PADCLIP: Pseudo-labeling with Adaptive Debiasing in CLIP for Unsupervised Domain Adaptation (Supplementary)

Zhengfeng Lai[1]* Noranart Vesdapunt[2]* Ning Zhou[2] Jun Wu[2]
Cong Phuoc Huynh[2] Xuelu Li[2] Kah Kuen Fu[2] Chen-Nee Chuah[1]
[1]University of California, Davis [2]Amazon
[1]{lzhengfeng, chuah}@ucdavis.edu [2] {solves, ningzho, jwum, conghuyn, xueluli, kahkuen}@amazon.com

1. Appendix

1.1. Full Result Tables

We include tables for VisDA-2017 [20] (Tab. 4), Office-31 [23] (Tab. 7), Office-Home [26] (Tab. 5), DomainNet [19] (Tab. 6) from section 5.2 (External Comparison). Note that PMTrans [36] was available after our submission.

1.2. Generalization of Our Method

Few-shot learning. Our method preserves CLIP's few-shot learning ability because we did not modify CLIP's architecture (no extra layers on top of any representations). To test our method, we setup a fair ablation study in the setting of ResNet-101 on ImageNet 16 shots (no UDA datasets were used). Tab. 1 shows that we outperform few-shot CLIP methods: Tip-Adapter-F [34] (+0.5%), CLIP-Adapter [8] (+3.6%), and CoOP [35] (+2.4%). For UDA, we perform DomainNet 16-shot (8 source, 8 target images) with CFM on CLIP and improve accuracy from zero-shot setting by +2.6% (Tab. 2).

Incremental learning. Our method works on incremental learning. We follow Split CIFAR-100 [1] (20 disjoint subsets and each subset was randomly sampled from 5 classes without a replacement from a total of 100 classes) to evaluate the incremental learning accuracy (average accuracy over each step of adding a subset for 20 subsets). Tab. 3 shows that CLIP suffers from catastrophic forgetting (accuracy -9.9%) and CFM can mitigate it (accuracy +5.3%).

Domain Generalization. Our methods prevent catastrophic forgetting in Domain Generalization. Tab. 8 shows that CLIP suffers from catastrophic forgetting issues (-14.0% accuracy when fine-tuning on all DomainNet domains except "Painting" and test on "Painting") and CFM can mitigate it (accuracy +2.4%).

*Equal contributions. This work was done at Amazon.

1.3. Fine-tuning the whole vs partial network

Fine-tuning partial EVACLIP can achieve a similar effect to lowering the learning rate which can both mitigate catastrophic forgetting. However, fine-tuning partial network also decreases model capacity (many parameters are not learnable), which leads to lower accuracy. Whole model fine-tuning is slower to train, but we chose the whole model to not limit ourselves to the lower model capacity and introduce CFM and DCM to mitigate catastrophic forgetting.

Table 1: Classification accuracy on few-shot ImageNet [3] with ResNet-101.

Methods	Shot	Accuracy
CLIP [21]	0	62.53%
CLIP + CoOp [35]	16	66.60%
CLIP-Adapter [8]	16	65.39%
Tip-Adapter [34]	16	64.78%
Tip-Adapter-F [34]	16	68.56%
Ours	16	69.01%

Table 2: Classification accuracy on few-shot DomainNet [3] (16-shot: 8 source and 8 target images) with ViT-B.

Methods	Accuracy
Zero-shot CLIP [21]	56.6%
Ours: 16-shot CLIP with CFM	59.2 %
Ours: fine-tuned full dataset (for reference)	63.7%

Table 3: Incremental learning accuracy on Split CIFAR-100 [1] with ResNet-50.

Methods	Accuracy
Zero-shot CLIP [21]	84.8%
Fine-tuned without CFM	74.9 %
Ours: Fine-tuned with CFM	90.1%

Table 4: Accuracies (%) on **VisDA-2017**. *CDTrans uses DeiT-base backbone. "-B" indicates ViT-B backbone.

