Supplementary Materials: Adversarial Feature Augmentation for Cross-domain Few-shot Classification

Anonymous ECCV submission

Paper ID 2507

In this supplementary document, we provide additional ablation experiments and detailed comparison on the mini-ImageNet, CUB, Cars, Places, Plantae, CropDiseases, EuroSAT, ISIC, ChestX datasets, in addition to the results in the manuscript. Moreover, source codes are provided as supplementary materials which will be made publicly available upon publication of this work.

1 Integrating with Other Meta-learning Methods

To further verify the effectiveness of the proposed model-agnostic adversarial feature augmentation (AFA) module, we insert it into the Prototype Network (PN) [3] and Relation Network (RN) [5] other than the meta-learning methods in the manuscript. Results of our method are compared with competitive recent works including FT [6], ATA [8] and LRP [4] in Table 1. From these results, we have the following observations:
i. Our method integrated with the two additional meta-learning methods outperforms the state of the art for almost all the datasets and different-shot settings.

Method/shot	CUB		Cars		Places		Planae	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PN [3]	$43.09_{\pm0.5}$	$64.29_{\pm0.4}$	$31.81_{\pm0.3}$	$50.48_{\pm0.4}$	$51.86_{\pm0.5}$	$71.69_{\pm0.4}$	$36.54_{\pm0.4}$	$48.04_{\pm0.4}$
w/ FT [6]	$42.71_{\pm0.4}$	$63.91_{\pm0.4}$	$31.38_{\pm0.3}$	$50.77_{\pm0.4}$	$51.10_{\pm0.5}$	$71.58_{\pm0.4}$	$35.59_{\pm0.4}$	$51.75_{\pm0.4}$
w/ ATA [8]	$42.43_{\pm0.4}$	$64.13_{\pm0.4}$	$31.16_{\pm0.3}$	**$51.37_{\pm0.4}$**	$50.80_{\pm0.5}$	$71.99_{\pm0.4}$	$34.93_{\pm0.4}$	$52.94_{\pm0.4}$
Ours	**$45.66_{\pm0.4}$**	**$64.66_{\pm0.4}$**	**$33.21_{\pm0.4}$**	$50.89_{\pm0.4}$	**$52.70_{\pm0.5}$**	**$73.19_{\pm0.4}$**	**$36.98_{\pm0.4}$**	**$54.27_{\pm0.4}$**
RN [5]	$41.27_{\pm0.4}$	$56.77_{\pm0.4}$	$30.09_{\pm0.3}$	$40.46_{\pm0.4}$	$48.16_{\pm0.5}$	$64.25_{\pm0.4}$	$31.23_{\pm0.3}$	$42.71_{\pm0.3}$
w/ FT [6]	$43.33_{\pm0.4}$	$59.77_{\pm0.4}$	$30.45_{\pm0.3}$	$40.18_{\pm0.4}$	$49.92_{\pm0.5}$	$65.55_{\pm0.4}$	$32.57_{\pm0.3}$	$44.29_{\pm0.3}$
w/ LRP [4]	$41.57_{\pm0.4}$	$57.70_{\pm0.4}$	$30.48_{\pm0.3}$	$41.21_{\pm0.4}$	$48.47_{\pm0.5}$	$65.35_{\pm0.4}$	$32.11_{\pm0.3}$	$43.70_{\pm0.3}$
w/ ATA [8]	$43.02_{\pm0.4}$	$59.36_{\pm0.4}$	$31.79_{\pm0.3}$	$42.95_{\pm0.4}$	$51.16_{\pm0.5}$	**$66.90_{\pm0.4}$**	$33.72_{\pm0.3}$	$45.32_{\pm0.3}$
Ours	**$44.47_{\pm0.4}$**	**$62.39_{\pm0.4}$**	**$31.86_{\pm0.3}$**	**$43.09_{\pm0.4}$**	**$52.01_{\pm0.5}$**	$66.65_{\pm0.4}$	**$34.53_{\pm0.4}$**	**$47.31_{\pm0.4}$**
	CropDiseases		EuroSAT		ISIC		ChestX	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PN [3]	$61.68_{\pm0.5}$	$89.91_{\pm0.3}$	$58.94_{\pm0.5}$	$79.20_{\pm0.4}$	$30.65_{\pm0.3}$	$43.40_{\pm0.3}$	$22.44_{\pm0.2}$	$24.74_{\pm0.2}$
w/ FT [6]	$60.82_{\pm0.5}$	$87.99_{\pm0.3}$	$57.77_{\pm0.5}$	**$79.86_{\pm0.4}$**	$30.37_{\pm0.3}$	$42.38_{\pm0.3}$	$22.11_{\pm0.2}$	$24.94_{\pm0.2}$
w/ ATA [8]	$62.63_{\pm0.5}$	$87.90_{\pm0.3}$	**$60.92_{\pm0.4}$**	$79.35_{\pm0.4}$	$31.10_{\pm0.3}$	$45.40_{\pm0.3}$	$22.04_{\pm0.2}$	$24.33_{\pm0.2}$
Ours	**$64.12_{\pm0.5}$**	**$91.83_{\pm0.4}$**	$59.37_{\pm0.5}$	**$79.86_{\pm0.4}$**	**$31.98_{\pm0.3}$**	**$46.42_{\pm0.3}$**	**$22.77_{\pm0.2}$**	**$26.82_{\pm0.2}$**
RN [5]	$53.58_{\pm0.4}$	$72.86_{\pm0.4}$	$49.08_{\pm0.4}$	$65.56_{\pm0.4}$	$30.53_{\pm0.3}$	$38.60_{\pm0.3}$	$21.95_{\pm0.2}$	$24.07_{\pm0.2}$
w/ FT [6]	$57.57_{\pm0.5}$	$75.78_{\pm0.4}$	$53.53_{\pm0.4}$	$69.13_{\pm0.4}$	$30.38_{\pm0.3}$	$38.68_{\pm0.3}$	$21.79_{\pm0.2}$	$23.95_{\pm0.2}$
w/ LRP [4]	$55.01_{\pm0.4}$	$74.21_{\pm0.4}$	$50.99_{\pm0.4}$	$67.54_{\pm0.4}$	$31.16_{\pm0.3}$	$39.97_{\pm0.3}$	$22.11_{\pm0.2}$	$24.28_{\pm0.2}$
w/ ATA [8]	$61.17_{\pm0.5}$	**$78.20_{\pm0.4}$**	$55.69_{\pm0.5}$	$71.02_{\pm0.4}$	$31.13_{\pm0.3}$	$40.38_{\pm0.3}$	$22.14_{\pm0.2}$	$24.43_{\pm0.2}$
Ours	**$66.17_{\pm0.5}$**	$77.41_{\pm0.4}$	**$59.80_{\pm0.5}$**	**$73.29_{\pm0.4}$**	**$31.77_{\pm0.3}$**	**$41.41_{\pm0.3}$**	**$22.82_{\pm0.2}$**	**$24.93_{\pm0.2}$**

