视觉与学习青年学者研讨会
Vision And Learning SEminar

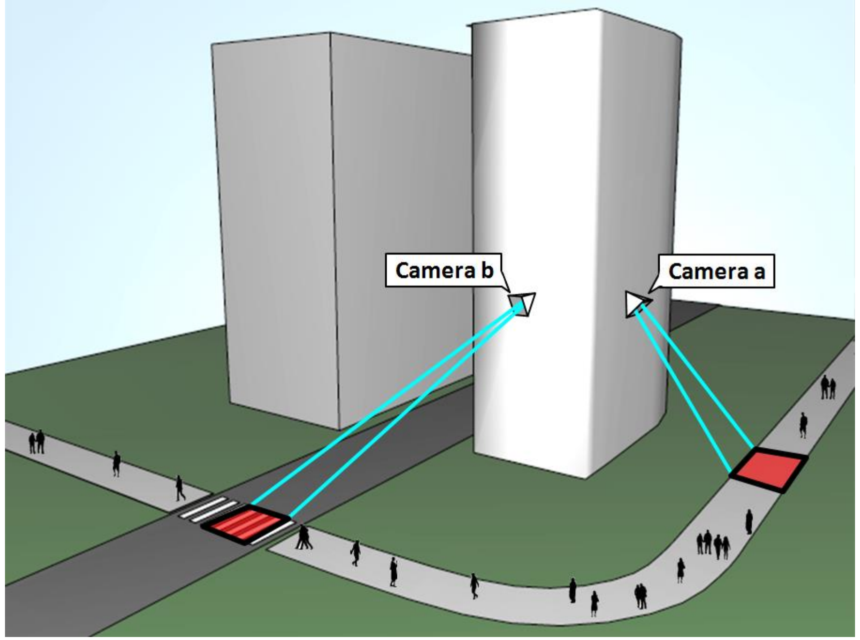
A Review of Person Re-identification with Deep Learning

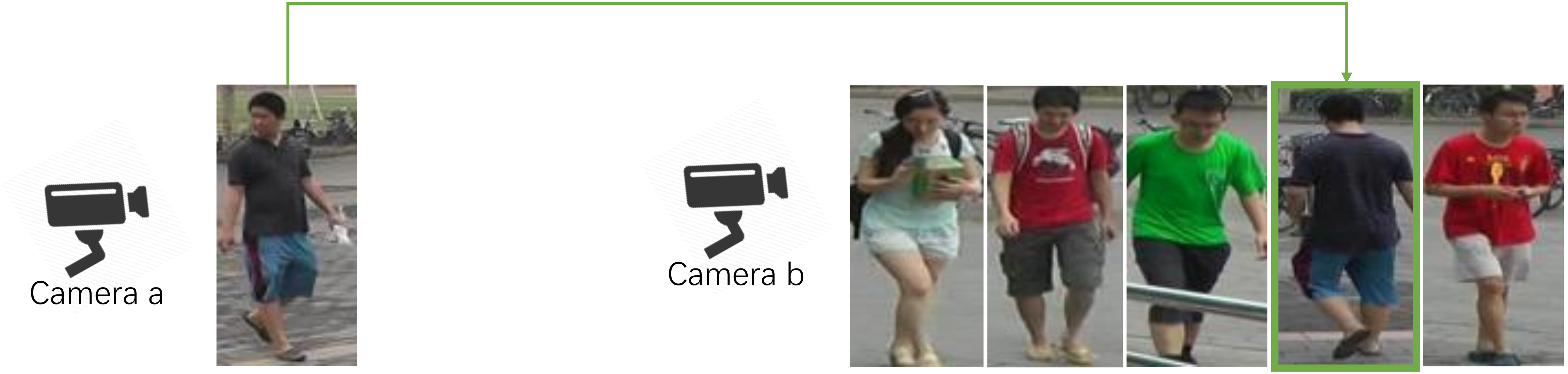
Xi Li
College of Computer Science, Zhejiang University
http://mypage.zju.edu.cn/xilics

Email: xilizju@zju.edu.cn

Person Re-identification

- Associate the person images across different cameras.





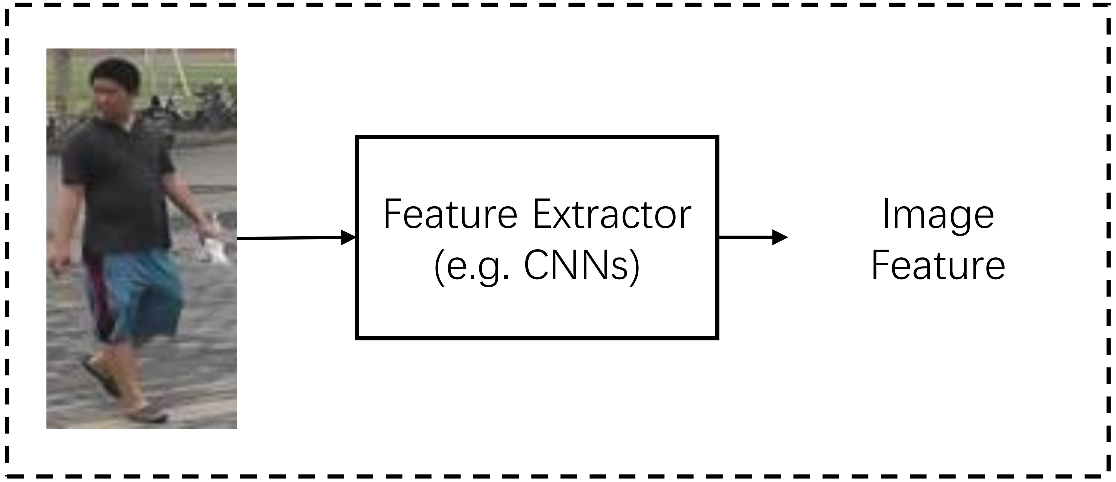
query

gallery

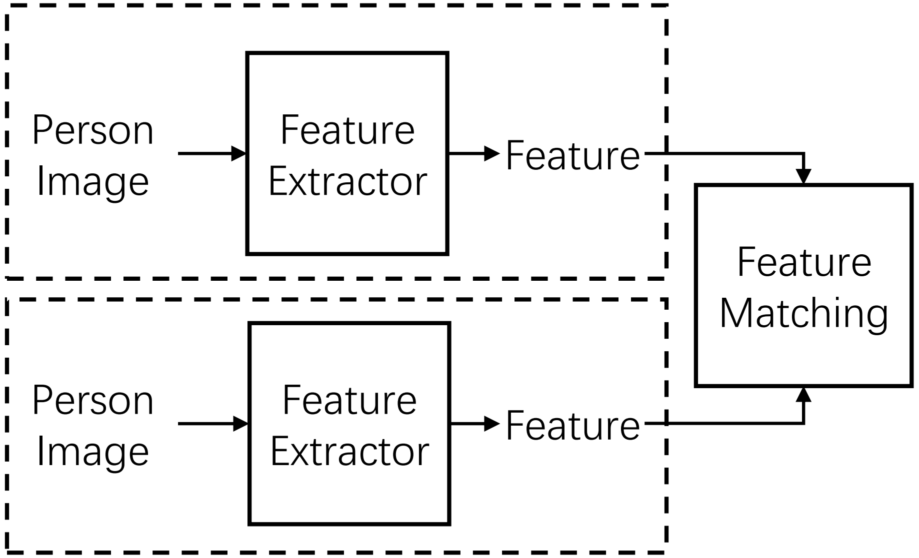
General Solutions

- Step 1: Feature Extraction
  - Extracting features for every person images
- Step 2: Feature Matching
  - Matching features to calculate the similarity score

