Deep Learning for Person ReID

Zhedong Zheng
University of Macau

Person Re-identification

Background

Learn pedestrian representations from

Multi-task learning <--->semantic
Part matching <--> syntax
Data augmentation

Conclusions and Future Works

Unsupervised / Semi-supervised Learning
Natural Language Based Retrieval
Video Based Person Re-ID

Person Re-identification

Background

Learn pedestrian representations from

- Multi-task learning
- Part matching
- Data augmentation

Conclusions and Future Works

- Unsupervised / Semi-supervised Learning
- Natural Language Based Retrieval
- Video Based Person Re-ID

Background

Background

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Input

Query

Camera Inputs

Large-scale

image pool (Gallery)

Retrieval

Output

Background

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

How the human do?

Input

Description in the brain

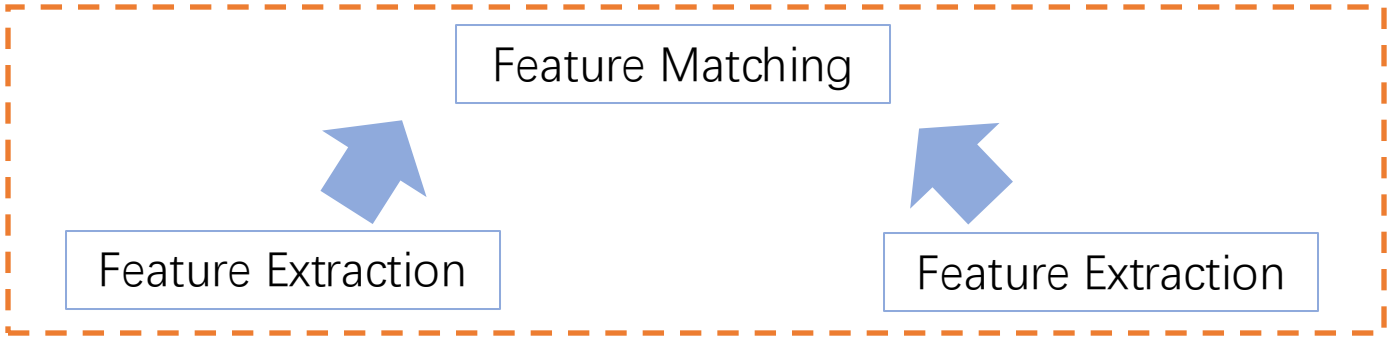
A man in dark coat and blue (red) pants are walking towards the camera.

Camera Inputs

Large-scale
image pool (Gallery)

Background

Zhedong Zheng
Centre for Artificial Intelligence (CAI)



How the machine do?

语义鸿沟 / 快速检索

Input

Camera Inputs

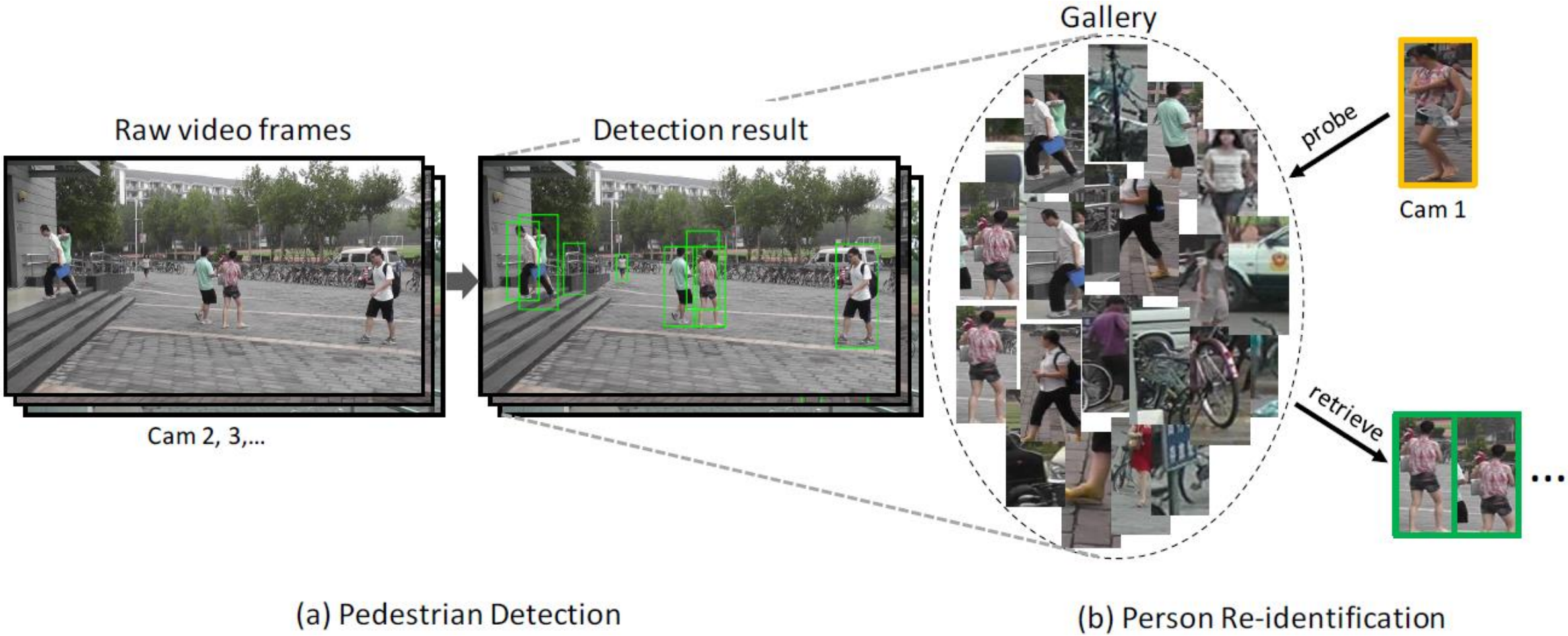
Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Only Color? It is not sufficient for hard samples.

（图片来自http://liangzheng.org/Project/project_reid.html）

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Face Recognition? Faces are small and front faces are rare.



（图片来自http://www.liangzheng.com.cn/Project/project_prw.html）

Background

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Cross-cameras

Detailed information

- Person Re-ID is mostly viewed as an image retrieval problem.
- The main challenge is the different view points. e.g., **camA <-> camB**
- We may face large-scale image pool. (detailed information)

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Deep Learning

1986 Rumelhart, Hinton and Williams propose back-propagating.
2006 Hinton etal., propose basic paper for 'deep learning'.
2012 Hinton etal., win the ImageNet competition by a large margin.

- Large-scale datasets (ImageNet)
- GPUs for faster training and testing
- New Improvements: Dropout, ReLu, BatchNorm and so on.

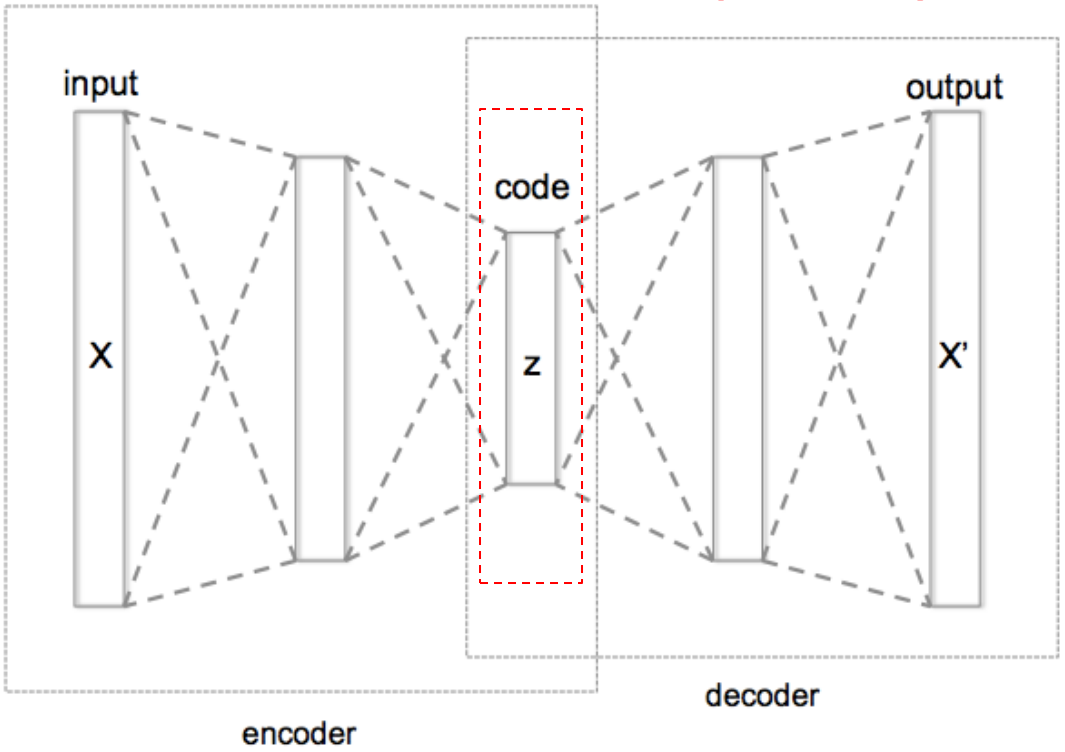
Face recognition competition (LFW) surpasses the human.

How to apply deep learning to Person re-ID?

