Representation Learning in Person Re-ID

Depu Meng

2019.01.19

Content

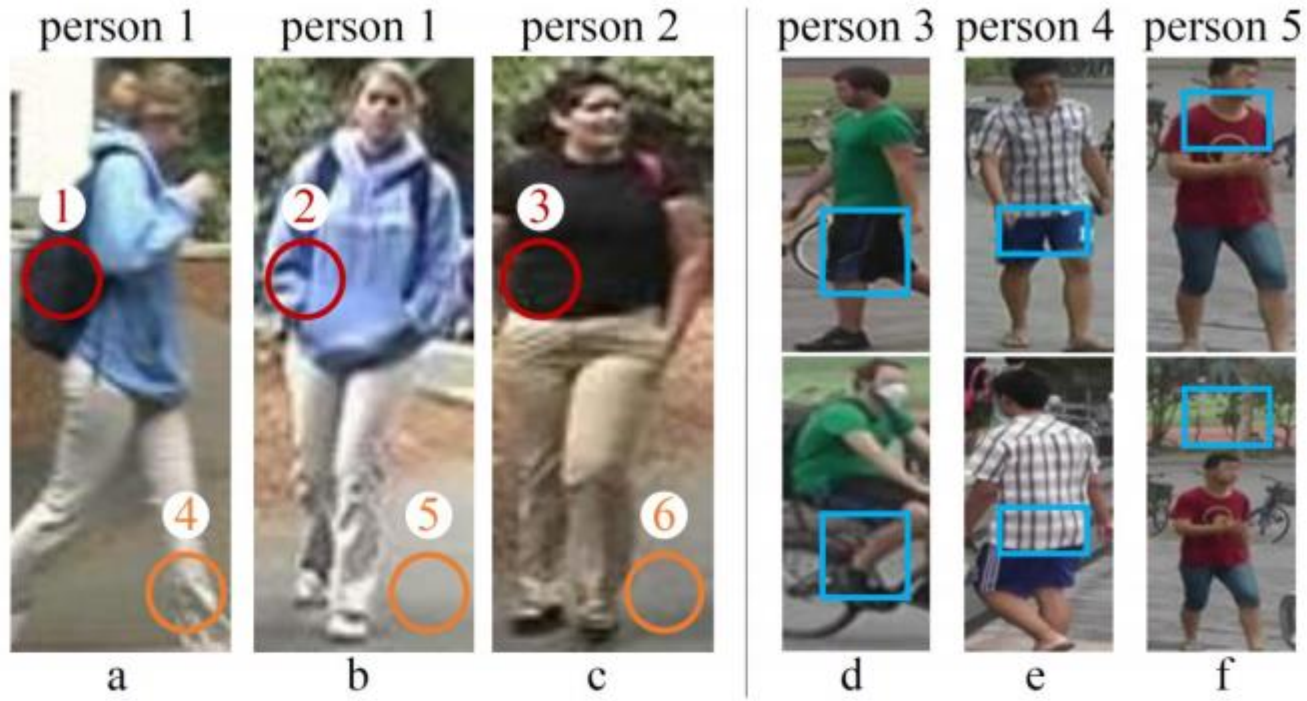
- Part-aligned Representation Learning
- Image generation in Re-ID
- GAN as supervisor

Content

- Part-aligned Representation Learning
- Image generation in Re-ID
- GAN as supervisor

Part-aligned representation

- Person Re-ID is based on comparison of body parts
- Body parts are misaligned in the most cases



Related Works

- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017
- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

Related Works

- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015 w/o pose
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017
- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

Related Works

- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017

w/o pose

- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

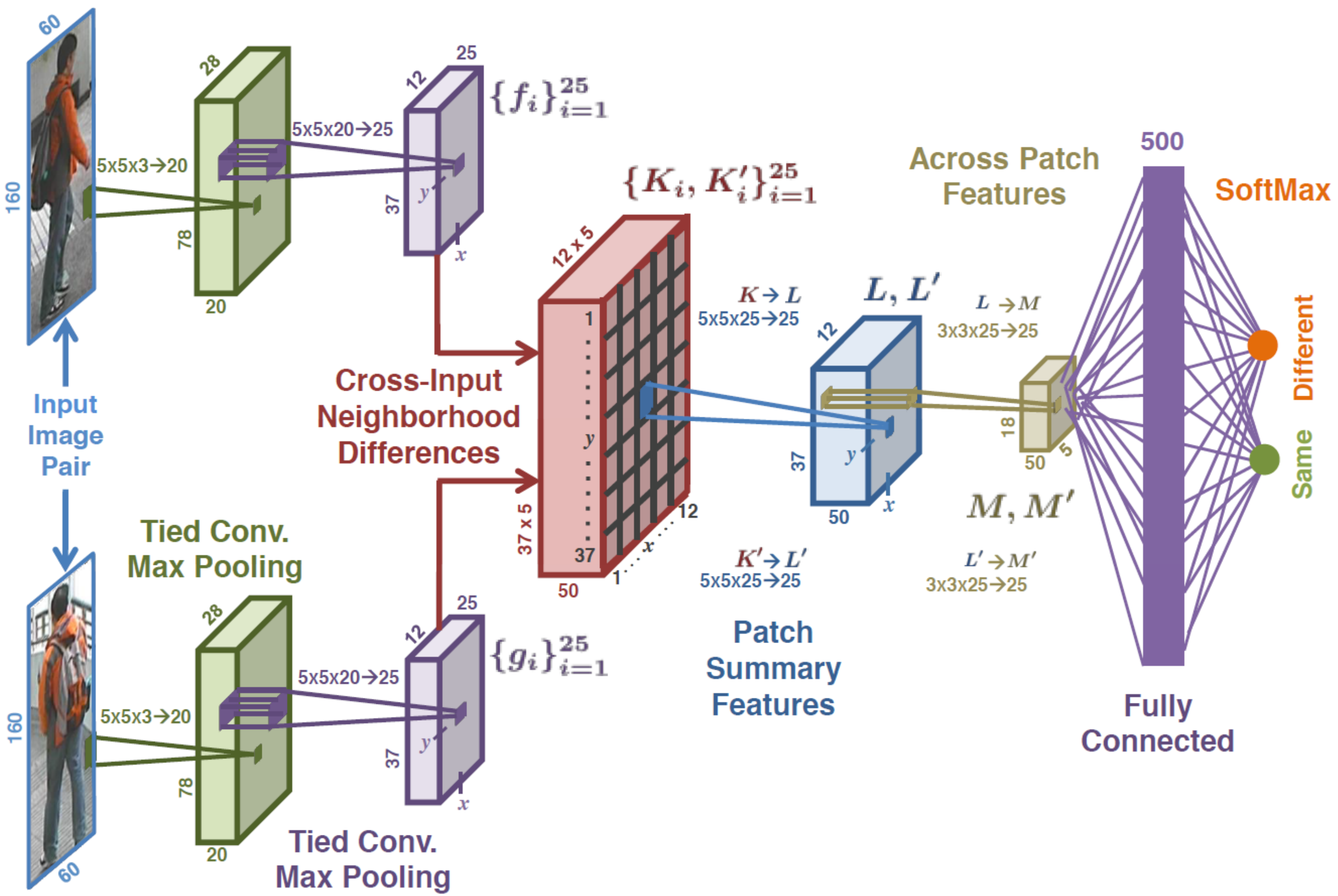
w/ pose

Related Works

- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015 w/o pose
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017
- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018 w/ pose
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

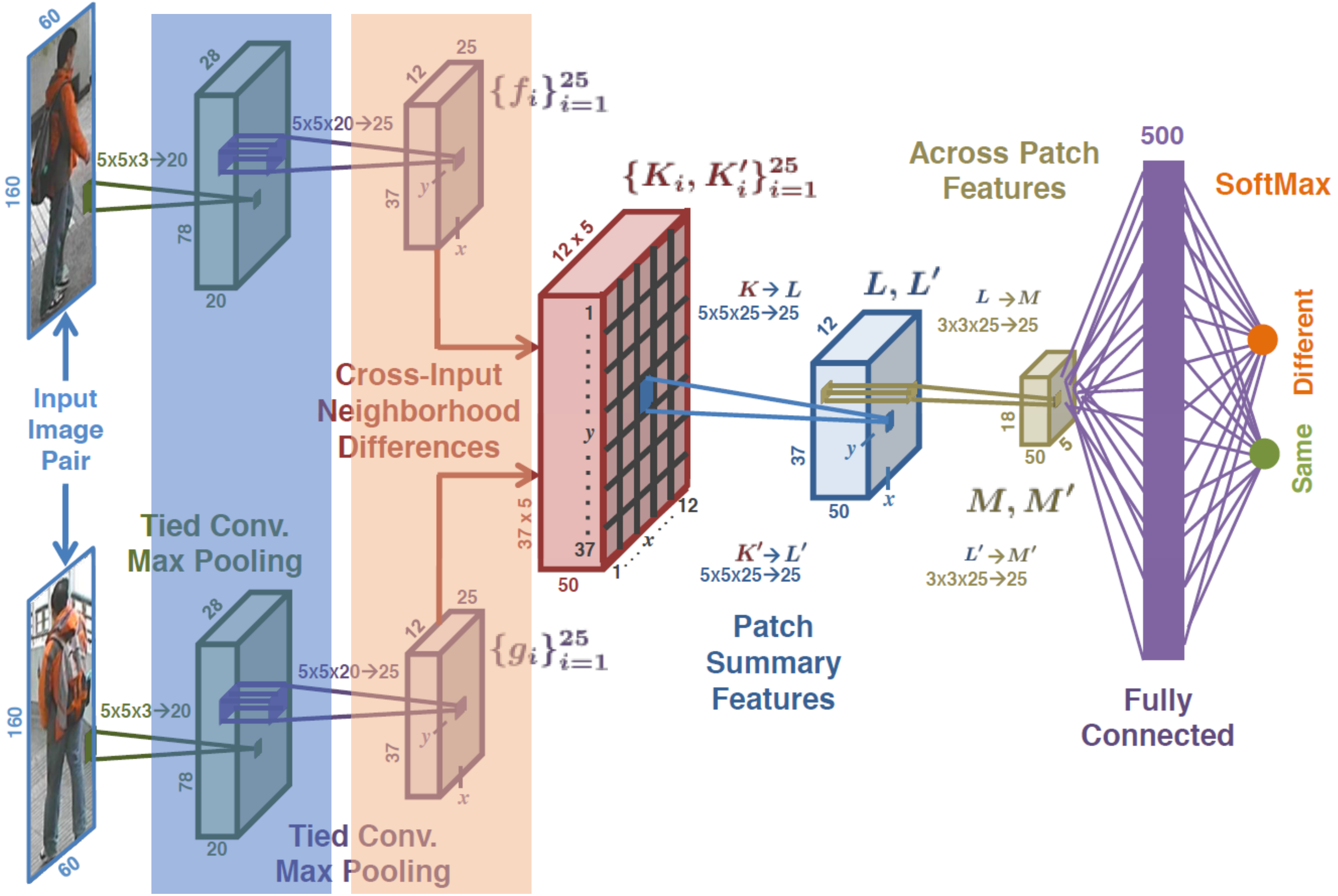
Architecture

Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015

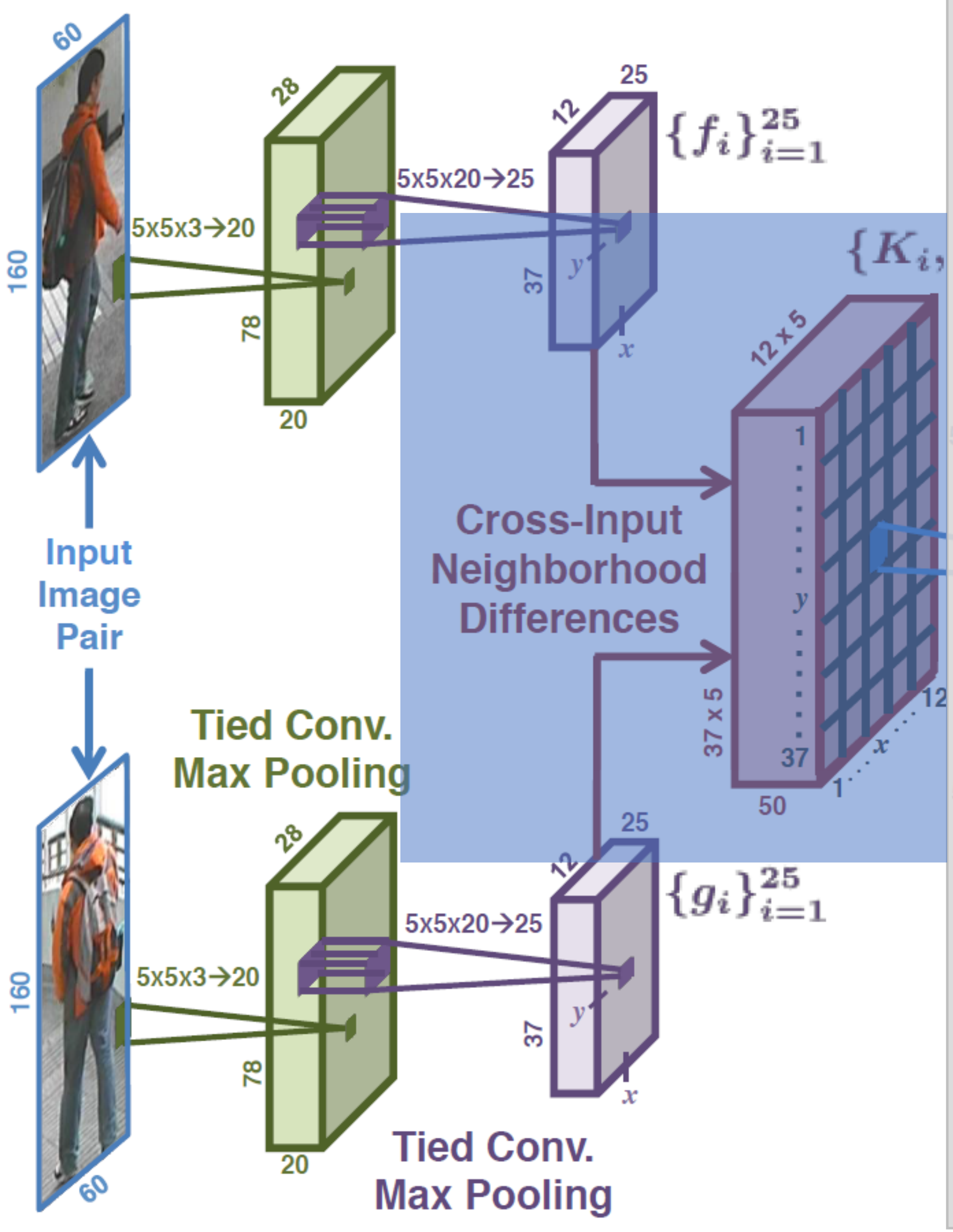


Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015

Shared weights respectively



Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015



$N[f_i(x,y)]$

12	63	65	53	76
22	43	76	21	43
21	57	68	67	78
56	47	65	27	23
78	87	58	24	24

$N[g_i(x,y)]$

65	32	24	98	113
42	43	76	87	43
52	57	68	97	34
52	47	65	43	23
24	54	54	65	24

68	68	68	68	68
68	68	68	68	68
68	68	68	68	68
68	68	68	68	68
68	68	68	68	68

$f_i(x,y)\mathbf{1}(5,5)$

65	32	24	98	113
42	43	76	87	43
52	57	68	97	34
52	47	65	43	23
24	54	54	65	24

$N[g_i(x,y)]$

$K_i(x,y)$

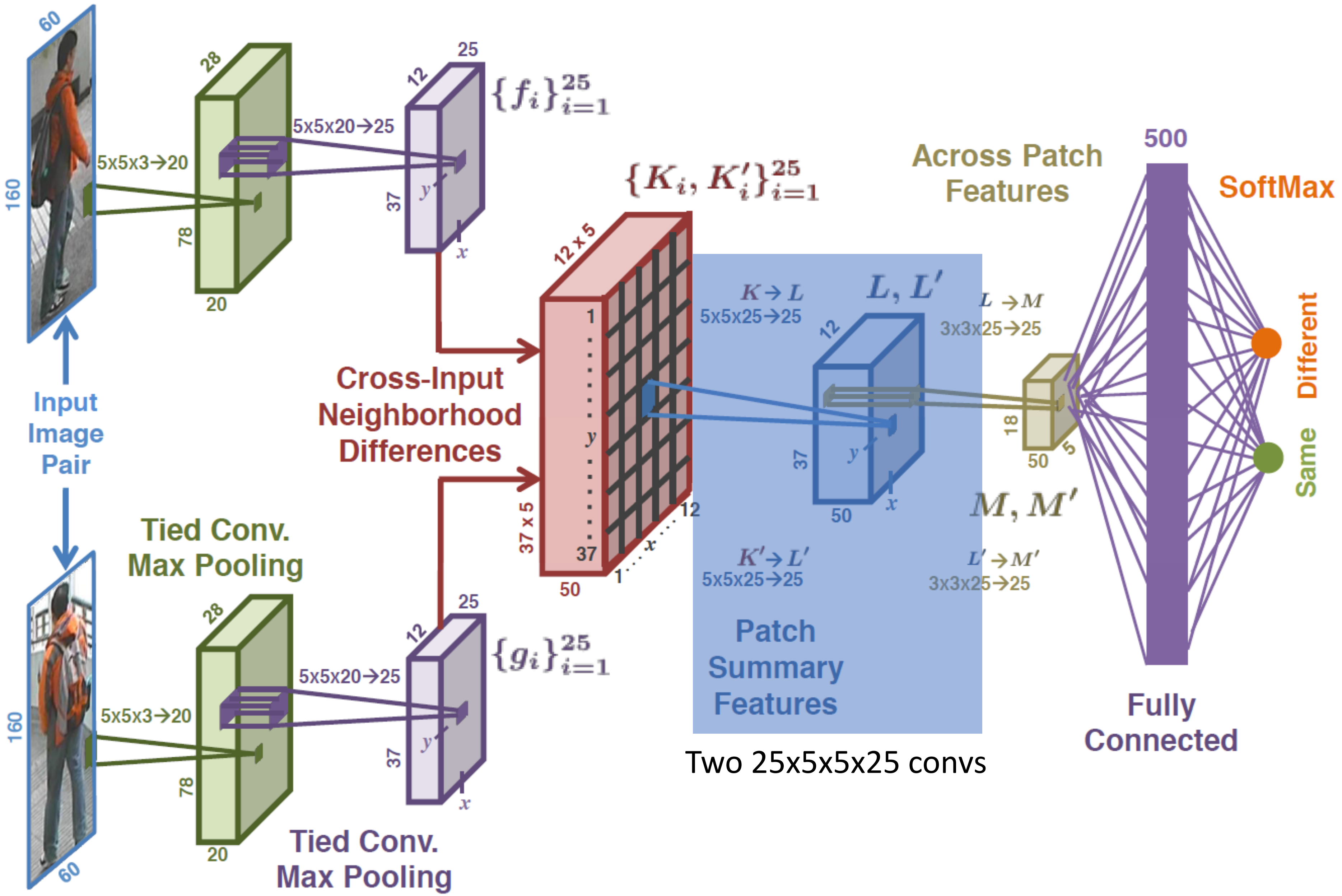
?				?
		?		
	?	0	?	
		?		
?				?

ReLU

Really make sense?