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Avg.
RN-101 [10]	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
DANN [6]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
CDAN [18]	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
SAFN [29]	93.6	61.3	84.1	70.6	94.1	79.0	91.8	79.6	89.9	55.6	89.0	24.4	76.1
SWD [13]	90.8	82.5	81.7	70.5	91.7	69.5	86.3	77.5	87.4	63.6	85.6	29.2	76.4
CaCo [11]	90.4	80.7	78.8	57.0	88.9	87.0	81.3	79.4	88.7	88.1	86.8	63.9	80.9
SUDA [33]	91.5	79.7	71.9	66.5	88.5	81.1	85.6	79.5	86.2	86.5	79.9	74.3	80.9
DTA [14]	93.7	82.2	85.6	83.8	93.0	81.0	90.7	82.0	95.1	78.1	86.4	32.1	81.5
CGDM [5]	93.4	82.7	73.2	68.4	92.9	94.5	88.7	82.1	93.4	82.5	86.8	49.2	82.3
SHOT [16]	94.3	88.5	80.1	57.3	93.1	94.9	80.7	80.3	91.5	89.1	86.3	58.2	82.9
MCC+NWD [2]	96.1	82.7	76.8	71.4	92.5	96.8	88.2	81.3	92.2	88.7	84.1	53.7	83.7
SDAT [22]	95.8	85.5	76.9	69.0	93.5	97.4	88.5	78.2	93.1	91.6	86.3	55.3	84.3
MSGD [28]	97.5	83.4	84.4	69.4	95.9	94.1	90.9	75.5	95.5	94.6	88.1	44.9	84.6
CAN [12]	97.0	87.2	82.5	74.3	97.8	96.2	90.8	80.7	96.6	96.3	87.5	59.9	87.2
AaD [32]	97.4	90.5	80.8	76.2	97.3	96.1	89.8	82.9	95.5	93.0	92.0	64.7	88.0
Ours (RN-101)	96.7	88.8	87.0	82.8	97.1	93.0	91.3	83.0	95.5	91.8	91.5	63.0	**88.5**
ViT-B [4]	99.1	60.7	70.6	82.7	96.5	73.1	97.1	19.7	64.5	94.7	97.2	15.4	72.6
TVT-B [31]	92.9	85.6	77.5	60.5	93.6	98.2	89.4	76.4	93.6	92.0	91.7	55.7	83.9
SHOT-B [30]	97.9	90.3	86.0	73.4	96.9	98.8	94.3	54.8	95.4	87.1	93.4	62.7	85.9
CDTrans* [30]	97.1	90.5	82.4	77.5	96.6	96.1	93.6	88.6	97.9	86.9	90.3	62.8	88.4
SSRT-B [25]	98.9	87.6	89.1	84.8	98.3	98.7	96.3	81.1	94.9	97.9	94.5	43.1	88.8
SDAT-B [22]	98.4	90.9	85.4	82.1	98.5	97.6	96.3	86.1	96.2	96.7	92.9	56.8	89.8
PMTrans [36]	98.9	93.7	84.5	73.3	99.0	98.0	96.2	67.8	94.2	98.4	96.6	49.0	87.5
Ours-B	98.1	93.8	87.1	85.5	98.0	96.0	94.4	86.0	94.9	93.3	93.5	70.2	**90.9**

Table 5: Accuracies (%) on **Office-Home**. *CDTrans uses DeiT-Base backbone. "-B" indicates ViT-B backbone.

