Image-Image Domain Adaptation with Preserved Self-Similarity and Domain-Dissimilarity for Person Re-identification

Weijian Deng[†], Liang Zheng[‡§], Qixiang Ye[†], Guoliang Kang[‡], Yi Yang[‡], Jianbin Jiao[†*]
[†]University of Chinese Academy of Sciences [‡]University of Technology Sydney
[§]Singapore University of Technology and Design

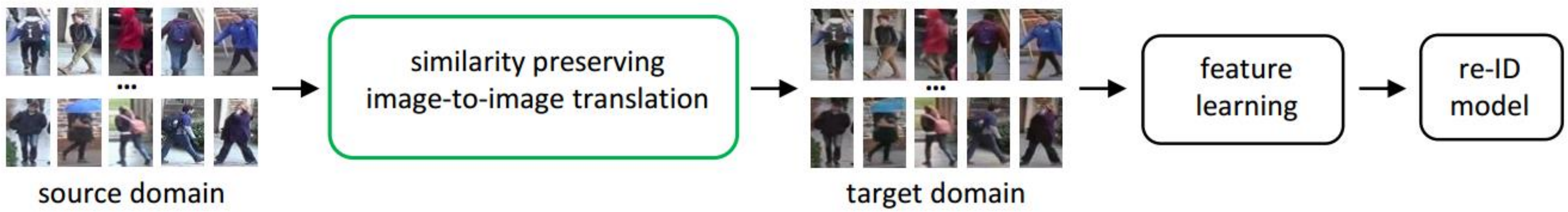
CVPR 2018

Motivation

- Poor performance while transferring Person Re-ID models to a new domain
  - Source domain: labeled data
  - Target domain: unlabeled data
  - Source → Target

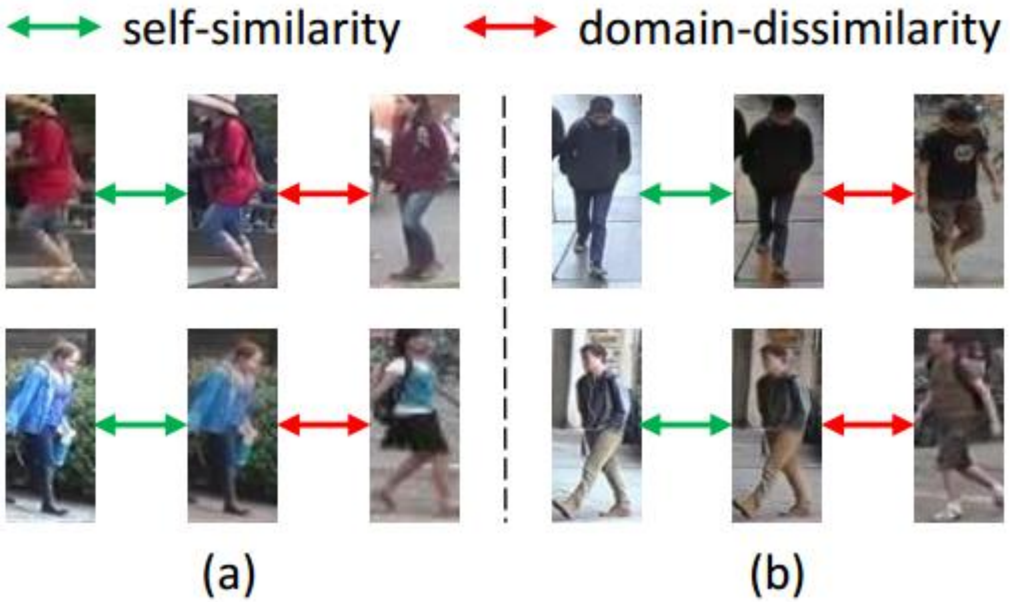
Transfer Learning by Image Translation

- Solve this problem by image translation:
  - Translate the source images to the ones with the style of target domains.
  - Achieve a collection of labeled generated images in target domains
  - Supervised Training on this generated collection.