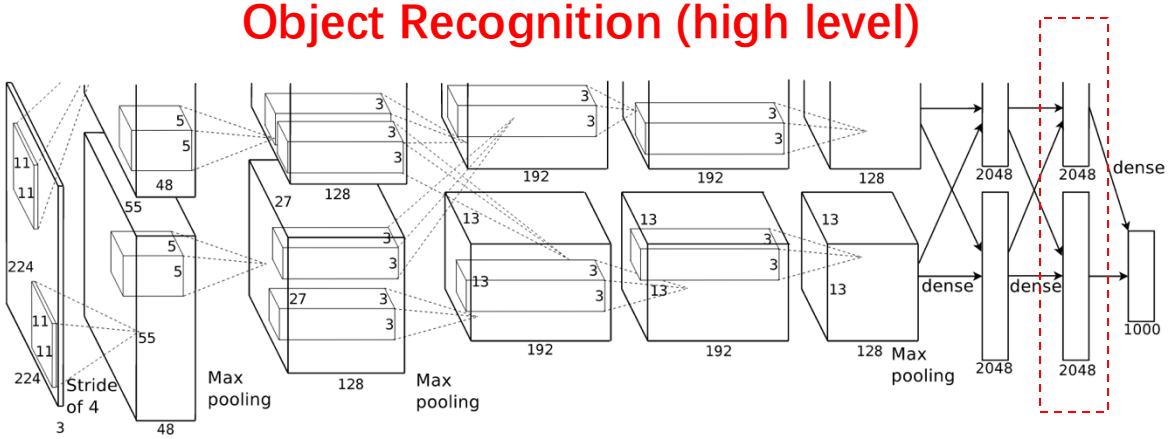
Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Representation Learning

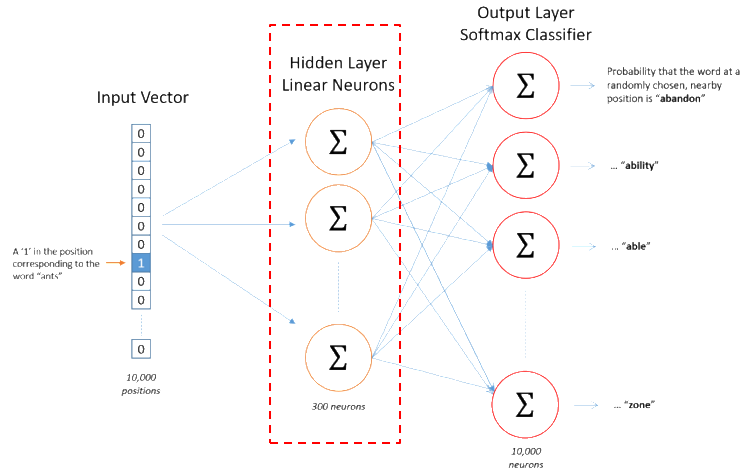
Auto-encoder (low level)



Object Recognition (high level)



Word2vec (context)



Person Re-identification

Background

Learn pedestrian representations from

- Multi-task learning <--->semantic
- Part matching
- Data augmentation

Conclusions and Future Works

- Unsupervised / Semi-supervised Learning
- Natural Language Based Retrieval
- Video Based Person Re-ID

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Multi-task learning

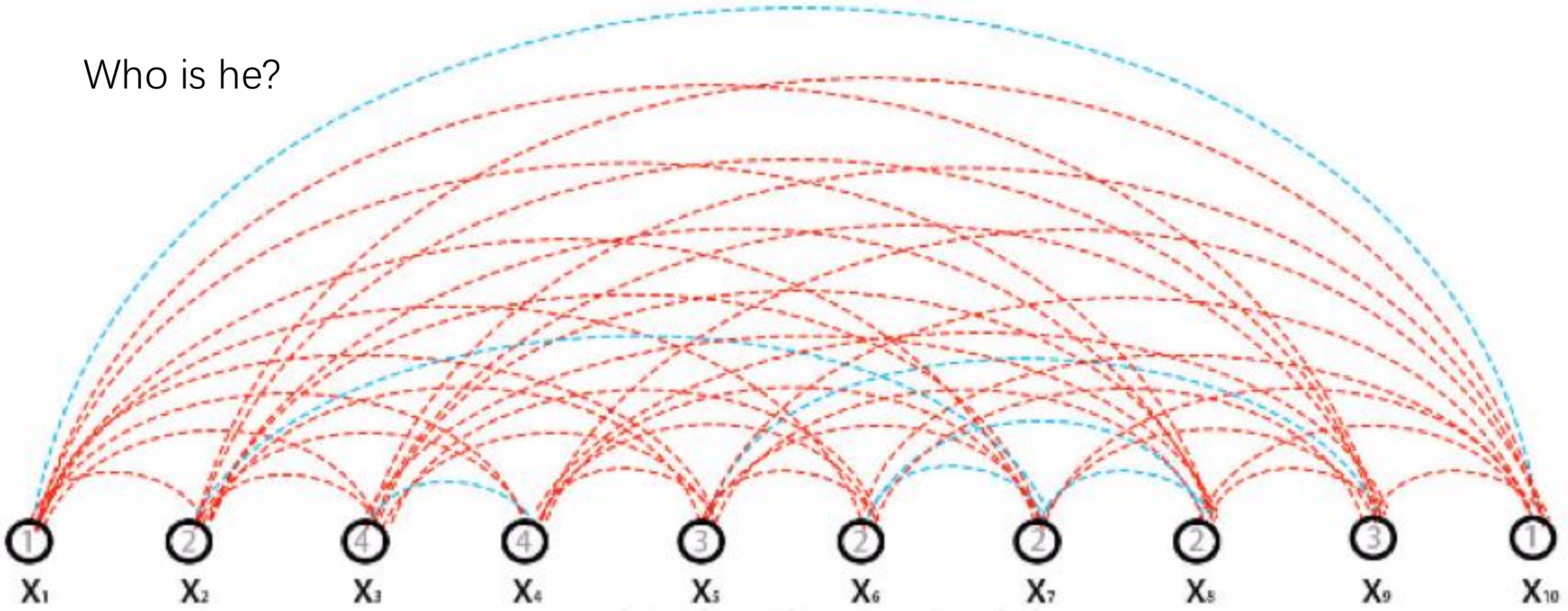
1. Identification + Verification Loss
2. Triplet Loss
3. Attribute Recognition

Multi-task Learning

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Identification Loss

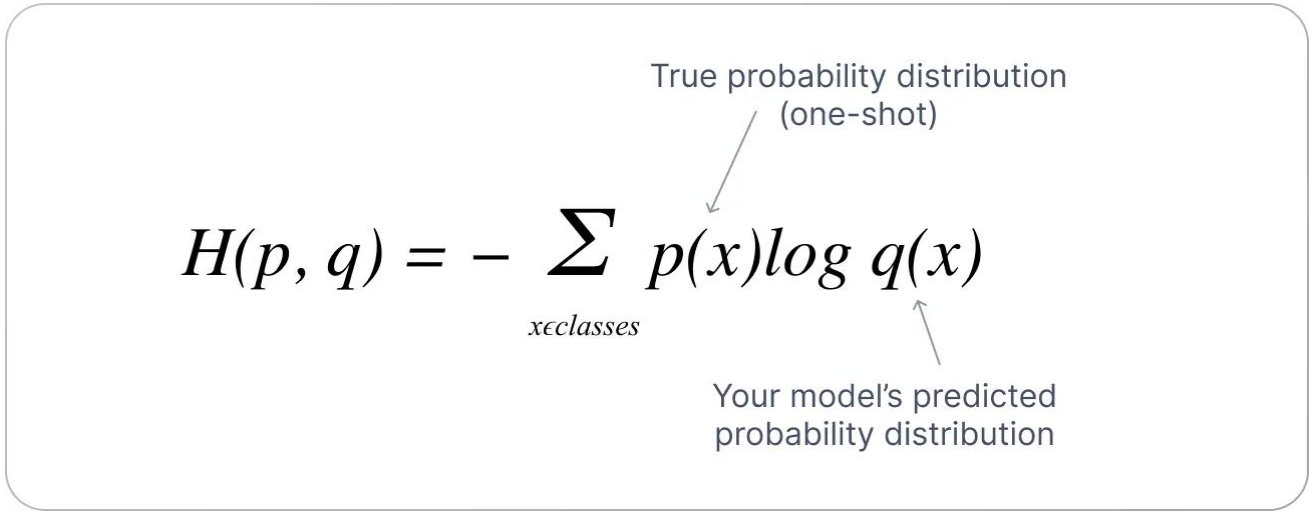
Who is he?



Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Identification Loss

How to calculate cross-entropy loss?



Cat -> 70% dog, 20% Cat, 10% bird
GT -> 0% dog, 100% Cat, 0% bird
-log(0.2) = 0.69897000433
-log(0.95) = 0.02227639471

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

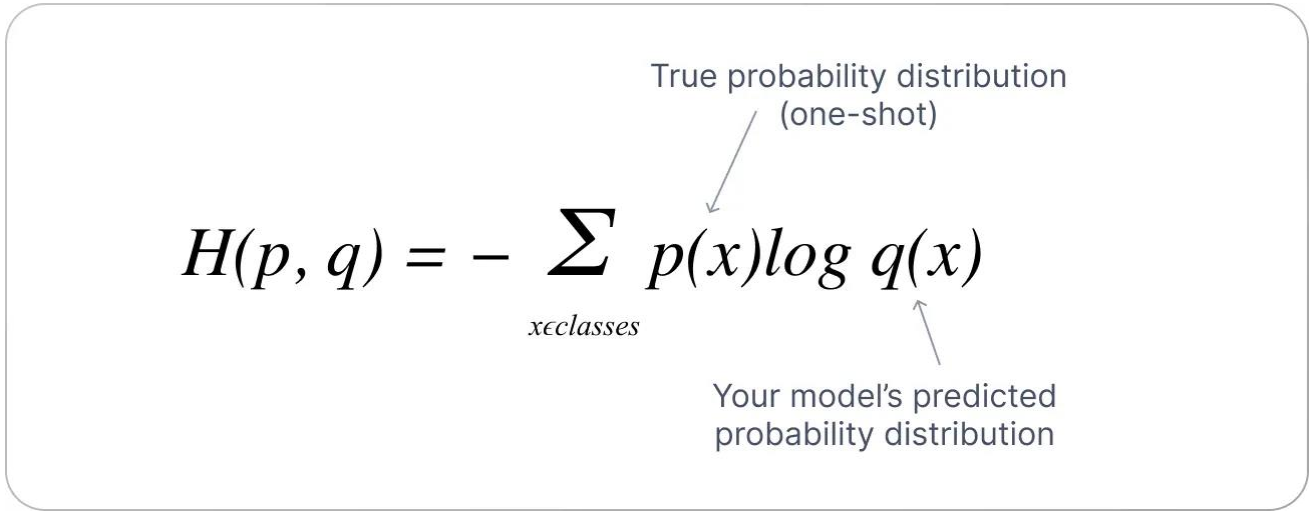
- Verification Loss

Are they the same person?

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Identification Loss

How to calculate cross-entropy loss?

