Joint Discriminative and Generative Learning for Person Re-identification

Zhedong Zheng[1,2] Xiaodong Yang[1] Zhiding Yu[1] Liang Zheng[3] Yi Yang[2] Jan Kautz[1]

[1]NVIDIA [2]CAI, University of Technology Sydney [3]Australian National University

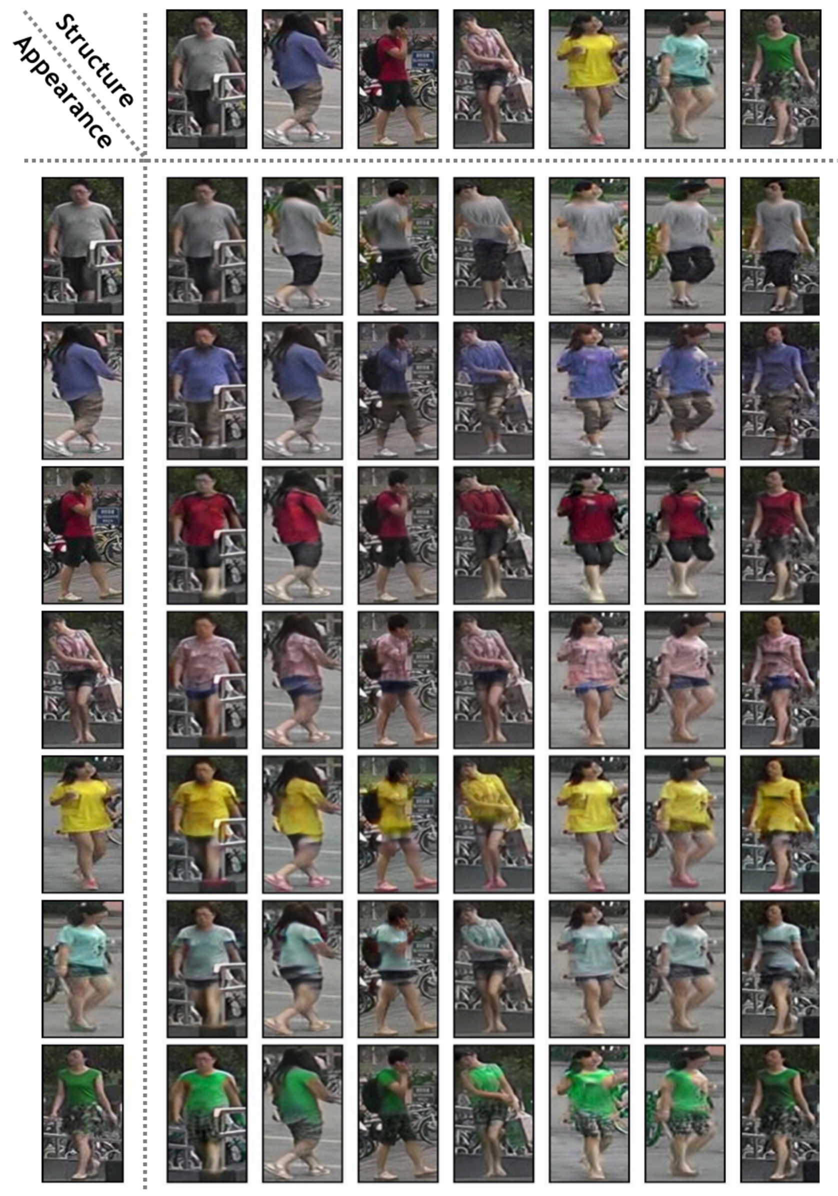
3-min video

1. Motivation

- Training data is one of the keys to deep learning. How to generate more high-fidelity images from the original data? How to better make use of the generated images for training?
- Image generation and discriminative learning are highly-related. Can we mutually benefit the discriminative and generative learning tasks?

