Supplementary of "Geometric Anchor Correspondence Mining with Uncertainty Modeling for Universal Domain Adaptation"

Liang Chen [1], Yihang Lou [2], Jianzhong He [2*], Tao Bai [2], Minghua Deng [1†]
[1] School of Mathematical Sciences, Peking University, [2] Intelligent Vision Dept, Huawei Technologies

1. Dataset split details

Office [7]. Following previous work [2], in the PDA setting, we select 10 categories ("backpack", "bike", "calculator", "headphones", "keyboard", "laptop computer", "monitor", "mouse", "mug" and "projector") as the common categories between the two domains, and the remaining 21 categories are used as source private categories. In the ODA setting, the same 10 categories are used as the common categories. Then we select other 11 categories ("tape dispenser", "ring binder", "stapler", "scissors", "punchers", "speaker", "pen", "trash can", "phone", "ruler" and "printer") as the target private categories. This setting is the same as [10]. In the OPDA setting, similar to [1], the same 10 categories are used as the common categories, and then, in alphabetical order, the next 10 categories are used as the source private categories, and the remaining 11 categories are used as the target private categories.

OfficeHome [11]. Following previous work [1], we use the same separation of categories in three DA settings. In the PDA setting, in alphabetical order, we select the first 25 categories as the common categories and use the remaining 40 categories as the source private categories. In the ODA setting, we select the first 25 categories as the common categories and use the rest 40 categories as the target private categories. In the OPDA setting, we select the first 10 categories as the common categories, the next 5 categories as the source private categories, and the remaining 50 categories as the target private categories.

VisDA [6]. We also use the same class split on VisDA as in previous work [1]. In the PDA setting, in alphabetical order, we select the first 6 categories as the common categories and use the remaining 6 categories as the source private categories. In the ODA setting, in alphabetical order, we use the first 6 categories as the common categories and the remaining 6 categories as the target private categories. In the OPDA setting, in alphabetical order, we select the first 6 categories as the common categories, the next 3 categories as the source private categories, and the remaining 3 categories as the target private categories.

ImageCLEF [4] & ExDark [3]. There are 8 common categories between these two datasets and 4 private categories for each dataset. The 8 common categories are "Bicycle", "Boat", "Bottle", "Bus", "Car", "Dog", "Motorbike" and "People", respectively. The 4 private categories of ImageCLEF are "Airplane", "Bird", "Horse" and "Monitor", respectively, while the 4 private categories of ExDark are "Cat", "Chair", "Cup" and "Table", respectively.

2. Implementation details

Network architecture. We denote $fc-k$ as a fully-connected layer with k-dimensional output. $drop$ represents the dropout layer and GRL refers to the gradient reverse layer. $relu$, $sigmoid$ and $softmax$ denote three kinds of activation functions. In our experiments, we apply ResNet50 as feature extractor $\mathcal{F}$ to obtain embedding feature z. The domain discriminator $\mathcal{D}$ inputs the feature z and outputs the domain predictions, i.e., $z \rightarrow fc-256 \rightarrow GRL \rightarrow fc-1024 \rightarrow relu \rightarrow drop \rightarrow fc-1024 \rightarrow relu \rightarrow drop \rightarrow fc-1 \rightarrow sigmoid$. The universal classifier $\mathcal{G}$ inputs the feature z and generates label predictions in C_s+1 categories, i.e., $z \rightarrow fc-256 \rightarrow fc-C_s+1 \rightarrow softmax$. Note that the $fc-256$ in $\mathcal{D}$ and $\mathcal{G}$ shares parameters.

Update strategy of the hybrid memory bank $\mathcal{Z}$. The bank $\mathcal{Z}$ contains both updated source and target features from the current mini-batch and the older features absent in the mini-batch. We update $\mathcal{Z}$ so that it simply stores features and provides global information, without utilizing the exponential moving average of features in previous epochs.

Training strategy. To handle the distribution shift in all datasets progressively, we set the initial 500 iterations as the warm-up stage where only the $\mathcal{L}_{sup}$ and $\mathcal{L}_{\mathcal{D}}^{rw}$ losses are applied on the source and target samples. After that, we add the $\mathcal{L}_{Contra}^{rw}$ and $\mathcal{L}_{ent}$ losses into training.

3. Supplemental results

CDA setting. Due to limited space in the text, we put the tested performance of each method in the CDA set-

*This work is completed in Huawei Technologies.
†Corresponding Author.

Table 1. Accuracy comparison in the CDA setting. Some results for previous methods are cited from DANCE [8] and DCC [1].

