Supplementary Material

Task Discrepancy Maximization for Fine-grained Few-Shot Classification

SuBeen Lee, WonJun Moon, Jae-Pil Heo*
Sungkyunkwan University
{leesb7426, wjun0830, jaepilheo}@skku.edu

1. Application to Baseline Methods

In this work, we apply our TDM to four baselines ProtoNet [3], DSN [2], CTX [1], and FRN [4] to validate the complementary benefits of our approach with existing methods. To train each model, we adopt hyperparameters and implementation details from FRN [4]. The reported 1-shot performances of DSN and CTX are different from [4], since those numbers reported in [4] are evaluated by models trained with 1-shot episodes while we train all the tested models with 5-shot episodes for a fair comparison. Note that, we found that utilizing 5-shot episodes provides higher accuracy than 1-shot episodes for all the methods.

To apply our TDM to the baseline models, we attach TDM at the end of the feature extractor except CTX. For CTX, we attach TDM after computing the query-aligned prototype.

Furthermore, TDM does not affect the closed-form solution of FRN. Specifically, the objective of FRN is to reconstruct the query instance $Q \in \mathbb{R}^{r\times d}$ as a weighted sum of rows of $S_c \in \mathbb{R}^{kr\times d}$ by finding a matrix $W \in \mathbb{R}^{r\times kr}$ such that $WS_c \approx Q$, where r is the spatial resolution (height $\times$ width) of the feature map, d is the number of channels, and c denotes the class index. They find the optimal $\bar{W}$ by solving the linear least-squares problem as follows:

$$\bar{W} = \underset{W}{\operatorname{argmin}} \parallel Q - WS_c \parallel^2 + \lambda \parallel W \parallel^2 \tag{1}$$

The above equation for a ridge regression has a widely-known closed-form solution for $\bar{W}$ described as follows:

$$\begin{aligned} \bar{W} &= QS_c^T(S_cS_c^T + \lambda I)^{-1} \\ \bar{Q}_c &= \bar{W}S_c \end{aligned} \tag{2}$$

When applying TDM to FRN, we need to reconstruct $A_c^Q \in \mathbb{R}^{r\times d}$ as a weighted sum of rows of $A_c^S \in \mathbb{R}^{kr\times d}$ by discovering a matrix $W \in \mathbb{R}^{r\times kr}$. Since A_c^S and A_c^Q are computed by multiplying $W_c^T = \operatorname{diag}\left(w_{c,1}^T, \cdots, w_{c,d}^T\right) \in \mathbb{R}^{r\times r}$, a diagonal matrix composed of channel weights from TDM, to S_c and Q, we can decompose it. Additionally, the weights from TDM are in a range of $(0, 2)$, so there is an inverse matrix. By putting the above together, we prove that the matrix W for $WA_c^S \approx A_c^Q$ is equivalent to $\bar{W}$ in Eq. (1) as follows:

$$\begin{aligned} WA_c^S &\approx A_c^Q \\ WS_cW_c^T &\approx QW_c^T, \\ WS_cW_c^T\left(W_c^T\right)^{-1} &\approx QW_c^T\left(W_c^T\right)^{-1}, \\ WS_c &\approx Q \end{aligned} \tag{3}$$

Therefore, the optimal matrix $\bar{W}$ of FRN with TDM is identical to $\bar{W}$ in Eq. (2). As a result, $\bar{W}$ and $\bar{Q}_c$ for the FRN with TDM are written as follows:

$$\begin{aligned} \bar{W} &= QS_c^T(S_cS_c^T + \lambda I)^{-1} \\ \bar{A}_c^Q &= \bar{W}A_c^S \end{aligned} \tag{4}$$

*Corresponding author

2. Split of Oxford-Pets

```
# cls_list: list of class name
import random

def split_dataset(cls_list)
    train_cls_num = 20
    val_cls_num = 7
    test_cls_num = 10
    random.seed(37)

    train_list = random.sample(cls_list, train_cls_num)
    remain_list = [rem for rem in cls_list if rem not in train_list]
    val_list = random.sample(remain_list, val_cls_num)
    test_list = [rem for rem in remain_list if rem not in val_list]

    return train_list, val_list, test_list
```

Listing 1. Python pseudo-code for data split of Oxford-Pets benchmark

split	class
train	american_bulldog, american_pit_bull_terrier, beagle, Bengal, Birman, British_Shorthair, english_cocker_spaniel, english_setter, german_shorthaired, leonberger, miniature_pinscher, Persian, pomeranian, pug, saint_bernard, samoyed, Siamese, Sphynx, staffordshire_bull_terrier, wheaten_terrier
val	great_pyrenees, keeshond, Maine_Coon, newfoundland, Russian_Blue, shiba_inu, yorkshire_terrier
test	Abyssinian, basset_hound, Bombay, boxer, chihuahua, Egyptian_Mau, havanese, japanese_chin, Ragdoll, scottish_terrier

Table 1. Split results of Oxford-Pets

3. Result Table of Fig. 7

Model	Standford-Cars		Standford-Dogs		Oxford-Pets	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
ProtoNet† [3]	47.60±0.21	72.81±0.18	45.92±0.21	67.50±0.17	42.53±0.19	65.66±0.15
+ TDM	55.29±0.23	78.48±0.17	53.47±0.23	72.23±0.17	48.38±0.19	69.64±0.14
DSN† [2]	61.51±0.22	80.21±0.15	54.74±0.22	69.63±0.17	49.51±0.19	64.63±0.15
+ TDM	64.31±0.23	81.29±0.15	56.51±0.22	71.61±0.16	**55.80±0.20**	70.95±0.14
CTX† [1]	65.67±0.22	84.48±0.13	55.66±0.22	73.78±0.16	49.36±0.19	66.33±0.14
+ TDM	**68.36±0.22**	**86.14±0.13**	57.50±0.22	**75.77±0.16**	54.54±0.19	**73.52±0.14**
FRN† [4]	62.07±0.22	83.18±0.14	55.49±0.21	74.54±0.16	50.78±0.19	70.07±0.14
+ TDM	67.10±0.22	86.05±0.12	**57.64±0.22**	75.03±0.16	55.23±0.20	71.34±0.14

Table 2. Performance on Standford-Cars, Standford-Dogs and Oxford-Pets.

In Tab. 2, we report the experimental results on three datasets, Stanford Cars, Stanford Dogs, and Oxford Pets, corresponding to Fig. 7 of the main paper for a precise comparison.

4. Hypothesis test

We perform a hypothesis test (a Wilcoxon signed-rank test with Bonferroni corrections for multiple testing of pairs) for each baseline to show statistical significance. For instance of FRN [4] and FRN with TDM, we validate all 20 experimental results conducted on 7 datasets. Therefore, the α value in the Wilcoxon test is 0.05 / 20 = 0.0025 by the Bonferroni corrections. To show that TDM improves the performances of FRN, the p-value should be lower than 0.0025, when the null hypothesis is that the median is negative. As a result, the p-value is about 1.8e-05, thus the null hypothesis is rejected and it is statistically verified that TDM boosts the performance of FRN. We observe the same results with other baselines. Therefore, we claim that TDM robustly improves all tested baselines.

References

[1] Carl Doersch, Ankush Gupta, and Andrew Zisserman. Crosstransformers: spatially-aware few-shot transfer. *In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[2] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4136–4145, 2020.

[3] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. *In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pages 4077–4087*, 2017.

[4] Davis Wertheimer, Luming Tang, and Bharath Hariharan. Few-shot classification with feature map reconstruction networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8012–8021, 2021.