MetaAlign: Coordinating Domain Alignment and Classification for Unsupervised Domain Adaptation (Supplementary Material)

Guoqiang Wei[1] Cuiling Lan[2] Wenjun Zeng[2] Zhibo Chen[1†]
[1] University of Science and Technology of China [2] Microsoft Research Asia
wgq7441@mail.ustc.edu.cn {culan,wezeng}@microsoft.com chenzhibo@ustc.edu.cn

1. More Details of Baselines

In our main manuscript, we have briefly described several representative alignment-based methods, which we use as our baselines for validating the effectiveness of our MetaAlign. Here, we present more details of some baselines.

DANNPE. As shown in Fig. 1, ***DANNPE*** differs from ***DANN*** in two key aspects: 1) Similar to [20, 5], the predicted object classification probability/likelihood $C(G(\cdot)) \in \mathbb{R}^K$ is treated as the input of domain discriminator D, instead of the output feature of $G(\cdot)$ in *DANN*. 2) Following [20], we prioritize the discriminator on those easy-to-transfer samples by re-weighting the samples based on the entropy of object class prediction, with the weight defined as $\omega(ent(\cdot)) = e^{-ent(\cdot)}$, where $ent(\cdot)$ denotes the entropy of the object class prediction. As shown in Table 1 in our main manuscript, *DANNPE* significantly outperforms *DANN*.

MMD. We directly add the MMD constraint [2] on the output of $G(\cdot)$ to encourage the feature alignment between source domain and target domain data (see Fig. 2 (b) in our main manuscript). The complete MMD loss (*i.e.*, Eq. (4) in the main manuscript) is formulated as:

$$\mathcal{L}_{dom} = \frac{1}{N_s}\sum_{i=1}^{N_s}\sum_{i'=1}^{N_s}\mathcal{K}(\mathbf{f}_i^s, \mathbf{f}_{i'}^s) + \frac{1}{N_t}\sum_{j=1}^{N_t}\sum_{j'=1}^{N_t}\mathcal{K}(\mathbf{f}_j^t, \mathbf{f}_{j'}^t) - \frac{2}{N_s N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t}\mathcal{K}(\mathbf{f}_i^s, \mathbf{f}_j^t). \tag{1}$$

where $\mathbf{f}_i = G(\mathbf{x}_i)$, and $\mathcal{K}(\mathbf{f}, \mathbf{f}')$ denotes a kernel function. Following [19], we use the well-known characteristic kernel RBF, *i.e.*, $\mathcal{K}(\mathbf{f}, \mathbf{f}') = \exp\left(-\frac{1}{2\sigma}||\mathbf{f} - \mathbf{f}'||^2\right)$, where σ is the bandwidth parameter [19].

For MMD-based UDA, similar to Eq. (8) in the main manuscript, the optimization objective of MetaAlign is:

$$\min_{\theta,\phi_c,\beta} \mathcal{L}_{dom}(\theta) + \mathcal{L}_{cls}\left(\{\theta_m - \alpha\beta_m\nabla_{\theta_m}\mathcal{L}_{dom}(\theta, \phi_d)\}_{m=1}^{M}, \phi_c\right) + \mathcal{L}_{\beta}(\beta). \tag{2}$$

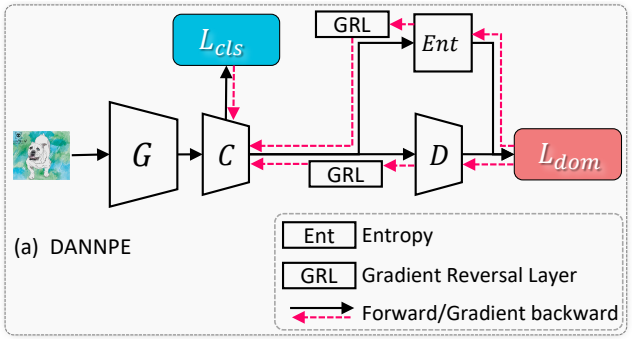

Figure 1: The pipeline of DANNPE.

2. Experiments

We describe more details on the implementation, datasets, settings, competitors, and present more experimental results.

2.1. UDA for Classification

Implementation Details. We adopt ResNet-50 [10] pretrained on ImageNet [15] as the feature extractor for all baselines. Following [5, 20], the domain classifier/discriminator is composed of three fully connected layers with inserted dropout and ReLU layers for stable training, followed by a sigmoid function to output the domain classification result. We divide the convolutional layers of the feature extractor G into 4 groups (*i.e.*, $M = 4$ in Eq. (8)): the conv1 and conv2_x as the first group, conv3_x, conv4_x, conv5_x as the second to fourth groups respectively for simplicity.