From 2014, deep models were used to improve these two parts



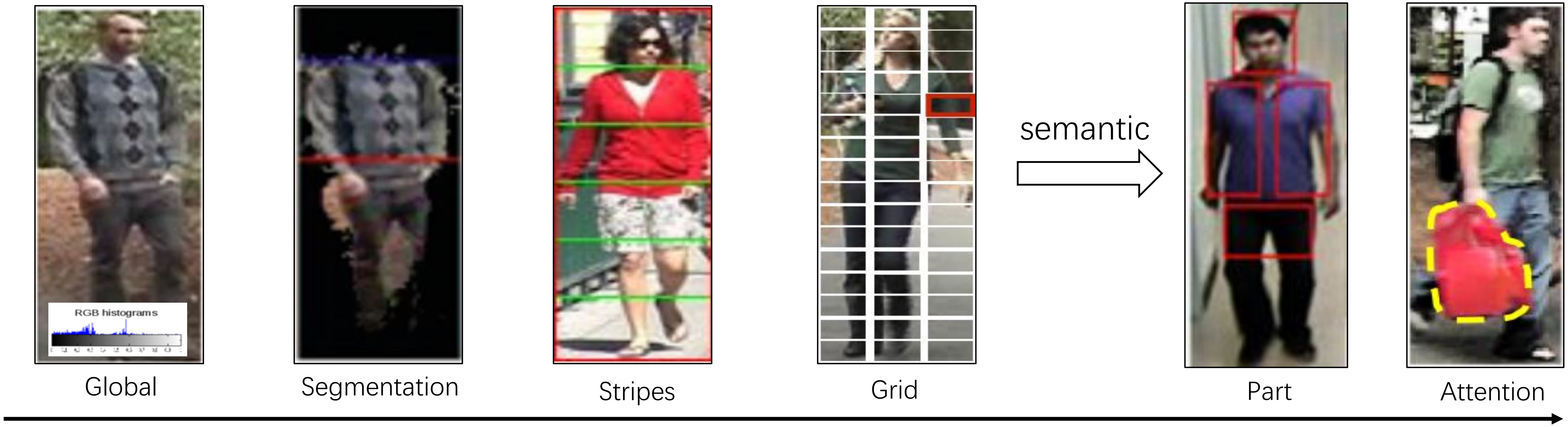
Feature Extraction



Feature Matching

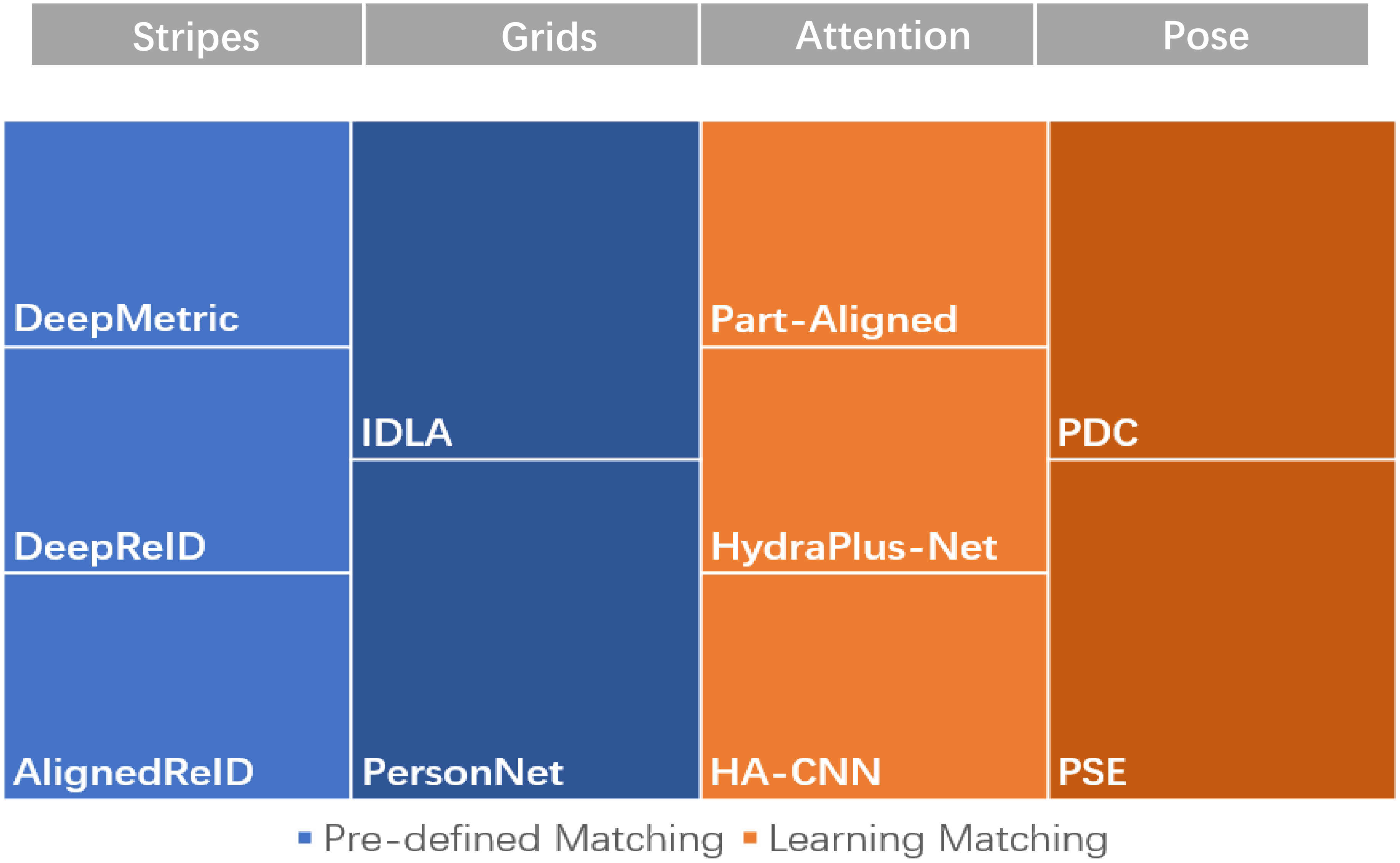
Feature Matching Methods

- Matching based on pre-defined locations
  - Global, local stripes, grid patches
- Matching based on semantic regions
  - Person parts, salient regions, attention regions