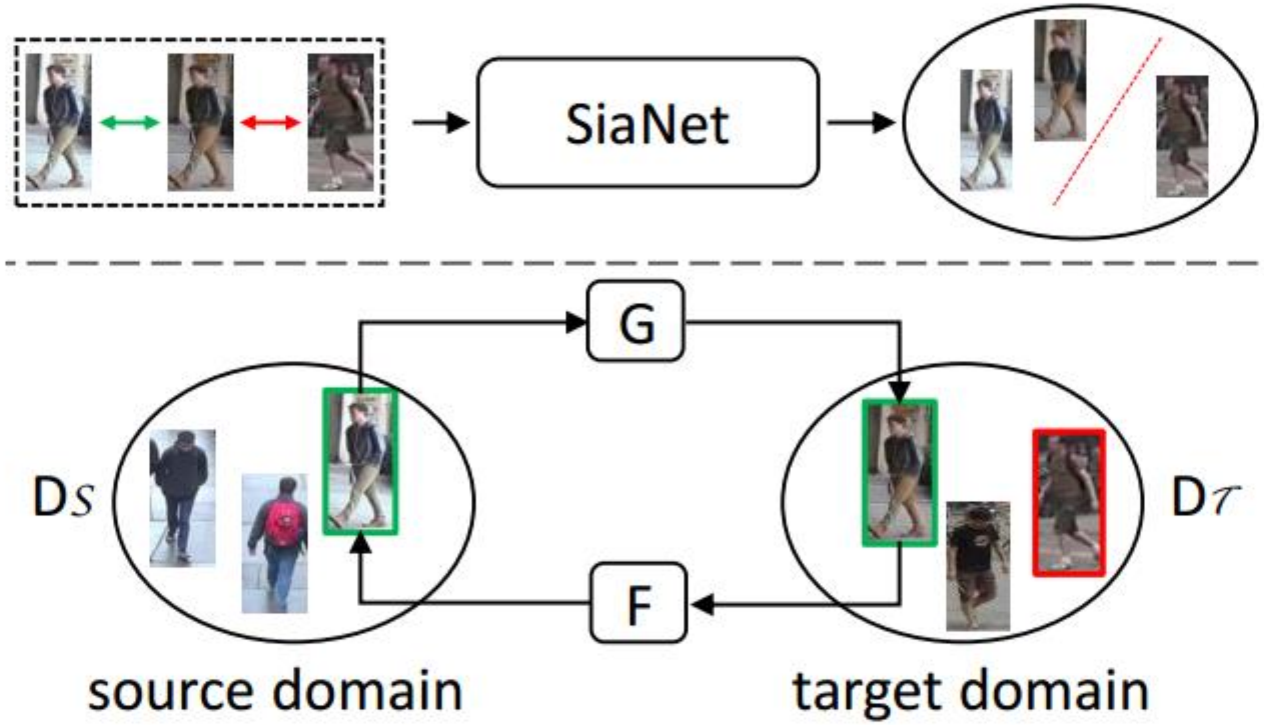
Advantages

- Compared with other translation based methods, two constraints are considered in this model:
  - Self-similarity of an image before and after translation.
  - Domain-dissimilarity of a translated source image and a target image.