Discriminative ⟷ Generative

2. Contributions



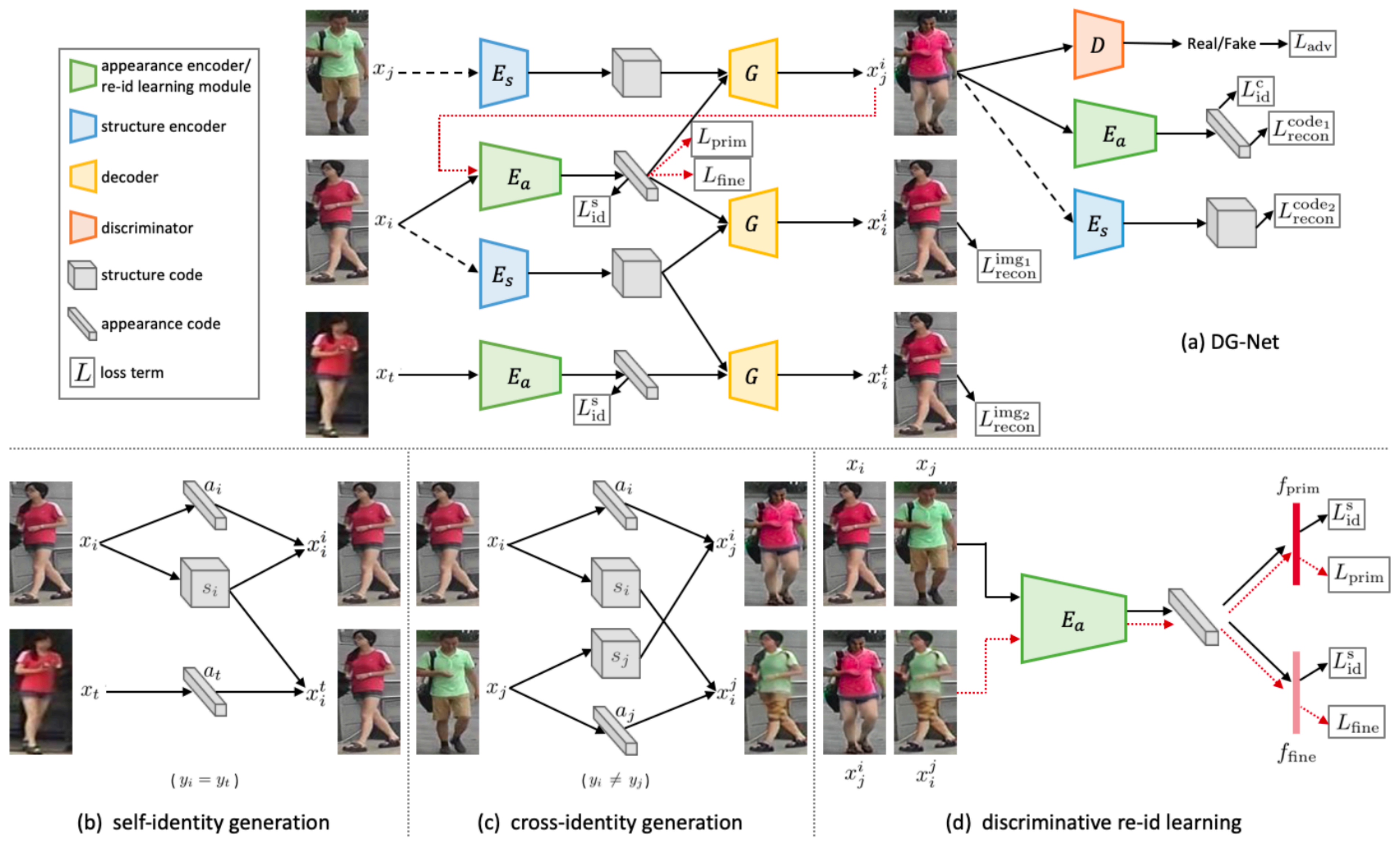
- Given N images, we can generate NxN high-fidelity images for training and therefore let the model see more realistic variants to boost re-id learning.
- We end-to-end couple image generation and re-id learning in a single unified network.

3. Method

- Define two spaces for pedestrian images

Appearance Space	Structure Space
clothing/shoes color, texture and style, other id-related cues, etc.	body size, hair, carrying, pose, background, position, viewpoint, etc.

