IMAGE-TO-VIDEO RE-IDENTIFICATION VIA MUTUAL DISCRIMINATIVE KNOWLEDGE TRANSFER

Pichao Wang, Fan Wang, Hao Li

Alibaba Group

icassp 2022 Singapore

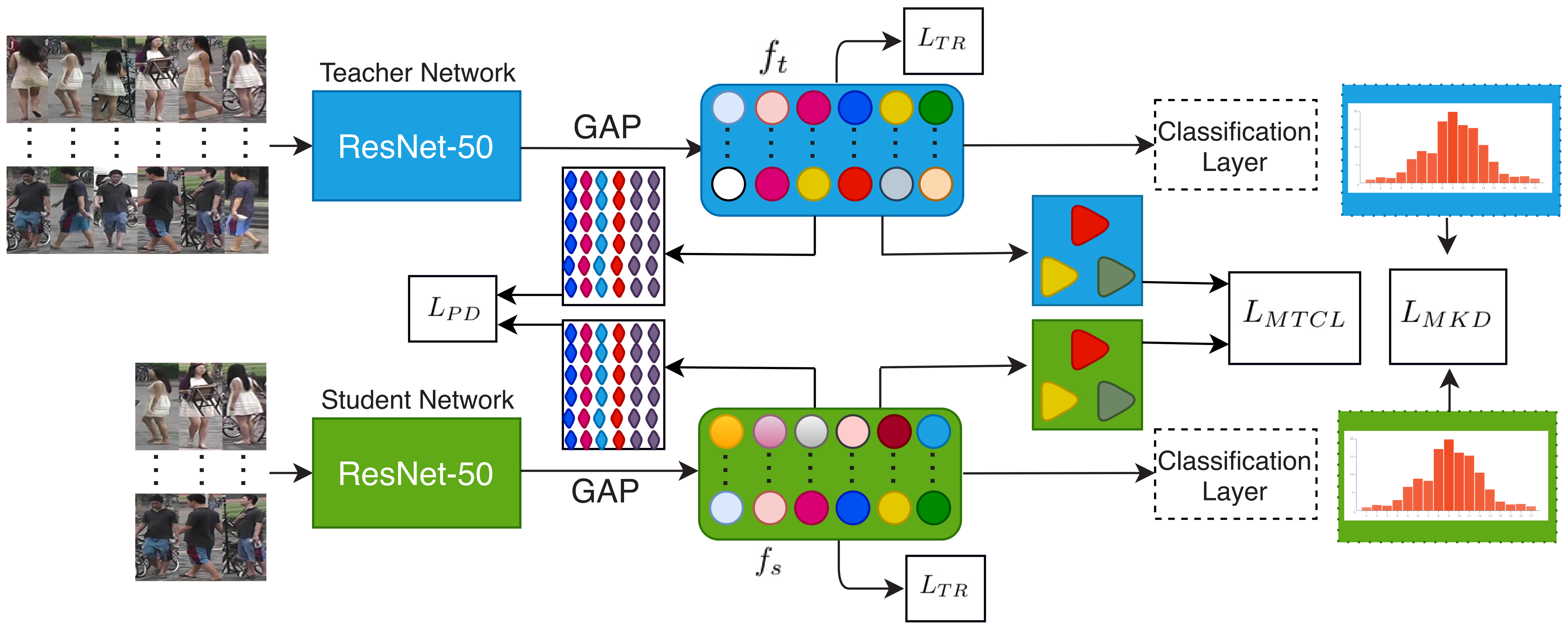
poster number:1148

Contribution

Key Contributions:

- We propose a novel triplet contrast loss that is specifically designed for discriminative feature transfer in the context of knowledge distillation, together with hard triplet sampling. The proposed loss is complementary to conventional KD losses and can be combined with them to further boost the performance.
- Mutual learning is adopted to regularize both the teacher and student networks training, and it shows its effectiveness under our proposed framework.
- State-of-the-art results are achieved on three commonly used datasets, and verify the generation of our proposed method on both person and vehicle re-identification.

Method



The framework of MDKT consists of two-stages: 1) the teacher network is trained using the standard V2V Re-ID setting. 2) we feed frames representing different numbers of views as input to the teacher and students networks for view KD using three level distillation losses, as depicted in Figure.