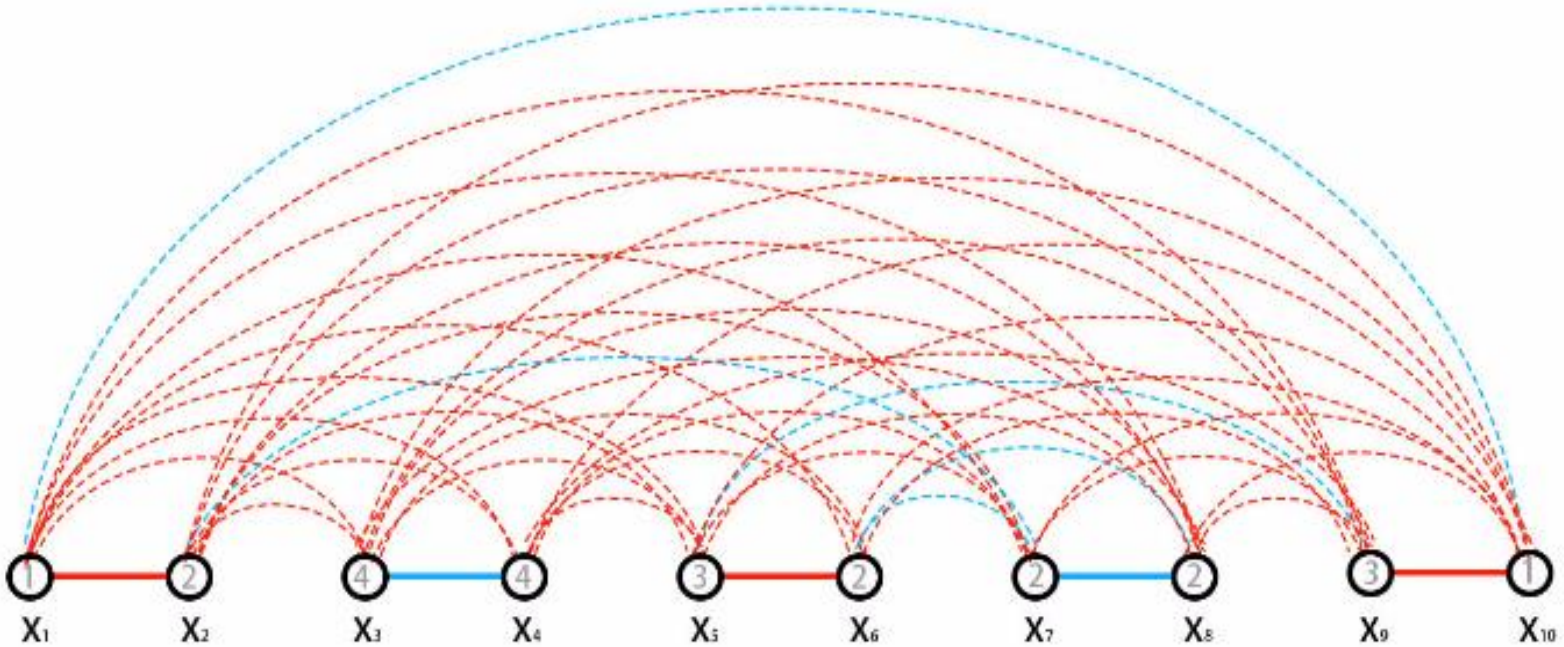


Cat -> 70% yes, 30% no,
GT -> 0% yes, 100% no,
-log(0.3)

Multi-task Learning

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Verification+Identification Loss



Zheng, Z., Zheng, L., & Yang, Y. (2016). **A discriminatively learned cnn embedding for person re-identification.** *arXiv preprint arXiv:1611.05666*.

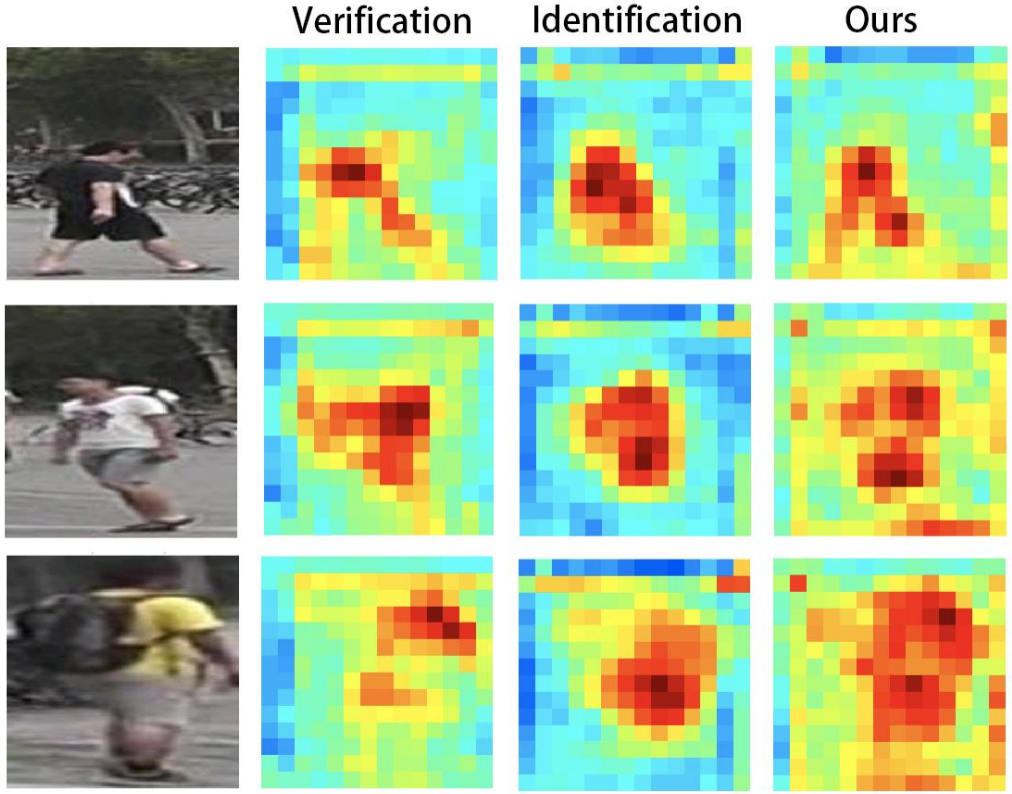
Geng, M., Wang, Y., Xiang, T., & Tian, Y. (2016). **Deep transfer learning for person re-identification.** *arXiv preprint arXiv:1611.05244*.

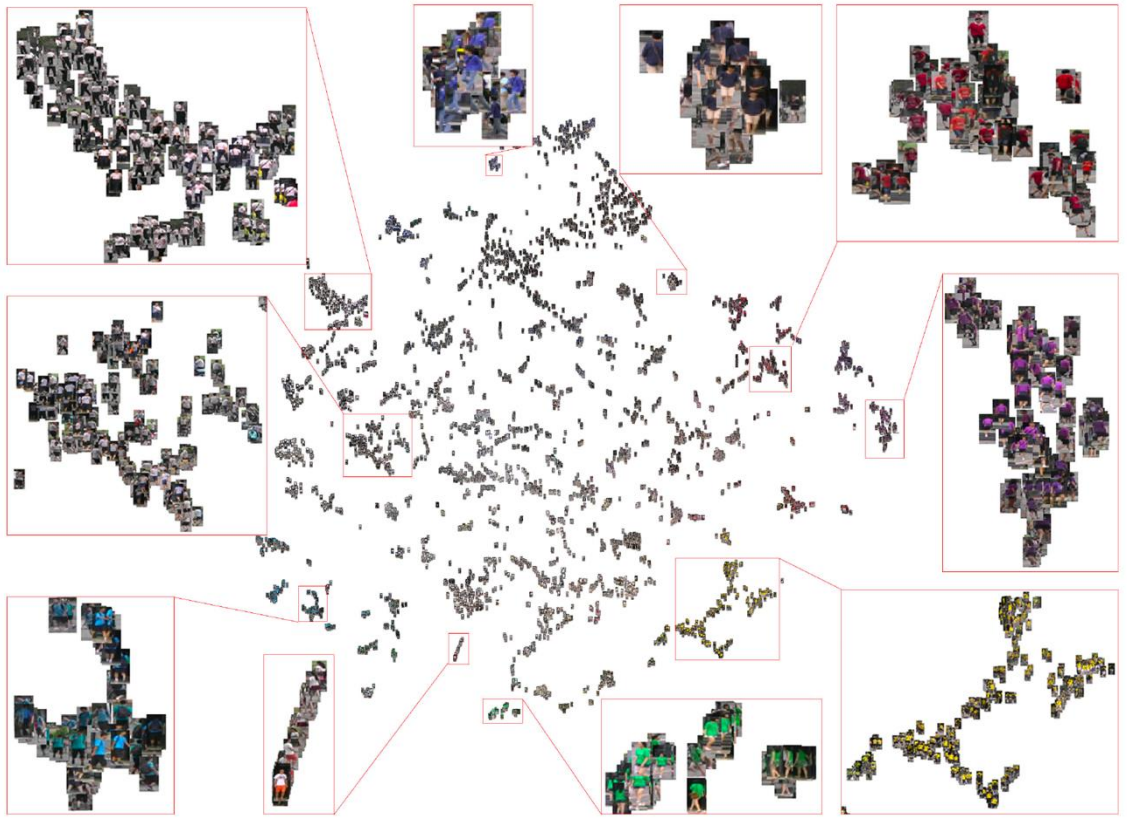
Github:https://github.com/layumi/2016_person_re-ID

Multi-task Learning

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Verification+Identification Loss

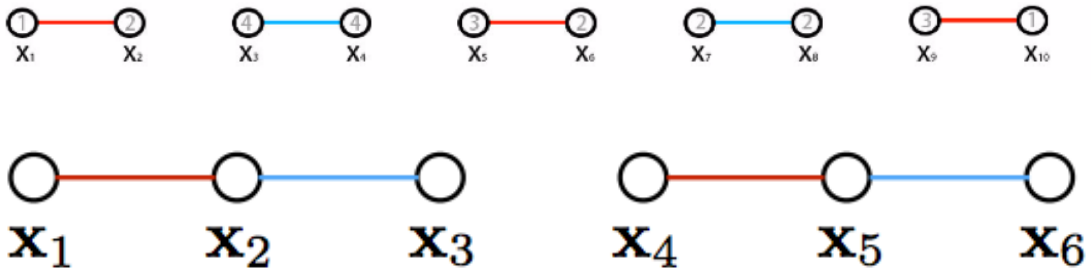




Github:https://github.com/layumi/2016_person_re-ID

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Triplet Loss



$$\mathcal{L}_{\text{tri}}(\theta) = \sum_{\substack{a,p,n \\ y_a=y_p\neq y_n}} [m + D_{a,p} - D_{a,n}]_+ .$$

Hermans, A., Beyer, L., & Leibe, B. (2017). **In Defense of the Triplet Loss for Person Re-Identification**. *arXiv preprint arXiv:1703.07737*.

Github:https://github.com/VisualComputingInstitute/triplet-reid

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Attribute Learning

Lin, Y., Zheng, L., Zheng, Z., Wu, Y., & Yang, Y. (2017). **Improving person re-identification by attribute and identity learning.** *arXiv preprint arXiv:1703.07220*.