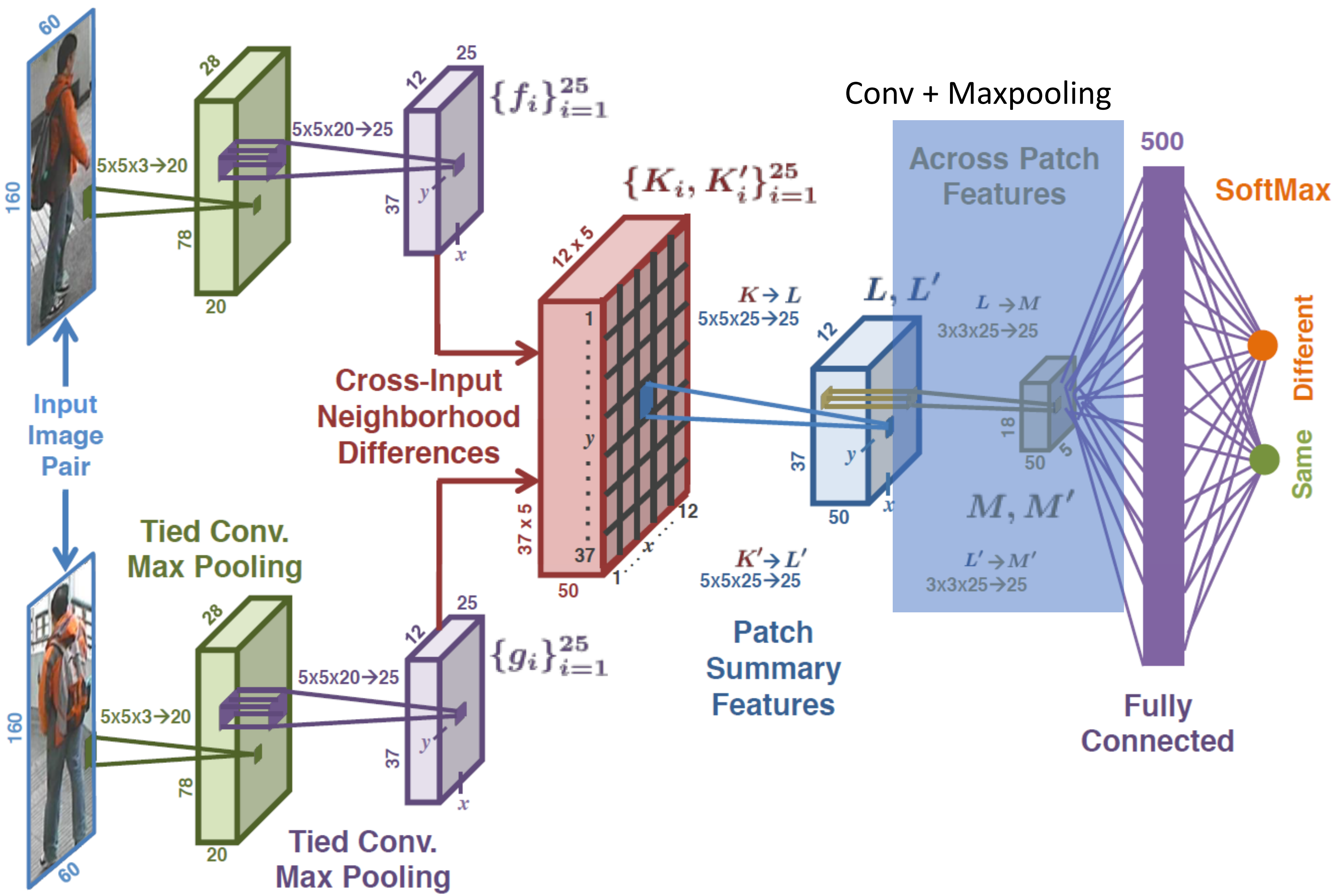
How about 1 extra conv layer?

Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015



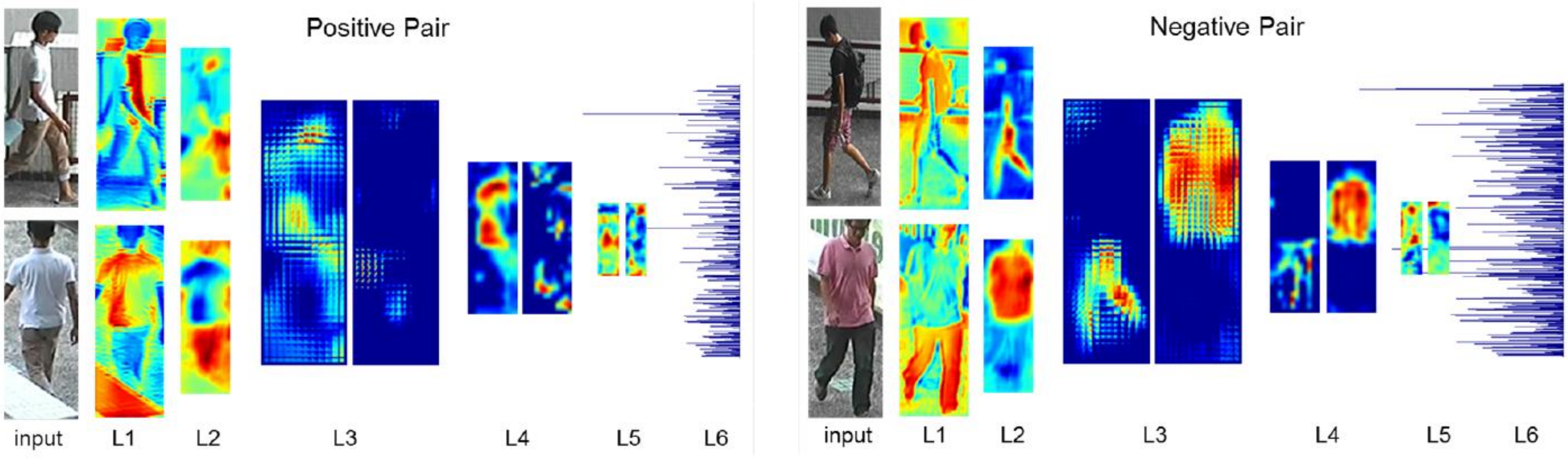
Two 25x5x5x25 convs

Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015



Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015

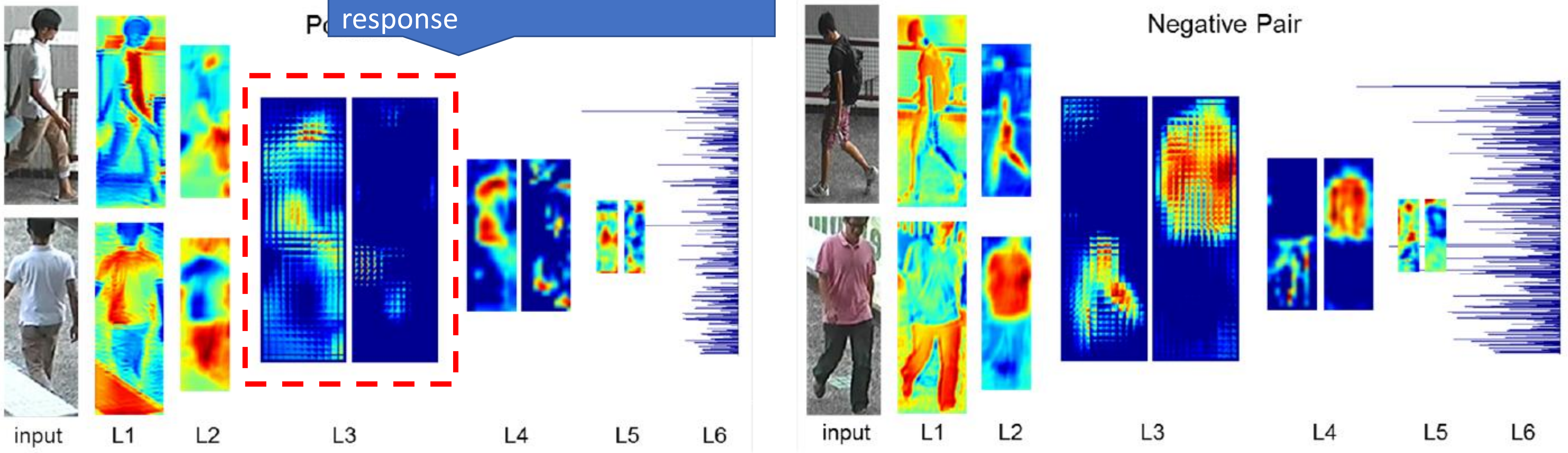
Response Maps



Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015

Response Maps

Q: Why these two maps look complementary?
A: ReLU removes the negative response



Ejaz Ahmed, Michael Jones, Tim K. Marks: An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015

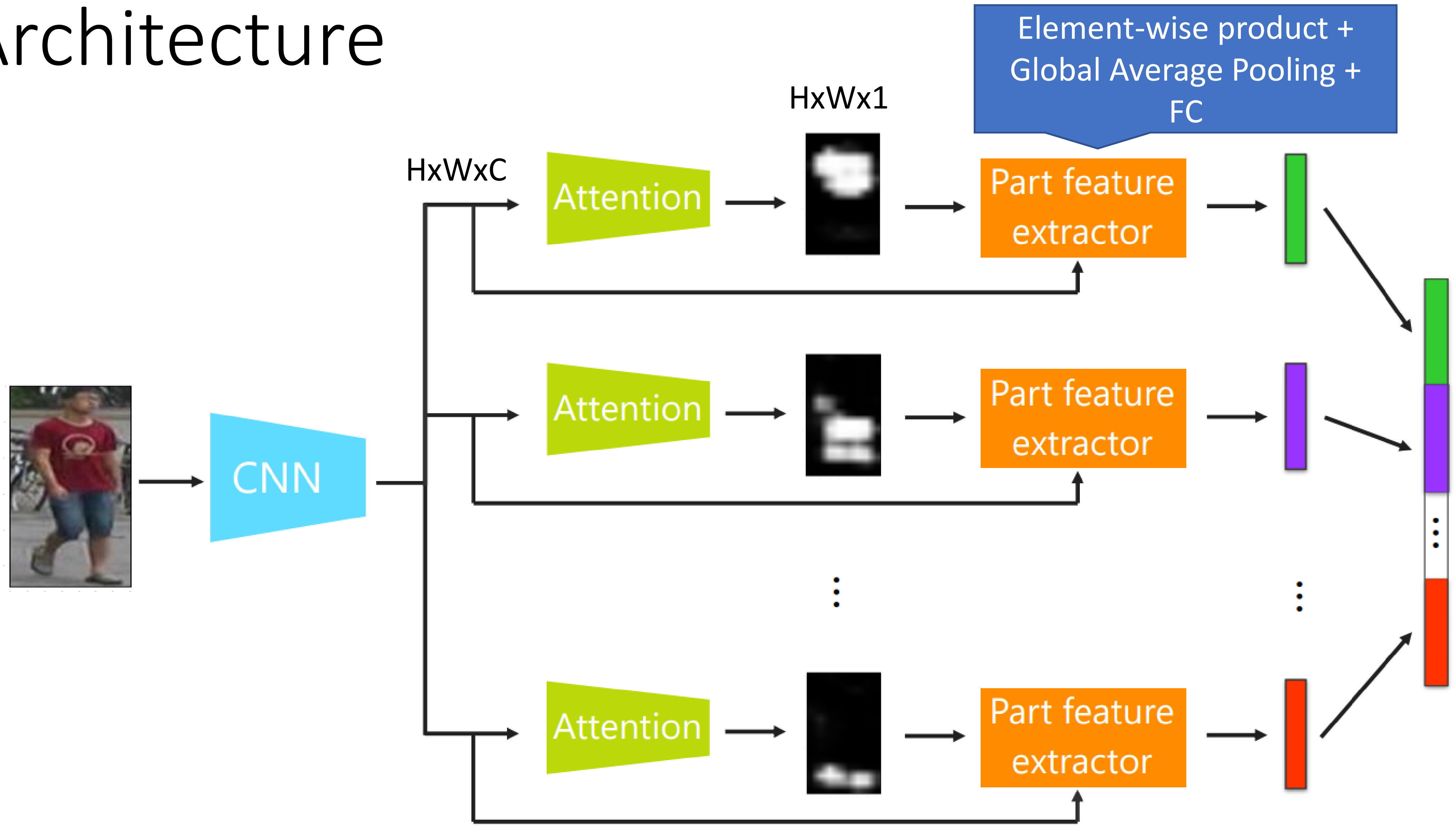
Overview

- CNN based model
- Considered misalignment of different parts
- A naïve solution

Related Works

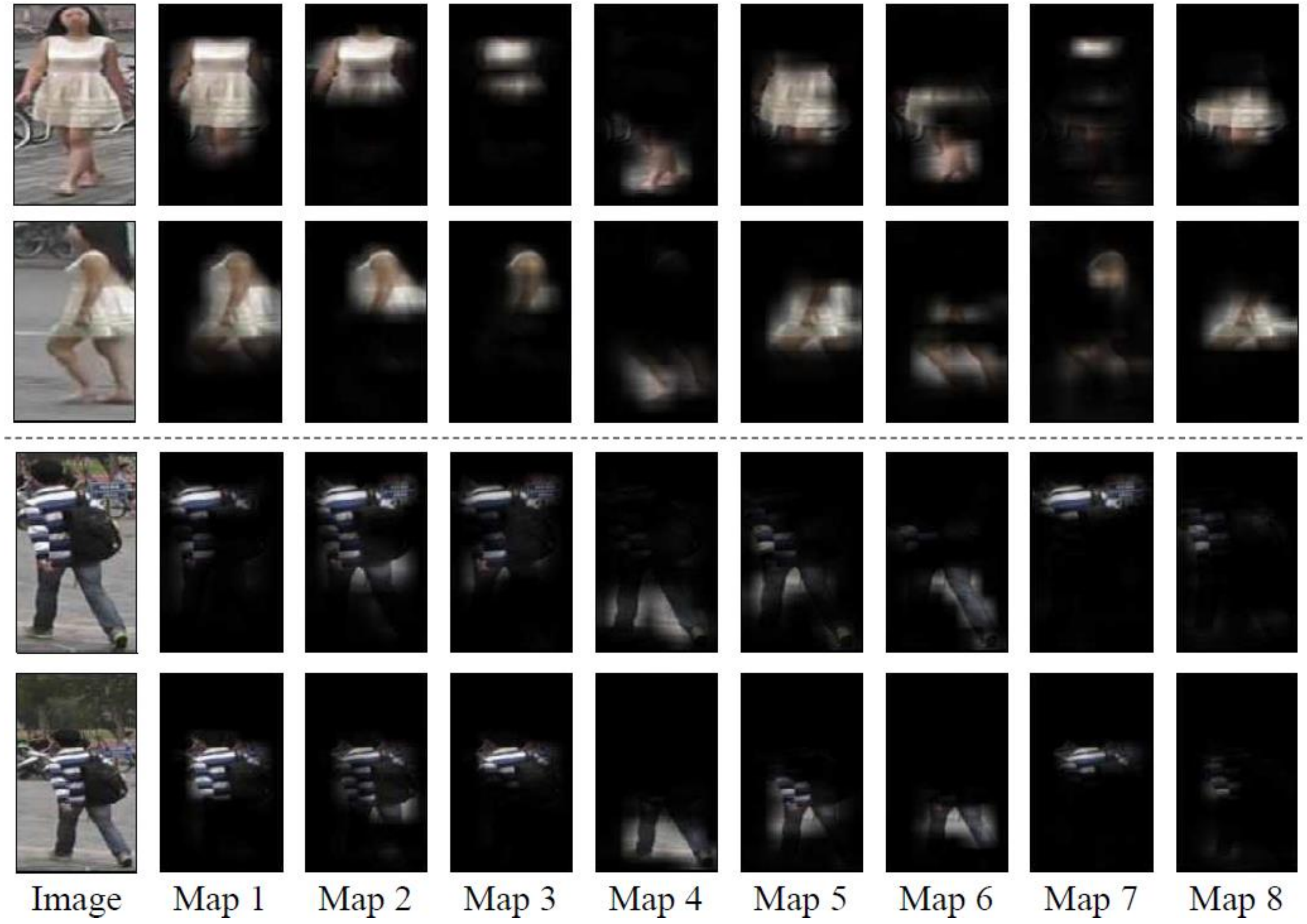
- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015 w/o pose
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017
- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018 w/ pose
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

Architecture



Liming Zhao, Xi Li, Yueting Zhuang, Jingdong Wang: Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017

Response Maps



Liming Zhao, Xi Li, Yueting Zhuang, Jingdong Wang: Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017

Overview

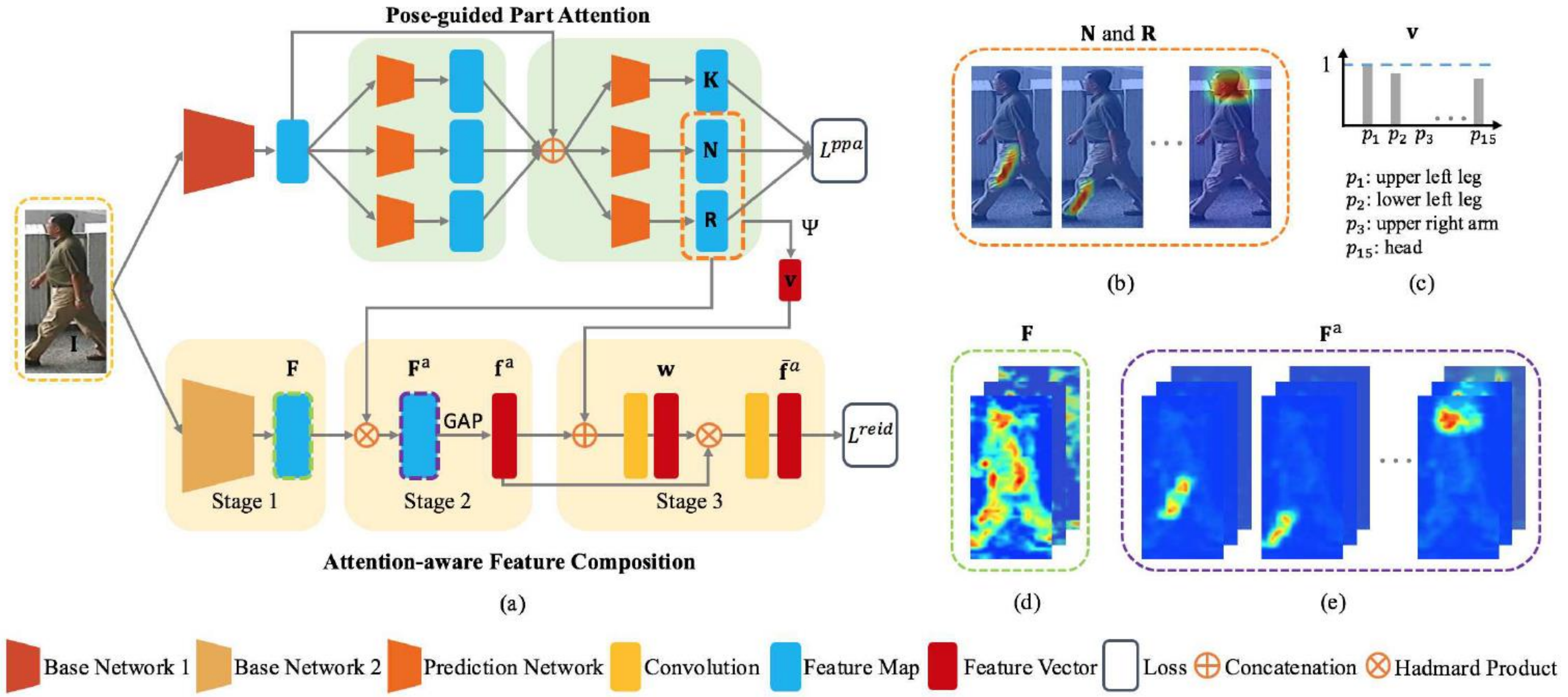
- Attention-based method
- Lack of guidance in splitting parts

Liming Zhao, Xi Li, Yueting Zhuang, Jingdong Wang: Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017

Related Works

- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015 w/o pose
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017
- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018 w/ pose
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