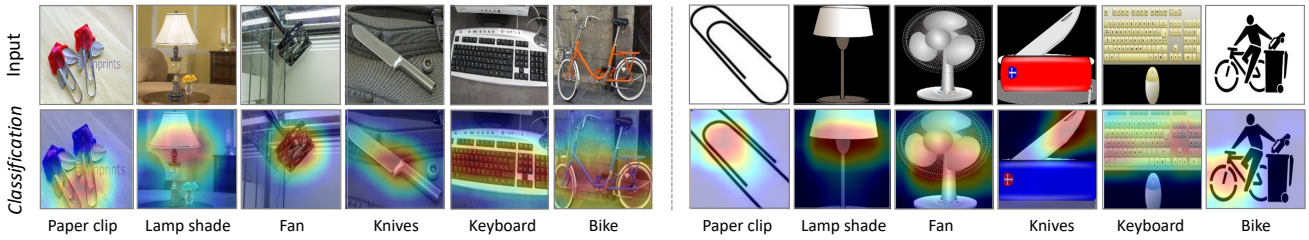

Figure 2: Visualization fo the Grad-CAMs [24] w.r.t. the object classification task. The first row of left/right panels show the samples from source (Rw)/target (Cl) domains on Office-Home, while the second and third rows show the Grad-CAMs. The object classification task always focuses on foreground objects, which is also claimed in [24, 26]

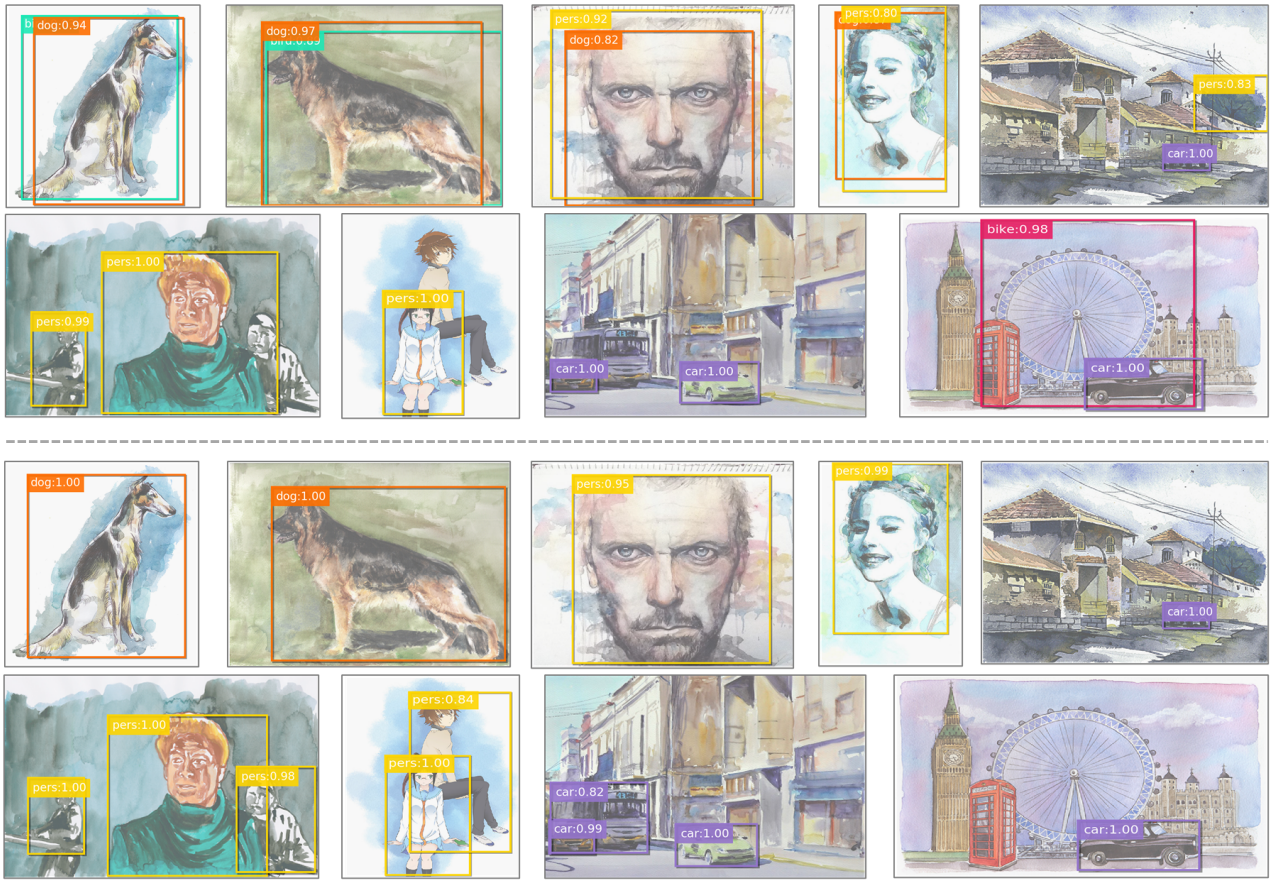

Figure 3: Object detection results on the target dataset Watercolor2k from (a) SW-DA (Baseline) (first two rows), and (b) SW-DA+MetaAlign (last two rows).

Grad-CAMs of Classification Task. We illustrate the Grad-CAMs [24] w.r.t. object classification task in Fig. 2. As can be seen, the object classification task always focuses on the foreground objects, which is also validated in [24, 26].