Project: https://vana77.github.io/

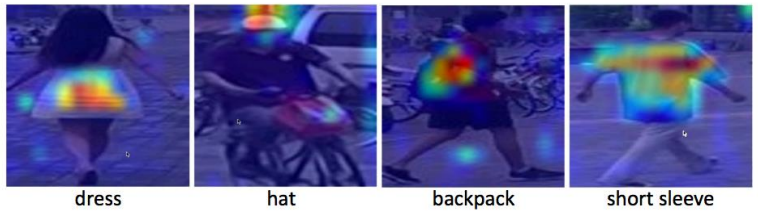


Figure 6. Intermediate features maps learned in our network correspond to certain attributes.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Attribute Learning

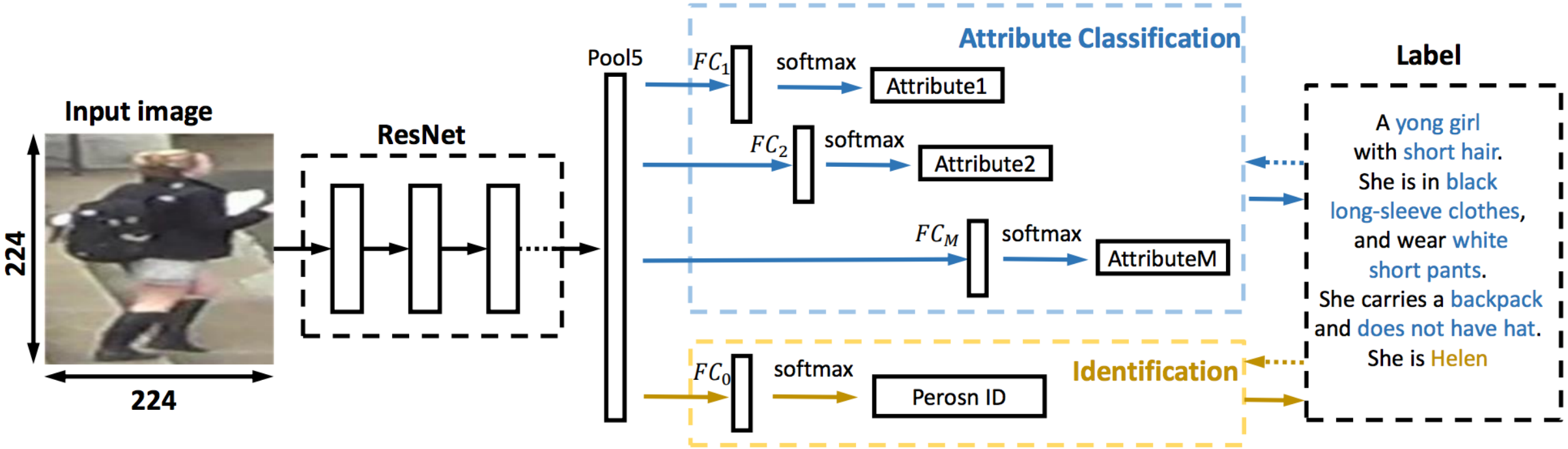


Figure 2. An overview of the APR network. During training, it predicts M attribute labels and an ID label. The weighted sum of the individual losses is back propagated. During testing, we extract the Pool5 (ResNet-50) or FC7 (CaffeNet) descriptors for retrieval.

Person Re-identification

Background

Learn pedestrian representations from

- Multi-task learning
- Part matching <--> syntax
- Data augmentation

Conclusions and Future Works

- Unsupervised / Semi-supervised Learning
- Natural Language Based Retrieval
- Video Based Person Re-ID

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Part Matching

1. Horizontal Split/Neighbor Comparison
2. Part Detection/Alignment
3. Detection + reID
4. Attention Matching

Part Matching

- Horizontal Split

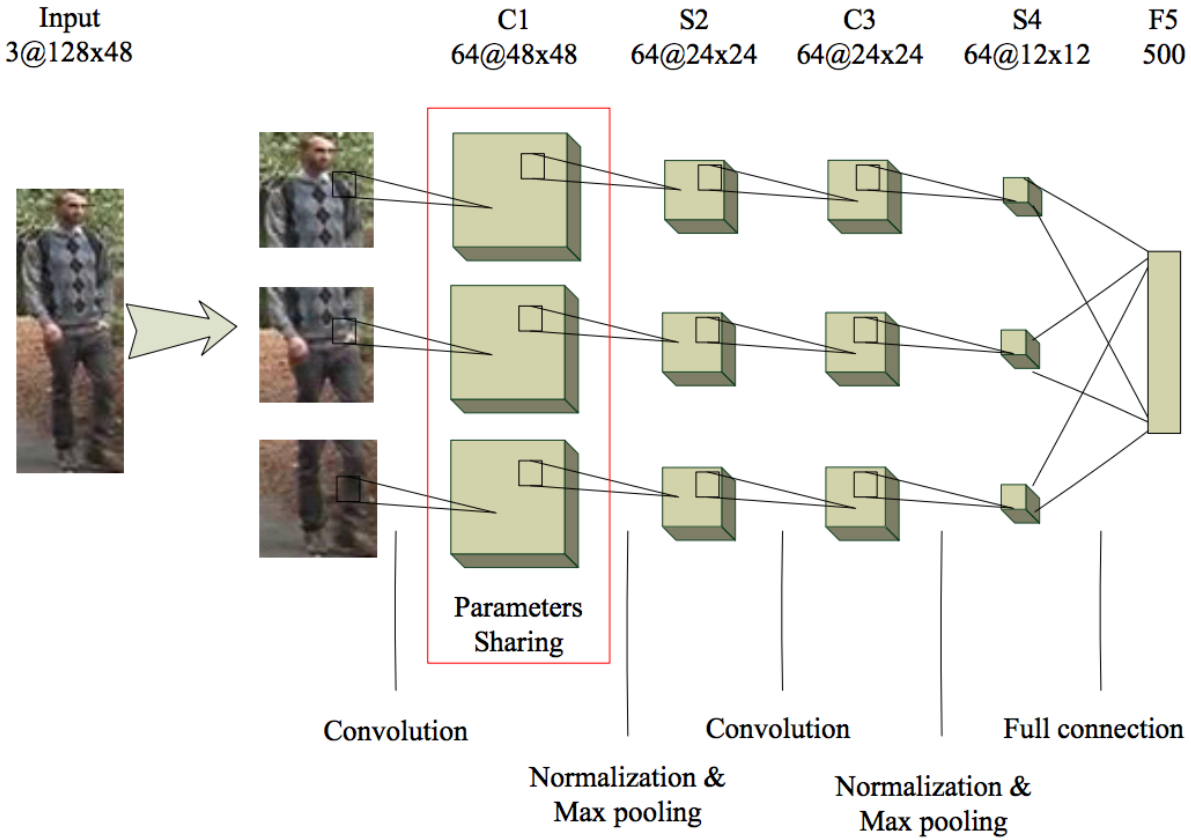


Fig. 2. The structure of the 5-layer CNN used in our method.

Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014, August). **Deep metric learning for person re-identification**. In *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 34-39). IEEE.

Part Matching

- Horizontal Split

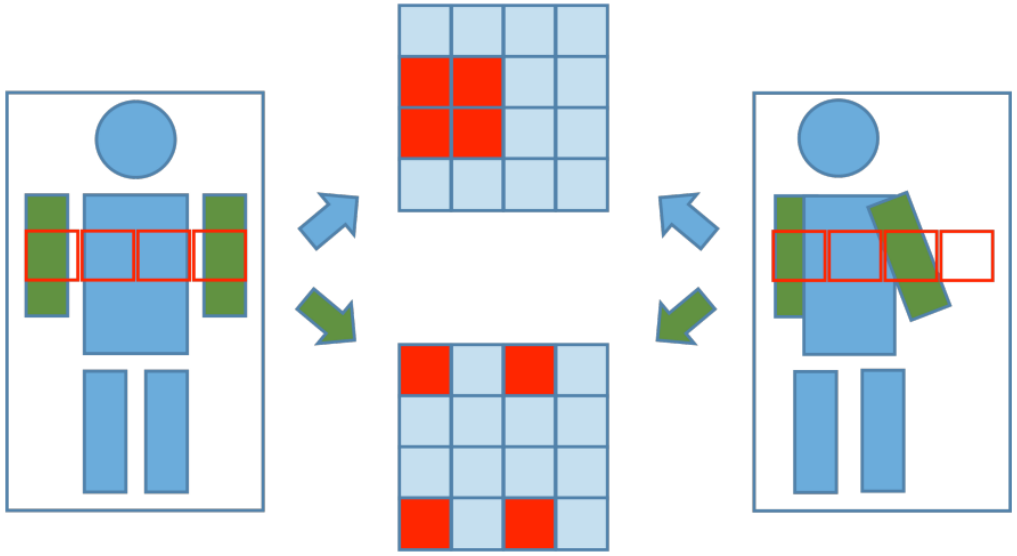


Figure 4. Illustration of patch matching in FPNN. One stripe generates two patch displacement matrices because there are two filter pairs. One detects blue color and the other detects green.

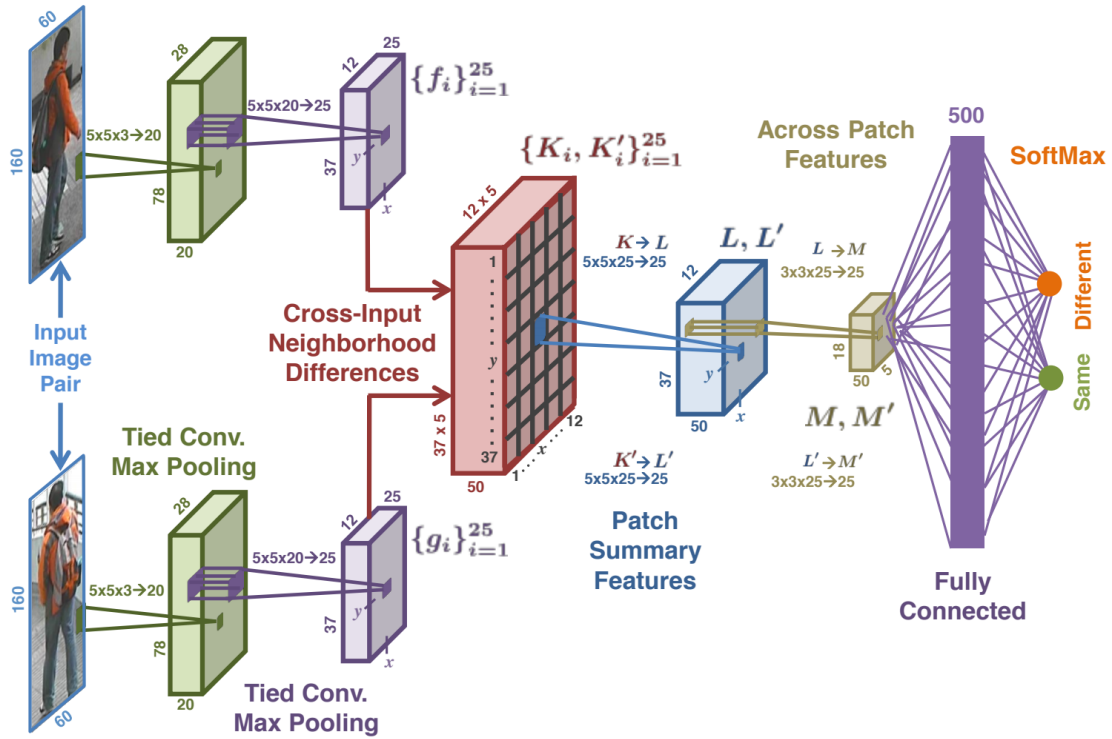
Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). **Deepreid: Deep filter pairing neural network for person re-identification**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 152-159).

$$S^k_{(i,j)(i',j')} = f^k_{ij} g^k_{i'j'},$$

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Neighbor Comparison



Ahmed, E., Jones, M., & Marks, T. K. (2015). **An improved deep learning architecture for person re-identification**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3908-3916).

$$K_i(x,y) = f_i(x,y)\mathbb{1}(5,5) - \mathcal{N}[g_i(x,y)] \qquad (1)$$

where

$\mathbb{1}(5,5) \in \mathbb{R}^{5\times5}$ is a 5×5 matrix of 1s,

$\mathcal{N}[g_i(x,y)] \in \mathbb{R}^{5\times5}$ is the 5×5 neighborhood of g_i centered at (x,y).