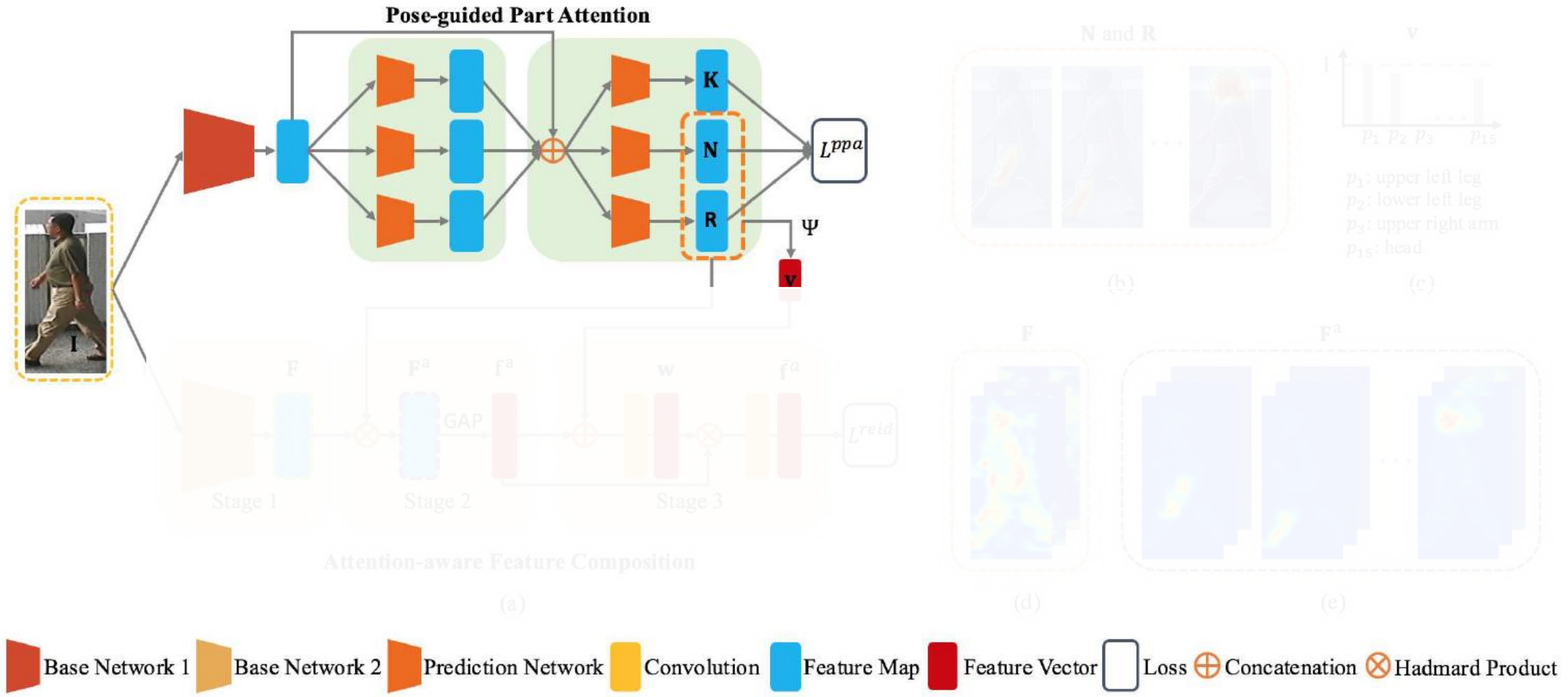
Architecture



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

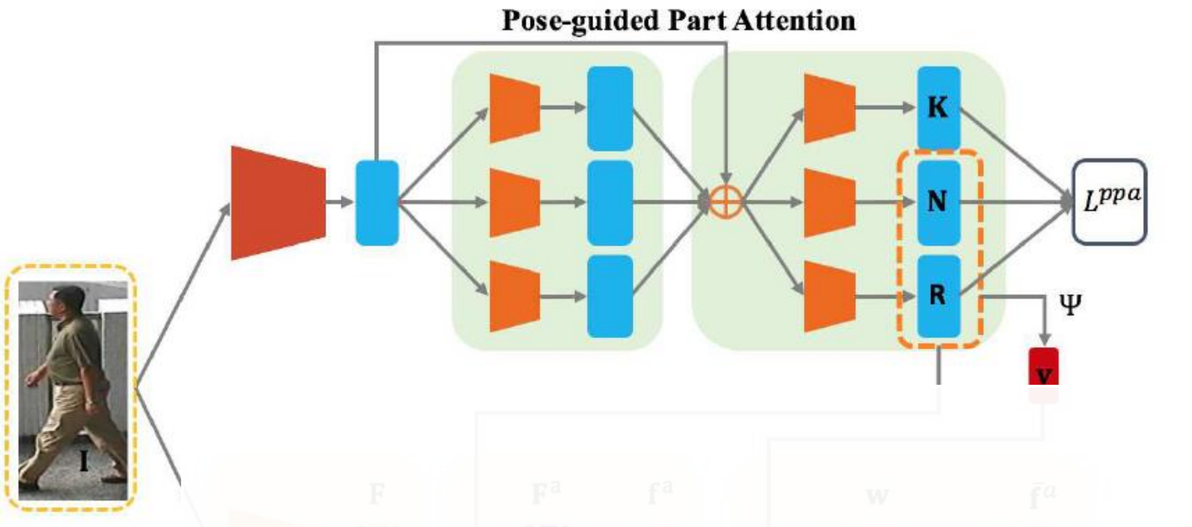
Step 1: Train PPA independently

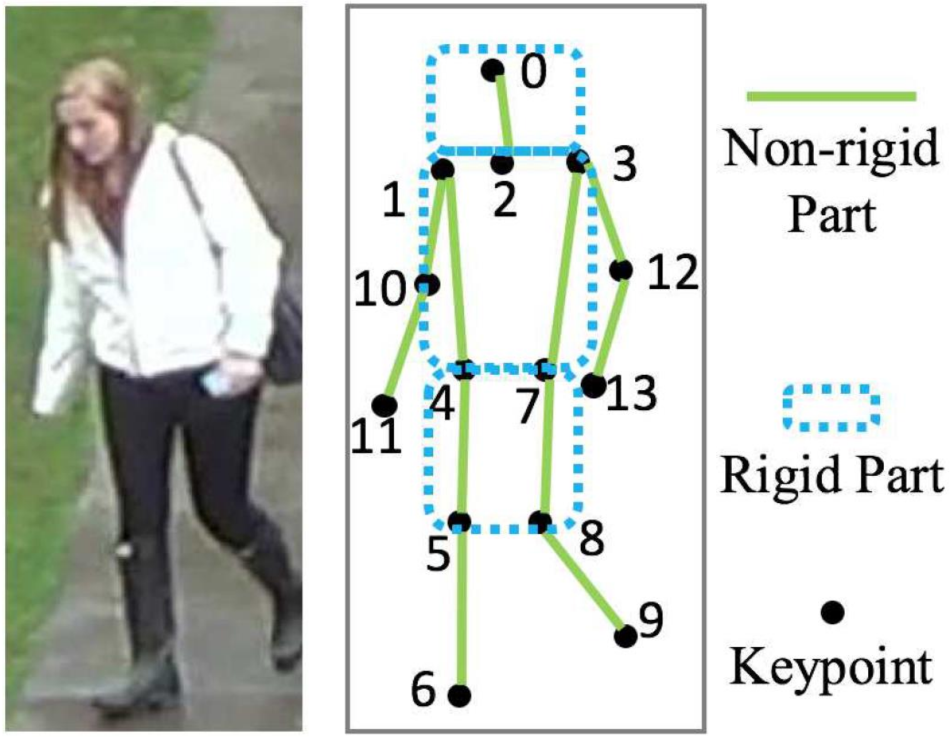
Architecture



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

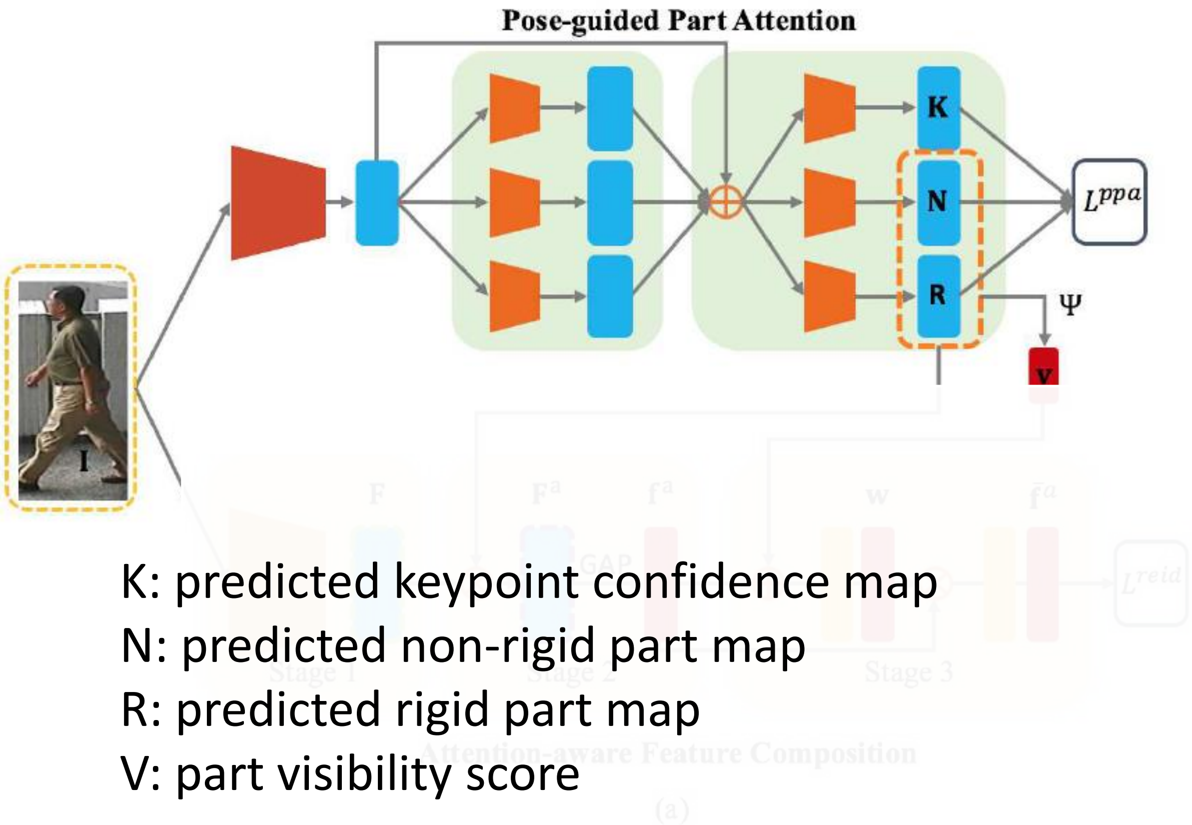




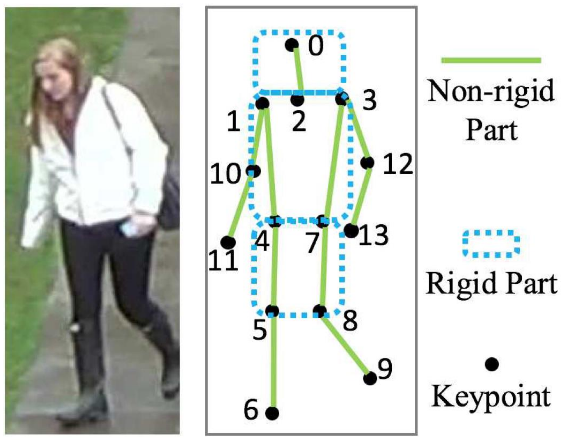
K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture



$$L^{ppa}(\rho, \phi, \psi) = \sum_{t=1,2} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$

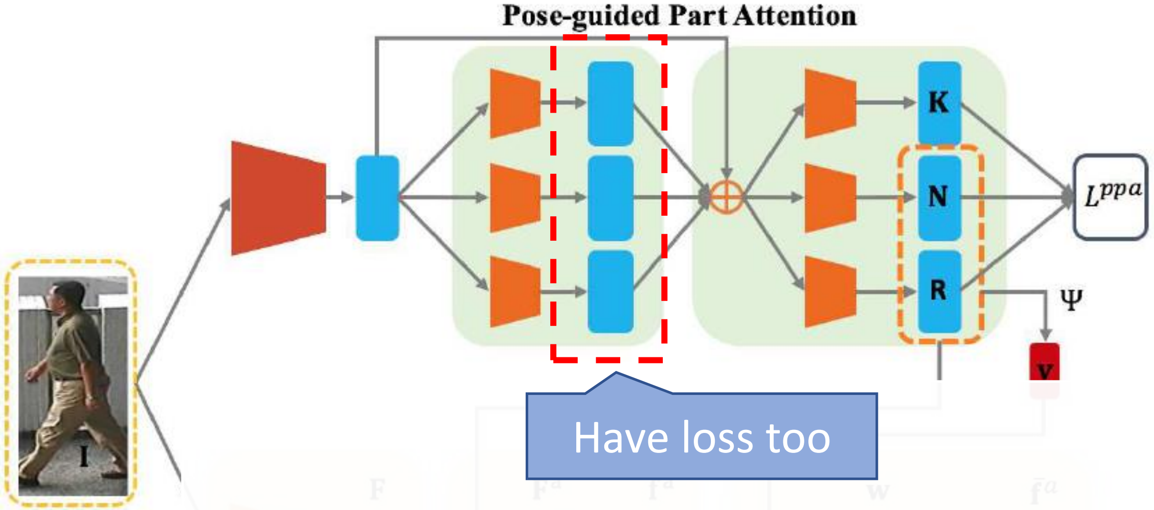


K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

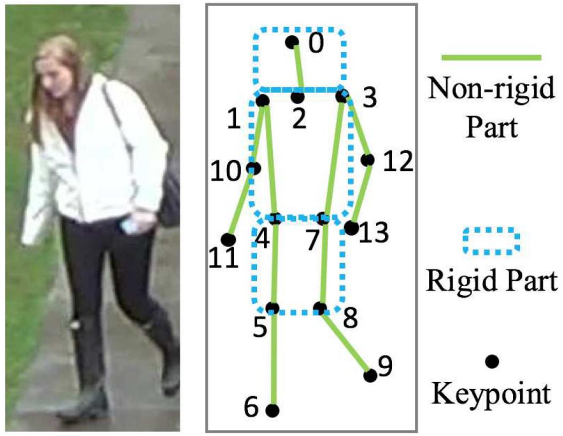
Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently

Architecture



$$L^{ppa}(\rho,\phi,\psi) = \boxed{\sum_{t=1,2}} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$



K: predicted keypoint confidence map

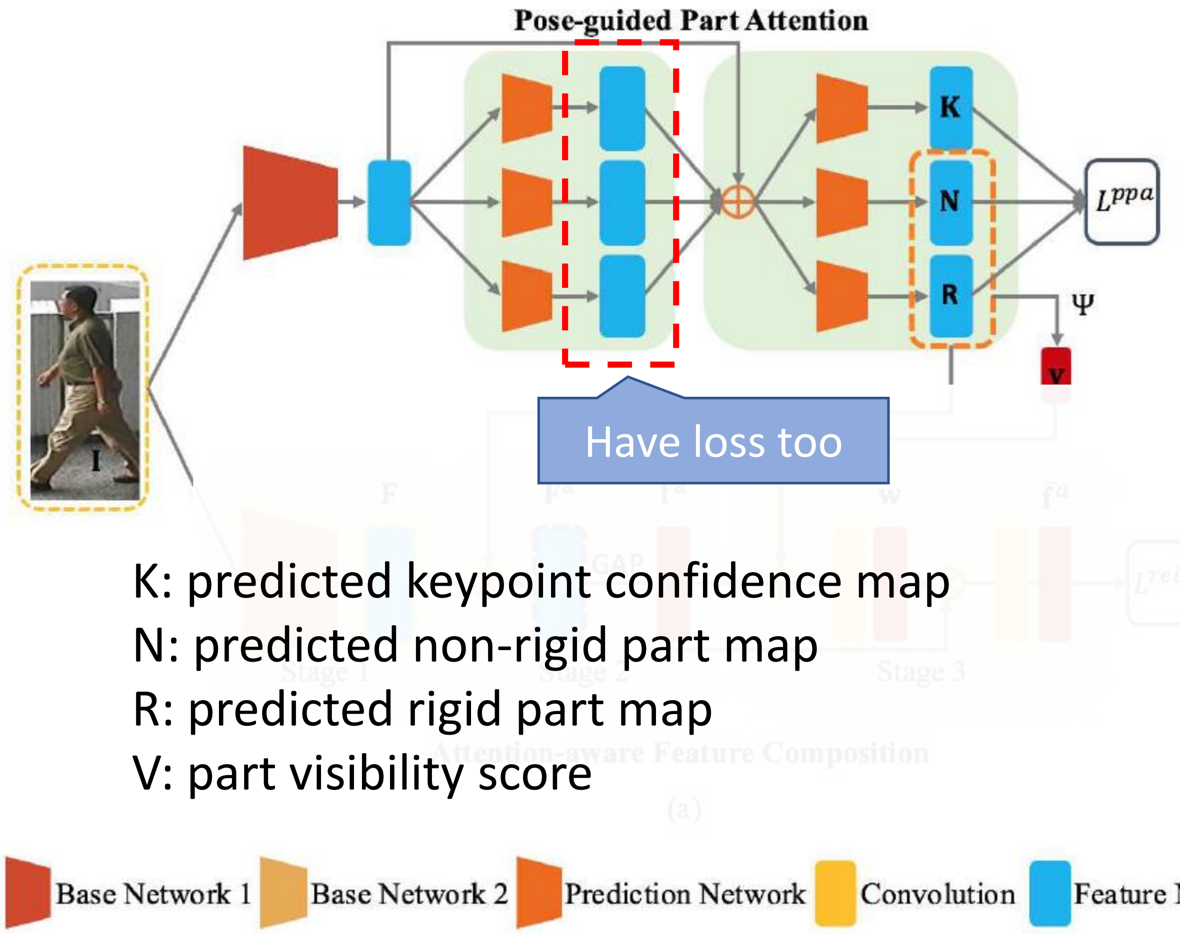
N: predicted non-rigid part map

R: predicted rigid part map

V: part visibility score

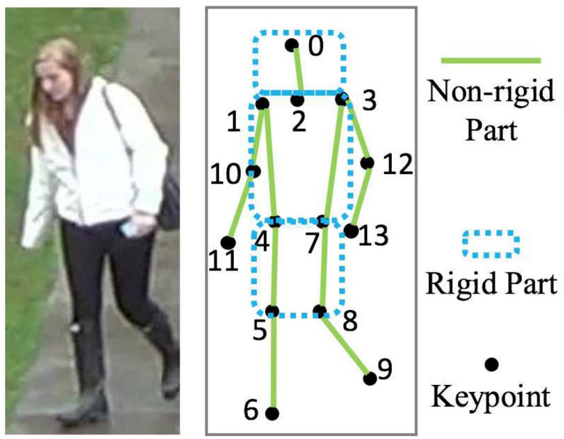
Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture



$$L^{ppa}(\rho, \phi, \psi) = \sum_{t=1,2} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$

$$L^k(\mathbf{K}) = \frac{1}{C^k} \sum_{i=1}^{C^k} ||\mathbf{K}_i^* - \mathbf{K}_i||^2$$