Table 1. Few-shot classification accuracy (%) of 5-way 5-shot/1-shot setting trained on the mini-ImageNet dataset, and tested on various datasets from target domains. The best results in different settings are in **Bold**.

For 1-shot classification, our method improves the baseline Relation Network (RN) by 4.69% averagely over the eight datasets. In 5-shot setting, the average improvement is 3.9% compared to the RN baseline.
ii. Compared to the competitive ATA with task augmentation, the proposed AFA integrated with the two baselines of PN and RN achieves an average improvement of about 1.3% by feature augmentation.

2 Classification with F_o instead of F_a

In our method, the augmented features F_a generated by the AFA module are used for classification with the class discriminator. Compared with F_o, the augmented features F_a vary from the source domain that can help to improve the robustness of the class discriminator. For experimental evaluation, results by changing F_a to F_o ($F_a \rightarrow F_o$) for classification in our method are reported in Table. 2. Based on these results, we can see that:
i. When using F_o for classification, the performance degrades (comparing the second row and the last row of Table 2). This indicates that the augmented features F_a varying from the source domain, can improve the performance on various datesets under the different-shot setting (e.g., an average 2.8% improvement on the 1-shot setting and about 2.7% on the 5-shot setting).
ii. Compared to the baseline meta-learning results in the first row of Table. 2, it cannot bring significant improvement or even perform worse in some datasets by training the class discriminator and encoder with the original features F_o. This indicates that it may lead to the overfitting problem, if the proposed AFA is not fully utilized by directly using the source domain features F_o for classification.

3 Training with Worst-case Feature Distribution

In [8], the ATA method attempts to enlarge the source task distribution space through "worst-case" tasks for generalization to the unseen target tasks. To determine the "worst-case" tasks, gradient ascent is applied on the original source tasks in each iteration. This means that the "worst-case" depends on the source tasks in each iteration, which is not a set of trainable model parameters. For comparison, we conduct an experiment to analyse if training with the "worst-case" feature distribution still works in our method. To identify the "worst-case" feature distribution, we update the parameters of the feature augmentation layers by the gradient ascent strategy as in the ATA. In each iteration, the parameters of the feature augmentation layers are first updated by only gradient ascent with the loss function L_c, then the updated parameters are set as fixed when training the models. Experimental results by modifying the proposed AFA with the learning strategy of the 'worst-case' feature distribution ('AFA'→'WC') are shown in the third row of Table 2.

These results show that it leads to worse performance when training with the "worst-case" (WC) feature distribution in some datasets, compared to the meta-learning baselines, e.g., PN on CropDiseases and EuroSAT datasets, or