Github:
1.https://github.com/ptran516/idla-person-reid
2.https://github.com/Ning-Ding/Implementation-CVPR2015-CNN-for-ReID

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Part Detect + Matching

Figure 1. Examples of misalignment correction by PoseBox. Row 1: original bounding boxes with detection errors/occlusions. Every consecutive two boxes correspond to a same person. Row 2: corresponding PoseBoxes. We observe that misalignment can be corrected to some extent.

Zheng, L., Huang, Y., Lu, H., & Yang, Y. (2017). **Pose invariant embedding for deep person re-identification.** *arXiv preprint arXiv:1701.07732*.

Github: https://github.com/erichuang0771/PoseBox-Reid

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Pedestrian Alignment

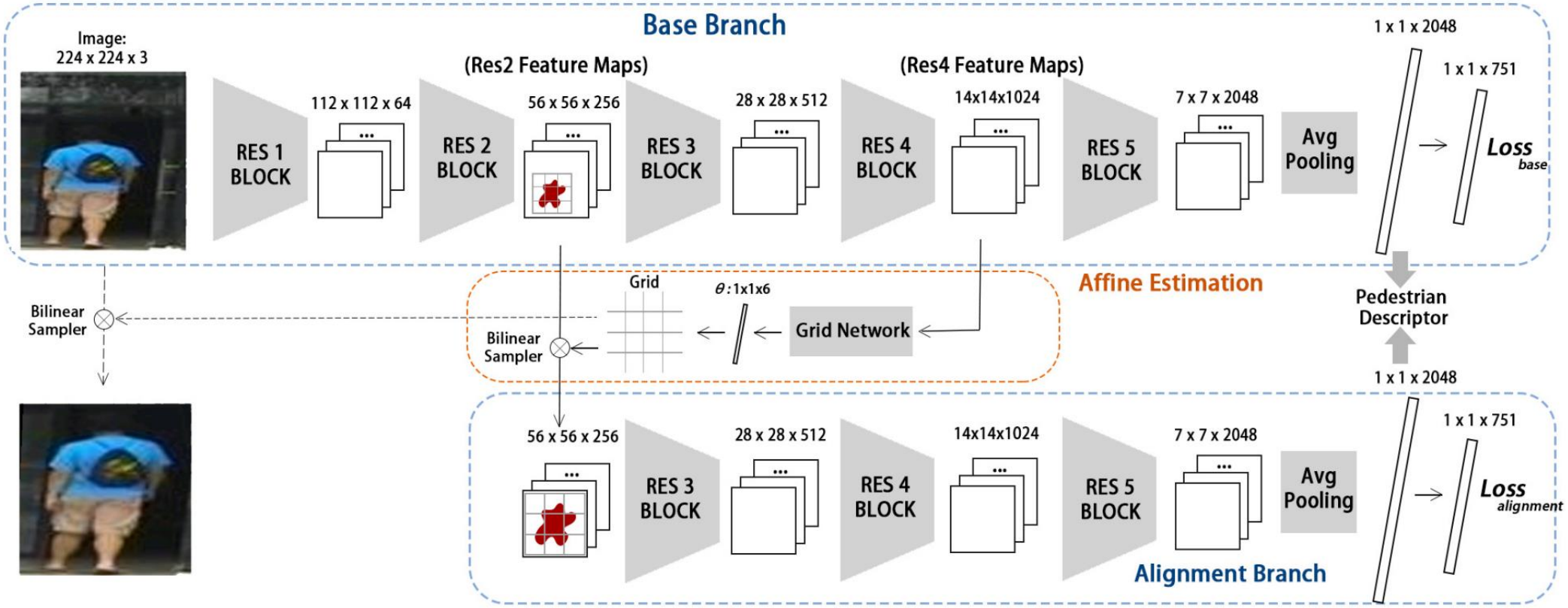
Zheng, Z., Zheng, L., & Yang, Y. (2017). **Pedestrian Alignment Network for Large-scale Person Re-identification**. *arXiv preprint arXiv:1707.00408*.

（Github: https://github.com/layumi/Pedestrian_Alignment ）

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Pedestrian Alignment

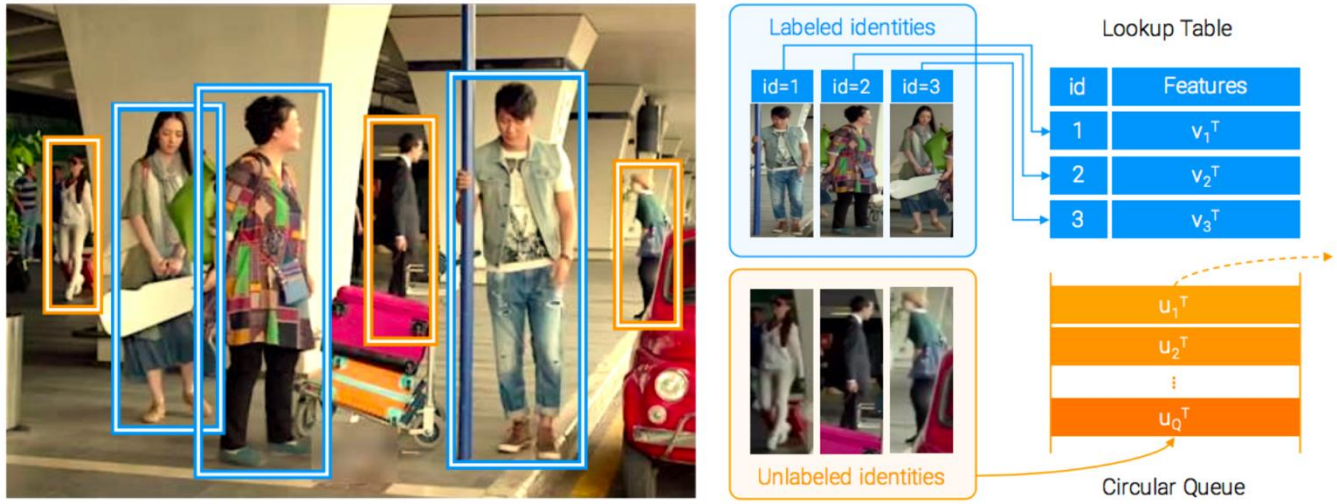


（Github: https://github.com/layumi/Pedestrian_Alignment ）

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Detection + reID



Xiao, T., Li, S., Wang, B., Lin, L., & Wang, X. (2017). **Joint detection and identification feature learning for person search.** In *Proc. CVPR*.

(Github: https://github.com/ShuangLI59/person_search)

Part Matching

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Attention Matching

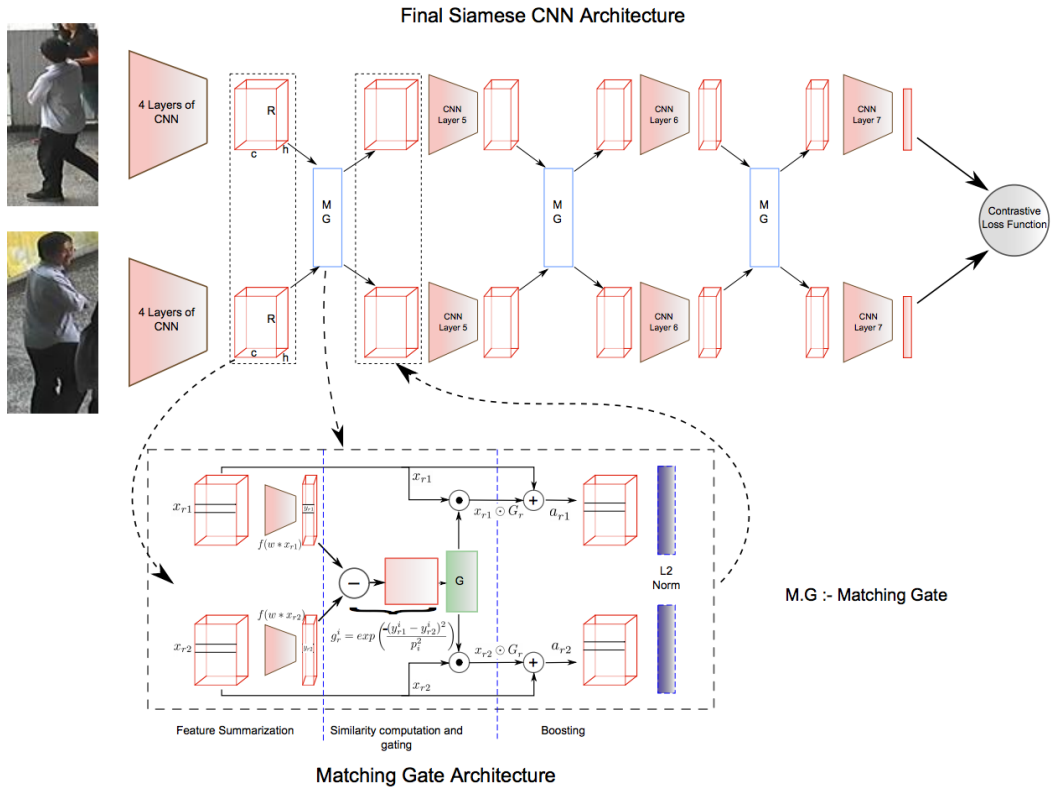


Fig. 2. Proposed architecture: The proposed architecture is a modified version of our baseline S-CNN proposed in Table 1. The matching gate is inserted between layers $4-5$, $5-6$ and $6-7$. The detailed architecture of the gating function is also shown in the figure. See text for details. **Best viewed in color**

Query: n Gallery: m
Original: n+m
Attention: n*m

Varior, R. R., Haloi, M., & Wang, G. (2016, October). **Gated siamese convolutional neural network architecture for human re-identification.** In *European Conference on Computer Vision* (pp. 791-808). Springer International Publishing.