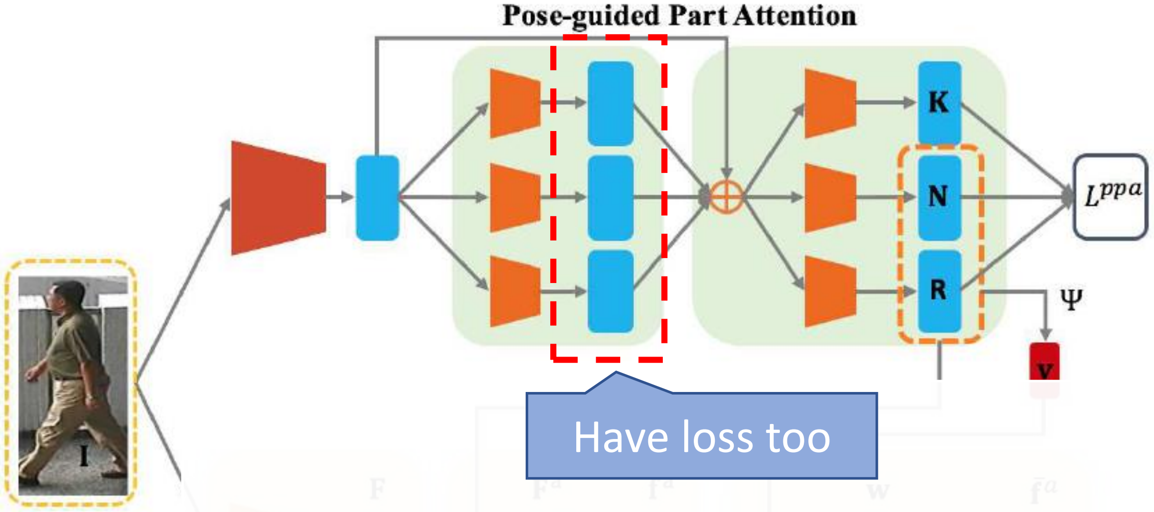
K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

K_i^* is generated by applying Gaussian kernel at the true location of keypoint i

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently

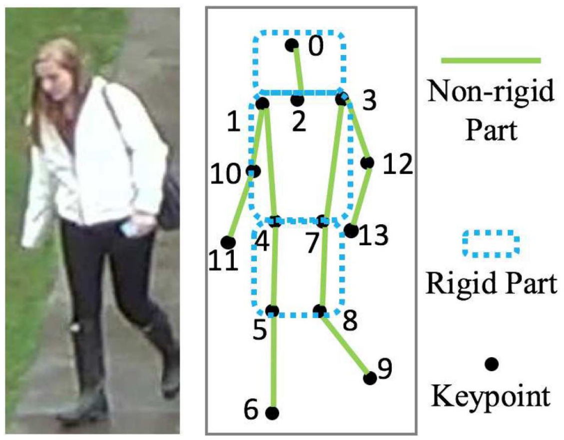
Architecture



$$L^{ppa}(\rho, \phi, \psi) = \boxed{\sum_{t=1,2}} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$

$$L^k(\mathbf{K}) = \frac{1}{C^k} \sum_{i=1}^{C^k} ||\mathbf{K}_i^* - \mathbf{K}_i||^2$$

$$L^n(\mathbf{N}) = \frac{1}{C^n} \sum_{p=1}^{C^n} ||\mathbf{N}_p^* - \mathbf{N}_p||^2$$

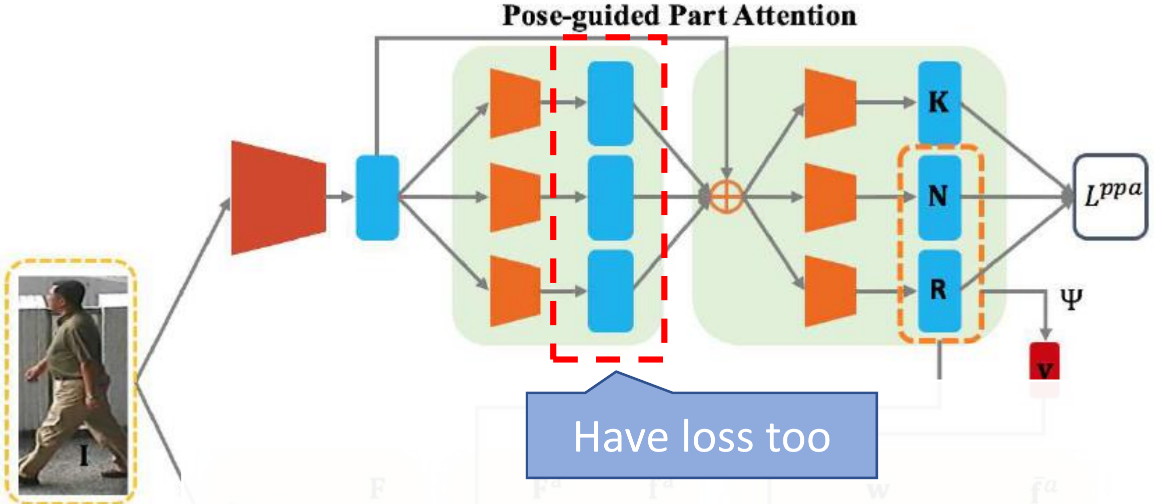


K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

N_p^* is the p-th non-grid part, defined as a rectangle area connecting two keypoints with bandwidth σ

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture



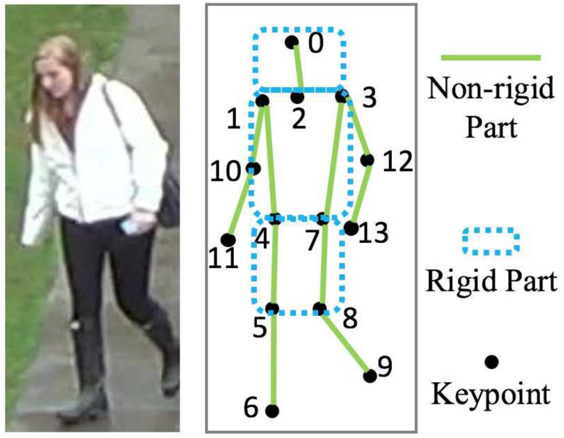
K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

$$L^{ppa}(\rho, \phi, \psi) = \boxed{\sum_{t=1,2}} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$

$$L^k(\mathbf{K}) = \frac{1}{C^k} \sum_{i=1}^{C^k} ||\mathbf{K}_i^* - \mathbf{K}_i||^2$$

$$L^n(\mathbf{N}) = \frac{1}{C^n} \sum_{p=1}^{C^n} ||\mathbf{N}_p^* - \mathbf{N}_p||^2$$

$$L^r(\mathbf{R}, \mathbf{N}) = \frac{1}{C^r} \sum_{p=1}^{C^r} ||\mathbf{R}_p^* - \hat{\mathbf{R}}_p||^2$$

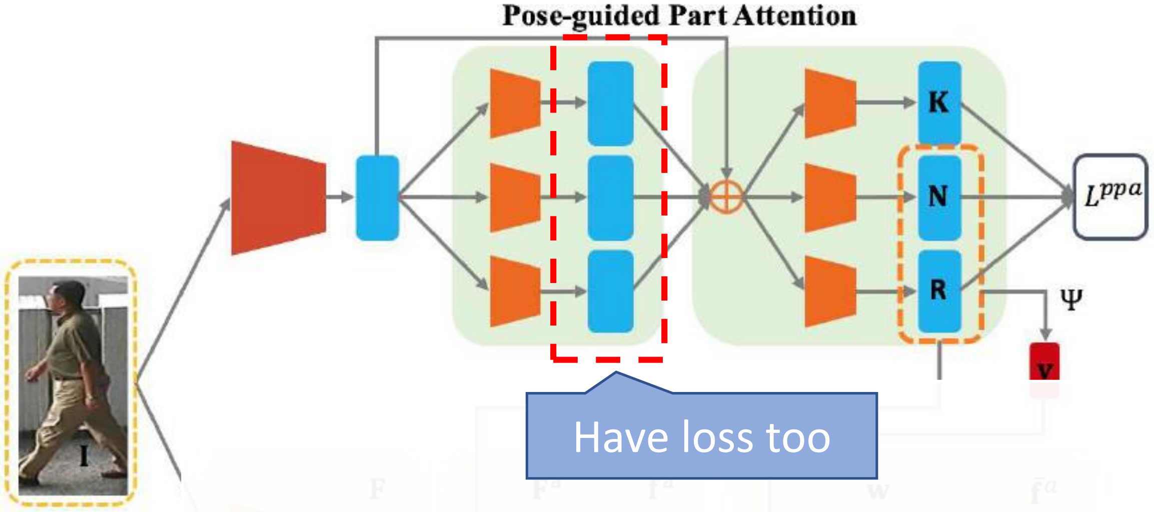


R_p^* is defined by a rectangle tightly contains specified keypoints

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently

Architecture



K: predicted keypoint confidence map
N: predicted non-rigid part map
R: predicted rigid part map
V: part visibility score

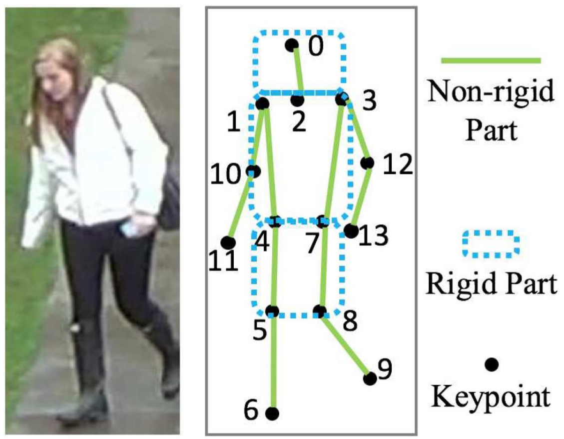
$$L^{ppa}(\rho,\phi,\psi) = \boxed{\sum_{t=1,2}} L^k(\mathbf{K}^t) + \mu_1 L^n(\mathbf{N}^t) + \mu_2 L^r(\mathbf{R}^t)$$

$$L^k(\mathbf{K}) = \frac{1}{C^k}\sum_{i=1}^{C^k} ||\mathbf{K}_i^* - \mathbf{K}_i||^2$$

$$L^n(\mathbf{N}) = \frac{1}{C^n}\sum_{p=1}^{C^n} ||\mathbf{N}_p^* - \mathbf{N}_p||^2$$

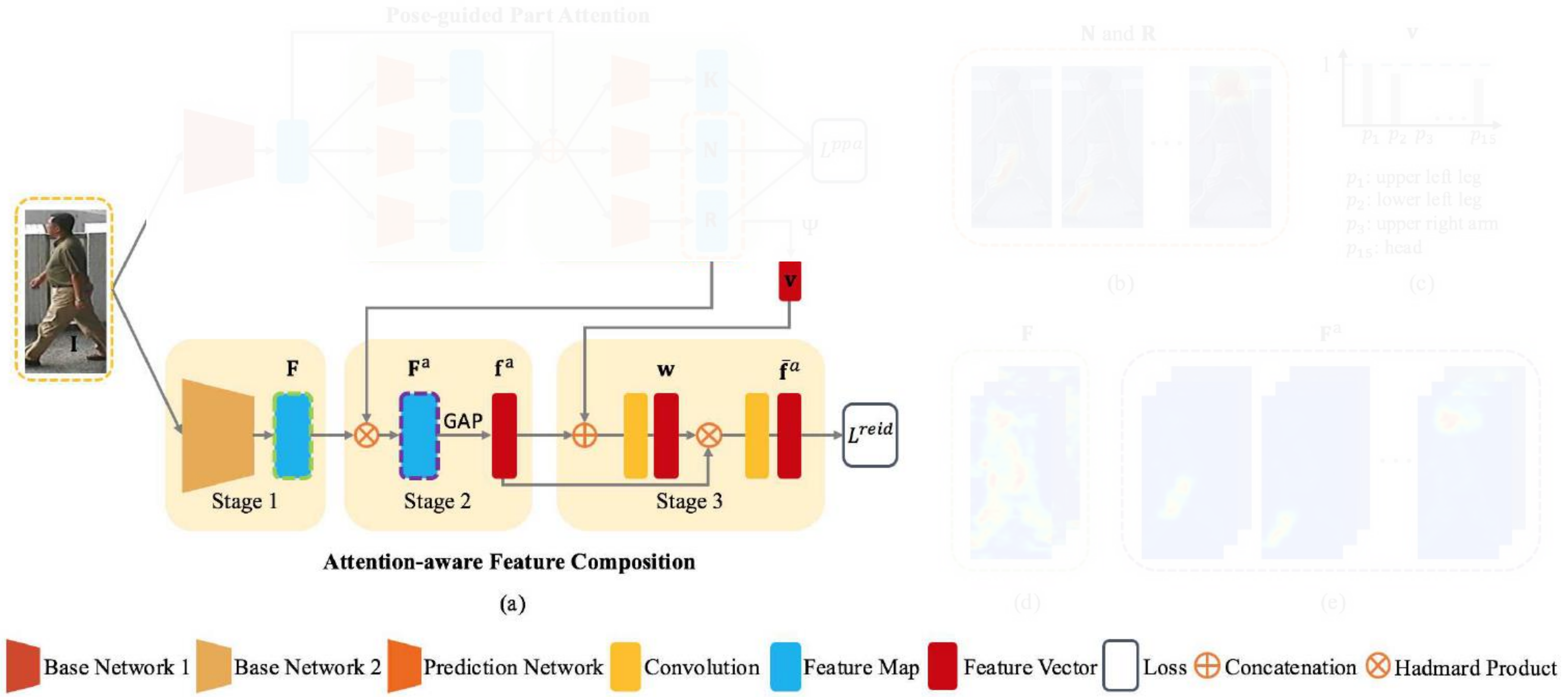
$$L^r(\mathbf{R},\mathbf{N}) = \frac{1}{C^r}\sum_{p=1}^{C^r} ||\mathbf{R}_p^* - \hat{\mathbf{R}}_p||^2$$

$$v_p = \sum_{x,y} |\mathbf{R}_p(x,y)|, \quad \text{or} \quad v_p = \sum_{x,y} |\mathbf{N}_p(x,y)|.$$



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture



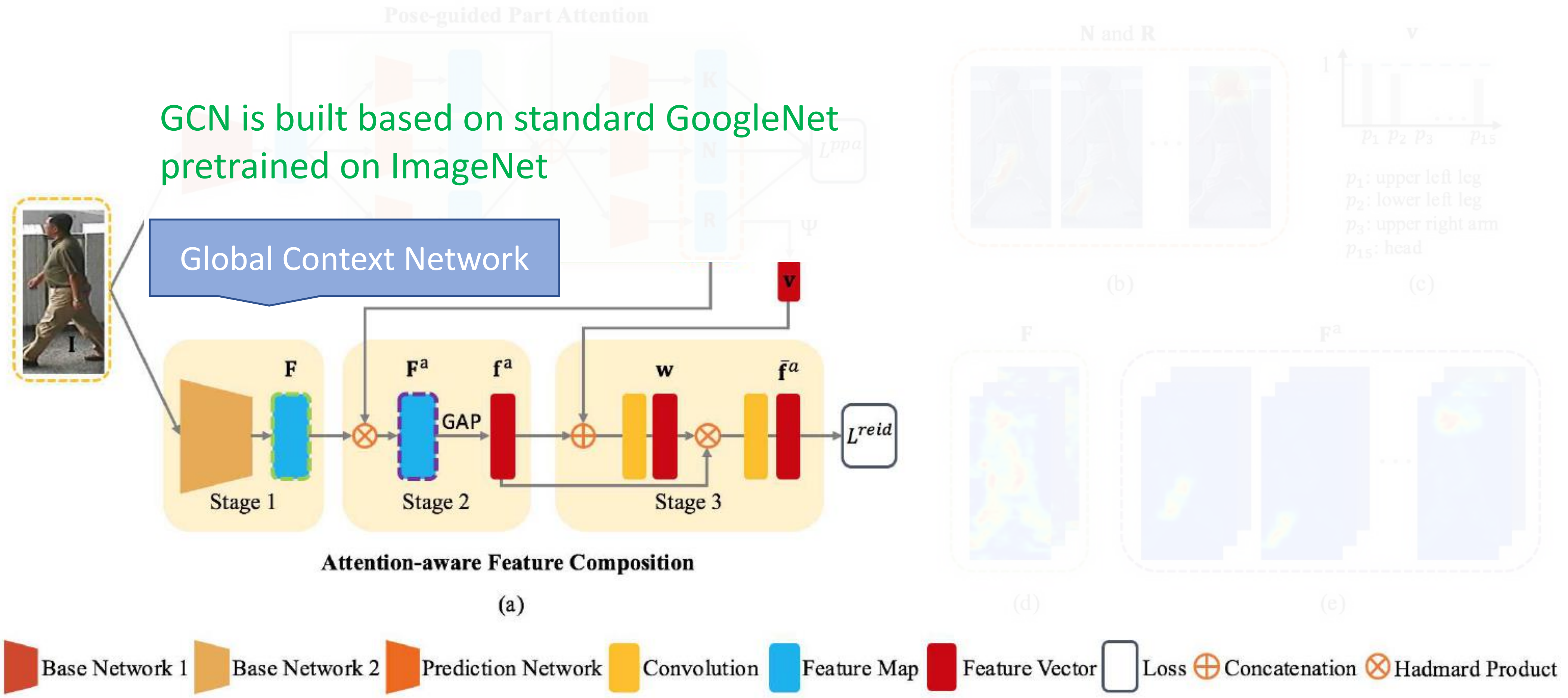
Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently

Step 2: Train GCN independently

Architecture

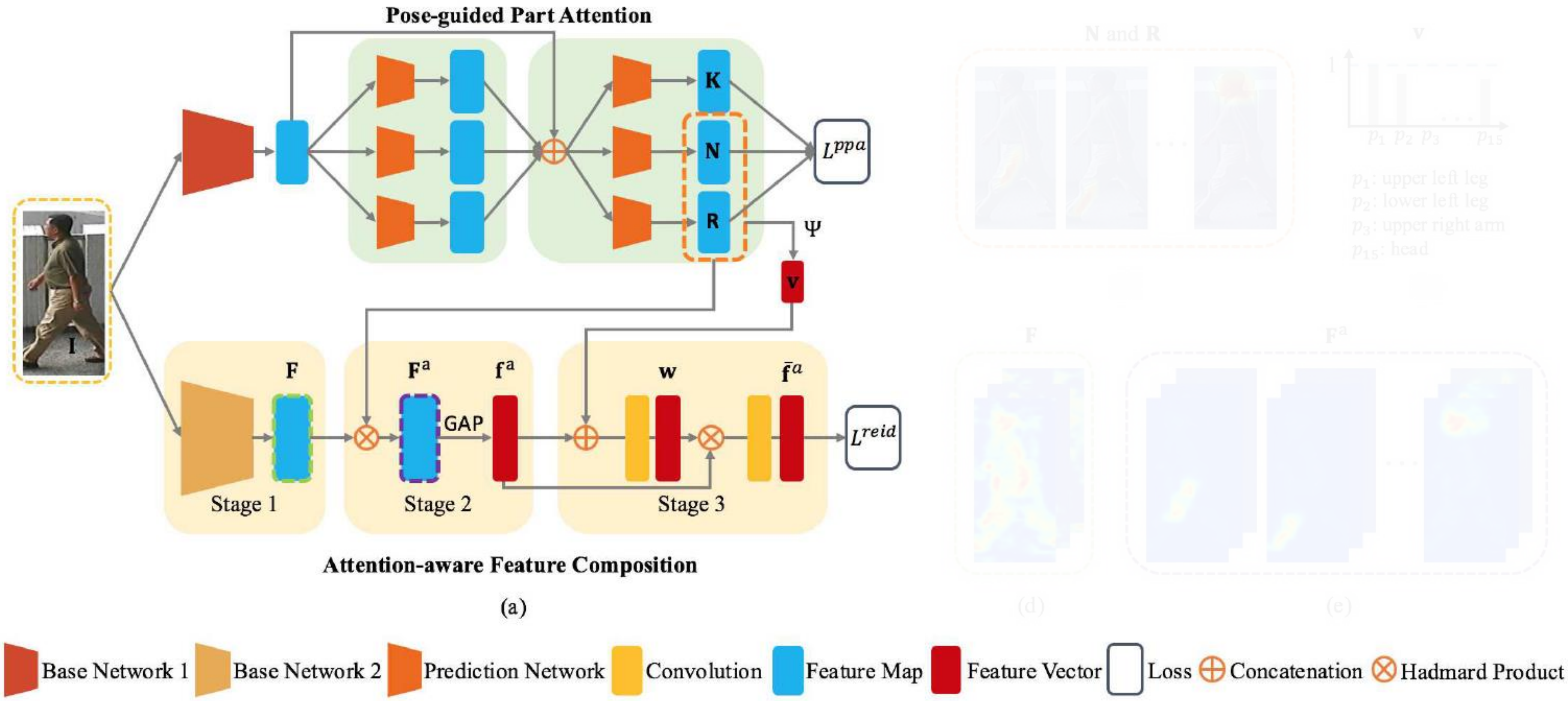
GCN is built based on standard GoogleNet pretrained on ImageNet