Method/shot	CUB		Cars		Places		Planae	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MN [7]	35.89$_{\pm0.5}$	51.37$_{\pm0.8}$	30.77$_{\pm0.5}$	38.99$_{\pm0.6}$	49.86$_{\pm0.8}$	63.16$_{\pm0.8}$	32.70$_{\pm0.6}$	46.53$_{\pm0.7}$
$F_a \rightarrow F_o$	38.48$_{\pm0.4}$	50.93$_{\pm0.4}$	30.84$_{\pm0.3}$	43.60$_{\pm0.4}$	51.60$_{\pm0.5}$	63.68$_{\pm0.4}$	36.64$_{\pm0.4}$	44.30$_{\pm0.4}$
'AFA'→'WC'	39.87$_{\pm0.4}$	55.09$_{\pm0.4}$	32.68$_{\pm0.4}$	44.12$_{\pm0.4}$	51.47$_{\pm0.5}$	66.63$_{\pm0.4}$	36.02$_{\pm0.4}$	49.40$_{\pm0.4}$
Ours	**41.02$_{\pm0.4}$**	**59.46$_{\pm0.4}$**	**33.52$_{\pm0.4}$**	**46.13$_{\pm0.4}$**	**54.66$_{\pm0.5}$**	**68.87$_{\pm0.4}$**	**37.60$_{\pm0.4}$**	**52.43$_{\pm0.4}$**
GNN [2]	44.40$_{\pm0.5}$	62.87$_{\pm0.5}$	31.72$_{\pm0.4}$	43.70$_{\pm0.4}$	52.42$_{\pm0.5}$	70.91$_{\pm0.5}$	33.60$_{\pm0.4}$	48.51$_{\pm0.4}$
$F_a \rightarrow F_o$	42.12$_{\pm0.5}$	63.96$_{\pm0.5}$	32.77$_{\pm0.3}$	46.34$_{\pm0.5}$	49.90$_{\pm0.4}$	71.87$_{\pm0.5}$	**37.48$_{\pm0.4}$**	51.08$_{\pm0.4}$
'AFA'→'WC'	42.91$_{\pm0.5}$	62.17$_{\pm0.5}$	32.30$_{\pm0.4}$	45.11$_{\pm0.4}$	53.47$_{\pm0.6}$	70.36$_{\pm0.5}$	35.76$_{\pm0.4}$	53.27$_{\pm0.4}$
Ours	**46.86$_{\pm0.5}$**	**68.25$_{\pm0.5}$**	**34.25$_{\pm0.4}$**	**49.28$_{\pm0.5}$**	**54.04$_{\pm0.6}$**	**76.21$_{\pm0.5}$**	36.76$_{\pm0.4}$	**54.26$_{\pm0.4}$**
TPN [1]	48.30$_{\pm0.4}$	63.52$_{\pm0.4}$	32.42$_{\pm0.4}$	44.54$_{\pm0.4}$	56.17$_{\pm0.5}$	71.39$_{\pm0.4}$	37.40$_{\pm0.4}$	50.96$_{\pm0.4}$
$F_a \rightarrow F_o$	49.37$_{\pm0.5}$	64.36$_{\pm0.4}$	34.30$_{\pm0.4}$	47.44$_{\pm0.4}$	58.09$_{\pm0.5}$	71.13$_{\pm0.4}$	38.39$_{\pm0.4}$	52.81$_{\pm0.4}$
'AFA'→'WC'	44.96$_{\pm0.5}$	60.18$_{\pm0.4}$	29.05$_{\pm0.4}$	39.85$_{\pm0.4}$	54.10$_{\pm0.5}$	69.89$_{\pm0.4}$	34.63$_{\pm0.4}$	51.77$_{\pm0.4}$
Ours	**50.85$_{\pm0.4}$**	**65.86$_{\pm0.4}$**	**38.43$_{\pm0.4}$**	**47.89$_{\pm0.4}$**	**60.29$_{\pm0.5}$**	**72.81$_{\pm0.4}$**	**40.27$_{\pm0.4}$**	**55.67$_{\pm0.4}$**
PN [3]	43.09$_{\pm0.5}$	64.29$_{\pm0.4}$	31.81$_{\pm0.3}$	50.48$_{\pm0.4}$	51.86$_{\pm0.5}$	71.69$_{\pm0.4}$	36.54$_{\pm0.4}$	48.04$_{\pm0.4}$
$F_a \rightarrow F_o$	44.42$_{\pm0.4}$	62.35$_{\pm0.4}$	32.05$_{\pm0.3}$	49.07$_{\pm0.4}$	50.44$_{\pm0.5}$	71.84$_{\pm0.4}$	36.23$_{\pm0.4}$	55.72$_{\pm0.4}$
'AFA'→'WC'	42.37$_{\pm0.4}$	**65.59$_{\pm0.4}$**	31.59$_{\pm0.4}$	**51.01$_{\pm0.4}$**	48.55$_{\pm0.5}$	71.93$_{\pm0.4}$	34.32$_{\pm0.4}$	**55.92$_{\pm0.4}$**
Ours	**45.66$_{\pm0.4}$**	64.66$_{\pm0.4}$	**33.21$_{\pm0.4}$**	50.89$_{\pm0.4}$	**52.70$_{\pm0.5}$**	**73.19$_{\pm0.4}$**	**36.98$_{\pm0.4}$**	54.27$_{\pm0.4}$
RN [5]	41.27$_{\pm0.4}$	56.77$_{\pm0.4}$	30.09$_{\pm0.3}$	40.46$_{\pm0.4}$	48.16$_{\pm0.5}$	64.25$_{\pm0.4}$	31.23$_{\pm0.3}$	42.71$_{\pm0.3}$
$F_a \rightarrow F_o$	41.20$_{\pm0.4}$	59.33$_{\pm0.4}$	**31.91$_{\pm0.3}$**	42.82$_{\pm0.4}$	50.53$_{\pm0.5}$	66.26$_{\pm0.4}$	33.55$_{\pm0.4}$	46.21$_{\pm0.3}$
'AFA'→'WC'	42.48$_{\pm0.4}$	59.80$_{\pm0.8}$	31.52$_{\pm0.4}$	42.15$_{\pm0.4}$	51.65$_{\pm0.5}$	**67.34$_{\pm0.4}$**	34.20$_{\pm0.4}$	47.14$_{\pm0.4}$
Ours	**44.47$_{\pm0.4}$**	**62.39$_{\pm0.4}$**	31.86$_{\pm0.3}$	**43.09$_{\pm0.4}$**	**52.01$_{\pm0.5}$**	66.65$_{\pm0.4}$	**34.53$_{\pm0.4}$**	**47.31$_{\pm0.4}$**