Deep Learning Based Methods

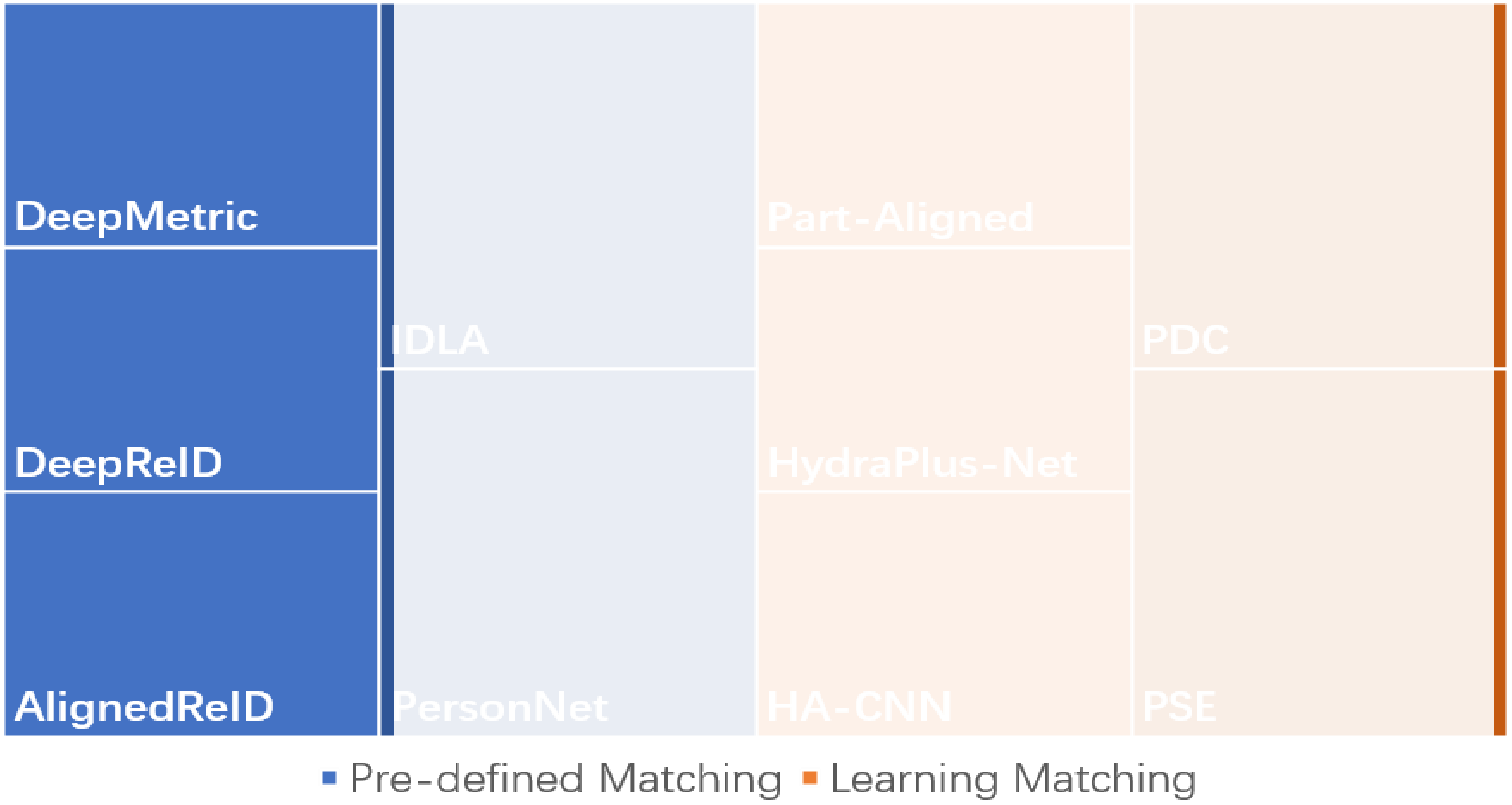
different matching or partitioning strategies



Deep Learning Based Methods

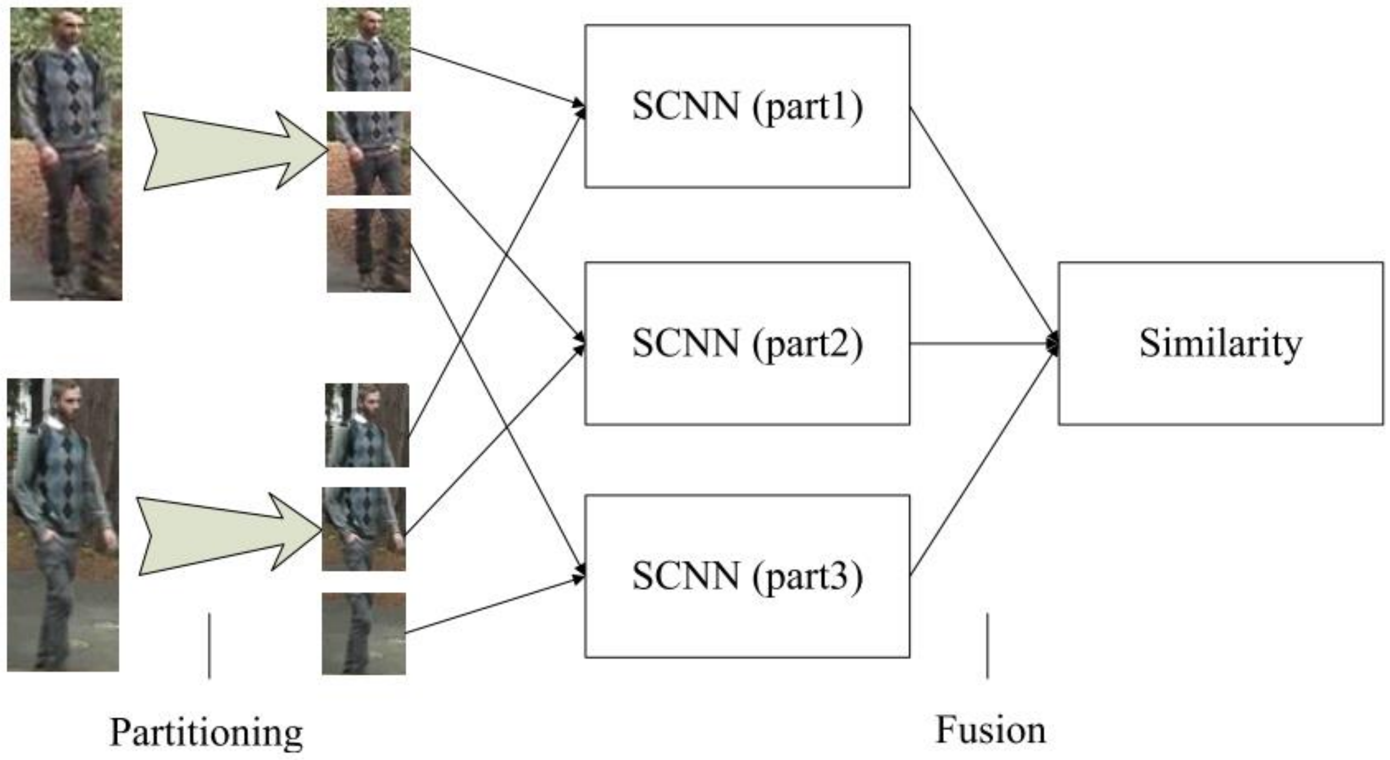
different matching or partitioning strategies

Stripes	Grids	Attention	Pose



Stripes Based Matching: DeepMetric (2014)

- Dividing person image into 3 horizontal stripes
- Extracting CNN features from a pair of images
- Combining features within each stripe
- Computing similarity scores with fused features