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment

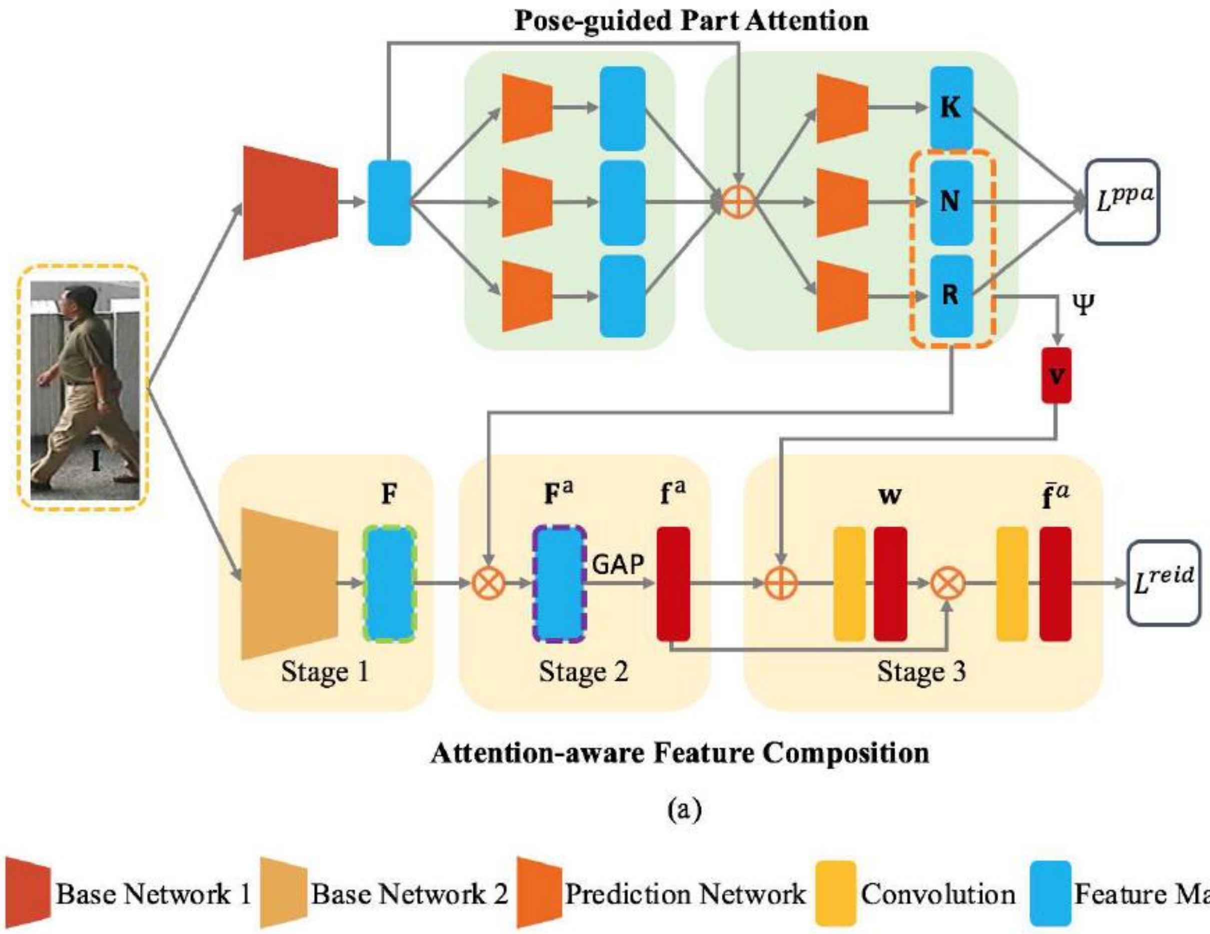
Architecture



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment



(a)

$\mathbf{M}_p \in \{\mathbf{N}_p, \mathbf{R}_p\}$ is the attention map for body parts,

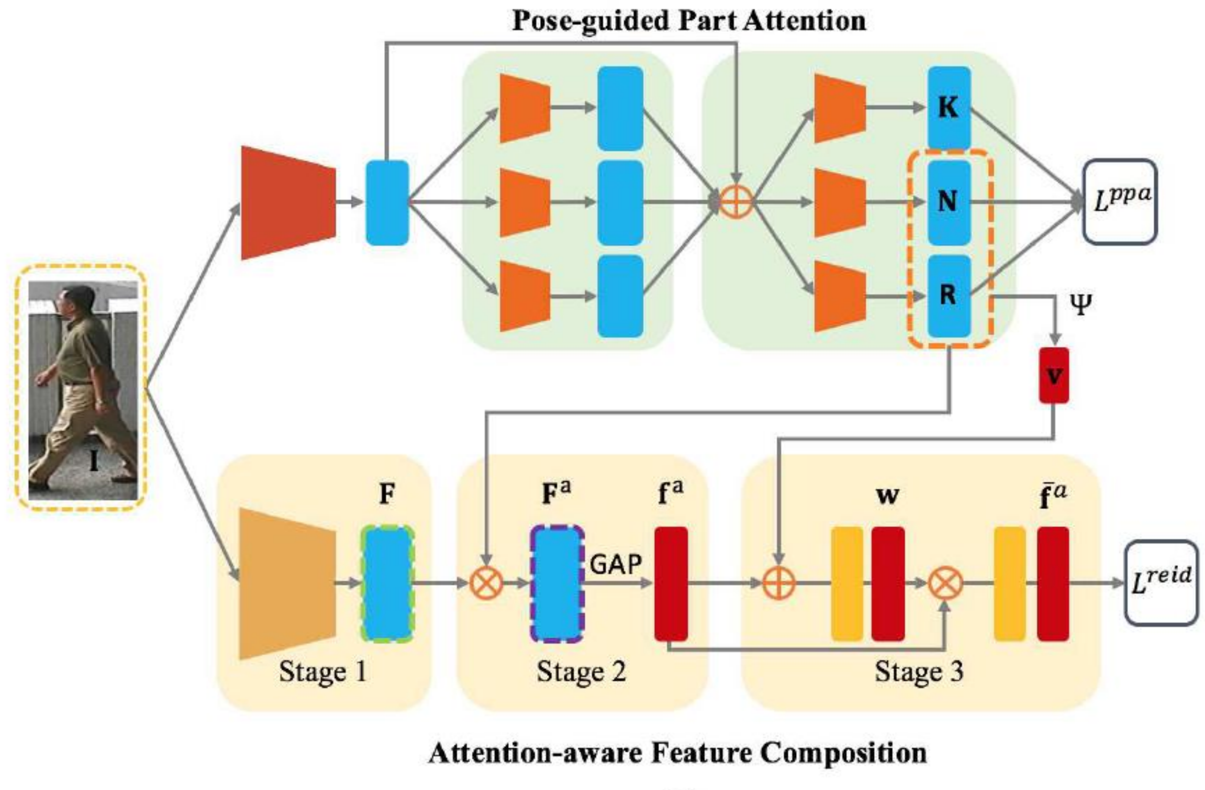
Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently

Step 2: Train GCN independently

Step 3: Attention-Aware Feature Alignment



(a)

$\mathbf{M}_p \in \{\mathbf{N}_p, \mathbf{R}_p\}$ is the attention map for body parts,

$$\bar{\mathbf{M}}_p = \frac{\mathbf{M}_p}{\max(\mathbf{M}_p)}$$

Normalize attention map

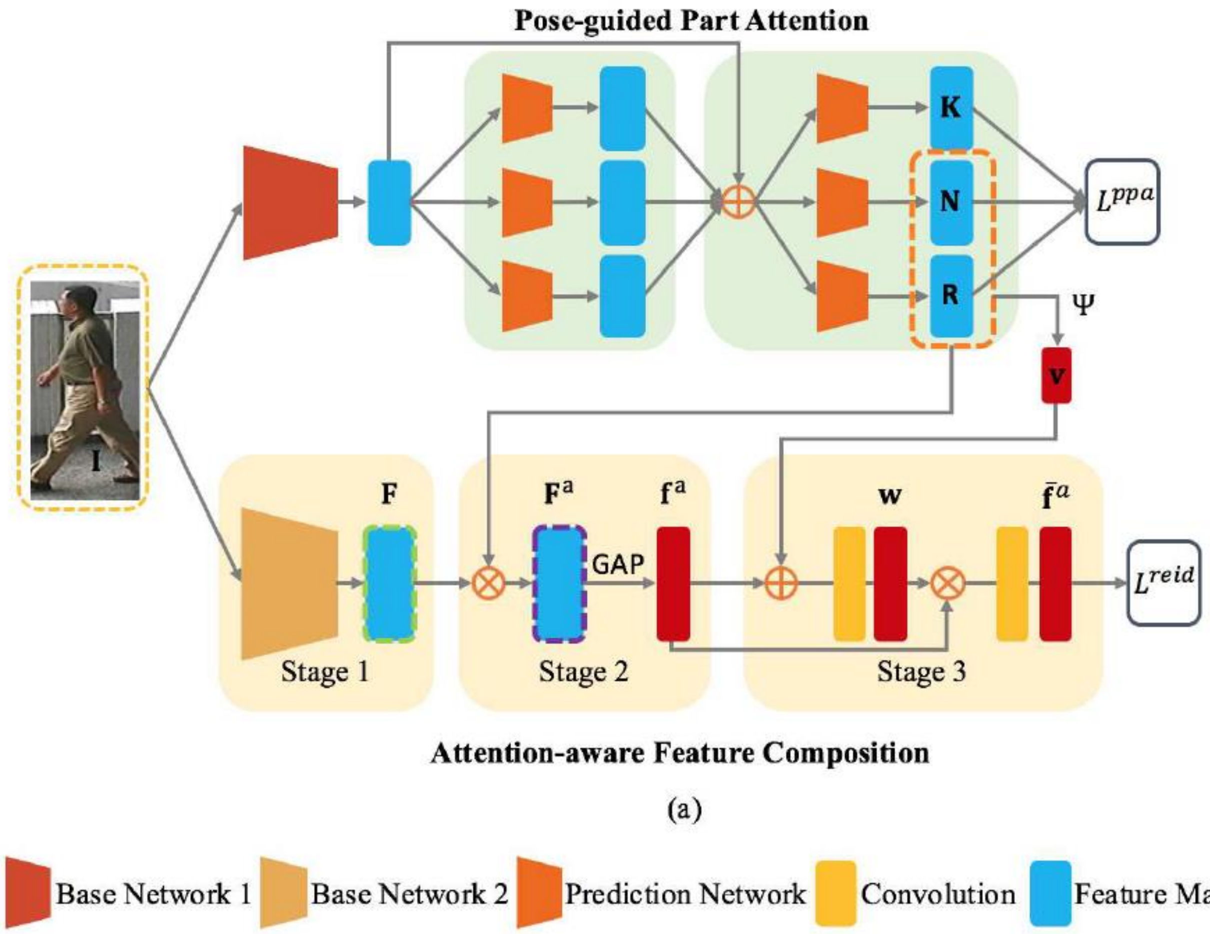
Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Step 1: Train PPA independently

Step 2: Train GCN independently

Step 3: Attention-Aware Feature Alignment

Architecture



$\mathbf{M}_p \in \{\mathbf{N}_p, \mathbf{R}_p\}$ is the attention map for body parts,

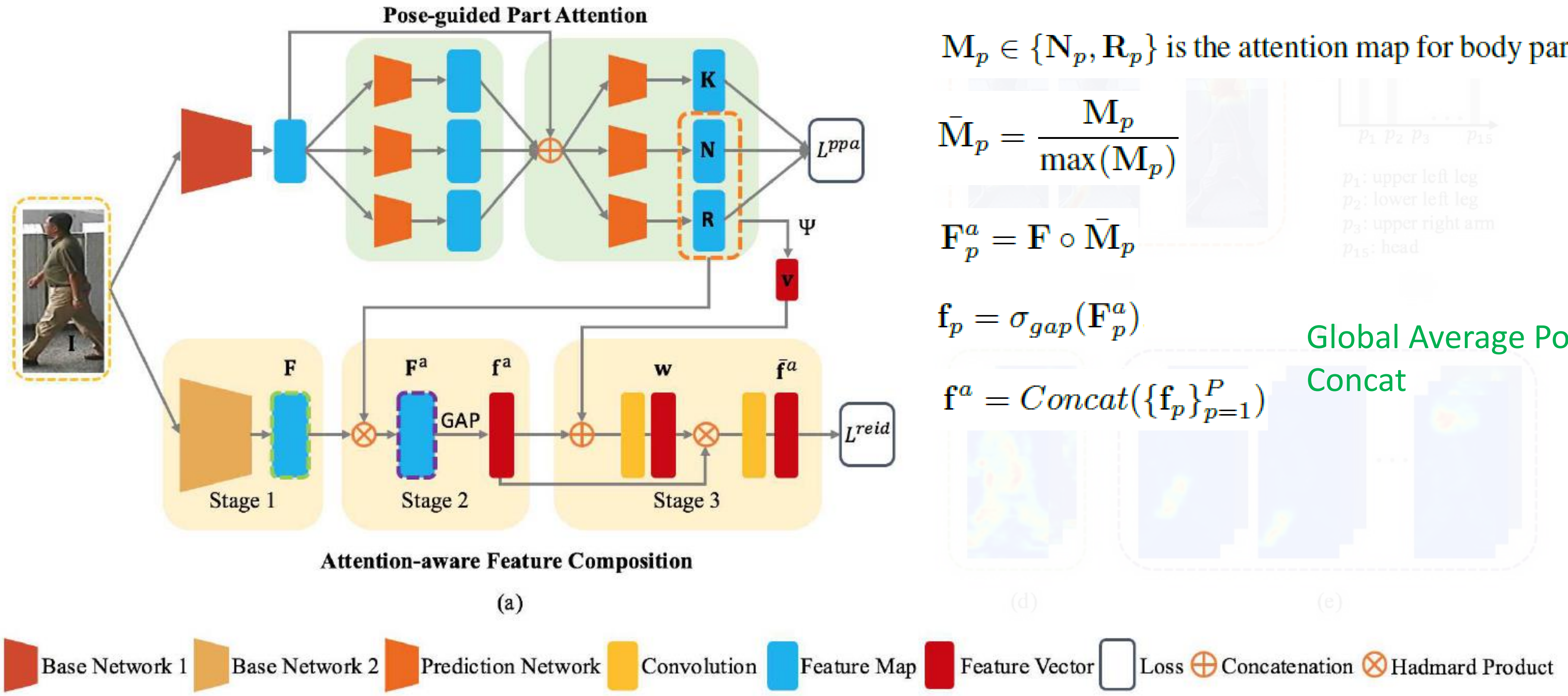
$$\bar{\mathbf{M}}_p = \frac{\mathbf{M}_p}{\max(\mathbf{M}_p)}$$

$$\mathbf{F}_p^a = \mathbf{F} \circ \bar{\mathbf{M}}_p$$ Element-wise Product

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment



$\mathbf{M}_p \in \{\mathbf{N}_p, \mathbf{R}_p\}$ is the attention map for body parts,

$$\bar{\mathbf{M}}_p = \frac{\mathbf{M}_p}{\max(\mathbf{M}_p)}$$

$$\mathbf{F}^a_p = \mathbf{F} \circ \bar{\mathbf{M}}_p$$

$$\mathbf{f}_p = \sigma_{gap}(\mathbf{F}^a_p)$$

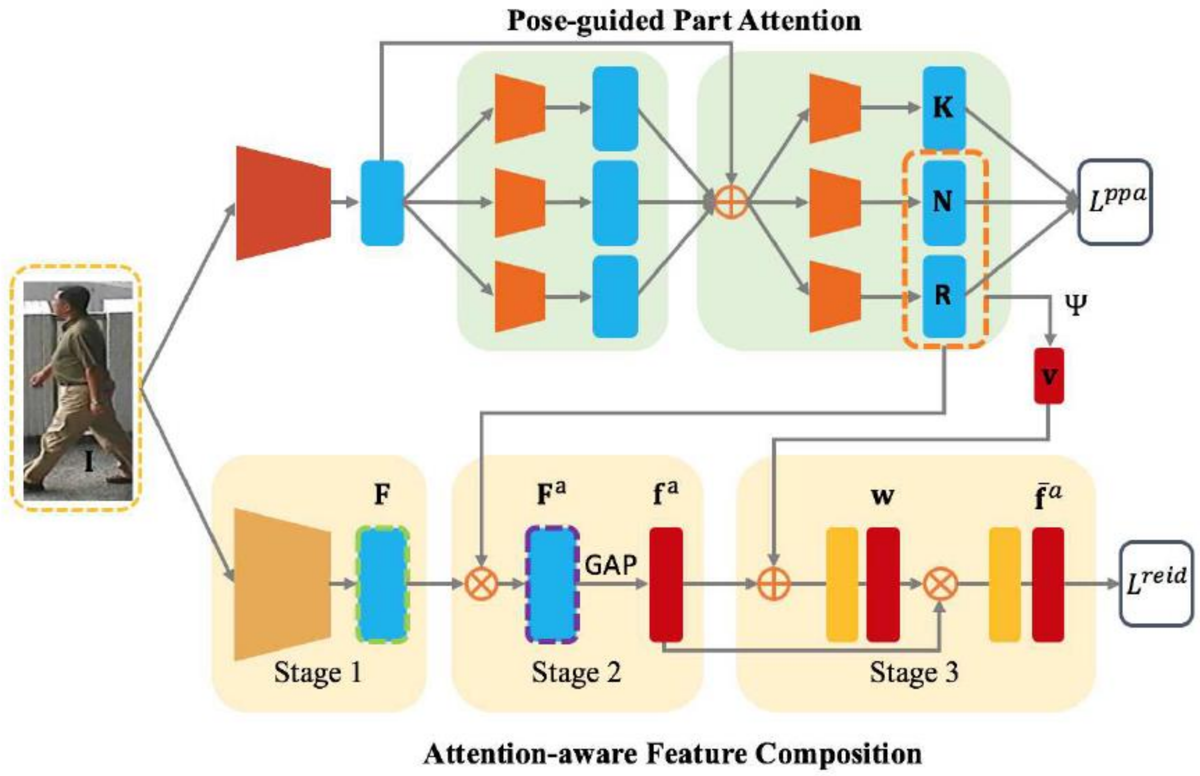
$$\mathbf{f}^a = Concat(\{\mathbf{f}_p\}^P_{p=1})$$

Global Average Pooling + Concat

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment
Step 4: Weighted Feature Composition