Models

- Translation model: **SPGAN** (CycleGAN + Similarity preserving loss)
- Supervised Training Model: SiaNet



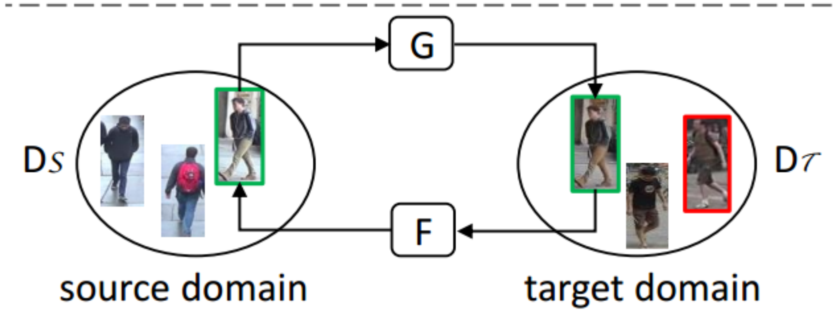
CycleGAN

$$\mathcal{L}_{\mathcal{T}adv}(G, D_{\mathcal{T}}, p_x, p_y) = \mathbb{E}_{y \sim p_y}[(D_{\mathcal{T}}(y) - 1)^2] + \mathbb{E}_{x \sim p_x}[(D_{\mathcal{T}}(G(x))^2],$$

$$\mathcal{L}_{\mathcal{S}adv}(F, D_{\mathcal{S}}, p_y, p_x) = \mathbb{E}_{x \sim p_x}[(D_{\mathcal{S}}(x) - 1)^2] + \mathbb{E}_{y \sim p_y}[(D_{\mathcal{S}}(F(y))^2].$$

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_x}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_y}[\|G(F(y)) - y\|_1].$$

$$\mathcal{L}_{ide}(G, F, p_x, p_y) = \mathbb{E}_{x \sim p_x}\|F(x) - x\|_1 + \mathbb{E}_{y \sim p_y}\|G(y) - y\|_1.$$



SPGAN

Similarity preserving loss function. We utilize the contrastive loss [16] to train SiaNet:

$$\mathcal{L}_{con}(i, x_1, x_2) = (1 - i)\{\max(0, m - d)\}^2 + id^2, \quad (5)$$

Overall objective function. The final SPGAN objective can be written as

$$\mathcal{L}_{sp} = \mathcal{L}_{\mathcal{T}adv} + \mathcal{L}_{\mathcal{S}adv} + \lambda_1 \mathcal{L}_{cyc} + \lambda_2 \mathcal{L}_{ide} + \lambda_3 \mathcal{L}_{con}, \quad (6)$$

SIANET

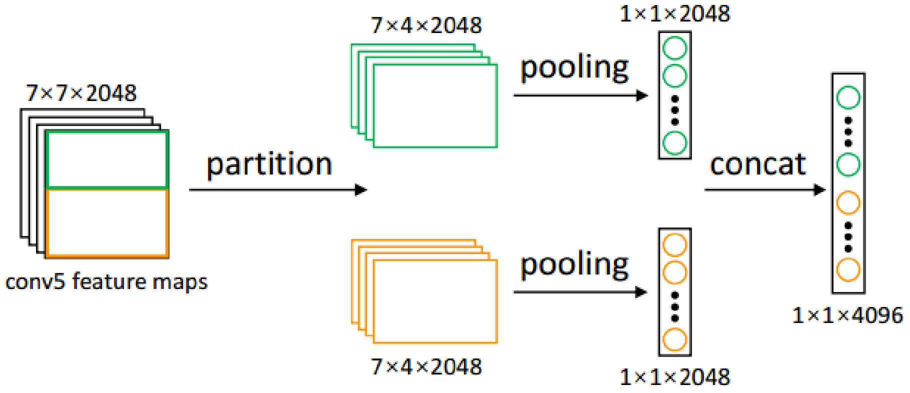


Figure 5: Illustration of LMP. We partition the feature map into P $(P = 2)$ parts horizontally. We conduct global max/avg pooling on each part and concatenate the feature vectors as the final representation.

- we introduce a feature pooling method named as local max pooling (LMP).
- It can reduce the impact of noisy signals incurred by the fake translated images.

Experiments

Figure 6: Sample images of (upper left:) DukeMTMC-reID dataset, (lower left:) Market-1501 dataset, (upper right:) Duke images which are translated to Market style, and (lower right:) Market images translated to Duke style. We use SPGAN for unpaired image-image translation.