	CropDiseases		EuroSAT		ISIC		ChestX	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MN [7]	57.57$_{\pm0.5}$	73.26$_{\pm0.5}$	54.19$_{\pm0.5}$	67.50$_{\pm0.5}$	29.62$_{\pm0.3}$	32.98$_{\pm0.3}$	**22.30$_{\pm0.2}$**	22.85$_{\pm0.2}$
$F_a \rightarrow F_o$	58.34$_{\pm0.5}$	75.02$_{\pm0.5}$	55.29$_{\pm0.5}$	66.19$_{\pm0.5}$	28.80$_{\pm0.3}$	33.24$_{\pm0.3}$	21.77$_{\pm0.2}$	23.05$_{\pm0.2}$
'AFA'→'WC'	52.15$_{\pm0.5}$	72.80$_{\pm0.4}$	47.55$_{\pm0.5}$	61.89$_{\pm0.5}$	29.17$_{\pm0.3}$	34.03$_{\pm0.3}$	21.64$_{\pm0.2}$	22.68$_{\pm0.2}$
Ours	**60.71$_{\pm0.5}$**	**80.07$_{\pm0.4}$**	**61.28$_{\pm0.5}$**	**69.63$_{\pm0.5}$**	**32.32$_{\pm0.3}$**	**39.88$_{\pm0.3}$**	22.11$_{\pm0.2}$	**23.18$_{\pm0.2}$**
GNN [2]	59.19$_{\pm0.5}$	83.12$_{\pm0.4}$	54.61$_{\pm0.5}$	78.69$_{\pm0.4}$	30.14$_{\pm0.3}$	42.54$_{\pm0.4}$	21.94$_{\pm0.2}$	23.87$_{\pm0.2}$
$F_a \rightarrow F_o$	59.53$_{\pm0.5}$	84.27$_{\pm0.4}$	59.33$_{\pm0.5}$	79.19$_{\pm0.4}$	30.02$_{\pm0.3}$	43.42$_{\pm0.4}$	21.88$_{\pm0.2}$	24.58$_{\pm0.2}$
'AFA'→'WC'	64.18$_{\pm0.6}$	86.08$_{\pm0.4}$	59.27$_{\pm0.6}$	79.33$_{\pm0.5}$	31.84$_{\pm0.4}$	44.35$_{\pm0.4}$	22.40$_{\pm0.2}$	24.82$_{\pm0.3}$
Ours	**67.61$_{\pm0.5}$**	**88.06$_{\pm0.3}$**	**63.12$_{\pm0.5}$**	**85.58$_{\pm0.4}$**	**33.21$_{\pm0.3}$**	**46.01$_{\pm0.4}$**	**22.92$_{\pm0.2}$**	**25.02$_{\pm0.2}$**
TPN [1]	68.39$_{\pm0.6}$	81.91$_{\pm0.5}$	63.90$_{\pm0.5}$	77.22$_{\pm0.4}$	**35.08$_{\pm0.4}$**	45.66$_{\pm0.3}$	21.05$_{\pm0.2}$	22.17$_{\pm0.2}$
$F_a \rightarrow F_o$	68.94$_{\pm0.6}$	82.28$_{\pm0.5}$	60.05$_{\pm0.5}$	78.16$_{\pm0.4}$	33.78$_{\pm0.3}$	47.06$_{\pm0.3}$	21.25$_{\pm0.2}$	22.21$_{\pm0.2}$
'AFA'→'WC'	56.02$_{\pm0.7}$	74.44$_{\pm0.7}$	46.75$_{\pm0.5}$	73.60$_{\pm0.4}$	31.29$_{\pm0.3}$	40.33$_{\pm0.3}$	20.57$_{\pm0.1}$	21.76$_{\pm0.1}$
Ours	**72.44$_{\pm0.6}$**	**85.69$_{\pm0.4}$**	**66.17$_{\pm0.4}$**	**80.12$_{\pm0.4}$**	34.25$_{\pm0.4}$	**46.29$_{\pm0.3}$**	**21.69$_{\pm0.1}$**	**23.47$_{\pm0.2}$**
PN [3]	61.68$_{\pm0.5}$	89.91$_{\pm0.3}$	58.94$_{\pm0.5}$	79.20$_{\pm0.4}$	30.65$_{\pm0.3}$	43.40$_{\pm0.3}$	22.44$_{\pm0.2}$	24.74$_{\pm0.2}$
$F_a \rightarrow F_o$	58.86$_{\pm0.5}$	86.28$_{\pm0.3}$	58.32$_{\pm0.5}$	76.28$_{\pm0.4}$	30.84$_{\pm0.3}$	43.75$_{\pm0.3}$	22.01$_{\pm0.2}$	26.52$_{\pm0.2}$
'AFA'→'WC'	57.60$_{\pm0.5}$	86.89$_{\pm0.3}$	56.29$_{\pm0.5}$	77.54$_{\pm0.4}$	29.25$_{\pm0.3}$	43.54$_{\pm0.3}$	22.01$_{\pm0.2}$	26.52$_{\pm0.2}$
Ours	**64.12$_{\pm0.5}$**	**91.83$_{\pm0.4}$**	**59.37$_{\pm0.5}$**	**79.86$_{\pm0.4}$**	**31.98$_{\pm0.3}$**	**46.42$_{\pm0.3}$**	**22.77$_{\pm0.2}$**	**26.82$_{\pm0.2}$**
RN [5]	53.58$_{\pm0.4}$	72.86$_{\pm0.4}$	49.08$_{\pm0.4}$	65.56$_{\pm0.4}$	30.53$_{\pm0.3}$	38.60$_{\pm0.3}$	21.95$_{\pm0.2}$	24.07$_{\pm0.2}$
$F_a \rightarrow F_o$	53.80$_{\pm0.4}$	74.56$_{\pm0.4}$	52.72$_{\pm0.5}$	69.39$_{\pm0.4}$	29.34$_{\pm0.3}$	40.21$_{\pm0.3}$	21.55$_{\pm0.2}$	24.84$_{\pm0.2}$
'AFA'→'WC'	57.64$_{\pm0.5}$	73.88$_{\pm0.4}$	51.42$_{\pm0.5}$	64.61$_{\pm0.4}$	29.89$_{\pm0.3}$	40.36$_{\pm0.3}$	22.19$_{\pm0.2}$	24.35$_{\pm0.2}$
Ours	**66.17$_{\pm0.5}$**	**77.41$_{\pm0.4}$**	**59.80$_{\pm0.5}$**	**73.29$_{\pm0.4}$**	**31.77$_{\pm0.3}$**	**41.41$_{\pm0.3}$**	**22.82$_{\pm0.2}$**	**24.93$_{\pm0.2}$**