Mutual Discriminative Knowledge Transfer:

- Mutual Logits Distillation:

$$L_{KD_{t2s}} = \tau_1^2 KL(y_t \| y_s) \quad L_{KD_{s2t}} = \tau_1^2 KL(y_s \| y_t) \quad L_{MKD} = L_{KD_{t2s}} + L_{KD_{s2t}} \tag{1}$$

- Pairwise Distance in Embedding:

$$L_{PD} = \sum_{(i,j)\in\binom{B}{2}} (D_t[i,j] - D_s[i,j])^2 \tag{2}$$

- Triplet Contrast Loss: we propose to measure the probability of the two distances.

$$p_{apn_{\tau_2}} = \frac{\exp(-d^t_{a2p}/\tau_2)}{\exp(-d^t_{a2p}/\tau_2) + \exp(-d^t_{a2n}/\tau_2)} \tag{3}$$

Define the distribution $P^t_{apn_{\tau_2}} = [p^t_{apn_{\tau_2}}, 1 - p^t_{apn_{\tau_2}}]$ and $P^s_{apn_{\tau_2}} = [p^s_{apn_{\tau_2}}, 1 - p^s_{apn_{\tau_2}}]$ and the TCL loss between teacher and student is formulated as:

$$L_{TCL_{t2s}} = \sum_{a,p,n}^{N} KL(P^t_{apn_{\tau_2}} \| P^s_{apn_{\tau_2}}) \tag{4}$$

Experiments & Results

Models	Losses				MARS				Duke-video			
					I2V		V2V		I2V		V2V	
	L_{TR}	L_{MKD}	L_{PD}	L_{MTCL}	cmc1	mAP	cmc1	mAP	cmc1	mAP	cmc1	mAP
TR	√	×	×	×	76.77	66.85	84.55	74.23	78.24	70.66	88.24	84.96
TCL	×	×	×	√	80.71	71.56	86.82	78.04	82.69	79.26	93.38	92.01
TR+TCL	√	×	×	√	81.16	72.91	86.36	78.68	83.65	80.32	95.01	93.22
KD+PD+TCL	×	√	√	√	84.70	77.56	89.19	82.53	86.32	84.57	**95.58**	93.94
TR+KD+PD	√	√	√	×	83.96	77.43	88.89	82.47	85.04	83.97	95.01	93.69
TR+KD+TCL	√	√	×	√	83.59	76.28	88.69	81.83	84.90	83.89	94.87	93.56
TR+PD+TCL	√	×	√	√	85.33	77.90	89.01	82.65	84.90	83.74	95.30	**93.94**
ALL	√	√	√	√	**85.65**	**78.02**	**89.48**	**82.90**	**86.78**	**84.82**	95.26	93.83

Table 1. Ablation study on the impact of loss terms on MARS and Duke-video datasets using ResNet-50.

Models	MARS			
	I2V		V2V	
	cmc1	mAP	cmc1	mAP
freeze teacher	85.10	77.65	89.44	82.79
without mutual	85.33	77.77	89.22	82.80
with mutual	**85.65**	**78.02**	**89.48**	**82.90**
	Duke-video			
	I2V		V2V	
	cmc1	mAP	cmc1	mAP
freeze teacher	86.65	84.58	95.09	93.70
without mutual	86.63	84.72	95.10	93.51
with mutual	**86.78**	**84.82**	**95.26**	**93.83**

Table 2. Ablation study on the impact of mutual learning.

Method	top1	top5	mAP
P2SNet [12]	55.3	72.9	-
TMSL [13]	56.5	70.6	-
TKP [1]	75.6	87.6	65.1
STE-NVAN [14]	80.3	-	68.8
NVAN [14]	80.1	-	70.2
MGAT [15]	81.1	92.2	71.8
READ [16]	81.5	92.1	70.4
VKD [2]	83.9	93.2	77.3
MDKT	**85.7**	**93.3**	**78.0**

Table 4. Comparison with SOTA methods on MARS dataset.

Method	top1	top5	mAP
PROVID [17]	76.8	91.4	48.5
VFL-LSTM [18]	88.0	94.6	59.2
RAM [19]	88.6	-	61.5
VANet [20]	89.8	96.0	66.3
PAMTRI [21]	92.9	92.9	71.9
SAN [22]	93.3	97.1	72.5
VKD [2]	95.2	98.0	82.2
MDKT	**96.0**	**99.3**	**83.4**

Table 6. Comparison with SOTA methods on VeRi dataset.

Method	top1	top5	mAP
STE-NVAN [14]	42.2	-	41.3
TKP [1]	77.9	-	75.9
NVAN [14]	78.4	-	76.7
VKD [2]	85.6	93.9	83.8
READ [16]	86.3	94.4	83.4
MDKT	**86.8**	**94.9**	**84.8**

Table 5. Comparison with SOTA methods on Duke-video dataset.

Conclusion & Contact Information

In this paper, we propose a mutual discriminative knowledge transfer method for I2V ReID. The proposed method takes advantage of triplet for local discriminative feature learning and aligns the heterogeneous outputs of teacher and student networks. Coupled with the mutual learning, the proposed method achieves state-of-the-art results on three datasets, covering person and vehicle re-identification.

Contact person:Pichao Wang; Email:pichao.wang@alibaba-inc.com;
Wechat:wangpichao; Homepage:https://wangpichao.github.io/.