Multi-directional Knowledge Transfer for Few-Shot Learning

ShuoWang, Xinyu Zhang, Yanbin Hao, Chengbing Wang, Xiangnan He*

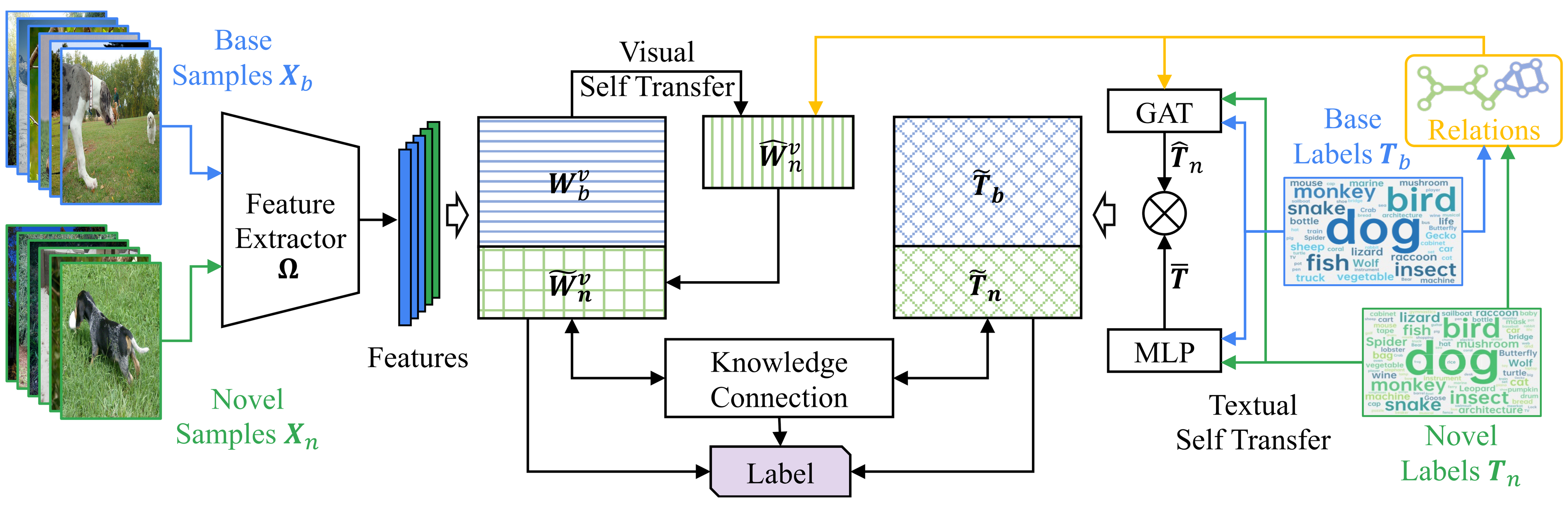
University of Science and Technology of China

shuowangcv@ustc.edu.cn, zhangxy21@mail.ustc.edu.cn, haoyanbin@hotmail.com, wwq197297@mail.ustc.edu.cn , xiangnanhe@gmail.com

Abstract

To reduce the influences of biases in knowledge transfer-based FSL, we propose a multi-directional knowledge transfer (MDKT). Specifically, (1) we use two independent unidirectional knowledge self-transfer strategies to calibrate the distributions of the novel categories from base categories in the visual and the textual space. It aims to yield transferable knowledge of the base categories to describe a novel category. (2) To reduce the inferences of semantic gaps, we first use a bidirectional knowledge connection to exchange the knowledge between the visual and the textual space. Then we adopt an online fusion strategy to enhance the expressions of the textual knowledge and improve the prediction accuracy of the novel categories by combining the knowledge from different modalities. Empirical studies on three FSL benchmark datasets demonstrate the effectiveness of MDKT, which improves the recognition accuracy on novel categories under limited samples.