Table 2. Accuracy (%) of different training strategies in our method under 5-way 5-shot/1-shot setting. $F_a \rightarrow F_o$ means that F_a is replaced by F_o for classification during training. **'AFA'→'WC'** refers to the experiments by modifying the update strategy of the feature augmentation layers by searching for the "worst-case" feature distribution with gradient ascent. The best results in different settings are in **Bold**.

TPN on CUB and Cars datasets. The main reason behind is that, the "worst-case" features of the source domain may not be able to represent the target features, so that the expanded source domain feature space is still far away from the target domain. Moreover, it also indicates that the proposed AFA module based on sufficient statistic disturbance can better generalize to different target domains, since various unseen domain feature distributions are generated for training with to extract domain-invariant features in classification.

methods	baseline	w/ FT	w/ ATA	**Ours**	w/o D_d	w/o L_g	Non-linear	'AFA' $\rightarrow$ 'WC'
Time	0.19	0.21	0.78	**0.25**	0.23	0.21	0.33	0.32

Table 3. Training of 5-way 5-shot classification on the CUB measured by s/100imgs.

4 Model Complexity

In this experiment, we compare the training time of the proposed method with the ATA [8], FT [6] and the Matching Network (MN) baseline in Table 3. Moreover, the training complexity of our method with different variations is also reported. These results show that the computational overhead of the competitive method ATA is much higher than other works as mentioned in the manuscript, which takes twice time longer than our method. The training time of our method with the variation of the "worse-cast" strategy for feature augmentation with inferior performance is shorter than the ATA but longer than the proposed AFA. Compared to the baseline (Matching Network), the training time of our method by adding the proposed adversarial feature augmentation module is slightly longer, but brings significant improvement on both the source and target domain. As illustrated in the manuscript, it takes longer time for training, if the linear perturbation in our method is replaced by the non-linear (convolution) transformation function for feature augmentation.

5 Additional Ablation Experiments

Besides the ablation experiments based on the Matching Network (MN) reported in the manuscript, additional ablation results by using other meta-learning methods including GNN, TPN, Prototype Network (PN) and Relation Network (RN) are shown in Table 4. From these results, we can observe that each component in our method can help to robustly improve the performance for almost all the target domain datasets when integrating in different meta-learning methods.

6 Source Domain Performance of Our Method with Different Variations

Since both the performance on the source and target domain dataset are important, we also report the results of the average performance on the source domain (i.e., the test set of the mini-ImageNet dataset) in Figure 1. These results show that our method (compared to different variations) achieves the best performance on the source domain when integrating in most meta-learning approaches. In addition, the performance of feature augmentation via non-linear transformation instead of linear perturbation and training with the "worst-case" feature distribution degrades remarkably. When using the Matching Network or GNN as the baseline, it could obtain very mild improvement for the source domain sometimes by training without the domain discriminator or gram-matrix