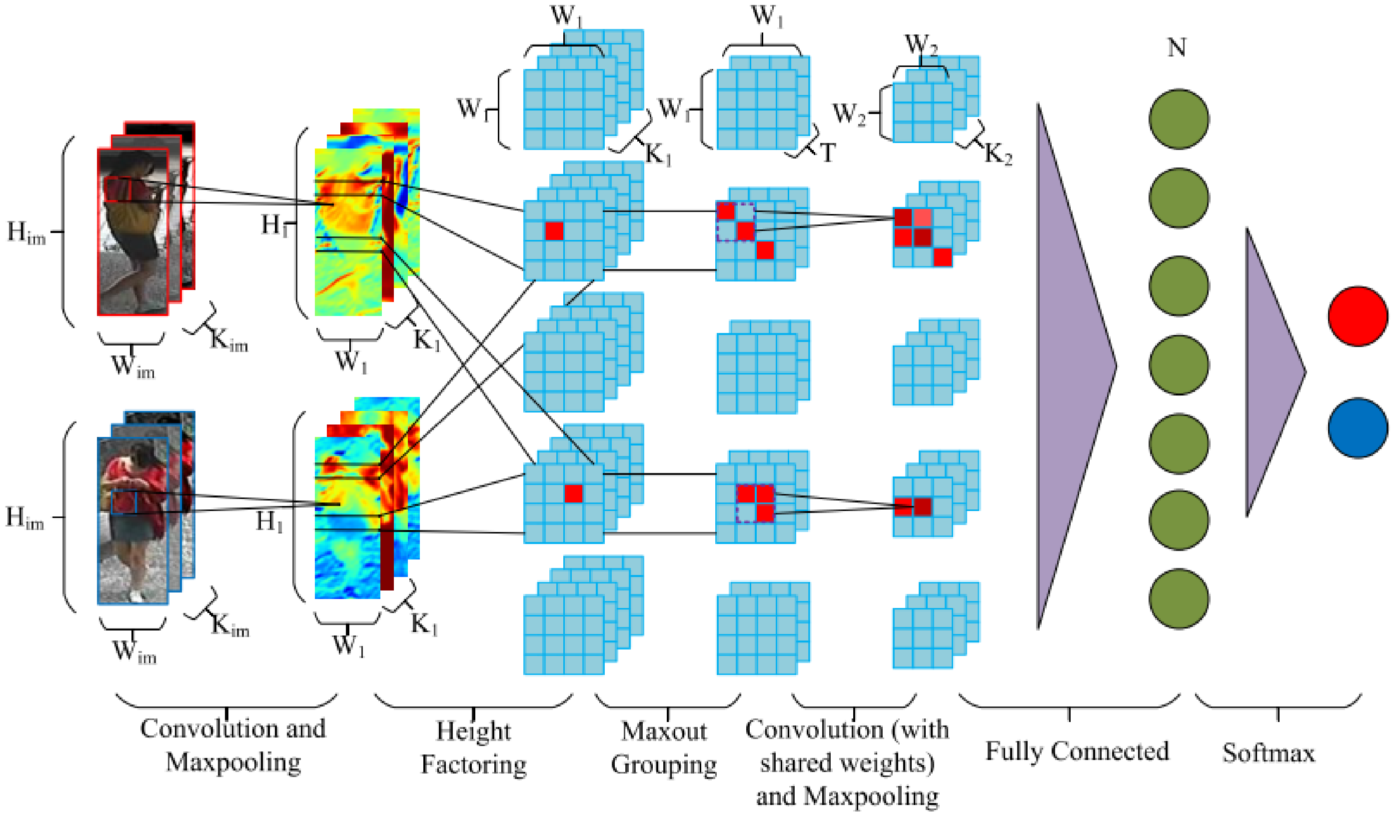


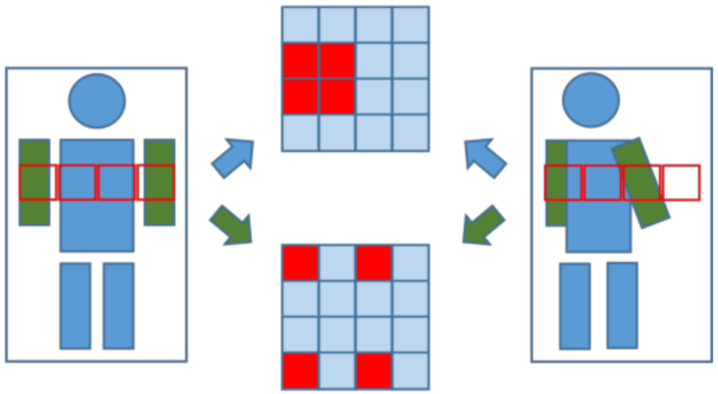
Dong Yi, Zhen Lei, and Stan Z. Li. Deep metric learning for practical person re-identification. ICPR, 2014.

Stripes Based Matching: DeepReID (2014)

- Dividing person image into horizontal stripes
- Extracting CNN features from a pair of images
- Patch matching within each stripe

	Rank1
	DeepReID
CUHK03	20.7%





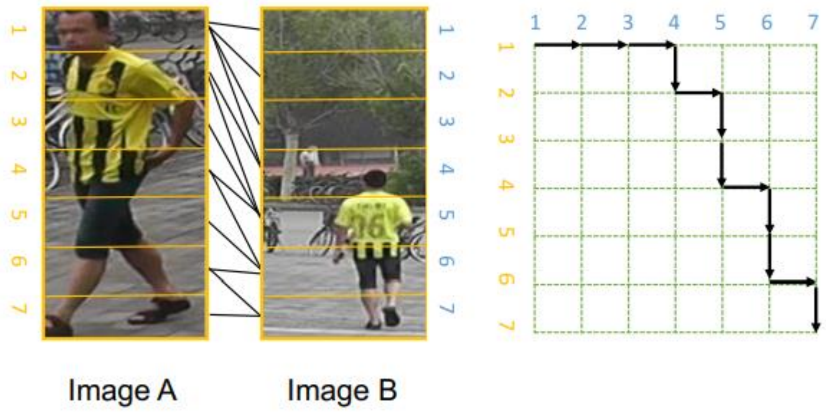
Patch matching:

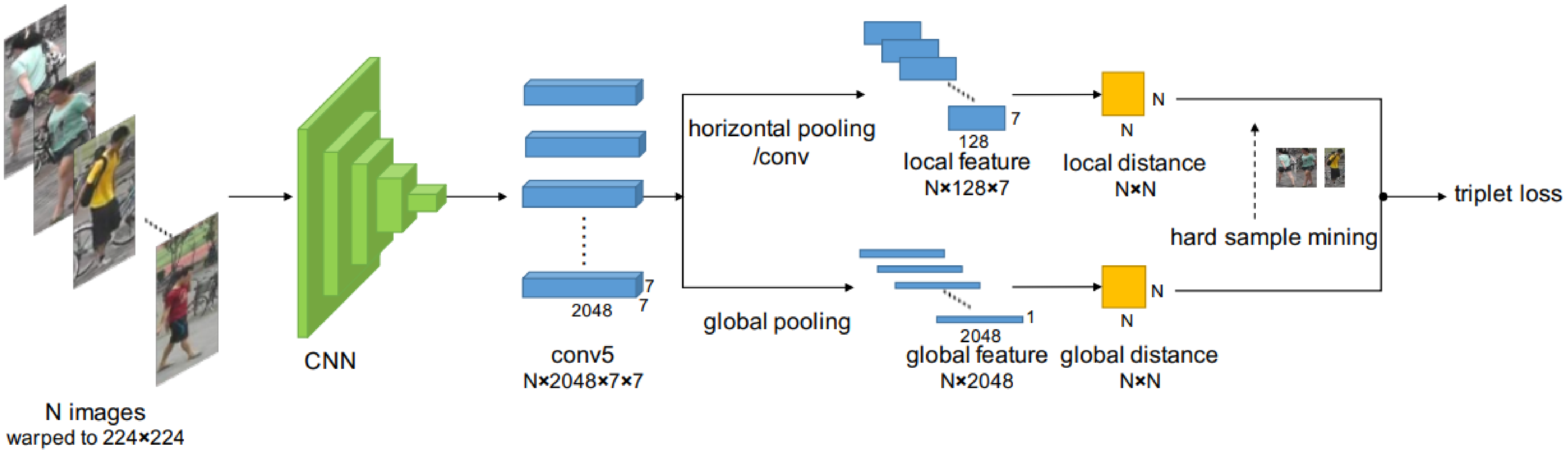
1. Suppose each stripe has 4 patches
2. Match a pair of feature within one stripe
3. Get 4x4 response map for one channel
4. First channel detects blue, another green

Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang.
DeepReID: Deep filter pairing neural network for person re-identification. In CVPR, 2014.