Experiments

Methods	DukeMTMC-reID					Market-1501				
	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP
Supervised Learning	66.7	79.1	83.8	88.7	46.3	75.8	89.6	92.8	95.4	52.2
Direct Transfer	33.1	49.3	55.6	61.9	16.7	43.1	60.8	68.1	74.7	17.0
CycleGAN (basel.)	38.1	54.4	60.5	65.9	19.6	45.6	63.8	71.3	77.8	19.1
CycleGAN (basel.) + L_{ide}	38.5	54.6	60.8	66.6	19.9	48.1	66.2	72.7	80.1	20.7
SPGAN ($m = 0$)	37.7	53.1	59.5	65.6	20.0	49.2	66.9	74.0	80.0	20.5
SPGAN ($m = 1$)	39.5	55.0	61.4	67.3	21.0	48.7	65.7	73.0	79.3	21.0
SPGAN ($m = 2$)	41.1	56.6	63.0	69.6	22.3	51.5	70.1	76.8	82.4	22.8
SPGAN ($m = 2$) + LMP	**46.9**	**62.6**	**68.5**	**74.0**	**26.4**	**58.1**	**76.0**	**82.7**	**87.9**	**26.9**

Experiments

Methods	Market-1501				
	Setting	Rank-1	Rank-5	Rank-10	mAP
Bow [51]	SQ	35.8	52.4	60.3	14.8
LOMO [26]	SQ	27.2	41.6	49.1	8.0
UMDL [35]	SQ	34.5	52.6	59.6	12.4
PUL [6]*	SQ	45.5	60.7	66.7	20.5
Direct transfer	SQ	43.1	60.8	68.1	17.0
Direct transfer	MQ	47.9	65.5	73.0	20.6
CAMEL [49]	MQ	54.5	-	-	26.3
SPGAN	SQ	51.5	70.1	76.8	22.8
SPGAN	MQ	57.0	73.9	80.3	27.1
SPGAN+LMP	SQ	**58.1**	**76.0**	**82.7**	**26.9**

Table 4: Comparison with state-of-the-art on Market-1501. * denotes unpublished papers. "SQ" and "MQ" are the single-query and multiple-query settings, respectively. The best results are in **bold**.

Methods	DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP
Bow [51]	17.1	28.8	34.9	8.3
LOMO [26]	12.3	21.3	26.6	4.8
UMDL [35]	18.5	31.4	37.6	7.3
PUL [6]*	30.0	43.4	48.5	16.4
Direct transfer	33.1	49.3	55.6	16.7
SPGAN	41.1	56.6	63.0	22.3
SPGAN+LMP	**46.9**	**62.6**	**68.5**	**26.4**

Table 5: Comparison with state-of-the-art on DukeMTMC-reID under the single-query setting. * denotes unpublished papers. The best results are in **bold**.