● Overview of the proposed MDKT



Details: We first employ two different format unidirectional knowledge self-transfer strategies both in the visual and the textual space to learn the potential novel knowledge from the base categories at first. Second, we connect such potential knowledge as bidirectional transfer to exchange information between the different modalities. Finally, the hard labels are used to optimize the whole parameters, which contain the weights of the classifier and the parameters of the transfer network. For the inference stage, we fuse the predictions from different modalities to improve the recognition accuracy of the novel samples.

● Unidirectional Knowledge Self-transfer

➢ Textual Self-Transfer

$$a_{k,m} = \frac{d(\boldsymbol{t}'_k, \boldsymbol{t}_m)}{\sum_{i \in \mathcal{M}^k_{\text{base}}} (d(\boldsymbol{t}'_k, \boldsymbol{t}_i))}, \tag{1}$$

$$\sigma^h_{k,m} = \frac{\exp(\text{LeakyRelu}([\boldsymbol{t}'_k \boldsymbol{W}^h || \boldsymbol{t}_m \boldsymbol{W}^h] \boldsymbol{W}^t_h))}{\sum_{\boldsymbol{t}_i \in \mathcal{M}^k_{\text{base}}} \exp(\text{LeakyRelu}([\boldsymbol{t}'_k \boldsymbol{W}^h || \boldsymbol{t}_i \boldsymbol{W}^h] \boldsymbol{W}^t_h))}, \tag{2}$$

$$\hat{\boldsymbol{T}}^h_n = (\boldsymbol{A} \odot \boldsymbol{\sigma}^h) \boldsymbol{T}^h_b \boldsymbol{W}^h, \tag{3}$$

$$\hat{\boldsymbol{T}}_n = \frac{1}{H} \sum \hat{\boldsymbol{T}}^h_n = \frac{1}{H} \sum (\boldsymbol{A} \odot \boldsymbol{\sigma}^h) \boldsymbol{T}^h_b \boldsymbol{W}^h, \tag{4}$$

$$\bar{\boldsymbol{T}} = \Phi_\theta([\boldsymbol{T}_b || \boldsymbol{T}_n]) = \delta(([\boldsymbol{T}_b || \boldsymbol{T}_n]) \boldsymbol{W}^t_\theta + \boldsymbol{b}^t_\theta), \tag{5}$$

$$\tilde{\boldsymbol{T}} = \text{Conv1D}([\bar{\boldsymbol{T}} || \hat{\boldsymbol{T}}_n]), \tag{6}$$

➢ Visual Self-Transfer

$$p_n = \text{Classifier}(\boldsymbol{x}_n) = \boldsymbol{W}^v \cdot \boldsymbol{x}_n^\top, \tag{8}$$

$$\hat{\boldsymbol{W}}^v_n = \boldsymbol{A} \cdot \boldsymbol{W}^v_b, \tag{9}$$

● Bidirectional Knowledge Connection

$$C = \sum_{c \in C_{\text{base}}} ||\boldsymbol{W}^v_c - \boldsymbol{T}_c||_2. \tag{11}$$

● Total Loss

$$\mathcal{L}^t = \frac{1}{B} \sum_{i=1}^{B} \text{CE}(\text{softmax}(\tilde{\boldsymbol{T}} \cdot \boldsymbol{x}_i^\top), l_i), \tag{7}$$

$$\mathcal{L}^v = \frac{1}{B} \sum_{i=1}^{B} \text{CE}(\text{softmax}(\tilde{\boldsymbol{W}}^v \cdot \boldsymbol{x}_i^\top), l_i), \tag{10}$$

$$\mathcal{L}^c = \frac{1}{|C_{\text{base}}| + |C_{\text{novel}}|} \Big(\sum_{c \in \{C_{\text{base}} + C_{\text{novel}}\}} ||\tilde{\boldsymbol{W}}^v_c - \tilde{\boldsymbol{T}}_c||_2 \Big). \tag{12}$$

$$\mathcal{L}^m = \frac{1}{B} \sum_{i=1}^{B} \text{CE}(\text{softmax}((\tilde{\boldsymbol{W}}^v + \lambda \tilde{\boldsymbol{T}})) \cdot \boldsymbol{x}_i^\top), l_i), \tag{13}$$

$$\mathcal{L} = \mathcal{L}^v + \mathcal{L}^t + \mathcal{L}^m + \mu \mathcal{L}^c, \tag{14}$$