Person Re-identification

Background

Learn pedestrian representations from

- Multi-task learning
- Part matching
- Data augmentation

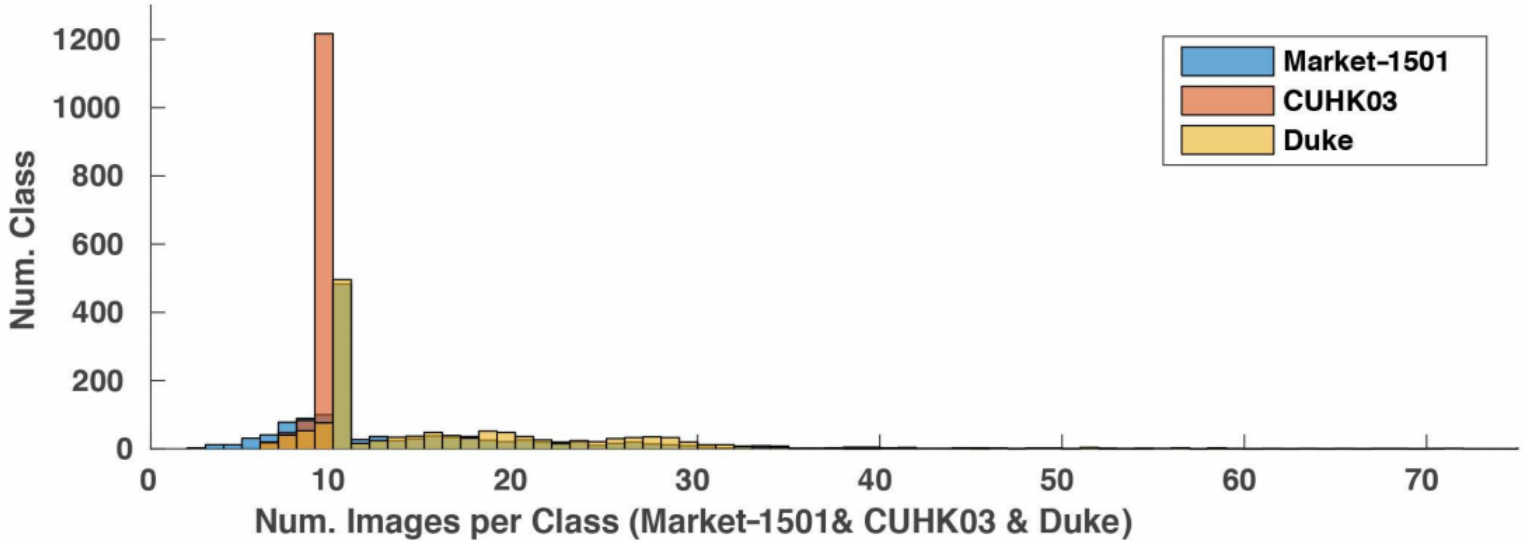
Conclusions and Future Works

- Unsupervised / Semi-supervised Learning
- Natural Language Based Retrieval
- Video Based Person Re-ID

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Data augmentation

1. Multi-dataset Fusion
2. GAN



Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Multi-dataset Fusion

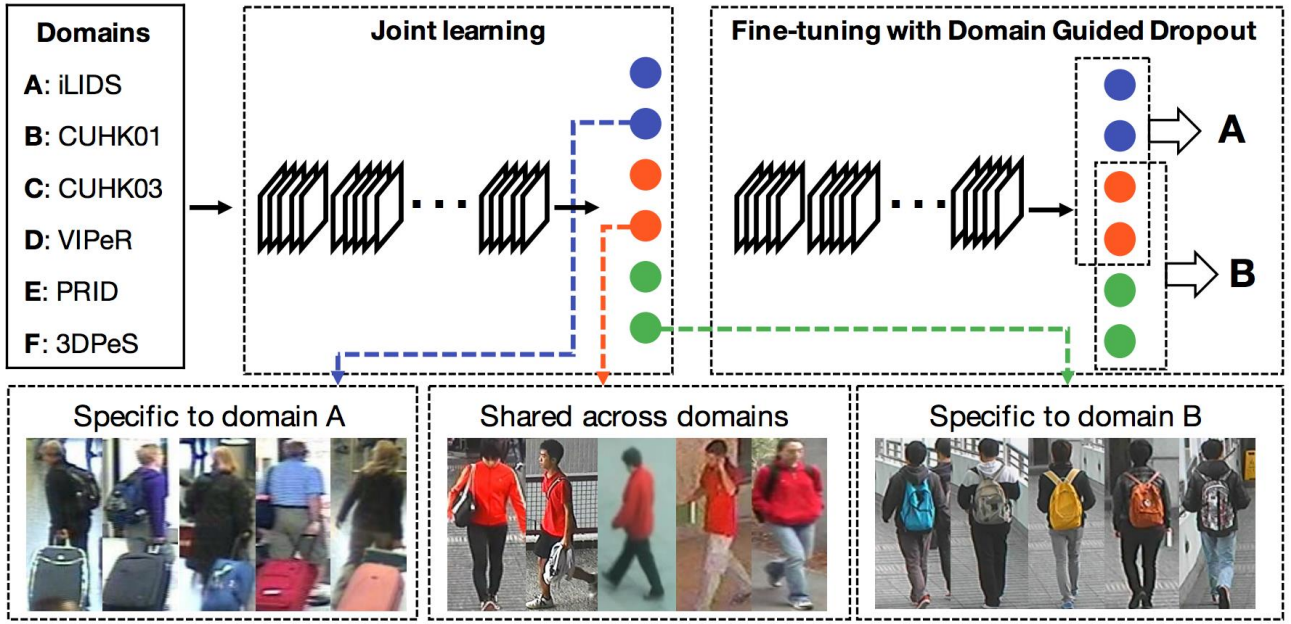


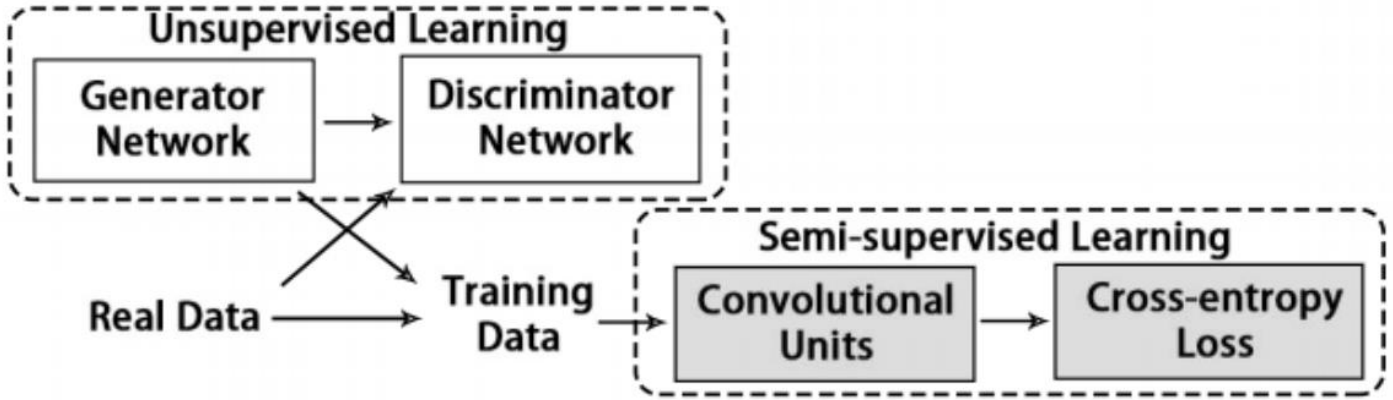
Figure 2. Overview of our pipeline. For the person re-identification problem, we first train a CNN jointly on all six domains. Then we analyze the effectiveness of each neuron on each domain. For example, some may capture the luggages that only appear in domain *A*, while some others may capture the red clothes shared across different domains. We propose a Domain Guided Dropout algorithm to discard useless neurons for each domain during the training process, which drives the CNN to learn better feature representations on all the domains simultaneously.

Xiao, T., Li, H., Ouyang, W., & Wang, X. (2016). **Learning deep feature representations with domain guided dropout for person re-identification.** In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1249-1258).

Github:https://github.com/Cysu/dgd_person_reid

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- GAN



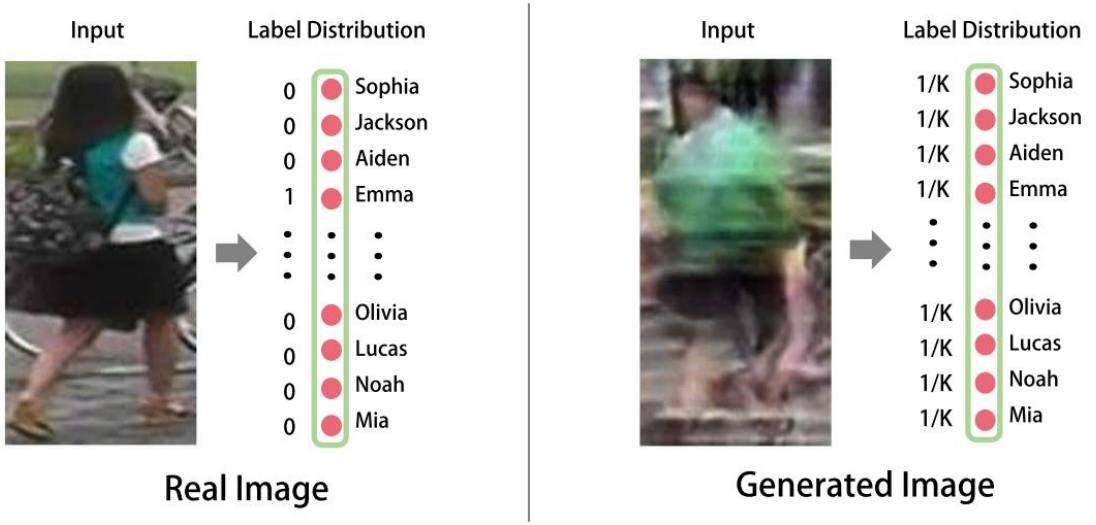
Zheng, Z., Zheng, L., & Yang, Y. (2017). **Unlabeled samples generated by gan improve the person re-identification baseline in vitro.** *arXiv preprint arXiv:1701.07717*.