Experiments

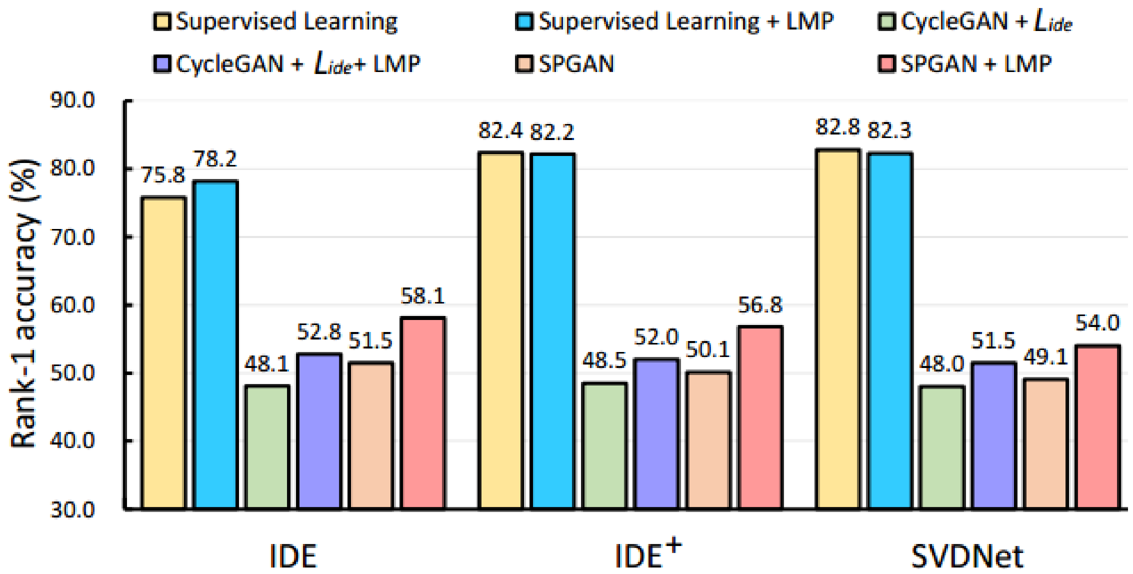


Figure 9: Experiment of LMP ($P = 7$) on scenarios of supervised learning and domain adaptation with SPGAN and Cycle + L_{ide}. Three feature learning methods are compared, *i.e.*, IDE [52], IDE$^+$ [55], and SVDNet [39]. The results are on Market-1501.

Experiments

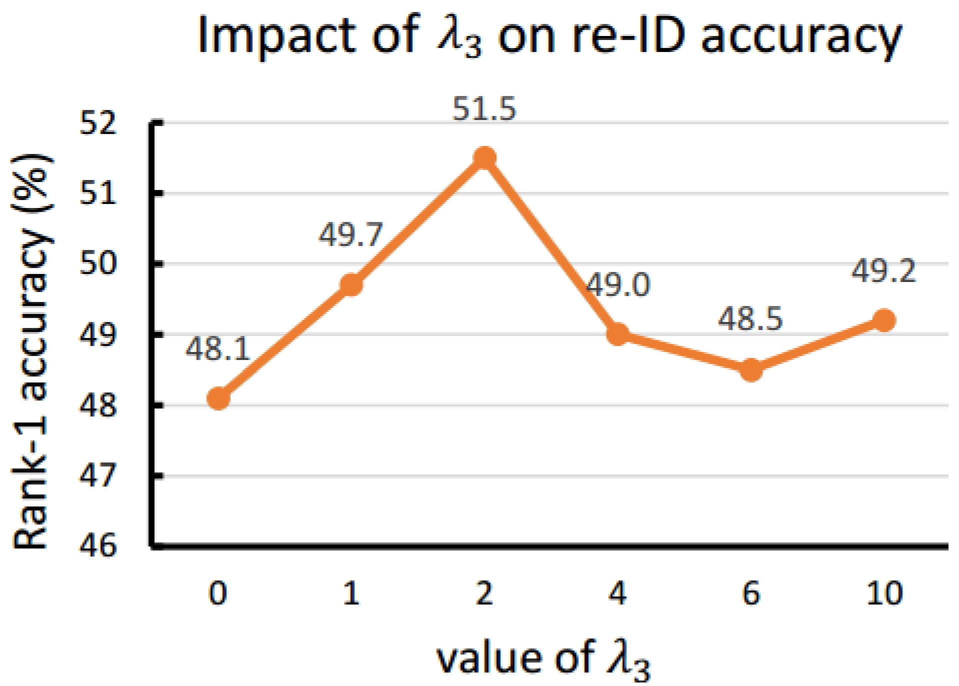


Figure 8: λ_3 (Eq. 6) *v.s* re-ID accuracy. A larger λ_3 means larger weight of similarity preserving constraint.

Thanks