Stripes Based Matching: AlignedReID (2017)

- Combing local feature and global feature
- Local using pre-defined stripes but dynamic matching
- Triplet loss by finding the shortest matching path.

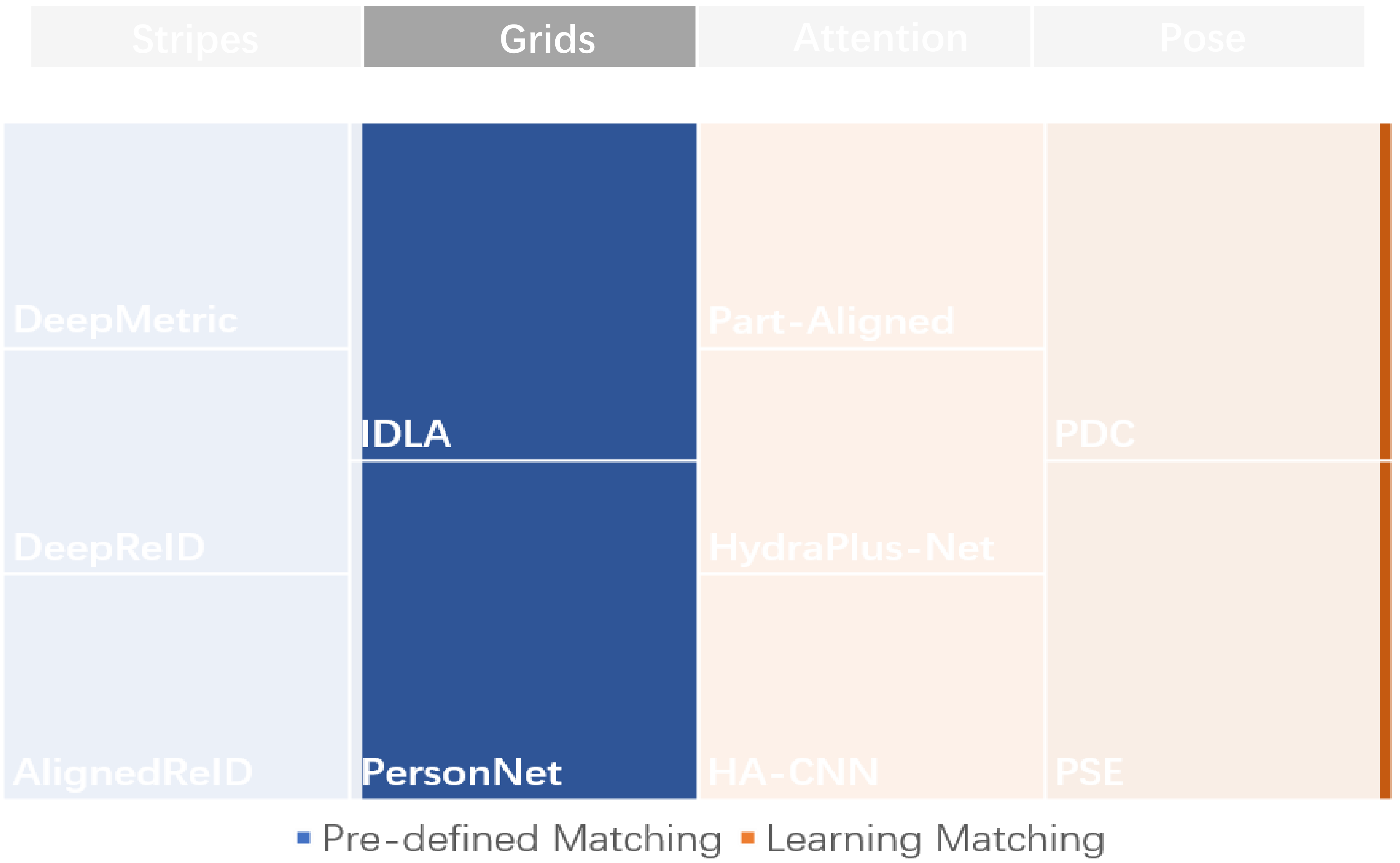




Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang.
AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. ArXiv, 2017.

Deep Learning Based Methods

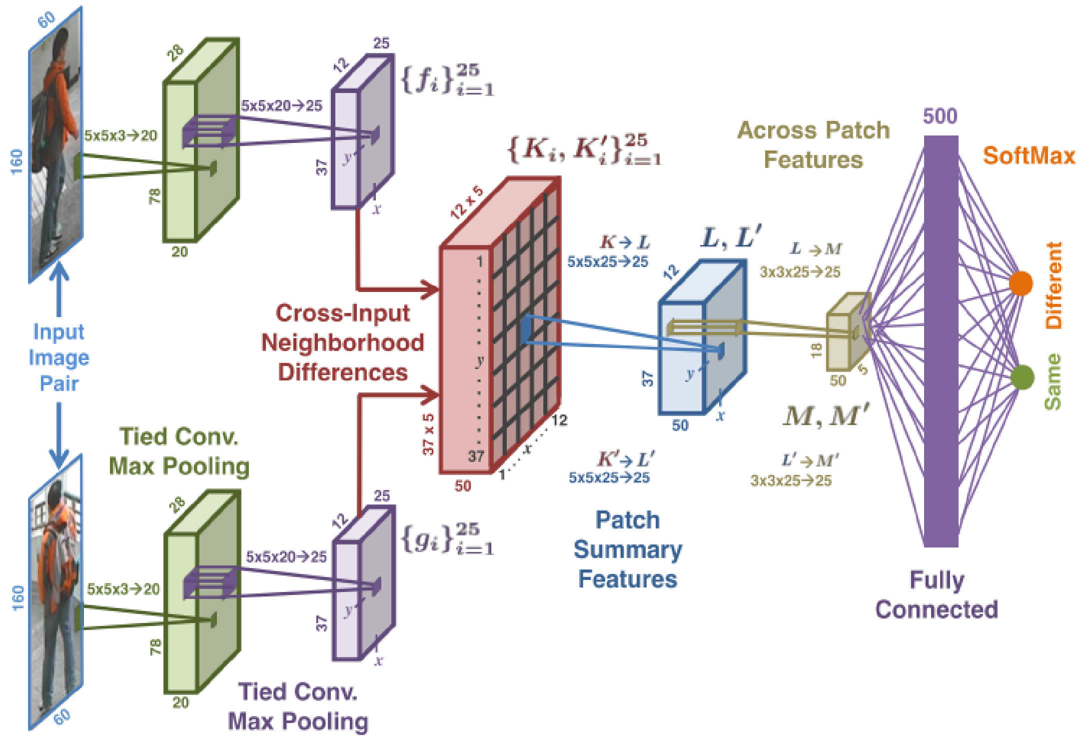
different matching or partitioning strategies



