Domain Generalization vs Data Augmentation: An Unbiased Perspective

Domain Generalization vs Data Augmentation: an Unbiased Perspective

Francesco Cappio Borlino[1], Antonio D'Innocente[2,3], and Tatiana Tommasi[1,3]

[1]Politecnico di Torino, Turin, Italy francesco.cappio@polito.it, tatiana.tommasi@polito.it
[2]University of Rome Sapienza, Rome, Italy dinnocente@diag.uniroma1.it
[3]Italian Institute of Technology, Turin, Italy

Abstract. In domain generalization the target domain is not known at training time. We show that a style transfer based data augmentation strategy can be implemented easily and outperforms the current state of the art domain generalization methods. Moreover, we observe that those methods, even if combined with the described data augmentation, do not take advantage of it, indicating the need of new generalization solutions.

Keywords: Domain Generalization · Data augmentation · Style transfer

1 Introduction

Domain Generalization (DG) research develops algorithms that are robust to domain shifts with the objective of obtaining good performance on a target domain that is not known at training time. Most of the existing DG strategies try to incorporate the observed data invariances, capturing them at feature [6] or model (meta-learning [5] and self-supervision [11]) level, in the hypothesis that analogous invariances hold for future test domains. An alternative solution consists in extending the source domains by synthesizing new images and including a larger variability in the training set. Some methods do this through generative models which are often difficult to train, but give quite effective results [14]. Still we noticed that newly introduced feature and model-based DG approaches avoid benchmarks against data augmentation strategies [10,3], probably considering them unfair competitors due to the extended training set.

We believe that the field needs some clarification and we dedicate our work on this topic. Specifically our main contributions are: **(1) The proposal of a simple and effective style transfer data augmentation approach for domain generalization** based on AdaIN [2]. **(2) The design of tailored strategies to integrate style transfer data augmentation with the current state of the art methods**. We show that the original advantage of those techniques almost always disappears when compared with the data augmented baseline. This suggests the need of rethinking domain generalization baselines. On one side simple data augmentation strategies should be envisaged to increase source data variability compatible with orthogonal feature and model generalization approaches. On the other, new cross-source adaptive strategies should be designed to build over images generated by style transfer.

2 Source Augmentation by Style Transfer

We focus on multi-source DG. Our strategy consists in the following two steps.

Training the Style Transfer Model. We use *AdaIN* [2] that allows style transfer in real time, by taking the style from a *style image* and applying it over a *content image*. We train AdaIN on source data: all the train splits of the source domains are used together both as *content dataset* and *style dataset*.

Training the Classification Model on the augmented source data. A standard classification model (AlexNet or ResNet18) is trained on the source domains, exploiting the style transfer model to apply data augmentation. For each image of a training minibatch, we decide with probability p if we want to apply the style transfer. In this case we randomly choose another image of the minibatch and borrow its style. The content image is then substituted with its augmented version. Considering that each batch contains equal parts coming from the different source domains, we obtain a high variability in image styles.

3 Experiments

Datasets. We consider three standard benchmark datasets which differ in number of classes and covered domains. **PACS [4]** contains images of 7 object classes spanning 4 visual domains: Photo, Art Painting, Cartoon, Sketch. **OfficeHome [9]** is similar to PACS, it covers 4 domains (Art, Clipart, Product and Real-World) but shows a much larger set of 65 object classes. **VLCS [8]** is built upon 4 different datasets: PASCAL VOC 2007, Labelme, Caltech and SUN and contains 5 object categories. All its domains are composed of real world photos with the shift mainly due to camera type, illumination conditions, point of view, *etc*. For all our experiments we used the same experimental protocols described in [1], train splits for model training and validation splits for model selection. All our results are average performance over 3 runs.