Method/shot	CUB 1-shot	CUB 5-shot	Cars 1-shot	Cars 5-shot	Places 1-shot	Places 5-shot	Planae 1-shot	Planae 5-shot
GNN [2]	44.40$_{\pm0.5}$	62.87$_{\pm0.5}$	31.72$_{\pm0.4}$	43.70$_{\pm0.4}$	52.42$_{\pm0.5}$	70.91$_{\pm0.5}$	33.60$_{\pm0.4}$	48.51$_{\pm0.4}$
w/o D_d	45.09$_{\pm0.5}$	65.12$_{\pm0.5}$	34.17$_{\pm0.4}$	48.22$_{\pm0.5}$	51.08$_{\pm0.5}$	74.88$_{\pm0.5}$	34.47$_{\pm0.4}$	52.58$_{\pm0.4}$
w/o L_g	46.57$_{\pm0.5}$	65.28$_{\pm0.5}$	33.91$_{\pm0.4}$	49.06$_{\pm0.5}$	**54.52$_{\pm0.6}$**	75.58$_{\pm0.5}$	36.69$_{\pm0.4}$	52.73$_{\pm0.4}$
Non-linear	44.59$_{\pm0.5}$	63.35$_{\pm0.5}$	32.81$_{\pm0.4}$	47.74$_{\pm0.4}$	53.64$_{\pm0.6}$	71.33$_{\pm0.5}$	36.02$_{\pm0.4}$	53.33$_{\pm0.4}$
Ours	**46.86$_{\pm0.5}$**	**68.25$_{\pm0.5}$**	**34.25$_{\pm0.4}$**	**49.28$_{\pm0.5}$**	54.04$_{\pm0.6}$	**76.21$_{\pm0.5}$**	**36.76$_{\pm0.4}$**	**54.26$_{\pm0.4}$**
TPN [1]	48.30$_{\pm0.4}$	63.52$_{\pm0.4}$	32.42$_{\pm0.4}$	44.54$_{\pm0.4}$	56.17$_{\pm0.5}$	71.39$_{\pm0.4}$	37.40$_{\pm0.4}$	50.96$_{\pm0.4}$
w/o D_d	45.56$_{\pm0.5}$	62.13$_{\pm0.4}$	30.79$_{\pm0.4}$	38.57$_{\pm0.5}$	55.25$_{\pm0.5}$	69.04$_{\pm0.4}$	35.79$_{\pm0.4}$	51.01$_{\pm0.5}$
w/o L_g	46.68$_{\pm0.5}$	65.65$_{\pm0.5}$	30.26$_{\pm0.4}$	46.47$_{\pm0.4}$	55.35$_{\pm0.5}$	69.50$_{\pm0.5}$	36.44$_{\pm0.4}$	52.58$_{\pm0.5}$
Non-linear	46.17$_{\pm0.5}$	62.89$_{\pm0.4}$	32.85$_{\pm0.4}$	46.79$_{\pm0.4}$	55.22$_{\pm0.5}$	70.64$_{\pm0.4}$	35.73$_{\pm0.4}$	50.70$_{\pm0.4}$
Ours	**50.85$_{\pm0.4}$**	**65.86$_{\pm0.4}$**	**38.43$_{\pm0.4}$**	**47.89$_{\pm0.4}$**	**60.29$_{\pm0.5}$**	**72.81$_{\pm0.4}$**	**40.27$_{\pm0.4}$**	**55.67$_{\pm0.4}$**
PN [3]	43.09$_{\pm0.5}$	64.29$_{\pm0.4}$	31.81$_{\pm0.3}$	50.48$_{\pm0.4}$	51.86$_{\pm0.5}$	71.69$_{\pm0.4}$	36.54$_{\pm0.4}$	48.04$_{\pm0.4}$
w/o D_d	44.14$_{\pm0.4}$	63.75$_{\pm0.4}$	32.10$_{\pm0.3}$	50.02$_{\pm0.4}$	50.88$_{\pm0.5}$	71.41$_{\pm0.4}$	36.66$_{\pm0.4}$	47.63$_{\pm0.4}$
w/o L_g	43.63$_{\pm0.4}$	62.58$_{\pm0.4}$	32.03$_{\pm0.4}$	48.53$_{\pm0.4}$	50.13$_{\pm0.5}$	71.35$_{\pm0.4}$	35.60$_{\pm0.4}$	54.17$_{\pm0.4}$
Non-linear	42.76$_{\pm0.4}$	61.54$_{\pm0.4}$	31.00$_{\pm0.3}$	50.13$_{\pm0.4}$	50.76$_{\pm0.5}$	71.44$_{\pm0.4}$	36.03$_{\pm0.4}$	**56.72$_{\pm0.4}$**
Ours	**45.66$_{\pm0.4}$**	**64.66$_{\pm0.4}$**	**33.21$_{\pm0.4}$**	**50.89$_{\pm0.4}$**	**52.70$_{\pm0.5}$**	**73.19$_{\pm0.4}$**	**36.98$_{\pm0.4}$**	54.27$_{\pm0.4}$
RN [5]	41.27$_{\pm0.4}$	56.77$_{\pm0.4}$	30.09$_{\pm0.3}$	40.46$_{\pm0.4}$	48.16$_{\pm0.5}$	64.25$_{\pm0.4}$	31.23$_{\pm0.3}$	42.71$_{\pm0.3}$
w/o D_d	41.13$_{\pm0.4}$	56.28$_{\pm0.4}$	31.50$_{\pm0.4}$	42.98$_{\pm0.4}$	49.95$_{\pm0.5}$	65.61$_{\pm0.4}$	32.57$_{\pm0.4}$	43.98$_{\pm0.4}$
w/o L_g	39.59$_{\pm0.4}$	58.66$_{\pm0.4}$	**32.75$_{\pm0.3}$**	42.69$_{\pm0.4}$	49.39$_{\pm0.5}$	**66.99$_{\pm0.4}$**	34.24$_{\pm0.4}$	47.30$_{\pm0.4}$
Non-linear	41.09$_{\pm0.4}$	58.28$_{\pm0.4}$	30.55$_{\pm0.3}$	40.47$_{\pm0.4}$	49.86$_{\pm0.5}$	65.22$_{\pm0.4}$	34.16$_{\pm0.4}$	45.33$_{\pm0.3}$
Ours	**44.47$_{\pm0.4}$**	**62.39$_{\pm0.4}$**	31.86$_{\pm0.3}$	**43.09$_{\pm0.4}$**	**52.01$_{\pm0.5}$**	66.65$_{\pm0.4}$	**34.53$_{\pm0.4}$**	**47.31$_{\pm0.4}$**