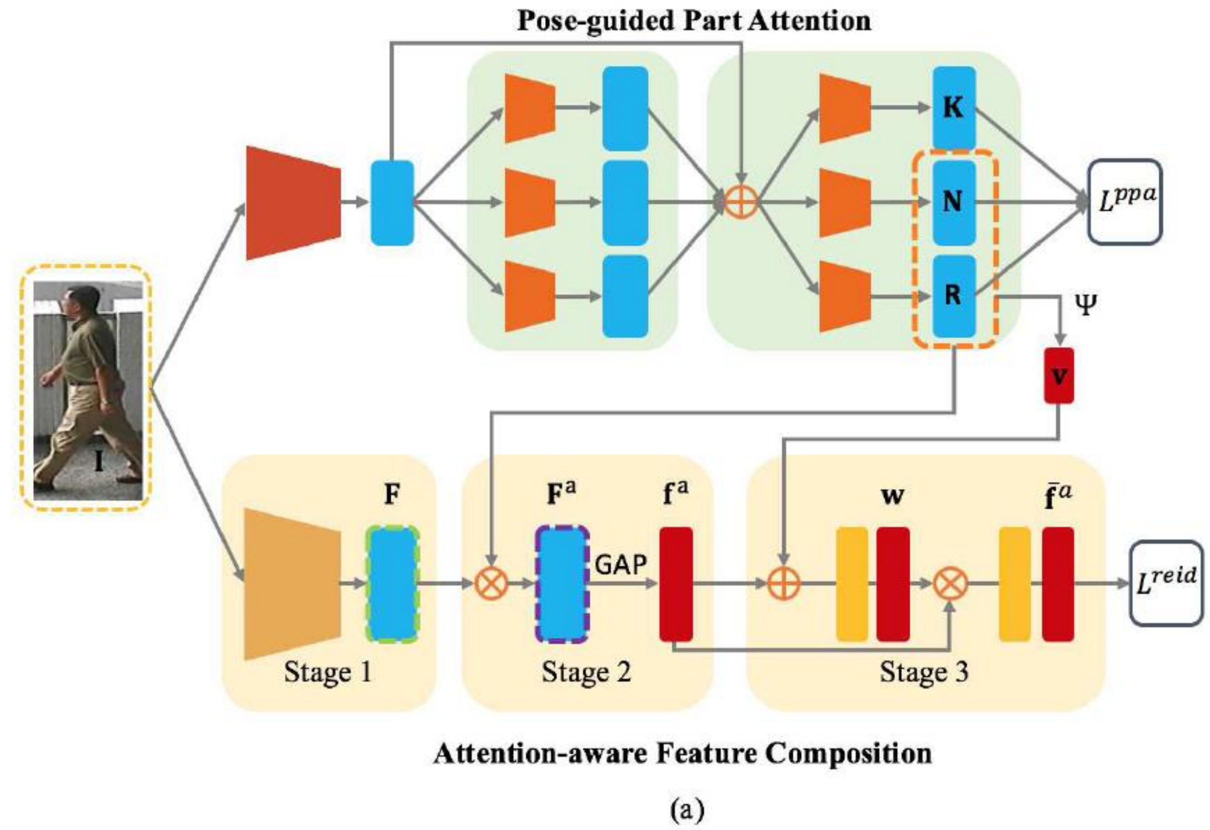
(a)

Generate a weight vector

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment
Step 4: Weighted Feature Composition



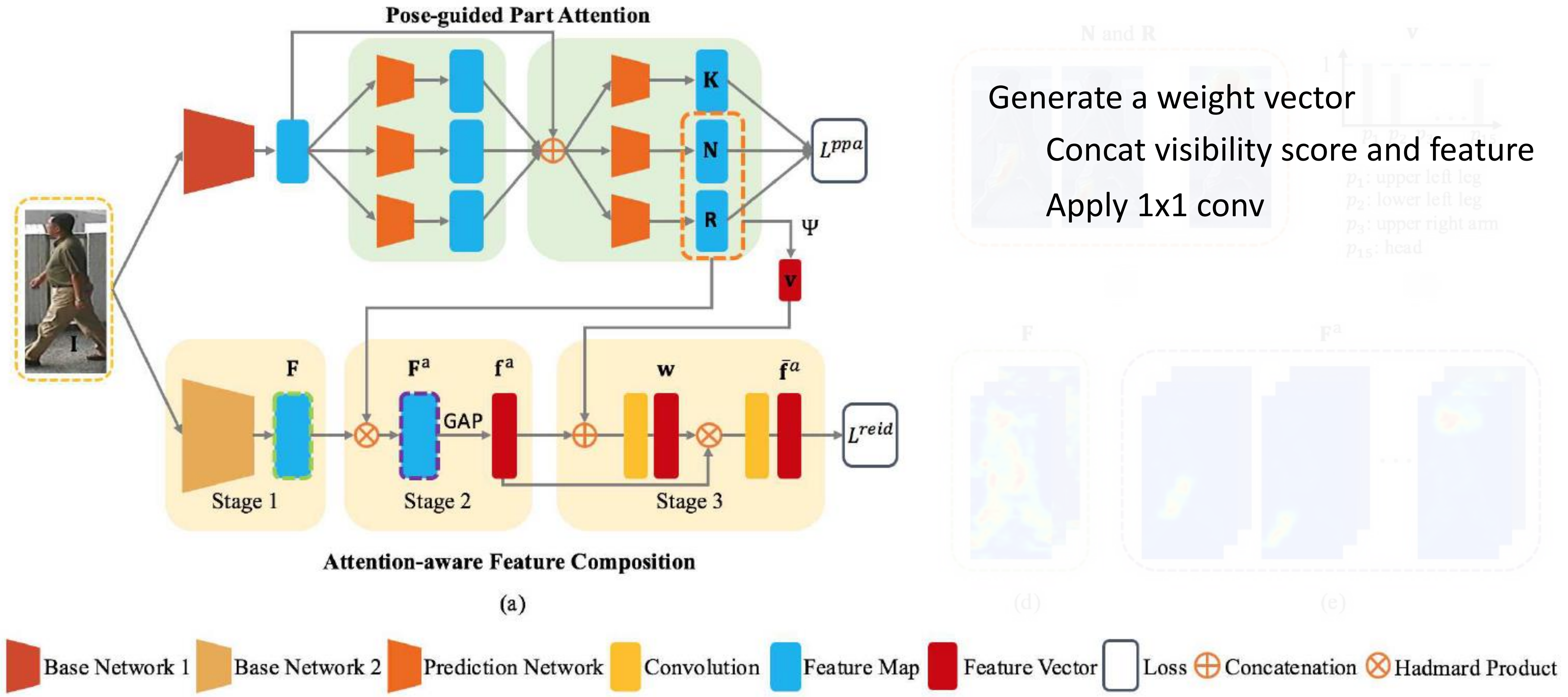
Generate a weight vector

- Concat visibility score and feature
- The weight should estimated by both visibility and feature salience

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment
Step 4: Weighted Feature Composition



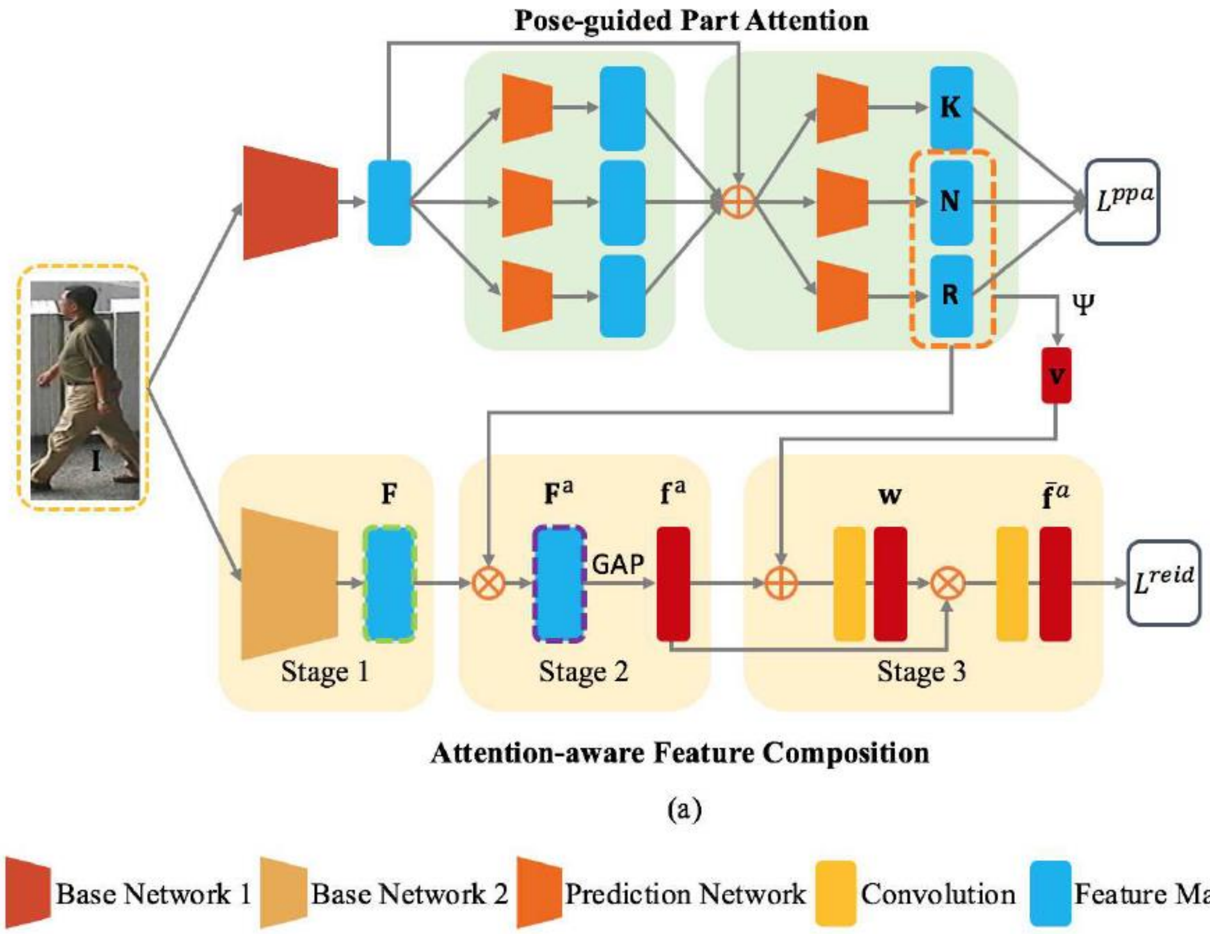
Generate a weight vector

- Concat visibility score and feature
- Apply 1x1 conv

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Architecture

Step 1: Train PPA independently
Step 2: Train GCN independently
Step 3: Attention-Aware Feature Alignment
Step 4: Weighted Feature Composition



Generate a weight vector

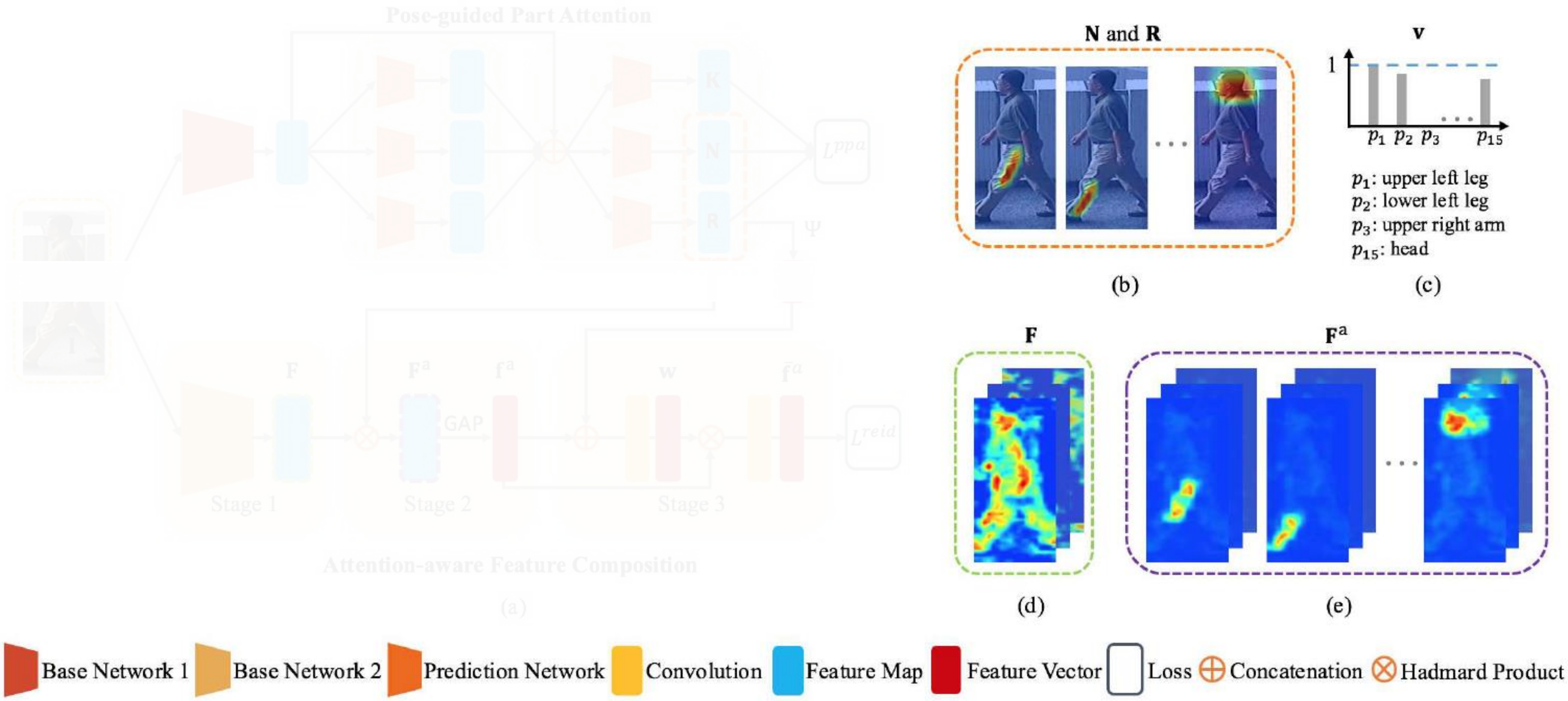
- Concat visibility score and feature
- Apply 1x1 conv

Element-wise product of weight vector and feature vector

$$\bar{f}^a = Conv(Concat(\{w_p \cdot f_p\}_{p=1}^{P})),$$

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Response Maps



Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Overview

- Good idea, this can avoid the noise in the RoI methods
- Too complex…

Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, Wanli Ouyang: Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018

Related Works

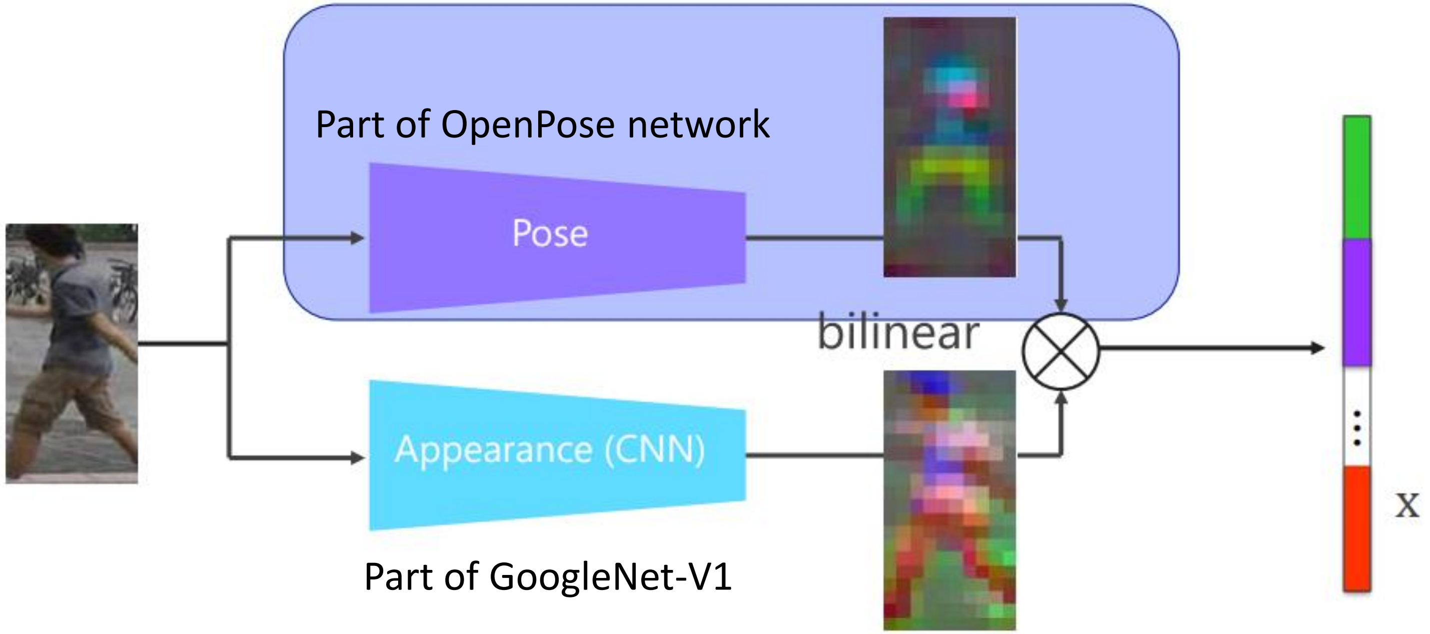
- An Improved Deep Learning Architecture for Person Re-Identification, CVPR 2015
- Deeply-Learned Part-Aligned Representations for Person Re-Identification, ICCV 2017

w/o pose

- Attention-Aware Compositional Network for Person Re-Identification, CVPR 2018
- Part-Aligned Bilinear Representations for Person Re-Identification, ECCV 2018

w/ pose

Architecture



Each subvector in $\mathbf{x}$ corresponds to a key point:

$$\mathrm{vec}(\mathbf{a}\otimes\mathbf{p}) = [(p_1\mathbf{a})^\top\ (p_2\mathbf{a})^\top\ \dots (p_{c_P}\mathbf{a})^\top]^\top$$

Pose estimator *pretrained on COCO*, and re-trained *only with the re-id loss*

Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, Kyoung Mu Lee: Part-Aligned Bilinear Representation for Person Re-Identification, ECCV 2018

Response Maps

Left: Image, middle: appearance map, right: part map

Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, Kyoung Mu Lee: Part-Aligned Bilinear Representation for Person Re-Identification, ECCV 2018