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
RN-50 [10]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
CDAN+E [18]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
SAFN [29]	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
CDAN+TN [27]	50.2	71.4	77.4	59.3	72.7	73.1	61.0	53.1	79.5	71.9	59.0	82.9	67.6
FGDA+MDD [9]	57.1	77.5	81.0	68.4	77.2	75.9	65.8	55.8	81.0	74.3	60.5	83.6	71.5
SHOT [16]	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
SDAT [22]	58.2	77.1	82.2	66.3	77.6	76.8	63.3	57.0	82.2	74.9	64.7	86.0	72.2
MSGD [28]	58.7	76.9	78.9	70.1	76.2	76.6	69.0	57.2	82.3	74.9	62.7	84.5	72.4
MCC+NWD [2]	58.1	79.6	83.7	67.7	77.9	78.7	66.8	56.0	81.9	73.9	60.9	86.1	72.6
AaD [32]	59.3	79.3	82.1	68.9	79.8	79.5	67.2	57.4	83.1	72.1	58.5	85.4	72.7
CST [17]	59.0	79.6	83.4	68.4	77.1	76.7	68.9	56.4	83.0	75.3	62.2	85.1	73.0
DCAN+SCDA [15]	60.7	76.4	82.8	69.8	77.5	78.4	68.9	59.0	82.7	74.9	61.8	84.5	73.1
KUDA [24]	58.2	80.0	82.9	71.1	80.3	80.7	71.3	56.8	83.2	75.5	60.3	86.6	73.9
Ours (RN-50)	57.5	84.0	83.8	77.8	85.5	84.7	76.3	59.2	85.4	78.1	60.2	86.7	**76.6**
ViT-B [4]	54.7	83.0	87.2	77.3	83.4	85.5	74.4	50.9	87.2	79.6	53.8	88.8	75.5
SHOT-B [30]	67.1	83.5	85.5	76.6	83.4	83.7	76.3	65.3	85.3	80.4	66.7	83.4	78.1
CDTrans* [30]	68.8	85.0	86.9	81.5	87.1	87.3	79.6	63.3	88.2	82.0	66.0	90.6	80.5
TVT-B [31]	74.9	86.8	89.5	82.8	88.0	88.3	79.8	71.9	90.1	85.5	74.6	90.6	83.6
SDAT-B [22]	70.8	87.0	90.5	85.2	87.3	89.7	94.1	70.7	90.6	88.3	75.5	92.1	84.3
SSRT-B [25]	75.2	89.0	91.1	85.1	88.3	89.9	85.0	74.2	91.3	85.7	78.6	91.8	85.4
PMTrans [36]	81.2	91.6	92.4	88.9	91.6	93.0	88.5	80.0	93.4	89.5	82.4	94.5	88.9
Ours-B	76.4	90.6	90.8	86.7	92.3	92.0	86.0	74.5	91.5	86.9	79.1	93.1	**86.7**

References

[1] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. In *ICLR*, 2019. 1

[2] Lin Chen, Huaian Chen, Zhixiang Wei, Xin Jin, Xiao Tan, Yi Jin, and Enhong Chen. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *CVPR*, pages 7181–7190, 2022. 2

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1, 3

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021. 2, 3

Table 6: Accuracies (%) on **DomainNet**. In each sub-table, the column-wise means source domain and the row-wise means target domain. "-B" indicates ViT-B (except CDTrans uses DeiT).

ResNet-101 [10]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	19.3	37.5	11.1	52.2	41.0	32.2
inf	30.2	-	31.2	3.6	44.0	27.9	27.4
pnt	39.6	18.7	-	4.9	54.5	36.3	30.8
qdr	7.0	0.9	1.4	-	4.1	8.3	4.3
rel	48.4	22.2	49.4	6.4	-	38.8	33.0
skt	46.9	15.4	37.0	10.9	47.0	-	31.4
Avg.	34.4	15.3	31.3	7.4	40.4	30.5	26.6

MIMTFL [7]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	15.1	35.6	10.7	51.5	43.1	31.2
inf	32.1	-	31.0	2.9	48.5	31.0	29.1
pnt	40.1	14.7	-	4.2	55.4	36.8	30.2
qdr	18.8	3.1	5.0	-	16.0	13.8	11.3
rel	48.5	19.0	47.6	5.8	-	39.4	32.1
skt	51.7	16.5	40.3	12.3	53.5	-	34.9
Avg.	38.2	13.7	31.9	7.2	45.0	32.8	28.1

CDAN [18]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	20.4	36.6	9.0	50.7	42.3	31.8
inf	27.5	-	25.7	1.8	34.7	20.1	22.0
pnt	42.6	20.0	-	2.5	55.6	38.5	31.8
qdr	21.0	4.5	8.1	-	14.3	15.7	12.7
rel	51.9	23.3	50.4	5.4	-	41.4	34.5
skt	50.8	20.3	43.0	2.9	50.8	-	33.6
Avg.	38.8	17.7	32.8	4.3	41.2	31.6	27.7

MDD+SCDA [15]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	20.4	43.3	15.2	59.3	46.5	36.9
inf	32.7	-	34.5	6.3	47.6	29.2	30.1
pnt	46.4	19.9	-	8.1	58.8	42.9	35.2
qdr	31.1	6.6	18.0	-	28.8	22.0	21.3
rel	55.5	23.7	52.9	9.5	-	45.2	37.4
skt	55.8	20.1	46.5	15.0	56.7	-	38.8
Avg.	44.3	18.1	39.0	10.8	50.2	37.2	33.3