	CropDiseases 1-shot	CropDiseases 5-shot	EuroSAT 1-shot	EuroSAT 5-shot	ISIC 1-shot	ISIC 5-shot	ChestX 1-shot	ChestX 5-shot
GNN [2]	59.19$_{\pm0.5}$	83.12$_{\pm0.4}$	54.61$_{\pm0.5}$	78.69$_{\pm0.4}$	30.14$_{\pm0.3}$	42.54$_{\pm0.4}$	21.94$_{\pm0.2}$	23.87$_{\pm0.2}$
w/o D_d	64.40$_{\pm0.5}$	87.62$_{\pm0.4}$	62.87$_{\pm0.5}$	81.54$_{\pm0.4}$	31.73$_{\pm0.4}$	44.88$_{\pm0.4}$	22.09$_{\pm0.2}$	25.40$_{\pm0.3}$
w/o L_g	66.48$_{\pm0.6}$	88.02$_{\pm0.4}$	60.08$_{\pm0.6}$	80.03$_{\pm0.5}$	31.22$_{\pm0.4}$	45.48$_{\pm0.4}$	21.87$_{\pm0.2}$	24.99$_{\pm0.3}$
Non-linear	65.78$_{\pm0.6}$	86.16$_{\pm0.4}$	63.01$_{\pm0.6}$	81.33$_{\pm0.4}$	31.34$_{\pm0.4}$	41.91$_{\pm0.3}$	22.39$_{\pm0.2}$	24.43$_{\pm0.2}$
Ours	**67.61$_{\pm0.5}$**	**88.06$_{\pm0.3}$**	**63.12$_{\pm0.5}$**	**85.58$_{\pm0.4}$**	**33.21$_{\pm0.3}$**	**46.01$_{\pm0.4}$**	**22.92$_{\pm0.2}$**	**25.02$_{\pm0.2}$**
TPN [1]	68.39$_{\pm0.6}$	81.91$_{\pm0.5}$	63.90$_{\pm0.5}$	77.22$_{\pm0.4}$	**35.08$_{\pm0.4}$**	45.66$_{\pm0.3}$	21.05$_{\pm0.2}$	22.17$_{\pm0.2}$
w/o D_d	60.25$_{\pm0.6}$	72.38$_{\pm0.7}$	58.33$_{\pm0.5}$	70.53$_{\pm0.5}$	32.21$_{\pm0.3}$	42.48$_{\pm0.3}$	20.46$_{\pm0.1}$	21.61$_{\pm0.1}$
w/o L_g	68.62$_{\pm0.7}$	83.47$_{\pm0.7}$	62.78$_{\pm0.5}$	77.32$_{\pm0.6}$	31.96$_{\pm0.3}$	38.56$_{\pm0.4}$	21.06$_{\pm0.2}$	22.27$_{\pm0.1}$
Non-linear	63.20$_{\pm0.6}$	85.32$_{\pm0.4}$	59.10$_{\pm0.5}$	74.28$_{\pm0.4}$	32.16$_{\pm0.3}$	**46.70$_{\pm0.3}$**	21.21$_{\pm0.2}$	22.84$_{\pm0.2}$
Ours	**72.44$_{\pm0.6}$**	**85.69$_{\pm0.4}$**	**66.17$_{\pm0.4}$**	**80.12$_{\pm0.4}$**	34.25$_{\pm0.4}$	46.29$_{\pm0.3}$	**21.69$_{\pm0.1}$**	**23.47$_{\pm0.2}$**
PN [3]	61.68$_{\pm0.5}$	89.91$_{\pm0.3}$	58.94$_{\pm0.5}$	79.20$_{\pm0.4}$	30.65$_{\pm0.3}$	43.40$_{\pm0.3}$	22.44$_{\pm0.2}$	24.74$_{\pm0.2}$
w/o D_d	59.73$_{\pm0.5}$	88.03$_{\pm0.3}$	57.72$_{\pm0.5}$	79.00$_{\pm0.4}$	30.37$_{\pm0.3}$	43.42$_{\pm0.3}$	22.32$_{\pm0.2}$	25.55$_{\pm0.2}$
w/o L_g	60.15$_{\pm0.5}$	87.11$_{\pm0.3}$	58.01$_{\pm0.5}$	77.78$_{\pm0.4}$	30.21$_{\pm0.3}$	43.78$_{\pm0.3}$	22.29$_{\pm0.2}$	25.99$_{\pm0.2}$
Non-linear	59.73$_{\pm0.5}$	87.97$_{\pm0.3}$	58.97$_{\pm0.5}$	77.79$_{\pm0.4}$	30.74$_{\pm0.3}$	44.21$_{\pm0.3}$	22.29$_{\pm0.2}$	26.29$_{\pm0.2}$
Ours	**64.12$_{\pm0.5}$**	**91.83$_{\pm0.4}$**	**59.37$_{\pm0.5}$**	**79.86$_{\pm0.4}$**	**31.98$_{\pm0.3}$**	**46.42$_{\pm0.3}$**	**22.77$_{\pm0.2}$**	**26.82$_{\pm0.2}$**
RN [5]	53.58$_{\pm0.4}$	72.86$_{\pm0.4}$	49.08$_{\pm0.4}$	65.56$_{\pm0.4}$	30.53$_{\pm0.3}$	38.60$_{\pm0.3}$	21.95$_{\pm0.2}$	24.07$_{\pm0.2}$
w/o D_d	51.44$_{\pm0.5}$	73.56$_{\pm0.4}$	51.10$_{\pm0.5}$	60.01$_{\pm0.4}$	29.79$_{\pm0.3}$	40.78$_{\pm0.3}$	21.59$_{\pm0.2}$	24.05$_{\pm0.2}$
w/o L_g	59.91$_{\pm0.4}$	74.58$_{\pm0.4}$	56.20$_{\pm0.5}$	70.33$_{\pm0.4}$	29.05$_{\pm0.3}$	39.96$_{\pm0.3}$	22.34$_{\pm0.2}$	24.76$_{\pm0.2}$
Non-linear	56.30$_{\pm0.4}$	73.30$_{\pm0.4}$	52.23$_{\pm0.5}$	67.28$_{\pm0.4}$	30.24$_{\pm0.3}$	37.81$_{\pm0.3}$	21.78$_{\pm0.2}$	24.08$_{\pm0.2}$
Ours	**66.17$_{\pm0.5}$**	**77.41$_{\pm0.4}$**	**59.80$_{\pm0.5}$**	**73.29$_{\pm0.4}$**	**31.77$_{\pm0.3}$**	**41.41$_{\pm0.3}$**	**22.82$_{\pm0.2}$**	**24.93$_{\pm0.2}$**