2.2. UDA for Object Detection

Datasets and Experimental Setting. To simulate dissimilar domains, Pascal VOC [7, 8] and Watercolor2k [13] are treated as source and target domain respectively. 1) **Pascal VOC** [7, 8] is a well-known benchmark for object detection in real world scenario. In this dataset, 20 object classes with their corresponding bounding boxes are annotated. Following [23], we employ the split setting which uses Pascal VOC 2007 and 2012 as training and validation. 2) **Watercolor2k** [13] is a collection of 2K watercolor images. It contains 6 categories in common with Pascal VOC. 1K images are used for training and the other 1K for testing.

As in previous works [4, 23], we set the shorter side of the image to 600 pixels following the implementation of

Faster RCNN[22] with ROI-alignment [9]. The meta learning rate α is set to 0.01, which is 10 times the learning rate η.

Competitors. We compare with the following methods: 1) **Source Only** trains model on source domain and directly tests on target domain. 2) **BDC-Faster** adopts the typical design of DANN, which takes the global features as input of the domain discriminator D for adversarial learning. 3) **WST+BSR** [14] constructs self-training on easy samples to reduce the negative effects of inaccurate pseudo-labels. 4) **MAF** [11] incorporates multiple domain discriminators on hierarchical features. 5) **DT-UDA** [13] performs training on style-translated target images with predicted pseudo-labels. 6) **ATF** [12] designs an asymmetric tri-way model to alleviate the collapse and out-of-control risk of the source domain. 7) **SW-DA** [23] aligns both global-level features and local-level features between the source and target domains by adversarial learning, which we take as our baseline for evaluating MetaAlign.

Visualization Results. We have shown the performance comparison in Table 5 in our main manuscript. Here, we show the visualization of object detection results on the target dataset Watercolor2k [13] in Fig. 3. We can see that for the baseline scheme SW-DA, there are many false detections and missing detections. Thanks to the coordination between the domain alignment and the object detection optimization from our MetaAlign, the scheme SW-DA+MetaAlign achieves more accurate detections, where the false detections and missing detections are largely reduced.

2.3. Domain Generalization

Dataset and Settings. PACS [16] is a widely used benchmark for domain generalization. It contains 7 object categories from 4 domains (Photo, Art Painting, Cartoon and Sketch). We evaluate on this dataset under a commonly-used experimental protocol of leave-one-out [16, 3, 18], where three domains are used for training and the remaining one is considered as the target domain. The domain discriminator D of DANNPE here is kept the same as that for UDA classification, except that the final layer is a FC layer with 3 neurons instead of 1 for distinguishing the three source domains.

Competitors. 1) ***AGG*** simply trains a model directly on the aggregation of all source domains. 2) ***MMD-AAE***[19] equips an autoencoder with a MMD loss to train a domain-invariant encoder. 3) ***CrossGrad***[25] is a typical data augmentation based DG method which perturbs in the input manifold to augment data. 4) ***MetaReg***[1], 5) ***MLDG***[17] and 6) ***MASF***[6] utilize meta-learning, which separate the samples into meta-train splits and meta-test splits, to mimic domain shift during training on source domains. 7) ***JiGen*** imposes an auxiliary task of solving the Jigsaw puzzle on top of *AGG*. 8) ***Epi-FCR***[18] introduces a new episodic training strategy. 9) ***MMLD***[21] predicts the pseudo domain labels and uses them for the adversarial domain learning.

References

[1] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In *NeurIPS*, pages 998–1008, 2018. 3

[2] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006. 1

[3] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, pages 2229–2238, 2019. 3

[4] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *CVPR*, pages 3339–3348, 2018. 2

[5] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Chi Su, Qingming Huang, and Qi Tian. Gradually vanishing bridge for adversarial domain adaptation. In *CVPR*, pages 12455–12464, 2020. 1

[6] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. In *NeurIPS*, pages 6450–6461, 2019. 3

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html. 2

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 2

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *ICCV*, pages 2980–2988, 2017. 3

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1

[11] Zhenwei He and Lei Zhang. Multi-adversarial faster-rcnn for unrestricted object detection. In *ICCV*, pages 6668–6677, 2019. 3

[12] Zhenwei He and Lei Zhang. Domain adaptive object detection via asymmetric tri-way faster-rcnn. *ECCV*, 2020. 3

[13] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *CVPR*, pages 5001–5009, 2018. 2, 3

[14] Seunghyeon Kim, Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In *ICCV*, pages 6092–6101, 2019. 3

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NeurIPS*, page 1097–1105, 2012. 1

[16] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, pages 5542–5550, 2017. 3

[17] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. *AAAI*, 2018. 3

[18] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *ICCV*, pages 1446–1455, 2019. 3

[19] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, pages 5400–5409, 2018. 1, 3

[20] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, pages 1645–1655, 2018. 1

[21] Toshihiko Matsuura and Tatsuya Harada. Domain generalization using a mixture of multiple latent domains. In *AAAI*, pages 11749–11756, 2020. 3

[22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, pages 91–99, 2015. 3

[23] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *CVPR*, pages 6956–6965, 2019. 2, 3

[24] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *ICCV*, pages 618–626, 2017. 2

[25] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. In *ICLR*, 2018. 3

[26] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *CVPR*, 2016. 2