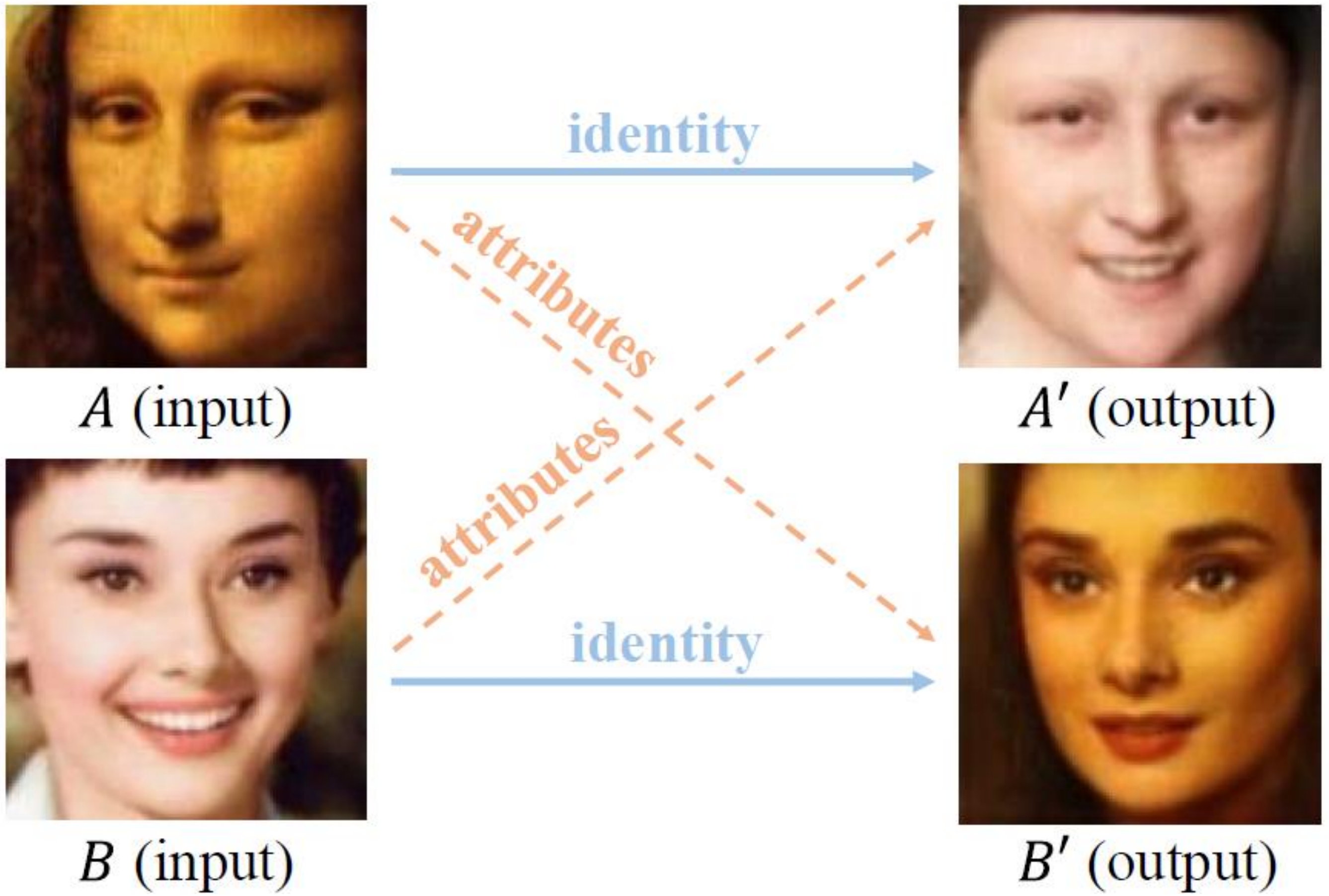
Overview

- Simple, effective
- Not take view changes into consideration

GANs methods

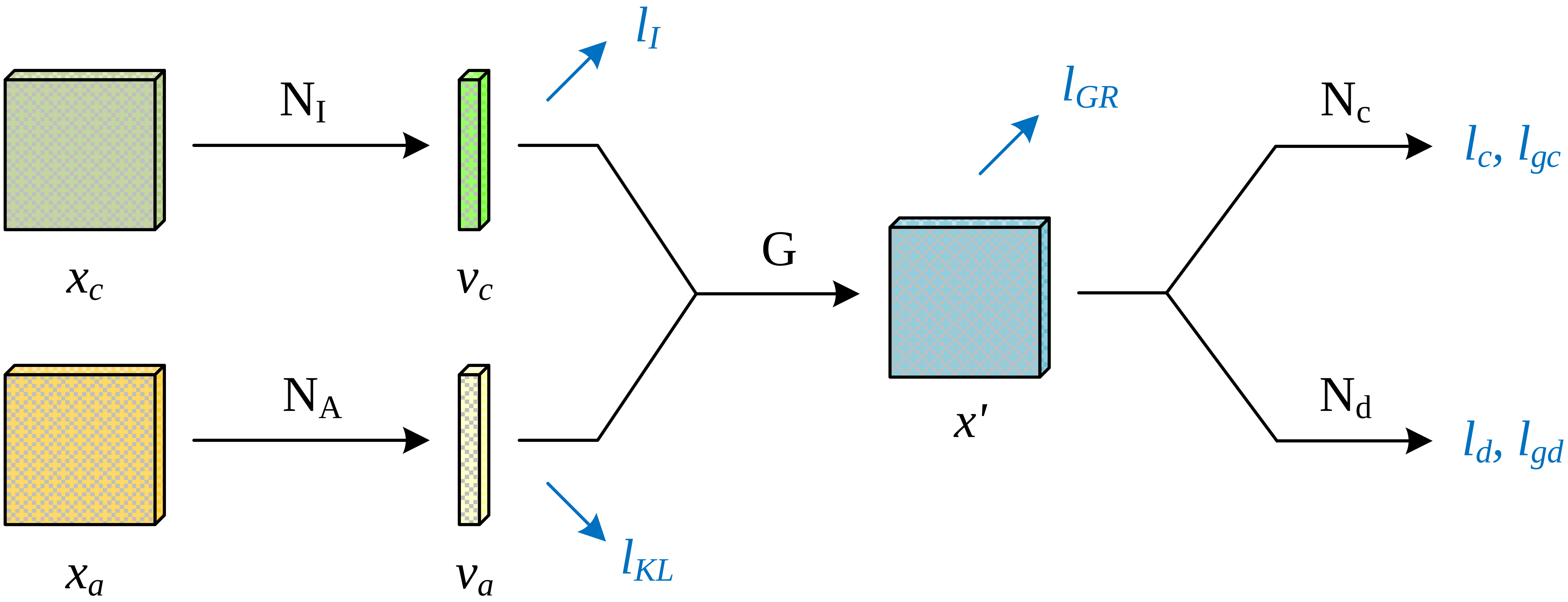
- IP-GAN for face synthesis, CVPR 2018
- PN-GAN for Re-ID, ECCV 2018
- FD-GAN for Re-ID, NIPS 2018

Identity Preserving GAN

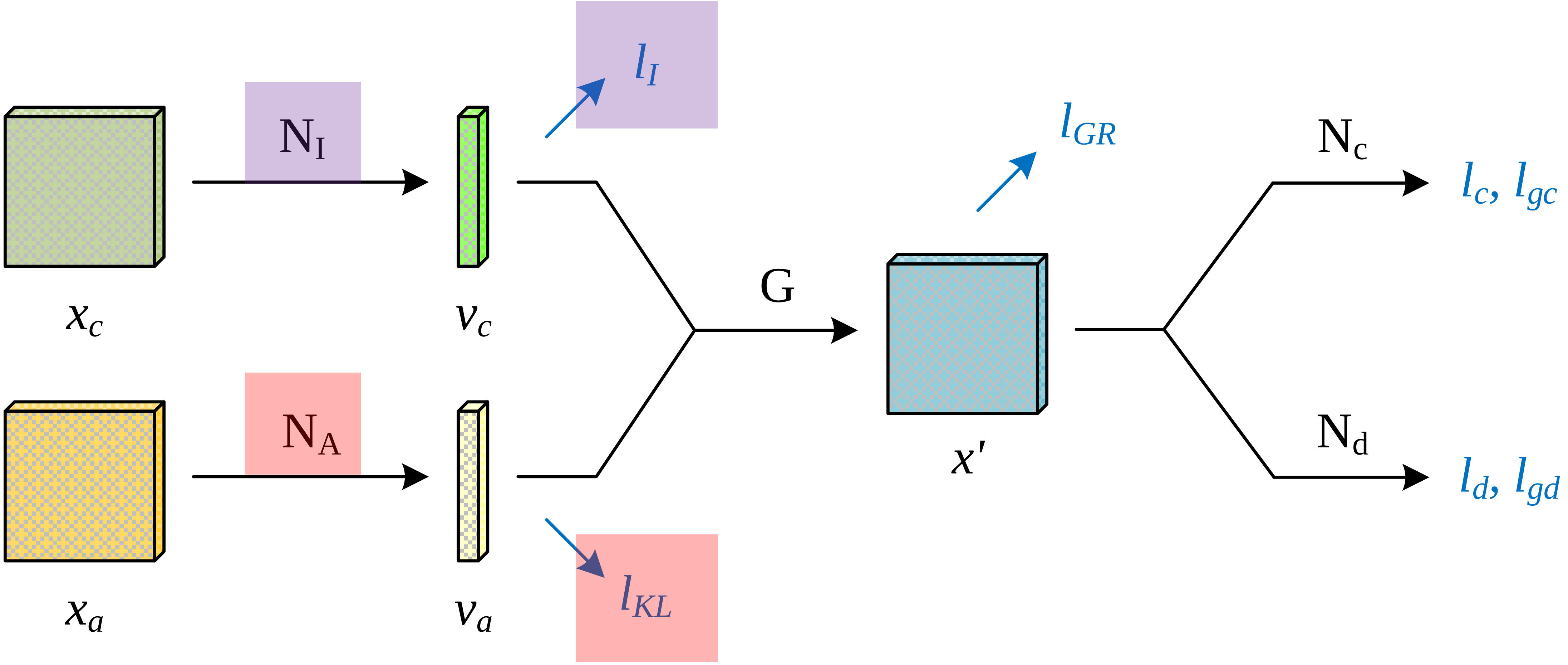


Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*

Identity Preserving GAN - Architecture



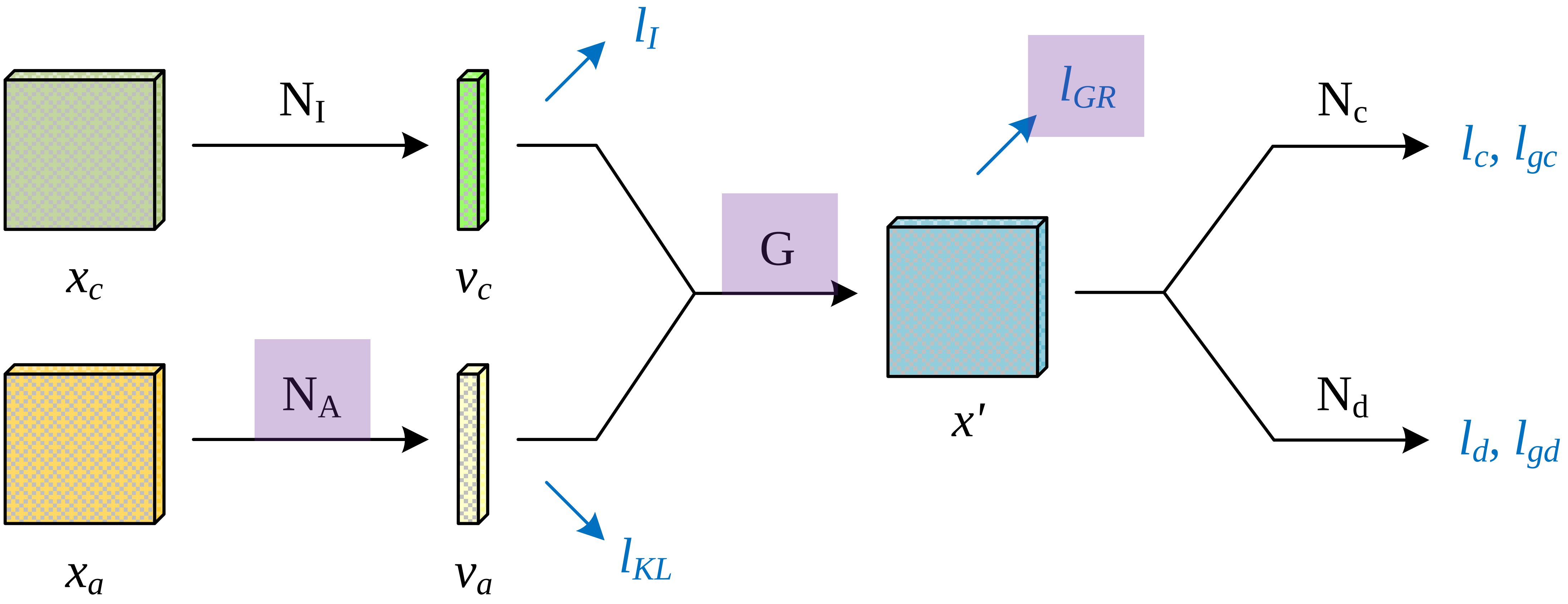
Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*



$$l_I = -E[\log P(c|x_c)]$$

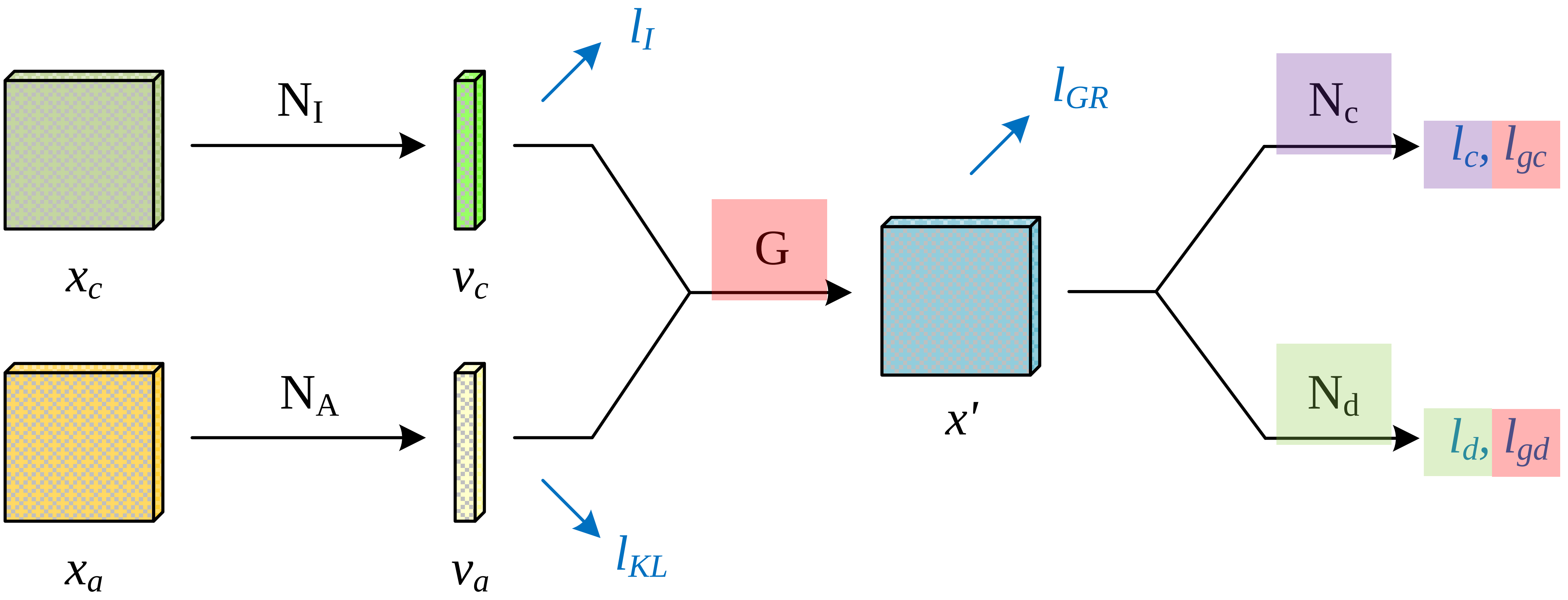
$$l_{KL} = \frac{1}{2}\left(\mu^T\mu + \sum_{j=1}^{J}(e^{\epsilon} - \epsilon - 1)\right)$$

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*



$$l_{GR} = \begin{cases} \frac{1}{2} \| (x_a) - (x') \|^2, & x_a = x_c \\ \frac{0.1}{2} \| (x_a) - (x') \|^2, & otherwise \end{cases}$$

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*



$$l_c = -E[\log P(c|x_c)]$$

$$l_d = -E[\log D(x_a)] - E[\log(1 - D(x'))]$$

$$l_{gc} = \frac{1}{2} \| f_c(x_c) - f_c(x') \|^2$$

$$l_{gd} = \frac{1}{2} \| f_d(x_a) - f_d(x') \|^2$$

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*

Identity Preserving GAN - Overview

- Objective: Generate faces with diff. id & attribute
- Weakness: Only can generate faces with known id
- Strength: g.t. for x' is not needed
- Training set: (x_c, x_a) only require x_c id is known
- Identity & attribute feature decoupled

Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua: Towards Open-Set Identity Preserving Face Synthesis, *CVPR 2018*

Content

- Part-aligned Representation Learning
- Image generation in Re-ID
- GAN as supervisor

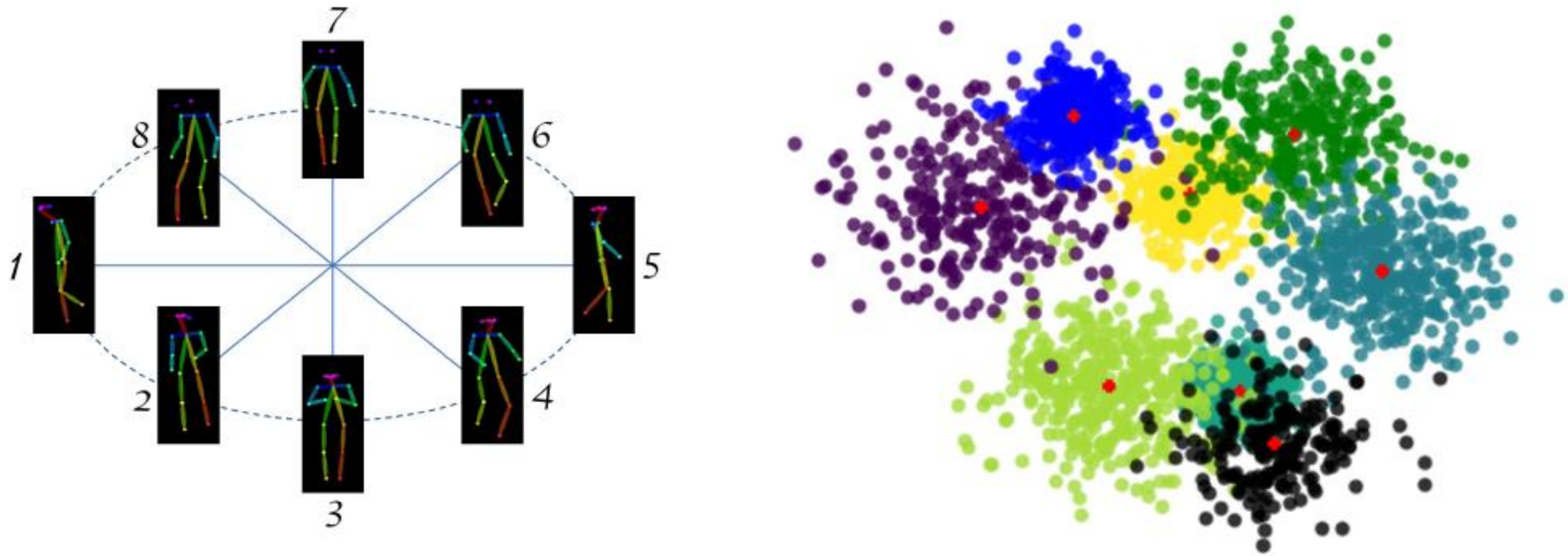
Pose Normalized GAN - Motivation

- Identity-Sensitive View-Insensitive (ISVI) features are needed
- → Part models
  - ~~Lack of scalability~~
  - ~~Lack of generalizability~~
  - ISVI features and IIVS features are not independent
  - $p(x, y) \Rightarrow p(x)$? Not easy
  - But we can have… $p(x, y) \Rightarrow p(x, y_1), p(x, y_2), p(x, y_3)$ …

Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Model

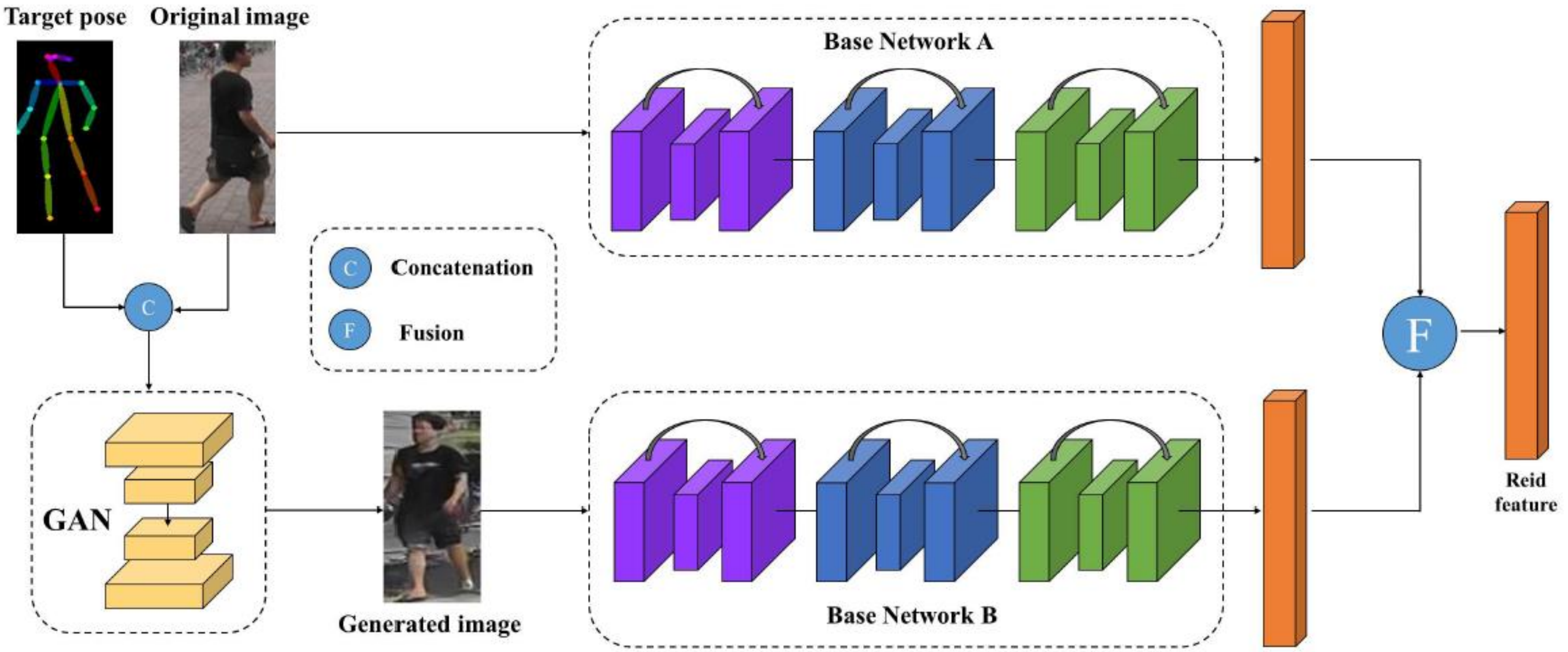
- At which poses we want to generate?