（Github: https://github.com/layumi/Person-reID_GAN ）

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- GAN



Hinton, G., Vinyals, O., & Dean, J. (2015). **Distilling the knowledge in a neural network**. *arXiv preprint arXiv:1503.02531*.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Result

method	Single Query		Multi. Query	
	rank-1	mAP	rank-1	mAP
BoW+kissme [45]	44.42	20.76	-	-
MR CNN [30]	45.58	26.11	56.59	32.26
FisherNet [38]	48.15	29.94	-	-
SL [4]	51.90	26.35	-	-
S-LSTM [32]	-	-	61.6	35.3
DNS [43]	55.43	29.87	71.56	46.03
Gate Reid [31]	65.88	39.55	76.04	48.45
SOMAnet [3]*	73.87	47.89	81.29	56.98
Verif.-Identif. [48]*	79.51	59.87	85.84	70.33
DeepTransfer [8]*	83.7	65.5	**89.6**	73.8
Basel. [46, 48]*	73.69	51.48	81.47	63.95
Basel. + LSRO	78.06	56.23	85.12	68.52
Verif-Identif. + LSRO	**83.97**	**66.07**	88.42	**76.10**

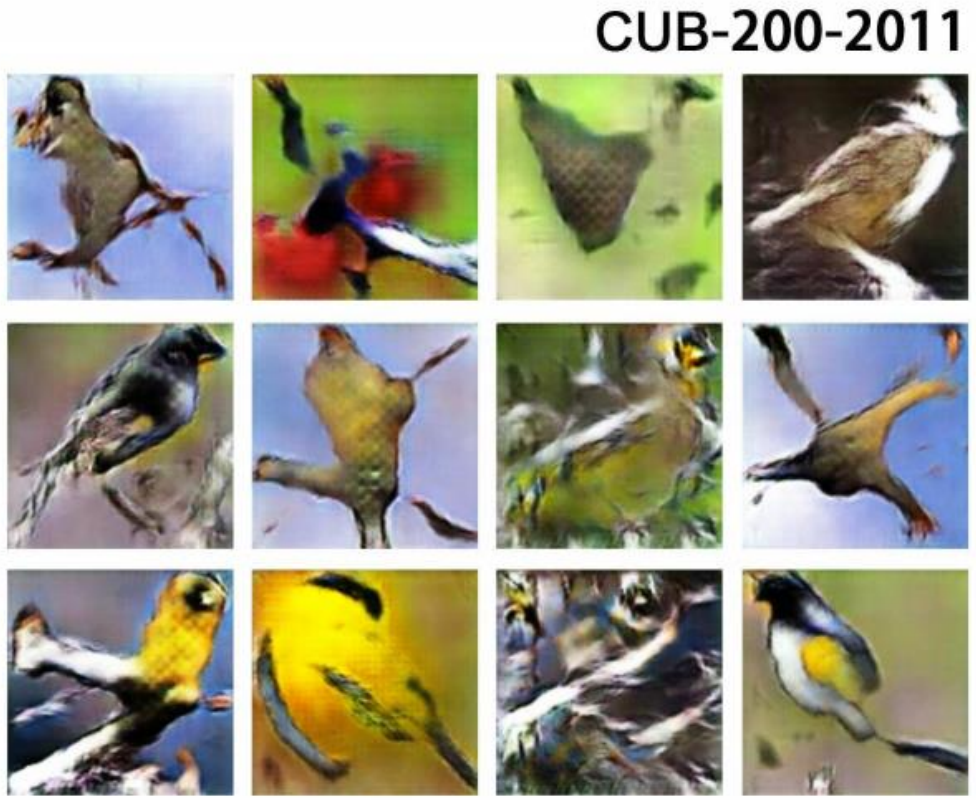
Market-1501 Dataset

method	rank-1	rank-5	rank-10	mAP
KISSME [14]	11.7	33.3	48.0	-
DeepReID [17]	19.9	49.3	64.7	-
BoW+HS [45]	24.3	-	-	-
LOMO+XQDA [18]	46.3	78.9	88.6	-
SI-CI [36]	52.2	84.3	94.8	-
DNS [43]	54.7	80.1	88.3	-
SOMAnet [3]*	72.4	92.1	95.8	-
Verif-Identif. [48]*	83.4	97.1	98.7	86.4
DeepTransfer [8]*	84.1	-	-	-
Basel. [46, 48]*	71.5	91.5	95.9	75.8
Basel.+LSRO	73.1	92.7	96.7	77.4
Verif-Identif. + LSRO	**84.6**	**97.6**	**98.9**	**87.4**

CUHK03 Dataset

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Other Applications



method	model	annotation	top-1
Zhang *et al*. [44]	AlexNet	2×part	76.7
Zhang *et al*. [44]	VGGNet	2×part	81.6
Liu *et al*. [19]	ResNet-50	attribute	82.9
Wang *et al*. [35]	3×VGGNet	×	83.0
Basel. [19]	ResNet-50	×	82.6
Basel.+LSRO	ResNet-50	×	83.2
Basel.+LSRO	2×ResNet-50	×	**84.4**

Table 6. We show the recognition accuracy (%) on CUB-200-2011. The proposed method has a 0.6% improvement over the competitive baseline. The two-model ensemble shows a competitive result.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Other Applications

(a) Script Images

(b) CycleGAN Images

(c) CycleWGAN Images

(d) Real Images

Wang, X., You, M., & Shen, C. (2017). **Adversarial Generation of Training Examples for Vehicle License Plate Recognition**. *arXiv preprint arXiv:1707.03124*.

Person Re-identification

Background

Learn pedestrian representations from

- Multi-task learning
- Part matching
- Data augmentation

Conclusions and Future Works

- Unsupervised / Semi-supervised Learning
- Natural Language Based Retrieval
- Video Based Person Re-ID

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Unsupervised / Semi-supervised Learning

If we do not have target dataset label, what can we do?

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Unsupervised / Semi-supervised Learning

TABLE II
PERSON RE-ID ACCURACY (RANK-1,5,10,20 PRECISION AND MAP) OF TWO METHODS: THE BASELINE (FINE-TUNED RESNET-50 ON ONE OF DATASETS AND TESTED ON ANOTHER ONE) AND PUL. NOTE THAT PUL WORK ON UNSUPERVISED SETTINGS WHICH NEED THE BASELINE MODEL FOR INITIALIZATION.

methods	Duke					Market					CUHK03				
	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP	rank-1	rank-5	rank-10	rank-20	mAP
supervised learning	63.4	78.9	83.3	87.4	42.6	76.2	89.5	93.3	95.8	53.1	24.8	41.1	52.4	64.2	23.2
baseline (Duke)	-	-	-	-	-	36.1	52.8	60.9	69.2	14.2	4.4	9.9	13.7	19.4	4.0
PUL (Duke)	-	-	-	-	-	44.7	59.1	65.6	71.7	20.1	5.6	11.2	15.8	22.7	5.2
baseline (Market)	21.9	38.3	44.4	50.5	10.9	-	-	-	-	-	5.5	10.7	14.4	19.2	4.4
PUL (Market)	30.4	44.5	50.7	56.0	16.4	-	-	-	-	-	7.6	13.8	18.4	25.1	7.3
baseline (CUHK03)	14.9	25.5	31.4	37.5	7.0	30.0	46.4	53.9	61.3	11.5	-	-	-	-	-
PUL (CUHK03)	23.0	34.0	39.5	44.2	12.0	41.9	57.3	64.3	70.5	18.0	-	-	-	-	-

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Unsupervised / Semi-supervised Learning

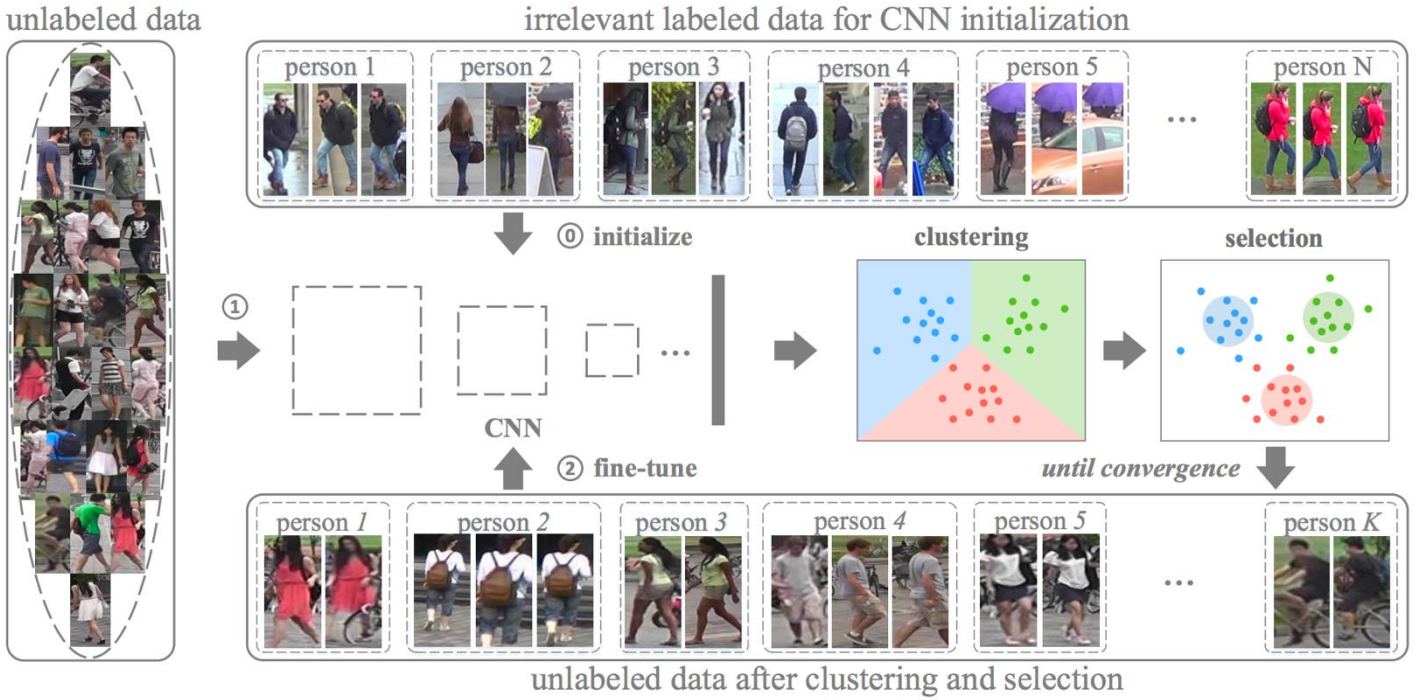


Fig. 1. Illustration of the PUL framework. In step 0, we initialize CNN on an irrelevant labeled dataset. Then we go into the iterations. During each iteration, we 1) extract CNN features for the unlabeled dataset, and perform clustering and sample selection, and 2) fine-tune the CNN model using the selected samples. Each cluster denotes a person.