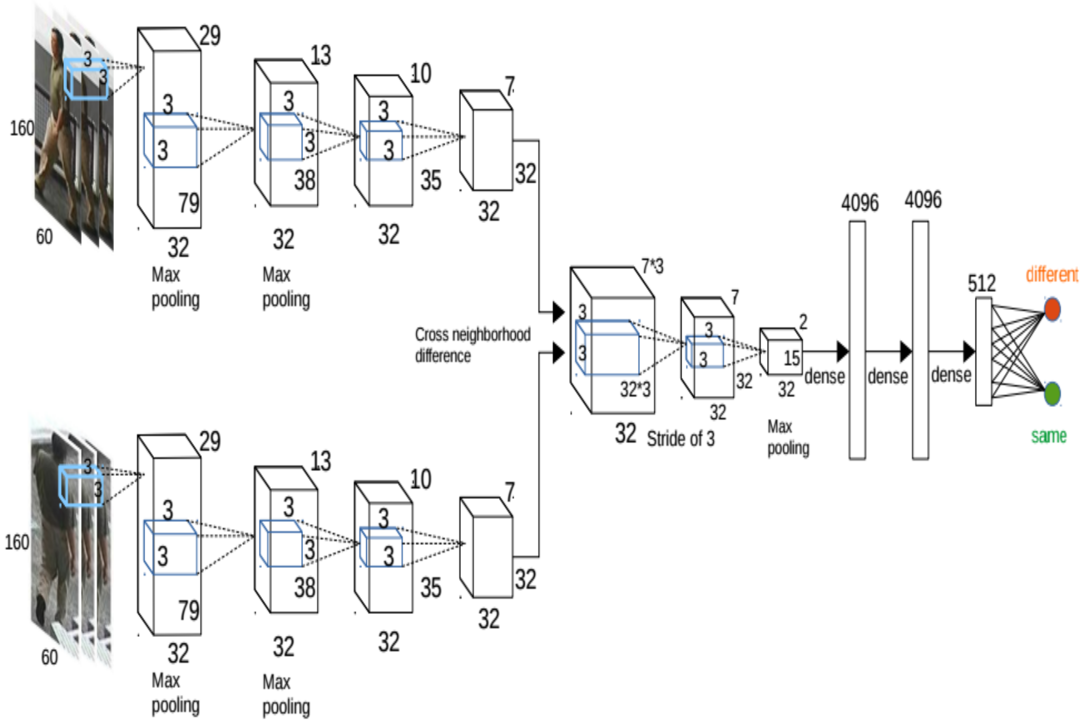
Grid Patches Based: IDLA (2015)

	DeepReID	IDLA
CUHK03	20.7%	54.7%

- Compute differences between each pixel and its 5x5 neighborhood pixels
- Concatenating the differences for similarity learning



[1] IDLA (2015)

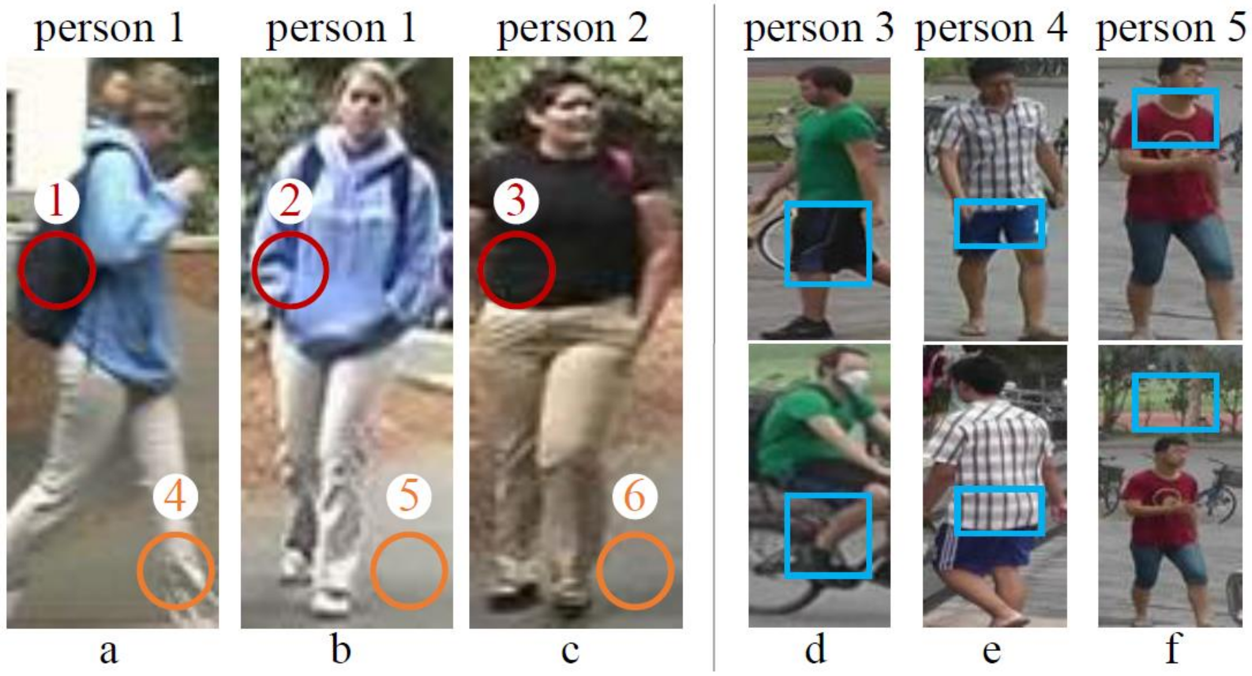


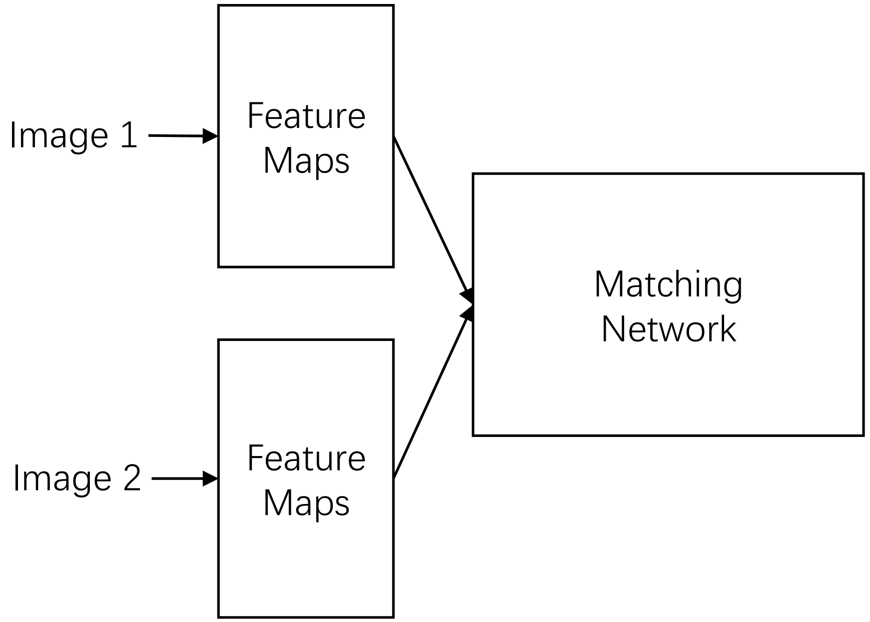
[2] PersonNet (2016) (deeper CNN)

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In CVPR, 2015.
[2] L. Wu, C. Shen, and A. van den Hengel. Personnet: Person re-identification with deep convolutional neural networks. CoRR, abs/1601.07255, 2016.