(a) Eight canonical poses on Market-1501 (b) t-SNE visualization of different poses.

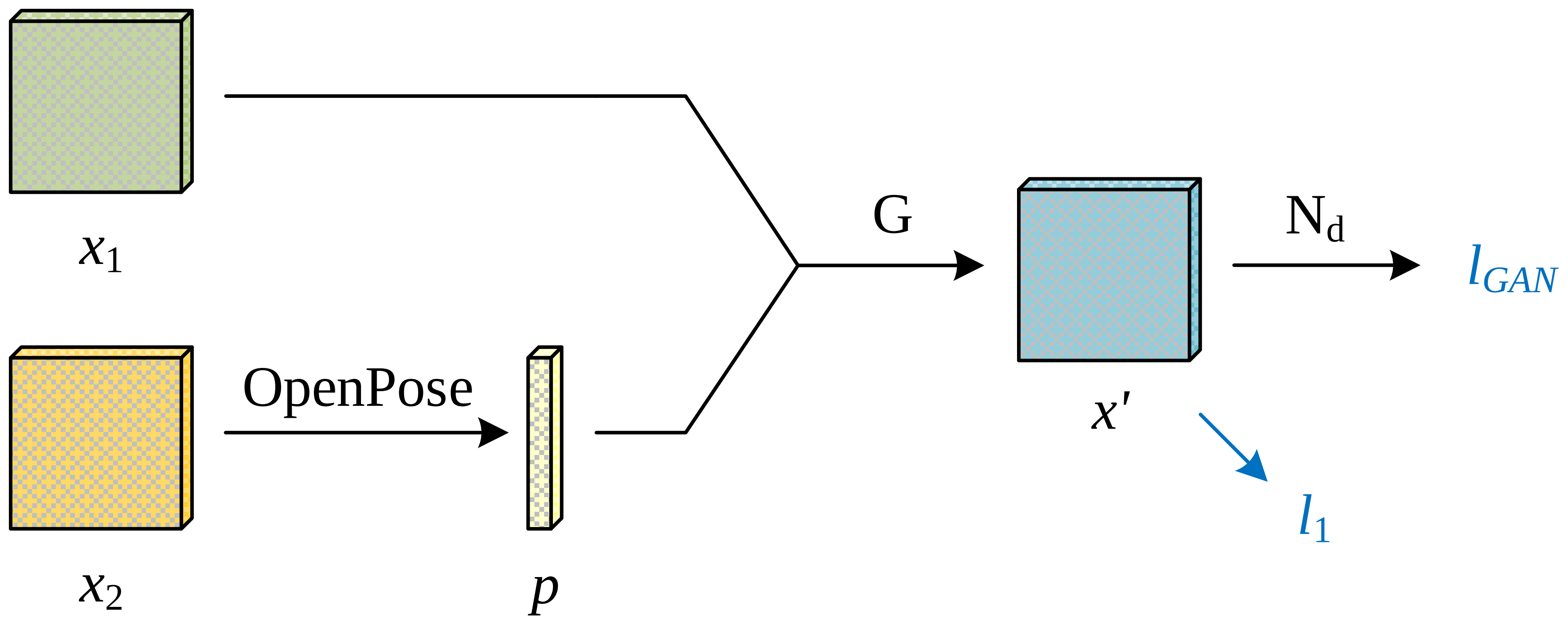
Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Architecture

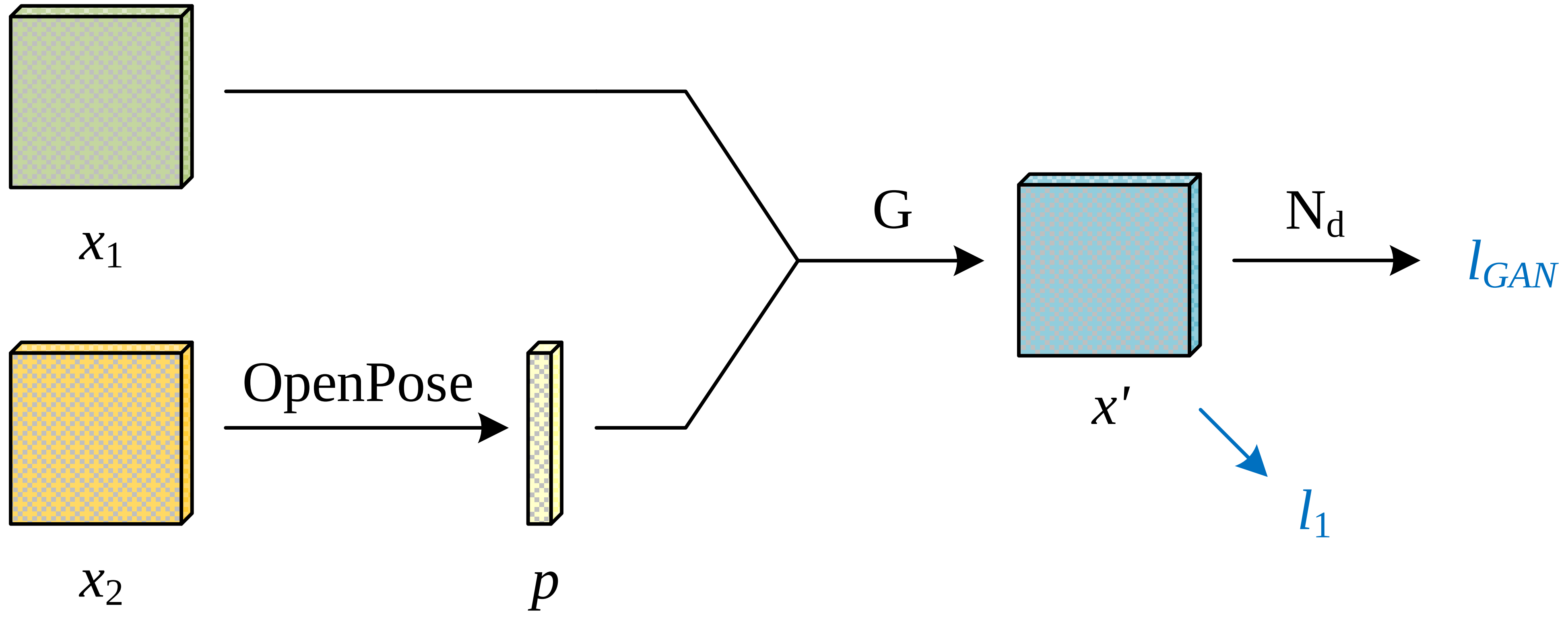


Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Architecture



Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*



$l_1 = \| x_2 - x' \|_1$

$l_{GAN} = -E[\log D(x_2)] - E[\log(1 - D(x'))]$

Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Training steps

- Train PN-GAN
- Do Re-ID using pretrained PN-GAN

Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Experimental Results

- Market-1501
  - Baseline (same structure without PN-GAN):
    - Top-1: 87.26
    - mAP: 69.32
  - PN-GAN
    - Top-1: 89.43
    - mAP: 72.58
- Performance boost is limited (+2 top-1, +3 mAP)
- Some part-models can achieve +10 mAP with similar setting

Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

Pose Normalized GAN - Overview

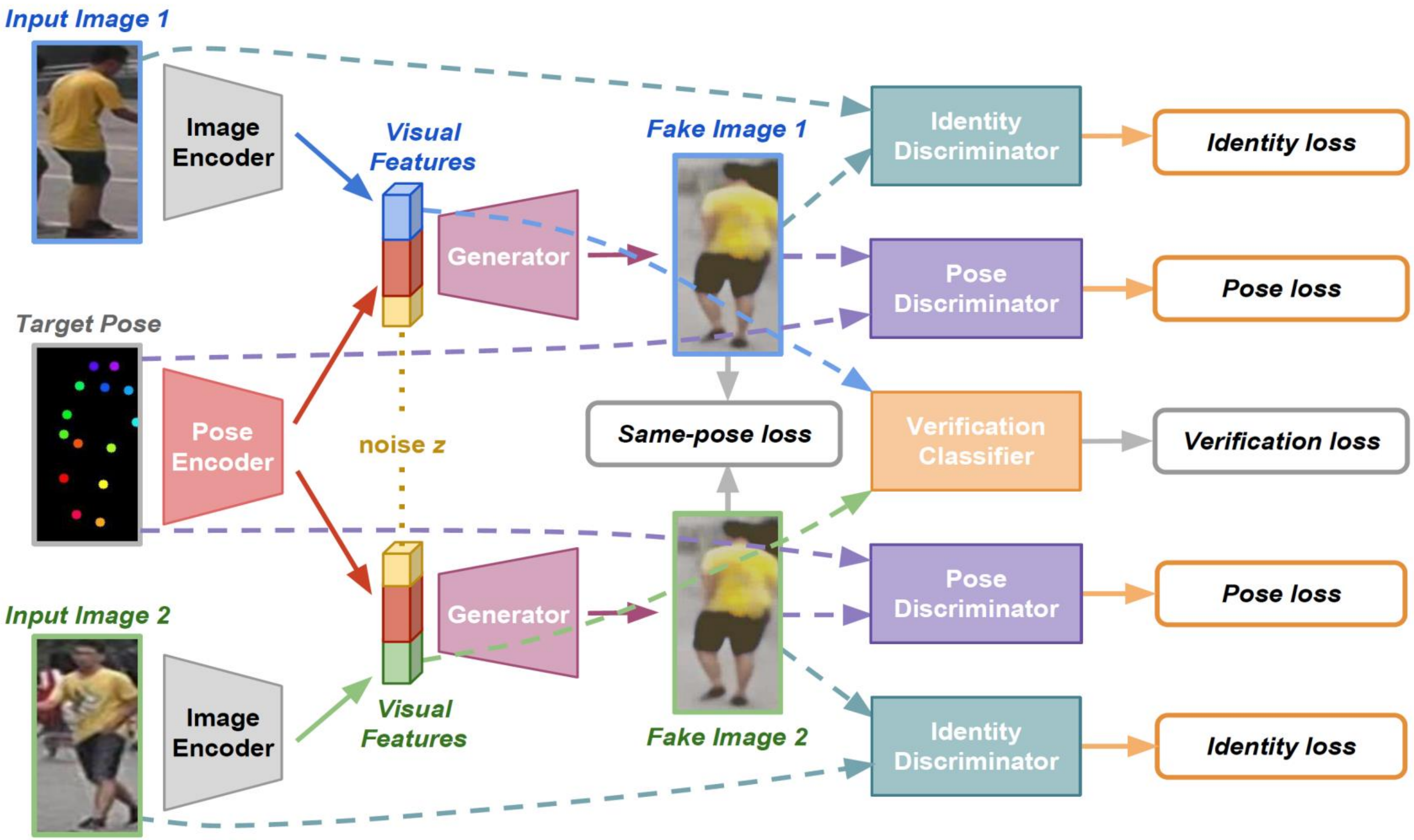
- Objective: Generate new images for better Re-ID
- Weakness: Small available data for training
  - x_1, x_2 required to be same person (g.t. for x' is needed)
- Implicitly decouple id & pose feature

Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yu-Gang Jiang, Xiangyang Xue: Pose-Normalized Image Generation for Person Re-identification, *ECCV 2018*

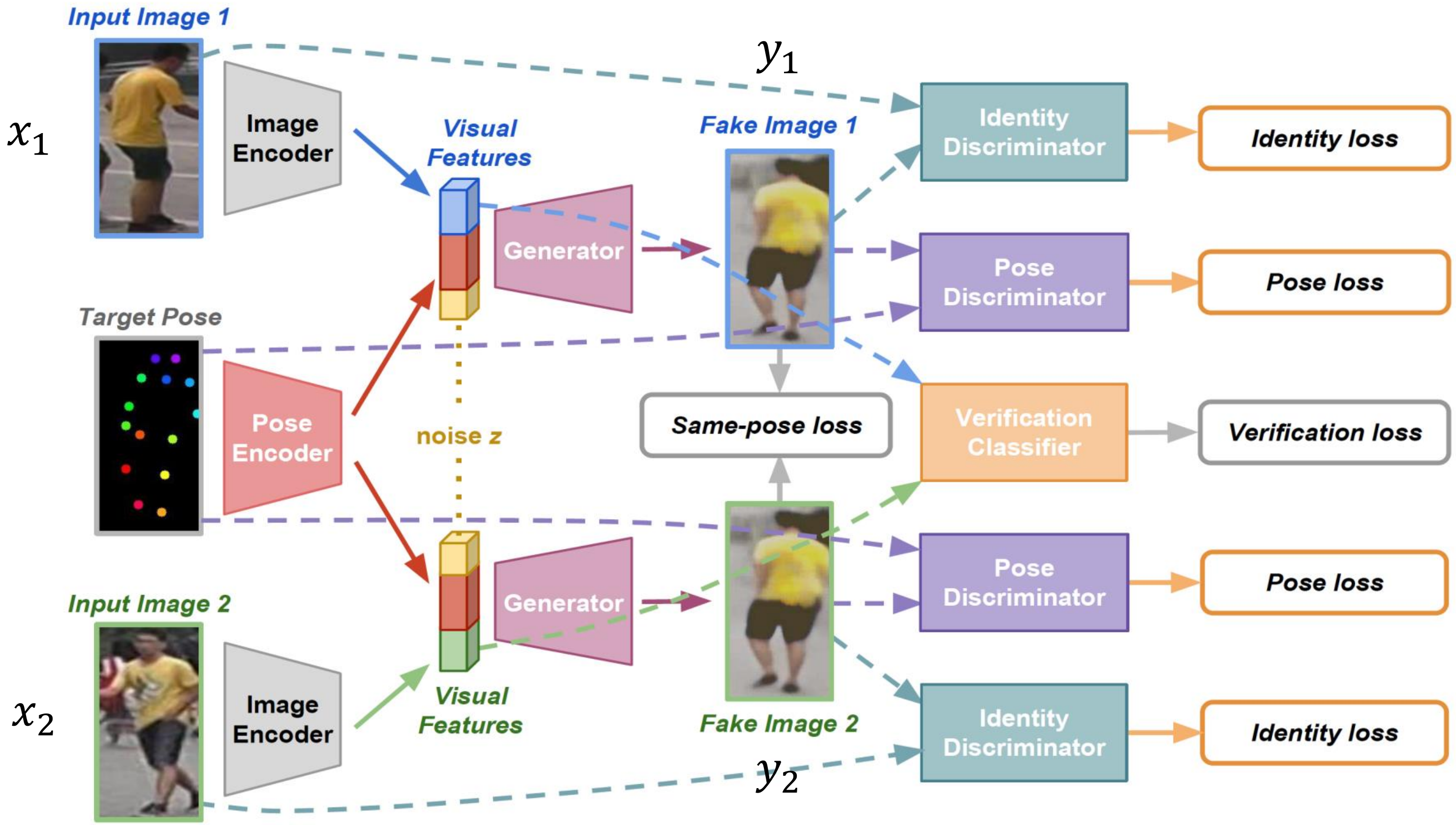
Content

- Part-aligned Representation Learning
- Image generation in Re-ID
- GAN as supervisor

Feature Distilling GAN



Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li: Pose-guided Feature Distilling GAN for Robust Person Re-identification, *NIPS 2018*



$$\mathcal{L}_v = -C \log d(x_1, x_2) - (1 - C)(1 - \log d(x_1, x_2)), \quad C \text{ is the ground-truth label}$$

$$\mathcal{L}_{id} = \max_{D_{id}} \quad \sum_{k=1}^{2} \left(\mathbb{E}_{y'_k \in \mathcal{Y}}[\log D_{id}(x_k, y'_k)] + \mathbb{E}_{y_k \in \mathcal{Z}}[\log(1 - D_{id}(x_k, y_k))] \right)$$

$$\mathcal{L}_{pd} = \max_{D_{pd}} \quad \sum_{k=1}^{2} \left(\mathbb{E}_{y'_k \in \mathcal{Y}}[\log D_{pd}([p, y'_k])] + \mathbb{E}_{y_k \in \mathcal{Z}}[\log(1 - D_{pd}([p, y_k]))] \right)$$

$$\mathcal{L}_r = \sum_{k=1}^{2} \frac{1}{mn} \|y_k - y'_k\|_1$$

$$\mathcal{L}_{sp} = \frac{1}{mn} \|y_1 - y_2\|_1$$

Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li: Pose-guided Feature Distilling GAN for Robust Person Re-identification, *NIPS 2018*

Feature Distilling GAN - Training steps

- Train feature encoder using Re-ID loss
- Train FD-GAN
- Global finetuning

Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li: Pose-guided Feature Distilling GAN for Robust Person Re-identification, *NIPS 2018*

Feature Distilling GAN - Experimental Results

- Market-1501
  - Baseline (same structure without FD-GAN):
    - Top-1: 88.2
    - mAP: 72.5
  - FD-GAN
    - Top-1: 90.5
    - mAP: 77.7
- Performance boost is limited (+2 top-1, +5 mAP)

Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li: Pose-guided Feature Distilling GAN for Robust Person Re-identification, *NIPS 2018*

Feature Distilling GAN - Overview

- Objective: Learn a better feature encoder
- Weakness: Small available data for training
  - x_1, x_2 required to be same person (g.t. for x' is needed)

Yixiao Ge, Zhuowan Li, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li: Pose-guided Feature Distilling GAN for Robust Person Re-identification, *NIPS 2018*

Overview

- Rethinking previous models
  - IP-GAN (face)
    - Objective: generate better face with specific id & attribute
    - Pro: g.t. for generated image is not required
    - Con: Can only generate known identities
  - PN-GAN (Re-ID)
    - Objective: generate better image with specific id & pose
    - Pro: can generate unknown id images
    - Con: g.t. is required during training, pose in not precious
  - FD-GAN (Re-ID)
    - Objective: GAN for better feature encoder learning
    - Pro: can generate unknown id images
    - Con: g.t. is required during training

Overview

- Rethinking previous models
  - IP-GAN (face)
    - Objective: generate better face with specific id & attribute
    - Pro: g.t. for generated image is not required
    - Con: Can only generate known identities
  - PN-GAN (Re-ID)
    - Objective: generate better image with specific id & pose
    - Pro: can generate unknown id images
    - Con: g.t. is required during training, pose in not precious
  - FD-GAN (Re-ID)
    - Objective: GAN for better feature encoder learning
    - Pro: can generate unknown id images
    - Con: g.t. is required during training

Thanks

Q&A