Reference methods. We consider as main *Baseline* a classification model learned on all the source data and naïvely applied on the target. We indicate with *Original* the standard data augmentation with horizontal flippling and random cropping, while we use *Stylized* to specify the cases where we add style transfer data augmentation. The behavior of four among the most recent DG methods is evaluated under both these augmentation settings. We integrate the style transfer in each of the considered approaches without undermining their nature: we carefully avoid to mix domains when methods require to access them separately. **DG-MMLD [7]** does not need the source domain labels, thus the style transfer data augmentation is applied exactly as done for the Baseline. **Epi-FCR [5]** is a meta-learning method which splits the network in two modules, each one is trained by pairing it with a partner that is badly tuned for the domain considered in the current learning episode. Since the network is also trained on all source data to build the classification ability, the style transfer data augmentation is applied here and not in the previous step. **DDAIG** [14] is a data augmentation strategy that uses a generator to produce augmented

Table 1. PACS classification accuracy (%). We used AdaIN with $p = 0.75$ for AlexNet-based experiments and $p = 0.90$ for those based on ResNet18.

		Painting	Cartoon	Sketch	Photo	Average
AlexNet						
Original	Baseline	66.83	70.85	59.75	89.78	71.80
	Rotation	65.66	71.89	62.15	89.88	72.39
	DG-MMLD	69.27	72.83	66.44	88.98	74.38
	Epi-FCR	64.70	72.30	65.00	86.10	72.03
	DDAIG*	62.77	67.06	58.90	86.82	68.89
Stylized	Baseline	71.96	72.47	76.47	88.34	77.31
	Rotation	71.74	73.39	75.98	89.22	77.59
	DG-MMLD	70.50	70.84	75.39	88.43	76.29
	Epi-FCR	65.19	69.54	71.97	83.43	72.53
	DDAIG	69.35	71.10	70.99	87.70	74.79
Mixup	pixel-level	66.03	68.00	51.18	88.90	68.53
	feature-level	67.04	69.10	55.40	88.88	70.11
ResNet18						
Original	Baseline	77.28	73.89	67.01	95.83	78.50
	Rotation	78.16	76.64	72.20	95.57	80.64
	DG-MMLD	81.28	77.16	72.29	96.06	81.83
	Epi-FCR	82.10	77.00	73.00	93.90	81.50
	DDAIG*	79.41	74.81	69.29	95.22	79.68
Stylized	Baseline	82.73	77.97	81.61	94.95	84.32
	Rotation	79.51	79.93	82.01	93.55	83.75
	DG-MMLD	80.85	77.10	77.69	95.11	82.69
	Epi-FCR	80.68	78.87	76.57	92.50	82.15
	DDAIG	81.02	78.75	79.67	95.07	83.63
Mixup	pixel-level	78.09	71.08	66.58	93.85	77.40
	feature-level	81.20	76.41	69.67	96.31	80.90

Table 2. OfficeHome classification accuracy (%). We used AdaIN with $p = 0.1$.

ResNet18		Art	Clipart	Product	Real World	Average
Original	Baseline	57.14	46.96	73.50	75.72	63.33
	Rotation	55.94	47.26	72.38	74.84	62.61
	DG-MMLD*	58.08	49.32	72.91	74.69	63.75
	Epi-FCR*	53.34	49.66	68.56	70.14	60.43
	DDAIG*	57.79	48.32	73.28	74.99	63.59
Stylized	Baseline	58.71	52.33	72.95	75.00	64.75
	Rotation	57.24	52.15	72.33	73.66	63.85
	DG-MMLD	59.24	49.30	73.56	75.85	64.49
	Epi-FCR	52.97	50.14	67.03	70.66	60.20
	DDAIG	58.21	50.26	73.81	74.99	64.32
Mixup	feature-level	58.33	39.76	70.96	72.07	60.28

Table 3. VLCS classification accuracy (%). We used AdaIN with $p = 0.75$.