Experiments

The evaluation of the textual self-transfer

Training under *	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$
$\mathcal{L}^t$ w/o Transfer	60.8	68.5	75.5	77.9	80.2
$\mathcal{L}^t$	**60.9**	**68.6**	**76.0**	**79.6**	**81.5**

The evaluation of the visual self-transfer

Training under *	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$
$\mathcal{L}^v$ w/o Transfer	52.2	64.6	75.6	**80.2**	**82.9**
$\mathcal{L}^v$	**56.1**	**66.4**	**75.8**	80.1	82.7

The evaluation of the bidirectional knowledge connection

Training under *		$K=1$	$K=2$	$K=5$	$K=10$	$K=20$
(1)	$\mathcal{L}^t + \mathcal{L}^v$	61.3	69.3	77.0	80.5	82.8
(2)	$\mathcal{L}^m$	61.4	68.9	76.9	80.5	82.9
(3)	$\mathcal{L}$	**61.8**	**69.6**	**77.1**	**80.7**	**83.5**

The evaluation of retraining strategy

Method	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$
First Training	61.8	69.6	77.1	80.7	83.4
Second Training	**62.5**	**70.0**	**77.2**	**80.8**	83.4

Evaluation under ImageNet-FS

	Method with ResNet-10					Method with ResNet-50				
	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$	$K=1$	$K=2$	$K=5$	$K=10$	$K=20$
Prototypical Nets [34]	39.3	54.4	66.3	71.2	73.9	49.5	59.9	70.1	75.1	77.6
Matching Networks [39]	43.6	54.0	66.0	72.5	76.9	49.6	64.0	74.4	78.1	80.0
SGM + Hallucination[16]	44.3	56.0	69.7	**75.3**	**78.6**	52.8	64.4	77.3	**82.0**	**84.9**
wDAE-GNN [11]	48.0	59.7	70.3	75.0	77.8	——	——	——	——	——
KTCH [23]	——	——	——	——	——	58.1	67.3	**77.6**	81.8	84.2
IDeMe-Net [3]	51.0	60.9	70.4	73.4	75.1	60.1	69.6	77.4	80.2	——
KTN [26]	54.7	61.7	70.4	75.0	77.9	61.9	68.7	76.4	80.1	82.4
Our MDKT	**55.2**	**63.2**	**70.8**	75.0	78.2	**62.6**	**70.1**	**77.6**	81.5	83.7

Evaluation under ImNet

Method	Novel Categories				
	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
NN (from [23])	34.2	43.6	48.7	52.3	54.0
PPA [27]	33.0	43.1	48.5	52.5	55.4
LSD [5]	33.2	44.7	50.2	53.4	57.6
KTCH [23]	39.0	48.9	54.9	58.7	60.5
KGTN [1]	42.5	50.3	55.4	58.4	60.7
Our Baseline	36.1	47.9	54.0	58.1	60.8
Our MDKT	**44.4**	**53.3**	**58.1**	**61.7**	**63.8**

Evaluation under Mini-ImageNet

Method	$K=1$	$K=5$
Meta-Baseline [2]	63.17 ± 0.23%	79.26 ± 0.17%
MetaFun [48]	64.13 ± 0.13%	80.82 ± 0.17%
P-Transfer [33]	64.21 ± 0.77%	80.38 ± 0.59%
MMKD [42]	64.40 ± 0.43%	83.05 ± 0.28%
IEPT [52]	67.05 ± 0.44%	82.90 ± 0.30%
FRN [44]	66.45 ± 0.19%	82.83 ± 0.13%
BML [54]	67.04 ± 0.63%	**83.63** ± 0.29%
Our MDKT	**67.39** ± 0.76%	82.25 ± 0.53%