ViT-B [4]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	27.2	53.1	13.2	71.2	53.3	43.6
inf	51.4	-	49.3	4.0	66.3	41.1	42.4
pnt	53.1	25.6	-	4.8	70.0	41.8	39.1
qdr	30.5	4.5	16.0	-	27.0	19.3	19.5
rel	58.4	29.0	60.0	6.0	-	45.8	39.9
skt	63.9	23.8	52.3	14.4	67.4	-	44.4
Avg.	51.5	22.0	46.1	8.5	60.4	40.3	38.1

CD-Trans [30]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	29.4	57.2	26.0	72.6	58.1	48.7
inf	57.0	-	54.4	12.8	69.5	48.4	48.4
pnt	62.9	27.4	-	15.8	72.1	53.9	46.4
qdr	44.6	8.9	29.0	-	42.6	28.5	30.7
rel	66.2	31.0	61.5	16.2	-	52.9	45.6
skt	69.0	29.6	59.0	27.2	72.5	-	51.5
Avg.	59.9	25.3	52.2	19.6	65.9	48.4	45.2

PMTrans [36]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	34.2	62.7	32.5	79.3	63.7	54.5
inf	67.4	-	61.1	22.2	78.0	57.6	57.3
pnt	69.7	33.5	-	23.9	79.8	61.2	53.6
qdr	54.6	17.4	38.9	-	49.5	41.0	40.3
rel	74.1	35.3	70.0	25.4	-	61.1	53.2
skt	73.8	33.0	62.6	30.9	77.5	-	55.6
Avg.	67.9	30.7	59.1	27.0	72.8	56.9	52.4

SSRT-B [25]	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	33.8	60.2	19.4	75.8	59.8	49.8
inf	55.5	-	54.0	9.0	68.2	44.7	46.3
pnt	61.7	28.5	-	8.4	71.4	55.2	45.0
qdr	42.5	8.8	24.2	-	37.6	33.6	29.3
rel	69.9	37.1	66.0	10.1	-	58.9	48.4
skt	70.6	32.8	62.2	21.7	73.2	-	52.1
Avg.	60.0	28.2	53.3	13.7	65.3	50.4	45.2

Ours-B	clp	inf	pnt	qdr	rel	skt	Avg.
clp	-	73.6	75.4	74.6	76.4	76.3	75.3
inf	55.1	-	54.3	53.6	54.9	54.9	54.6
pnt	71.1	70.6	-	70.0	72.7	71.7	71.2
qdr	36.8	18.0	32.0	-	31.7	34.9	30.7
rel	84.2	83.5	83.5	83.1	-	83.6	83.6
skt	68.1	66.6	67.2	66.1	67.5	-	67.1
Avg.	63.1	62.5	62.5	69.5	60.6	64.3	63.7

Table 7: Accuracies (%) on **Office-31**.

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg.
RN-50 [10]	68.4	96.7	99.3	68.9	62.5	60.7	76.1
DANN [6]	82.0	96.9	99.1	79.7	68.2	67.4	82.2
SAFN+ENT [29]	90.1	98.6	99.8	90.7	73.0	70.2	87.1
SUDA [33]	90.8	98.7	100.0	91.2	72.2	71.4	87.4
CaCo [11]	89.7	98.4	100.0	91.7	73.1	72.8	87.6
SHOT [16]	90.1	98.4	99.9	94.0	74.7	74.3	88.6
CDAN+TN [27]	95.7	98.7	100.	94.0	73.4	74.2	89.3
MDD+SCDA [15]	95.3	99.0	100.	95.4	77.2	75.9	90.5
ViT-B [4]	91.2	99.2	100.	90.4	81.1	80.6	90.4
SHOT-B [16]	94.3	99.0	100.	95.3	79.4	80.2	91.4
CDTrans* [30]	96.7	99.0	100.	97.0	81.1	81.9	92.6
SSRT-B [25]	97.7	99.2	100.	98.6	83.5	82.2	93.5
TVT-B [31]	96.4	99.4	100.	96.4	84.9	86.1	93.8
PMTrans [36]	99.1	99.6	100.0	99.4	85.7	86.3	95.0
Ours-B	97.9	99.2	100.	98.5	84.6	85.3	94.3

Table 8: Classification accuracy on DomainNet [3] with ViT-B for Domain Generalization (train on all domains except "Painting", and test on "Painting").