- Overview of DG-Net



- Objectives

self-identity generation

$$L^{\text{img}_1}_{\text{recon}} = \mathbb{E}[\|x_i - G(a_i, s_i)\|_1]. \qquad L^{\text{img}_2}_{\text{recon}} = \mathbb{E}[\|x_i - G(a_t, s_i)\|_1].$$

cross-identity generation

$$L^{\text{code}_1}_{\text{recon}} = \mathbb{E}[\|a_i - E_a(G(a_i, s_j))\|_1], \quad L^{\text{code}_2}_{\text{recon}} = \mathbb{E}[\|s_j - E_s(G(a_i, s_j))\|_1].$$

$$L_{\text{adv}} = \mathbb{E}[\log D(x_i) + \log(1 - D(G(a_i, s_j)))].$$

discriminative learning

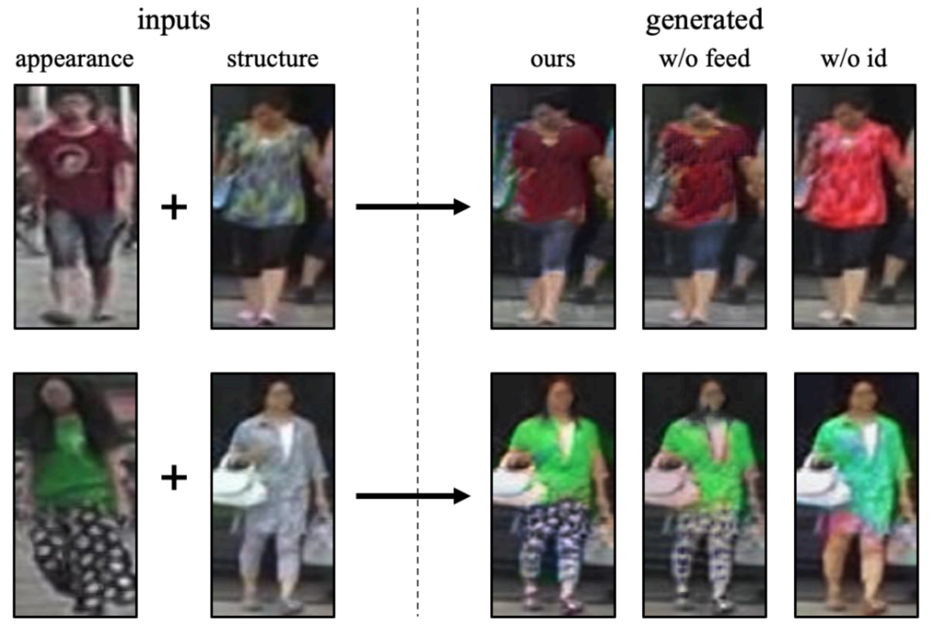
$$L^{\text{s}}_{\text{id}} = \mathbb{E}[-\log(p(y_i|x_i))], \qquad L_{\text{fine}} = \mathbb{E}[-\log(p(y_j|x^i_j))].$$

$$L_{\text{prim}} = \mathbb{E}[-\sum_{k=1}^{K} q(k|x^i_j) \log(\frac{p(k|x^i_j)}{q(k|x^i_j)})],$$