AlexNet		CALTECH	LABELME	PASCAL	SUN	Average
Original	Baseline	94.89	59.14	71.31	64.64	72.49
	Rotation	94.50	61.27	68.94	63.28	72.00
	DG-MMLD*	96.94	59.10	68.48	62.06	71.64
	Epi-FCR*	91.43	61.36	63.44	60.07	69.07
	DDAIG*	95.75	60.18	65.48	60.78	70.55
Stylized	Baseline	96.86	60.77	68.18	63.42	72.31
	Rotation	96.86	60.77	68.18	63.42	72.31
	DG-MMLD	97.49	61.02	64.23	62.37	71.28
	Epi-FCR	92.69	58.18	62.59	57.87	67.83
	DDAIG	97.48	60.48	65.19	62.57	71.43
Mixup	feature-level	94.73	62.15	69.82	62.98	72.42

samples. In this method the label classifier is trained on all the source data, both original and synthetic: we further extended this set with style transfer augmented data. **Rotation [11]** exploits self-supervised learning: rotation recognition is combined with classification in a multi-task model. Once again the style transfer data augmentation application is trivial because no domain labels are used. We also experiment with *Mixup* [13] as an alternative to AdaIN for interpolation of source data. We tested data mixing both at pixel and at feature level [12].
We implemented the *Baseline*, *Rotation* and *Mixup*, while we used for the others the code provided by the authors. We report the previously published results whenever possible. We will indicate with a star (*) the results we obtained by running the authors' code.

Results analysis. Table 1 shows results on PACS. We get two main outcomes. (1) There is an evident improvement in the Baseline performance when using the stylized augmented source data with respect to the original case. (2) All the considered state of the art DG methods benefit from the source augmentation. Indeed in absolute terms their performance grows, but at the same time they lose in effectiveness as they cannot outperform the Baseline any more. Table 2 shows results on OfficeHome. Even if in this case the improvement produced by the source augmentation by style transfer is more limited, the results confirm what already observed for PACS. Table 3 reports results on VLCS. This dataset is particularly challenging and shows a fundamental limit of tackling DG through style transfer data augmentation. Since the domain shift is not originally due to style differences, source augmentation by style transfer does not support generalization. Finally, the results of Mixup show that it is not able to generalize across domains and it might perform even worse than the Original Baseline.

Only the feature variant shows some advantage on PACS, so we focused on it in the other tests. Still, its results remain lower than those obtained by the DG methods both with and without style based data augmentation.

4 Conclusions

Among current DG methods some are based on data augmentation and use complex generative approaches, while other propose source feature adaptation and meta-learning strategies. Despite being orthogonal among each other, no previous work tried to integrate them. We investigated here a simple and effective style transfer data augmentation strategy for DG, showing how it overcomes its competitors. However when combined with the the most relevant existing DG approaches they lose their original effectiveness, not producing any improvement over the new data augmented baseline. Our work suggests the need of a shading new light on DG problems and calls for novel strategies able to take advantage of the data variability introduced by cross-domain style transfer.
Acknowledgements Computational resources provided by hpc@polito: (http://hpc.polito.it).

References

1. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: CVPR (2019)
2. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: ICCV (2017)
3. Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: ECCV (2020)
4. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: ICCV (2017)
5. Li, D., Zhang, J., Yang, Y., Liu, C., Song, Y.Z., Hospedales, T.M.: Episodic training for domain generalization. In: ICCV (2019)
6. Li, H., Jialin Pan, S., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: CVPR (2018)
7. Matsuura, T., Harada, T.: Domain generalization using a mixture of multiple latent domains. In: AAAI (2020)
8. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR (2011)
9. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: CVPR (2017)
10. Wang, H., Ge, S., Lipton, Z., Xing, E.P.: Learning robust global representations by penalizing local predictive power. In: NeurIPS (2019)
11. Xu, J., Xiao, L., López, A.M.: Self-supervised domain adaptation for computer vision tasks. IEEE Access **7**, 156694–156706 (2019)
12. Xu, M., Zhang, J., Ni, B., Li, T., Wang, C., Tian, Q., Zhang, W.: Adversarial domain adaptation with domain mixup. In: AAAI (2020)
13. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: ICLR (2018)
14. Zhou, K., Yang, Y., Hospedales, T., Xiang, T.: Deep domain-adversarial image generation for domain generalisation. AAAI (2020)