes the task-specific information. By using this simple loss, we expect that the model can learn domain-invariant but semantic features.

3. Proposed Method

3.1. Problem Setting and Notation

In DG, we assume that there are n source domains $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n\}$ where $\mathcal{D}_i$ indicates i-th source domain which contains sample-label pairs $\{x_i^j, y_i^j\}$. Using the source domains, our goal is to train a model h that performs well on unseen target domain $\mathcal{D}_t$. We assume the model h consists of a feature extractor f and a classifier c.

3.2. Deep-All Method

The Deep-All Method is a simple but effective baseline. In this method, we just aggregate all examples of source domains and train the model using the aggregated samples. If we work on a classification task, we can use cross-entropy loss as follows:

$$L_{all} = \mathbb{E}_{\mathcal{D}_s \sim \mathcal{D}} \left[\mathbb{E}_{\mathbf{x}_i, y_i \sim \mathcal{D}_s} \left[\mathbf{y}_i{}^T \log h(\mathbf{x}_i) \right] \right] \quad (1)$$

where $\mathbf{y}_i$ is a one-hot vector representation of y_i

3.3. Class-wise Feature Distribution Matching Regularization

We propose a consistency loss called Class-wise Feature Distribution Matching Regularization (*FDM*), which tries to make a model generate similar features when the labels are the same. The FDM loss is calculated as follows: For each iteration, the feature extractor produces features of the source domains. After that, we calculate the averages of the features by classes. Lastly, the FDM loss is calculated as a consistency loss between the averages and the features that share the same label. The FDM loss is defined as follows:

$$L_{fdm} = \mathbb{E}_{\mathcal{D}_s \sim \mathcal{D}} \left[\mathbb{E}_{\mathbf{x}_i, y_i \sim \mathcal{D}_s} \left[\mathbb{D}_{KL}(f(\mathbf{x}_i) || \mathbf{m}_{y_i}) \right] \right] \quad (2)$$

where $\mathbb{D}_{KL}(\cdot||\cdot)$ represents Kullback–Leibler divergence and $\mathbf{m}_{y_i}$ is the average feature of class y_i. For simplicity, $\mathbf{m}_{y_i}$ is calculated for each batch. The L_{fdm} regularize the feature extractor f to extract similar features when the classes are the same.

Finally, overall loss function is defined as a weighted sum of L_{all} and L_{fdm} as follows:

$$L = L_{all} + \lambda L_{fdm} \quad (3)$$

where λ is a hyperparameter that controlls the magnitude between the two losses.

4. Experiments

To demonstrate the effectiveness of the proposed method, we conduct an experiment on PACS dataset [15] using AlexNet [14]. We follow the experiment protocol of Li *et al.* [15]. We use a batch size 512, and λ is set to 0.1 for all tests. The results are shown in Table 1. As we can see, the Deep-All works better than many previous works, as Li *et al.* [17] mentioned before.

Nevertheless, the proposed method improves the performance of the Deep-All baseline. Furthermore, the proposed method shows better performance than the methods using MMD contraints [9, 22]. However, the improvement is quite marginal and sub-optimal compared to state-of-the-art, which indicates the proposed regularization is quite strong.

Src.	Trgt.	D-MTAE [9]	DSN [3]	CrossGrad [29]	DICA [22]	DANN [8]	TF-CNN [15]	MetaReg [1]	MLDG [16]	Epi-FCR [17]	JiGen [4]	Deep-All	FDM
C,P,S	A	60.3	61.1	61.0	64.6	63.2	62.9	63.5	66.2	64.7	**67.6**	60.4	61.7
A,P,S	C	58.7	66.5	67.2	64.5	67.5	67.0	69.5	66.9	72.3	71.7	70.2	**73.1**
A,C,S	P	91.1	83.3	87.6	**91.8**	88.1	89.5	87.4	88.0	86.1	89.0	85.3	87.4
A,C,P	S	47.9	58.6	55.9	51.1	57.0	57.5	59.1	59.0	65.0	**65.2**	61.6	62.3
Ave.		64.5	67.4	67.9	68.0	69.0	69.2	69.9	70.0	72.0	**73.4**	69.4	71.1

Table 1: Cross-domain object classification results (accuracy. %) on PACS using AlexNet.