		Office (31/0/0)							OfficeHome (65/0/0)													VisDA (12/0/0)
Methods	**Type**	A2W	A2D	D2W	W2D	D2A	W2A	Avg	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg	S2R
CDAN	C	93.1	89.8	98.2	**100.0**	70.1	68.0	86.6	49.0	69.3	74.5	54.4	66.0	68.4	55.6	48.3	75.9	68.4	55.4	80.5	63.8	70.0
MDD	C	94.5	93.5	98.4	**100.0**	74.6	72.2	88.9	**54.9**	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	**60.2**	82.3	68.1	74.6
SRDC	C	**95.7**	95.8	99.2	**100.0**	**76.7**	**77.1**	**90.8**	52.3	76.3	**81.0**	**69.5**	**76.2**	**78.0**	**68.7**	53.8	**81.7**	**76.3**	57.1	**85.0**	**71.3**	**81.9**
UAN	U	86.5	**97.0**	**100.0**	84.5	69.6	68.7	84.4	45.0	63.6	71.2	51.4	58.2	63.2	52.6	40.9	71.0	63.3	48.2	75.4	58.7	66.4
CMU	U	79.6	78.3	98.1	97.6	62.3	63.4	79.9	42.8	65.6	74.3	58.1	63.1	67.4	54.2	41.2	73.8	66.9	48.0	78.7	61.2	56.9
DANCE	U	88.6	89.4	97.5	**100.0**	69.5	68.2	85.5	54.3	75.9	78.4	64.8	72.1	73.4	63.2	53.0	79.4	73.0	58.2	82.9	69.1	70.2
DCC	U	89.1	87.2	96.8	**100.0**	74.4	76.8	87.4	35.4	61.4	75.2	45.7	59.1	62.7	43.9	30.9	70.2	57.8	41.0	77.9	55.1	69.3
OVANet	U	67.3	72.5	94.8	99.6	43.4	44.9	70.4	34.5	55.8	67.1	40.9	52.8	56.9	35.4	26.2	61.8	53.8	35.4	70.8	49.3	38.5
GATE	U	90.5	91.3	98.7	**100.0**	73.4	75.9	88.3	54.6	**76.9**	79.8	66.1	73.5	74.2	65.3	**54.8**	80.6	73.9	59.5	83.7	70.2	74.8

ting here. The results in Table 1 show that GATE outperforms other state-of-the-art UniDA methods on three datasets. Even compared to those methods specialized in CDA setting, GATE achieves comparable performance to some of them, such as only inferior to SRDC on Office, OfficeHome and VisDA datasets. But such methods customized for the CDA setting cannot adapt to the situations where "unknown" samples exist, thereby limiting their application in real-world scenarios.

OPDA setting on large-scale DomainNet dataset. DomainNet [5] is the largest domain adaptation dataset so far, with about 600K images covering 345 categories. Similar to DCC [1] and OVANet [9], we perform the OPDA experiment on three sub-domains in it, i.e., Painting (P), Real (R) and Sketch (S). From the results in Table 2, GATE yields consistent improvement over previous UniDA methods, verifying its effectiveness on large-scale DA dataset.

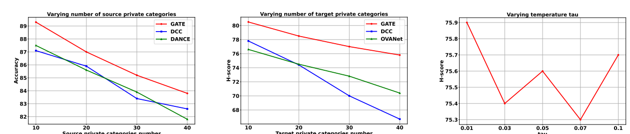
Table 2. H-score comparison on DomainNet dataset under the OPDA setting. Some results for previous methods are cited from DCC [1] and OVANet [9].

		DomainNet (150/50/145)						
Methods	**Type**	P2R	R2P	P2S	S2P	R2S	S2R	Avg
UAN	U	41.9	43.6	39.1	39.0	38.7	43.7	41.0
CMU	U	50.8	52.2	45.1	44.8	45.6	51.0	48.3
DANCE	U	55.7	51.1	47.0	47.9	46.4	55.7	50.6
DCC	U	56.9	50.3	43.7	44.9	43.3	56.2	49.2
OVANet	U	56.1	51.9	47.9	47.7	45.1	**56.4**	50.9
GATE	U	**57.4**	**52.8**	**48.7**	**49.5**	**47.6**	56.3	**52.1**

Number of source private categories. We compare the behavior of GATE with DANCE and DCC under the different number of source private categories in the PDA setting. In this analysis, we use "R2P" in OfficeHome to conduct experiments, where there are 25 common categories between two domains. We vary the category number present only in the source domain from 10 to 40. The accuracy result is shown in Figure 1a. With the appearance of more unshared private categories in the source domain, the performance of the three methods degrades. However, GATE consistently outperforms DANCE and DCC, indicating that it is robust to the change of source private category number.

Number of target private categories. We also analyze the behavior of GATE under the different "unknown" categories number. Here we perform the ODA experiment on the "A2R" task in OfficeHome, which has 25 shared categories. We increase the number of "unknown" categories only in the target domain from 10 to 40. Figure 1b shows the H-score comparison among three methods. As we add more target private categories, the H-score of all methods decreases. However, GATE consistently performs better than DCC and OVANet, validating its stability with respect to the "unknown" categories number.

Hyperparameter sensitivity of temperature τ. We also perform control experiments for temperature parameter τ in contrastive learning. We use the OfficeHome dataset under the OPDA setting to conduct this analysis. For τ in Figure 1c, within a wide range in $[0.01, 0.1]$, the H-score changes no more than 1%, showing that GATE is robust to the selection of τ.



(a) Real-World to Product (b) Art to Real-World (c) Temperature τ

Figure 1. Various case studies, including source and target private categories number, and temperature τ.

References

[1] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9757–9766, 2021. 1, 2

[2] Jian Liang, Yunbo Wang, Dapeng Hu, Ran He, and Jiashi Feng. A balanced and uncertainty-aware approach for partial domain adaptation. *In Computer Vision–ECCV 2020: 16th European Conference*, pages 123–140, 2020. 1

[3] Yuen Peng Loh and Chee Seng. Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 1

[4] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Deep transfer learning with joint adaptation networks. *In International conference on machine learning*, pages 2208–2217, 2017. 1

[5] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 1406–1415, 2019. 2

[6] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 1

[7] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. *In Euro-*

pean conference on computer vision, pages 213–226, 2010. 1

[8] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Advances in Neural Information Processing Systems*, 33, 2020. 2

[9] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9000–9009, 2021. 2

[10] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. *In Proceedings of the European Conference on Computer Vision*, pages 153–168, 2018. 1

[11] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017. 1