Challenge of Pre-defined Matching

- Spatial location misalignment due to detection or pose changes
- CNN based online feature matching is expensive for searching

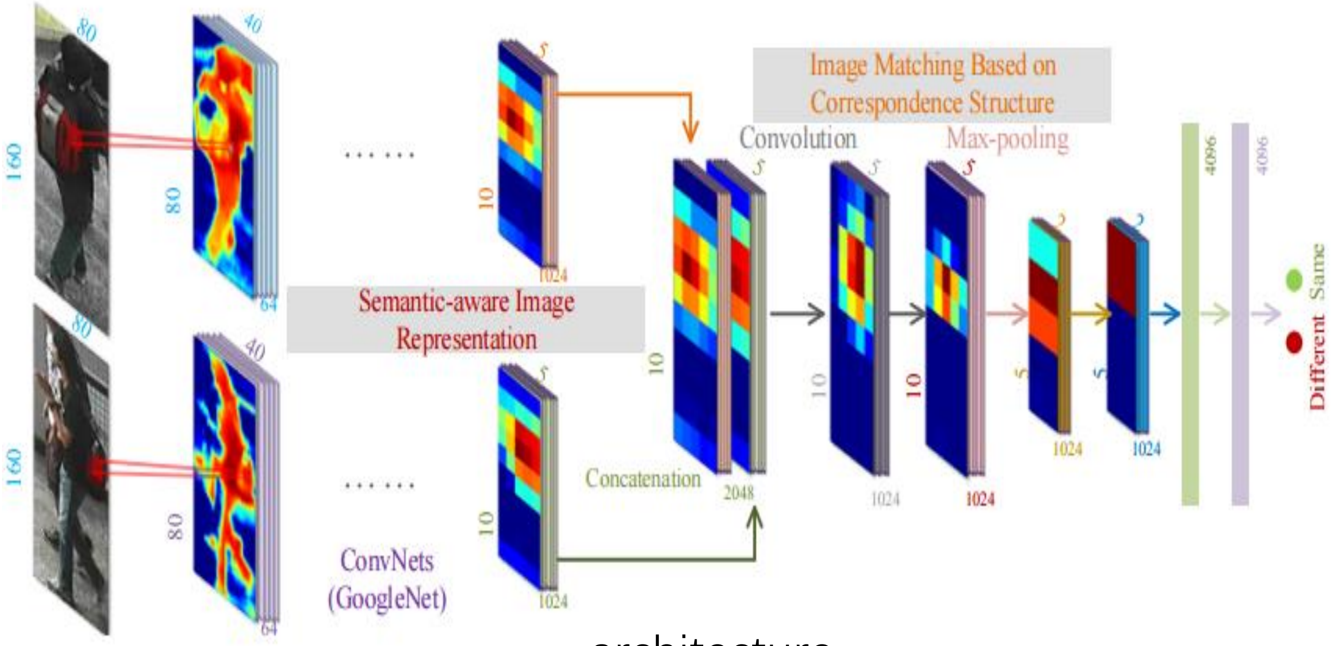




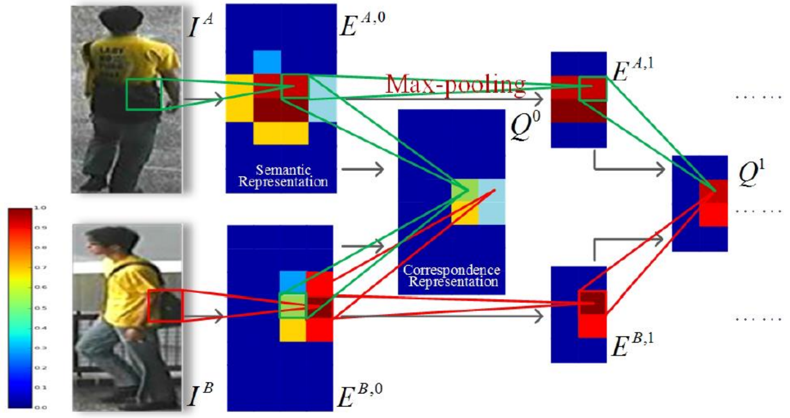
Online Feature Matching: DCSL (2016)

- Deep Correspondence Learning instead of manually defining the matching grid patches
- Adaptively learn a hierarchical data-driven feature matching function