Fan, H., Zheng, L., & Yang, Y. (2017). **Unsupervised Person Re-identification: Clustering and Fine-tuning**. *arXiv preprint arXiv:1705.10444*.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Unsupervised / Semi-supervised Learning

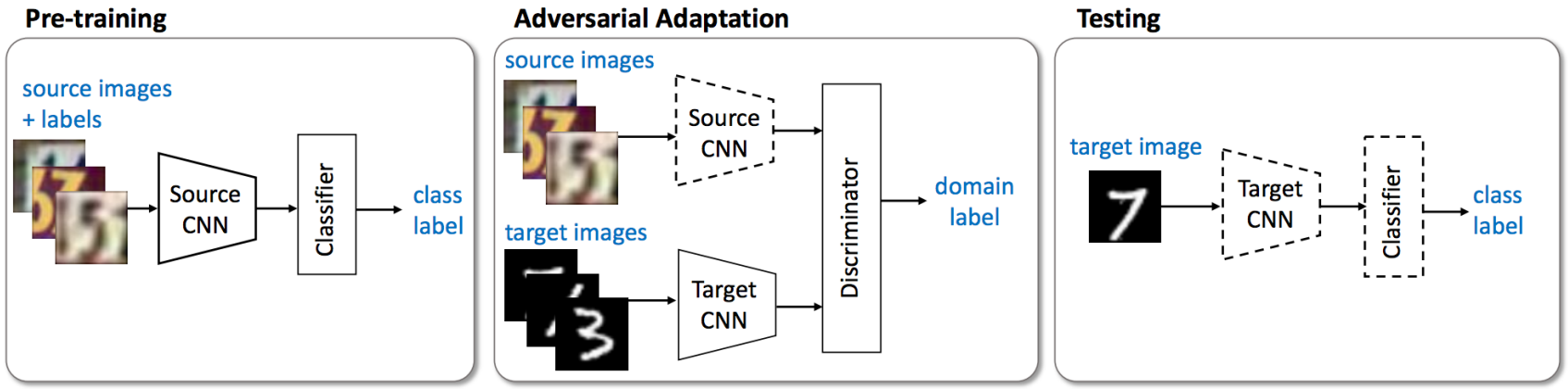


Figure 3: An overview of our proposed Adversarial Discriminative Domain Adaptation (ADDA) approach. We first pre-train a source encoder CNN using labeled source image examples. Next, we perform adversarial adaptation by learning a target encoder CNN such that a discriminator that sees encoded source and target examples cannot reliably predict their domain label. During testing, target images are mapped with the target encoder to the shared feature space and classified by the source classifier. Dashed lines indicate fixed network parameters.

Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). **Adversarial discriminative domain adaptation.** *arXiv preprint arXiv:1702.05464*.

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Natural Language Based Retrieval

The woman is dressed up like Marilyn Monroe, with a white dress that is blowing upward in the wind, short curly blonde hair, and high heels.	The man is wearing yellow sneakers, white socks with blue stripes on the top of them, black athletic shorts and a yellow with blue t-shirt. He has short black hair.	The man has dark hair and is wearing glasses. He has on a pink shirt, blue shorts, and white tennis shoes. He has on a blue backpack and is carrying a re-useable tote.	The girl is wearing a pink shirt with white shorts, she is wearing black converse, with her hair in a pony tail.	The woman has long light brown hair, is wearing a black business suit with white low-cut blouse with large, white cuffs, a gold ring, and is talking on a cellphone.	The man is wearing blue scrubs with a white lab coat on top. He is holding paperwork in his hand and has a name badge on the left side of his coat.

Figure 2. Example sentence descriptions from our dataset that describe persons' appearances in detail.

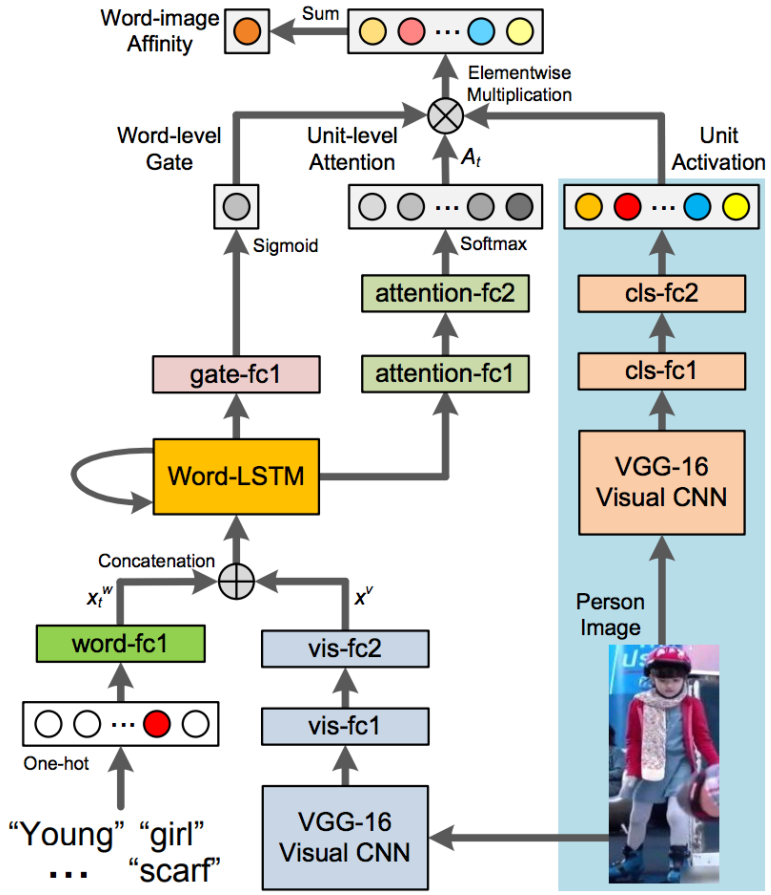
Li, S., Xiao, T., Li, H., Zhou, B., Yue, D., & Wang, X. (2017). **Person search with natural language description.** *arXiv preprint arXiv:1702.05729*.

Project: http://xiaotong.me/static/projects/person-search-language/dataset.html

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Natural Language Based Retrieval



Li, S., Xiao, T., Li, H., Zhou, B., Yue, D., & Wang, X. (2017). **Person search with natural language description.** *arXiv preprint arXiv:1702.05729*.

1. Generative model

Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., & Yuille, A. (2014**). Deep captioning with multimodal recurrent neural networks (m-rnn)**. *arXiv preprint arXiv:1412.6632*.

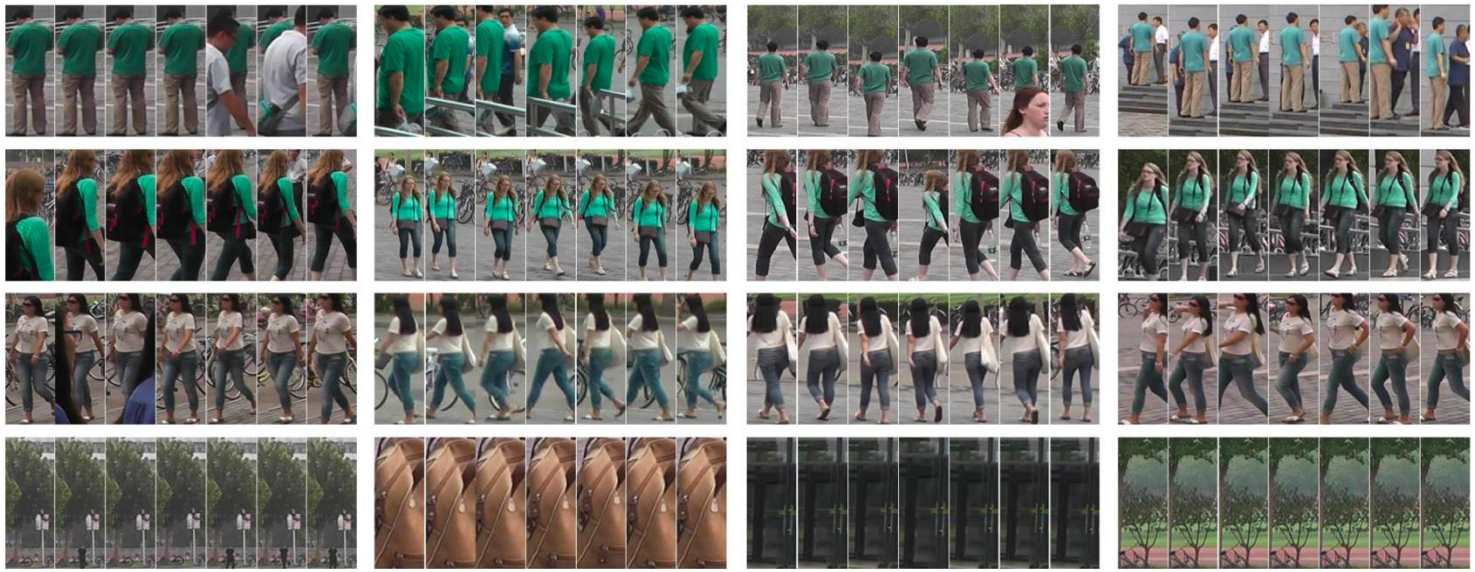
2. Natural Language <---> Attribute

Person Re-identification

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Video Based Person Re-ID

MARS (Motion Analysis and Re-identification Set) Dataset



Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., & Tian, Q. (2016, October). **Mars: A video benchmark for large-scale person re-identification.** In *European Conference on Computer Vision* (pp. 868-884). Springer International Publishing.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

- Video Based Person Re-ID

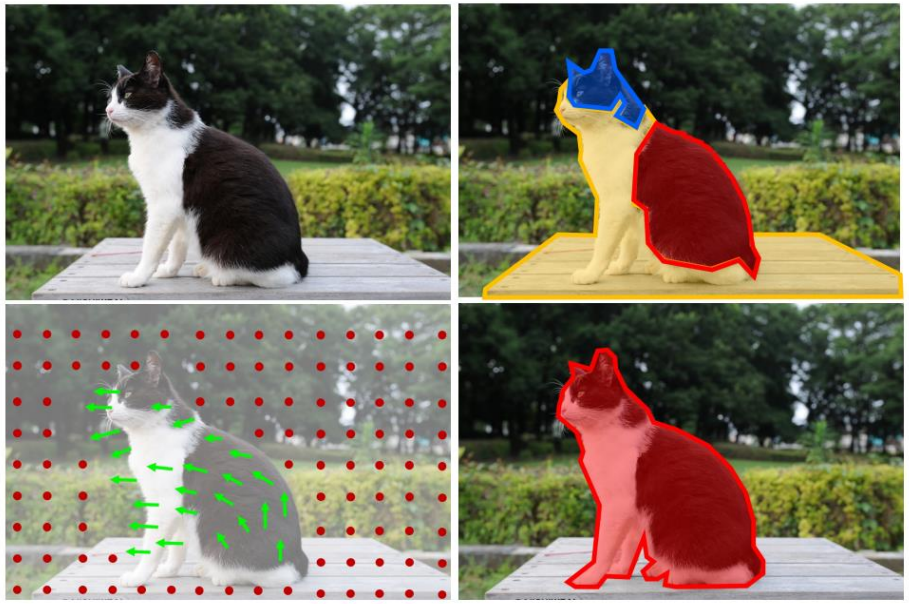


Figure 1. Low-level appearance cues lead to incorrect grouping (top right). Motion helps us to correctly group pixels that move together (bottom left) and identify this group as a single object (bottom right). We use unsupervised motion-based grouping to train a ConvNet to segment objects in *static images* and show that the network learns strong features that transfer well to other tasks.

Unsupervised semantic segmentation <---> Re-ID

Pathak, D., Girshick, R., Dollár, P., Darrell, T., & Hariharan, B. (2016). **Learning features by watching objects move**. *arXiv preprint arXiv:1612.06370*.

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Collaborators

Liang Zheng (ANU)

Yutian Lin (Wuhan University)

Yi Yang (Zhejiang University)

Zhedong Zheng
Centre for Artificial Intelligence (CAI)

Thank you!
Q&A