Table 4. Accuracy (%) of ablation experiments under 1-shot/5-shot 5-way few-shot classification on the target domains datasets. **w/o** D_d refers to our method trained without the domain discriminator. **w/o** L_g is the ablation of the gram-matrix loss. **Non-linear** means that the linear transformation in the adversarial feature augmentation is replaced by convolution layer as the non-linear transformation.

loss. Our method with the default training strategies suppresses all the other variations and the baselines by the average performance on the source domain.

References

1. Liu, Y., Lee, J., Park, M., Kim, S., Yang, E., Hwang, S.J., Yang, Y.: Learning to propagate labels: Transductive propagation network for few-shot learning. In: ICLR (2019)

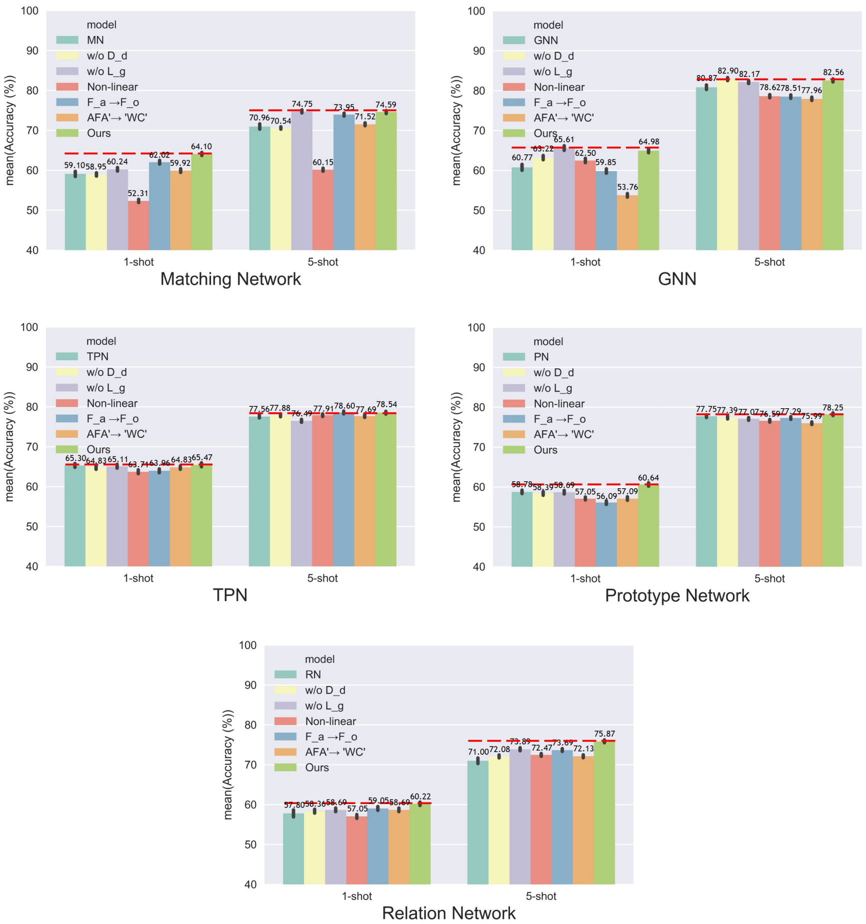

Fig. 1. Accuracy (%) of our method with different variations integrated in different meta-learning baselines for 1/5-shot classification on the source domain mini-ImageNet dataset.

2. Satorras, V.G., Estrach, J.B.: Few-shot learning with graph neural networks. In: ICLR (2018)
3. Snell, J., Swersky, K., Zemel, R.S.: Prototypical networks for few-shot learning. In: NeurIPS. pp. 4077–4087 (2017)
4. Sun, J., Lapuschkin, S., Samek, W., Zhao, Y., Cheung, N., Binder, A.: Explanation-guided training for cross-domain few-shot classification. In: ICPR. pp. 7609–7616 (2020)
5. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: CVPR. pp. 1199–1208 (June 2018)

6. Tseng, H., Lee, H., Huang, J., Yang, M.: Cross-domain few-shot classification via learned feature-wise transformation. In: ICLR (2020)
7. Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., Wierstra, D.: Matching networks for one shot learning. In: NeurIPS. pp. 3630–3638 (2016)
8. Wang, H., Deng, Z.: Cross-domain few-shot classification via adversarial task augmentation. In: Zhou, Z. (ed.) IJCAI. pp. 1075–1081 (2021)