Src.	Trgt.	D-MTAE [9]	LRE-SVM [33]	CrossGrad [29]	DICA [22]	DANN [8]	CCSA [21]	MetaReg [1]	MLDG [16]	Epi-FCR [17]	Deep-All ([17])	Deep-All	FDM
L,C,S	V	63.9	60.6	65.5	63.7	66.4	67.1	65.0	67.7	67.1	65.4	59.7	59.3
V,C,S	L	60.1	59.7	60.0	58.2	64.0	62.1	60.2	61.3	64.3	60.6	54.1	54.3
V,L,S	C	89.1	88.1	92.0	79.7	92.6	92.3	92.3	94.4	94.1	93.1	87.7	88.0
V,L,C	S	61.3	54.9	64.7	61.0	63.6	59.1	64.2	65.9	65.9	65.8	61.4	64.0
Ave.		68.6	65.8	70.5	65.7	71.7	70.2	70.4	72.3	72.9	71.2	65.7	66.4

Table 2: Cross-domain object classification results (accuracy. %) on VLCS using AlexNet.

5. Future Works

We are planning to improve the proposed method further. According to the results, the proposed regularization seems too hard. Specifically, the proposed constraint prevents the model from learning more semantic features. Therefore, We are planning to find a way to soften or stabilize the regularization. Additionally, we will conduct more experiments and analysis using other datasets like VLCS [6] and other models such as ResNet [12].

6. Conclusion

We suggested a simple but effective regularization method called Class-wise Feature Distribution Matching Regularization. The experiments on the PACS dataset showed the proposed method can improve the performance of the baseline. Additionally, even though the FDM is sub-optimal compared to state-of-the-art, FDM is comparable or better than the majority of the previous works.

References

[1] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In *NeurIPS*, 2018.

[2] Hakan Bilen and Andrea Vedaldi. Universal representations: The missing link between faces, text, planktons, and cat breeds. Technical report, 2017.

[3] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *NeurIPS*, 2016.

[4] Fabio Maria Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, 2019.

[5] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain specific batch normalization for unsupervised domain adaptation. In *CVPR*, 2019.

[6] Chen Fang, Ye Xu, and Daniel N. Rockmore. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *ICCV*, 2013.

[7] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.

[8] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *JMLR*, 2016.

[9] Muhammad Ghifary, W. Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *ICCV*, 2015.

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[11] Jiang Guo, Darsh Shah, and Regina Barzilay. Multi-source domain adaptation with mixture of experts. In *EMNLP*, 2018.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[13] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei Efros, and Antonio Torralba. Undoing the damage of dataset bias. In *ECCV*, 2012.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolution neural networks. In *NeurIPS*, 2012.

[15] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, 2017.

[16] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018.

[17] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M. Hospedales. Episodic training for domain generalization, 2019.

[18] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, 2018.

[19] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE TPAMI*, 2017.

[20] Saeid Motiian, Marco Piccirilli, Donald A. Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *ICCV*, 2017.

[21] Saeid Motiian, Marco Piccirilli, Donald A. Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *ICCV*, 2017.

Src.	Trgt.	CrossGrad [29]	DANN [8]	MetaReg [1]	MLDG [16]	Epi-FCR [17]	Deep-All	FDM
P,C,S	A	78.7	81.3	79.5	79.5	82.1	77.3	77.7
P,A,S	C	73.3	73.8	75.4	77.3	77.0	73.8	74.2
A,C,S	P	94.0	94.0	94.3	94.3	93.9	94.1	94.3
P,A,C	S	65.1	74.3	72.2	71.5	73.0	70.8	71.7
Ave.		77.8	80.8	80.4	80.7	81.5	79.0	79.5

Table 3: Cross-domain object classification results (accuracy. %) on PACS using ResNet-18.

[22] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, 2013.

[23] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2018.

[24] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. In *NeurIPS*, 2017.

[25] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Efficient parametrization of multi-domain deep neural networks. In *CVPR*, 2018.

[26] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE TPAMI*, 2017.

[27] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Adversarial dropout regularization. In *ICLR*, 2018.

[28] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018.

[29] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. In *ICLR*, 2018.

[30] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *JSPI*, 2000.

[31] Yongxin Yang and Timothy M. Hospedales. A unified perspective on multi-domain and multi-task learning. In *ICLR*, 2015.

[32] Han Zhao, Shanghang Zhang, Guanhang Wu, José M. F. Moura, Joao P Costeira, and Geoffrey J Gordon. Adversarial multiple source domain adaptation. In *NeurIPS*. 2018.

[33] Li Niu Zheng Xu, Wen Li and Dong Xu. Exploiting low-rank structure from latent domains for domain generalization. In *ECCV*, 2014.