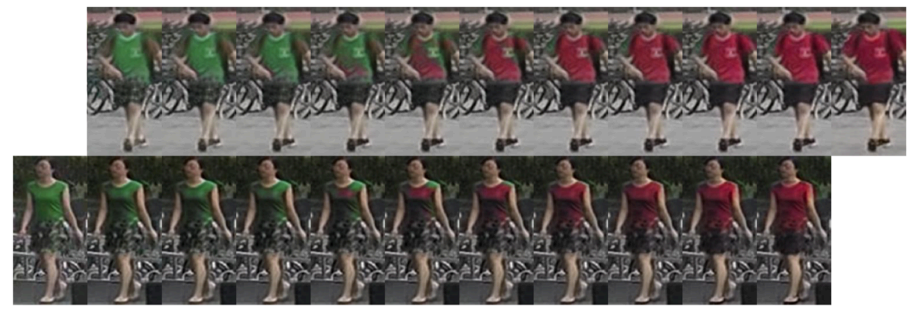
4. Experiments

- Generative evaluations

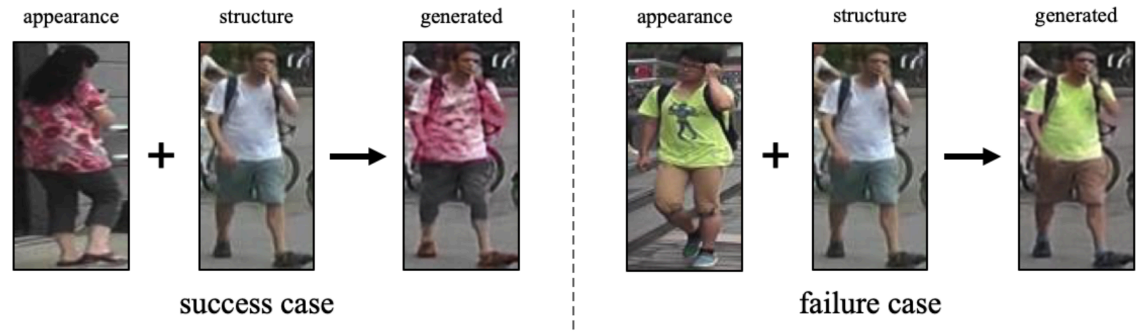
Comparison of the generated and real images on Market-1501 across different methods.



Comparison of the generated images by our full model, removing online feeding (w/o feed), and further removing identity supervision (w/o id).



Example of image generation by linear interpolation between two appearance codes.



Example of success and failure cases.

- Discriminative evaluations

Methods	Market-1501		DukeMTMC-reID		MSMT17	
	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP
Baseline	89.6	74.5	82.0	65.3	68.8	36.2
f_{prim}	94.0	84.4	85.6	72.7	76.0	49.7
f_{fine}	91.6	75.3	78.7	61.2	71.5	43.5
$f_{\text{prim}}, f_{\text{fine}}$	**94.8**	**86.0**	**86.6**	**74.8**	**77.2**	**52.3**

Comparison of the baseline and learned features.

Methods	Rank@1	Rank@5	Rank@10	mAP
Deep [40]	47.6	65.0	71.8	23.0
PDC [35]	58.0	73.6	79.4	29.7
Verif-Identif [55]	60.5	76.2	81.6	31.6
GLAD [47]	61.4	76.8	81.6	34.0
PCB [39]	68.2	81.2	85.5	40.4
Ours	**77.2**	**87.4**	**90.5**	**52.3**

Comparison with the state-of-the-arts on MSMT17.

Methods	Market-1501		DukeMTMC-reID	
	Rank@1	mAP	Rank@1	mAP
Verif-Identif [55]	79.5	59.9	68.9	49.3
DCF [22]	80.3	57.5	-	-
SSM [2]	82.2	68.8	-	-
SVDNet [38]	82.3	62.1	76.7	56.8
PAN [57]	82.8	63.4	71.6	51.5
GLAD [47]	89.9	73.9	-	-
HA-CNN [24]	91.2	75.7	80.5	63.8
MLFN [4]	90.0	74.3	81.0	62.8
Part-aligned [37]	91.7	79.6	84.4	69.3
PCB [39]	93.8	81.6	83.3	69.2
Mancs [43]	93.1	82.3	84.9	71.8
DeformGAN [34]	80.6	61.3	-	-
LSRO [56]	84.0	66.1	67.7	47.1
Multi-pseudo [17]	85.8	67.5	76.8	58.6
PT [27]	87.7	68.9	78.5	56.9
PN-GAN [31]	89.4	72.6	73.6	53.2
FD-GAN [10]	90.5	77.7	80.0	64.5
Ours	**94.8**	**86.0**	**86.6**	**74.8**

Comparison with the state-of-the-arts on Market and Duke.