Methods	Accuracy
Zero-shot CLIP [21]	69.6%
Fine-tuned without CFM	55.6%
Ours: Fine-tuned with CFM	72.0%

[5] Zhekai Du, Jingjing Li, Hongzu Su, Lei Zhu, and Ke Lu. Cross-domain gradient discrepancy minimization for unsupervised domain adaptation. In *CVPR*, pages 3937–3946, 2021. 2

[6] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, pages 1180–1189, 2015. 2, 3

[7] Jian Gao, Yang Hua, Guosheng Hu, Chi Wang, and Neil M Robertson. Reducing distributional uncertainty by mutual information maximisation and transferable feature learning. In *ECCV*, pages 587–605. Springer, 2020. 3

[8] Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *arXiv preprint arXiv:2110.04544*, 2021. 1

[9] Zhiqiang Gao, Shufei Zhang, Kaizhu Huang, Qiufeng Wang, and Chaoliang Zhong. Gradient distribution alignment certificates better adversarial domain adaptation. In *ICCV*, pages 8937–8946, 2021. 2

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2, 3

[11] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *CVPR*, pages 1203–1214, 2022. 2, 3

[12] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4893–4902, 2019. 2

[13] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *CVPR*, pages 10285–10295, 2019. 2

[14] Seungmin Lee, Dongwan Kim, Namil Kim, and Seong-Gyun Jeong. Drop to adapt: Learning discriminative features for unsupervised domain adaptation. In *Proceedings*

of the IEEE/CVF International Conference on Computer Vision, pages 91–100, 2019. 2

[15] Shuang Li, Mixue Xie, Fangrui Lv, Chi Harold Liu, Jian Liang, Chen Qin, and Wei Li. Semantic concentration for domain adaptation. In *ICCV*, pages 9102–9111, 2021. 2, 3

[16] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*, pages 6028–6039, 2020. 2, 3

[17] Hong Liu, Jianmin Wang, and Mingsheng Long. Cycle self-training for domain adaptation. *Advances in Neural Information Processing Systems*, 34:22968–22981, 2021. 2

[18] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, pages 1645–1655, 2018. 2, 3

[19] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, pages 1406–1415, 2019. 1

[20] 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

[21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763, 2021. 1, 3

[22] Harsh Rangwani, Sumukh K Aithal, Mayank Mishra, Arihant Jain, and Venkatesh Babu Radhakrishnan. A closer look at smoothness in domain adversarial training. In *ICML*, pages 18378–18399, 2022. 2

[23] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, pages 213–226, 2010. 1

[24] Tao Sun, Cheng Lu, and Haibin Ling. Prior knowledge guided unsupervised domain adaptation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, pages 639–655. Springer, 2022. 2

[25] Tao Sun, Cheng Lu, Tianshuo Zhang, and Haibin Ling. Safe self-refinement for transformer-based domain adaptation. In *CVPR*, pages 7191–7200, 2022. 2, 3

[26] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, pages 5018–5027, 2017. 1

[27] Ximei Wang, Ying Jin, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable normalization: Towards improving transferability of deep neural networks. In *NeurIPS*, 2019. 2, 3

[28] Haifeng Xia, Taotao Jing, and Zhengming Ding. Maximum structural generation discrepancy for unsupervised domain adaptation. *PAMI*, 2022. 2

[29] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *ICCV*, pages 1426–1435, 2019. 2, 3

[30] Tongkun Xu, Weihua Chen, Pichao Wang, Fan Wang, Hao Li, and Rong Jin. Cdtrans: Cross-domain transformer for unsupervised domain adaptation. In *ICLR*, 2022. 2, 3

[31] Jinyu Yang, Jingjing Liu, Ning Xu, and Junzhou Huang. Tvt: Transferable vision transformer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 520–530, 2023. 2, 3

[32] Shiqi Yang, Yaxing Wang, Kai Wang, Shangling Jui, et al. Attracting and dispersing: A simple approach for source-free domain adaptation. In *Advances in Neural Information Processing Systems*, 2022. 2

[33] Jingyi Zhang, Jiaxing Huang, Zichen Tian, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. In *CVPR*, pages 9829–9840, 2022. 2, 3

[34] Renrui Zhang, Wei Zhang, Rongyao Fang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free adaption of clip for few-shot classification. In *ECCV*, pages 493–510. Springer, 2022. 1

[35] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *Int. J. Comput. Vis.*, pages 1–12, 2022. 1

[36] Jinjing Zhu, Haotian Bai, and Lin Wang. Patch-mix transformer for unsupervised domain adaptation: A game perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3561–3571, 2023. 1, 2, 3