	IDLA	DCSL
CUHK03	54.7%	80.2%



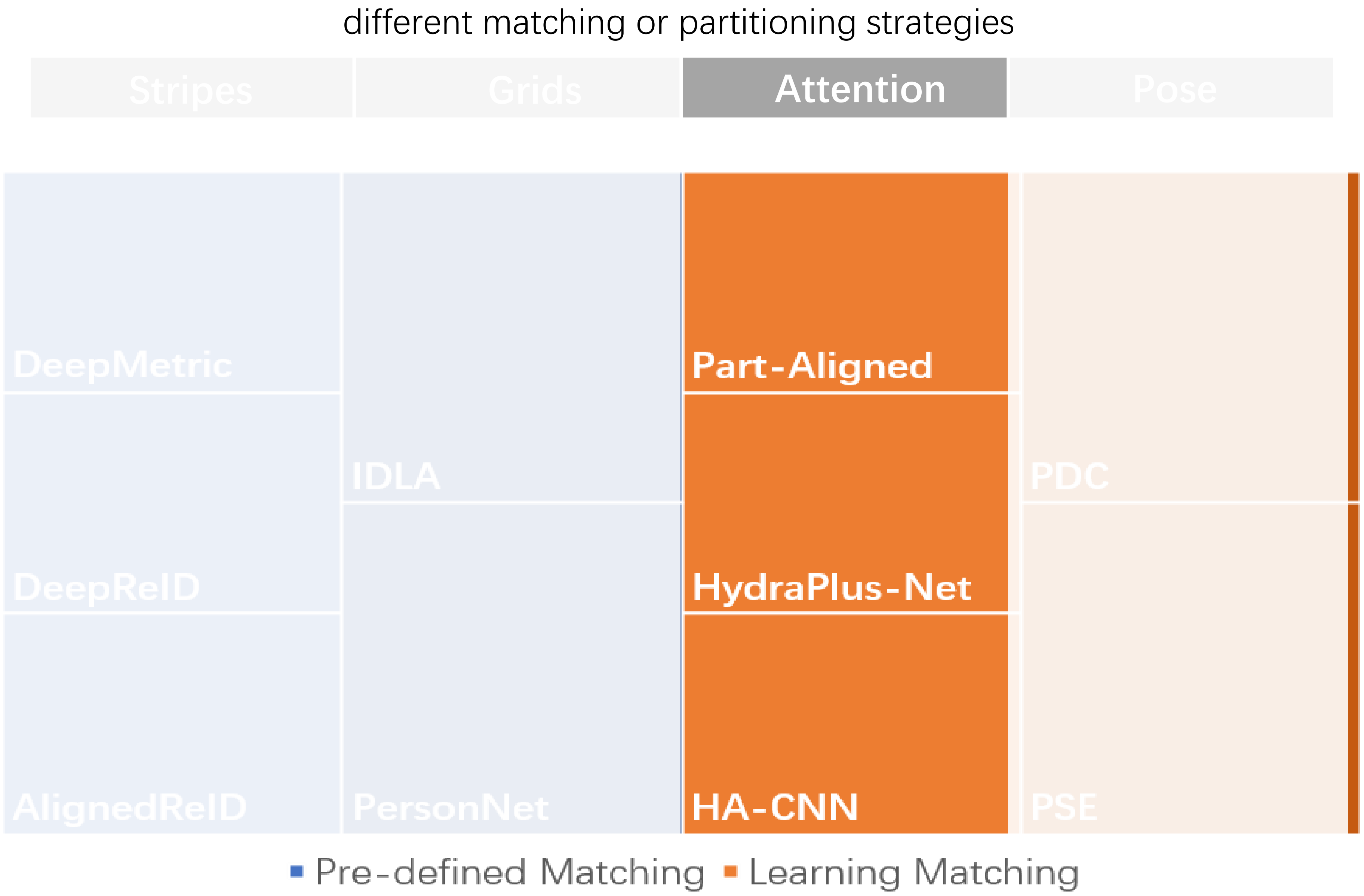
architecture



pyramid matching

Yaqing Zhang, Xi Li, Liming Zhao, Zhongfei Zhang.
Semantics-Aware **D**eep **C**orrespondence **S**tructure **L**earning for Robust Person Re-identification. IJCAI, 2016.

Deep Learning Based Methods

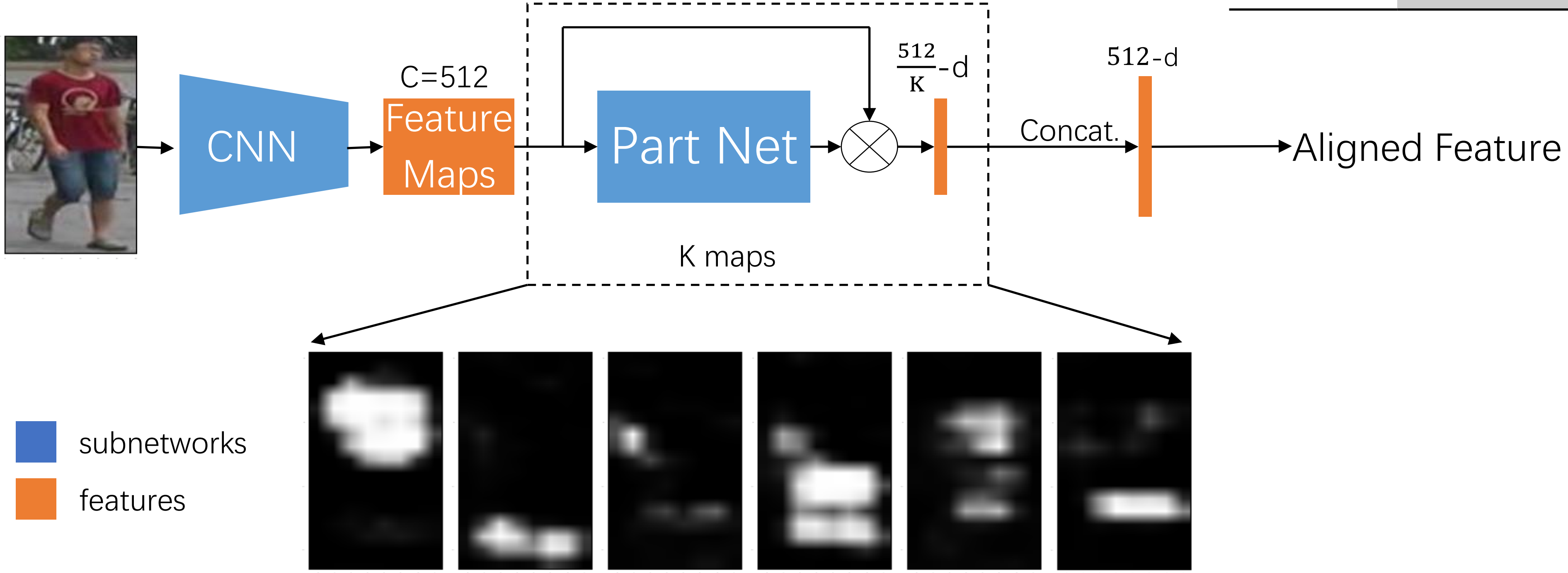


Part Regions Learning: Part-Aligned (2017)

- Learn the key regions for **Embedding** instead of **Matching**
- Align the feature with the learnt region maps
- Extracting features and then calculating Euclidean distances

Rank1 on CUHK03: **85.4%**
Rank1 on Market: **81.0%**

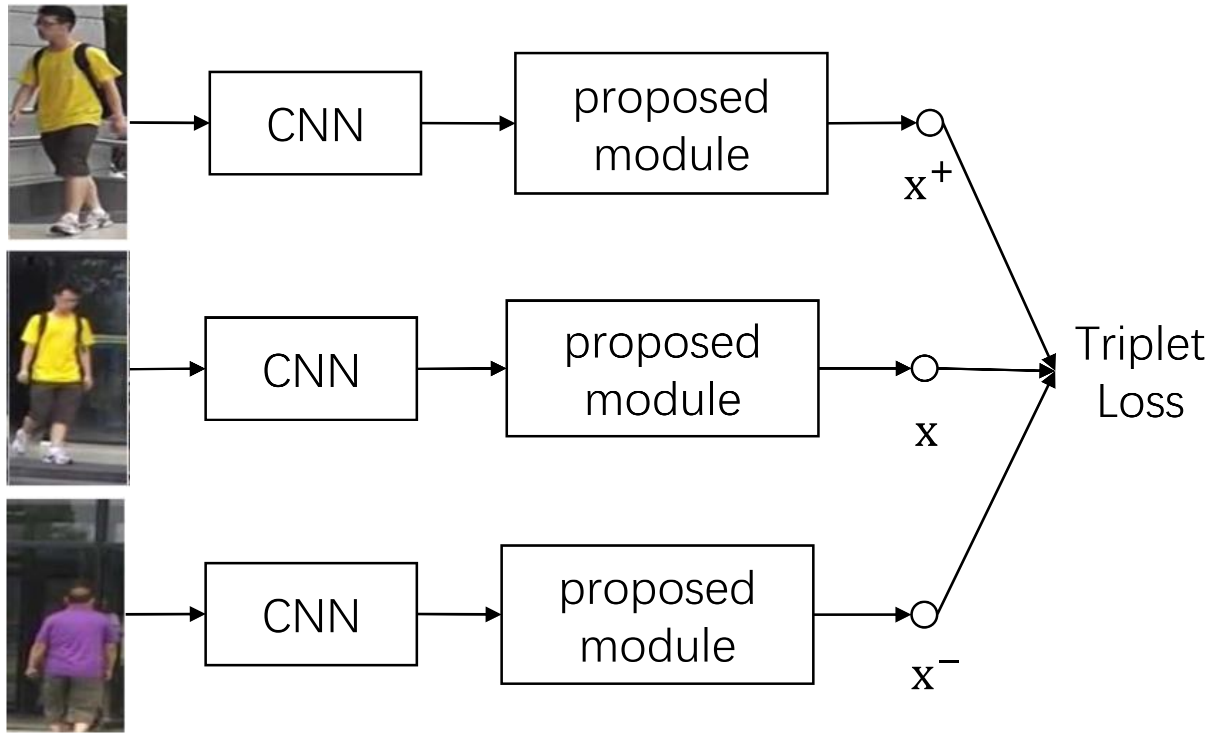
	DCSL	**DLPAR**
CUHK03	80.2%	85.4%



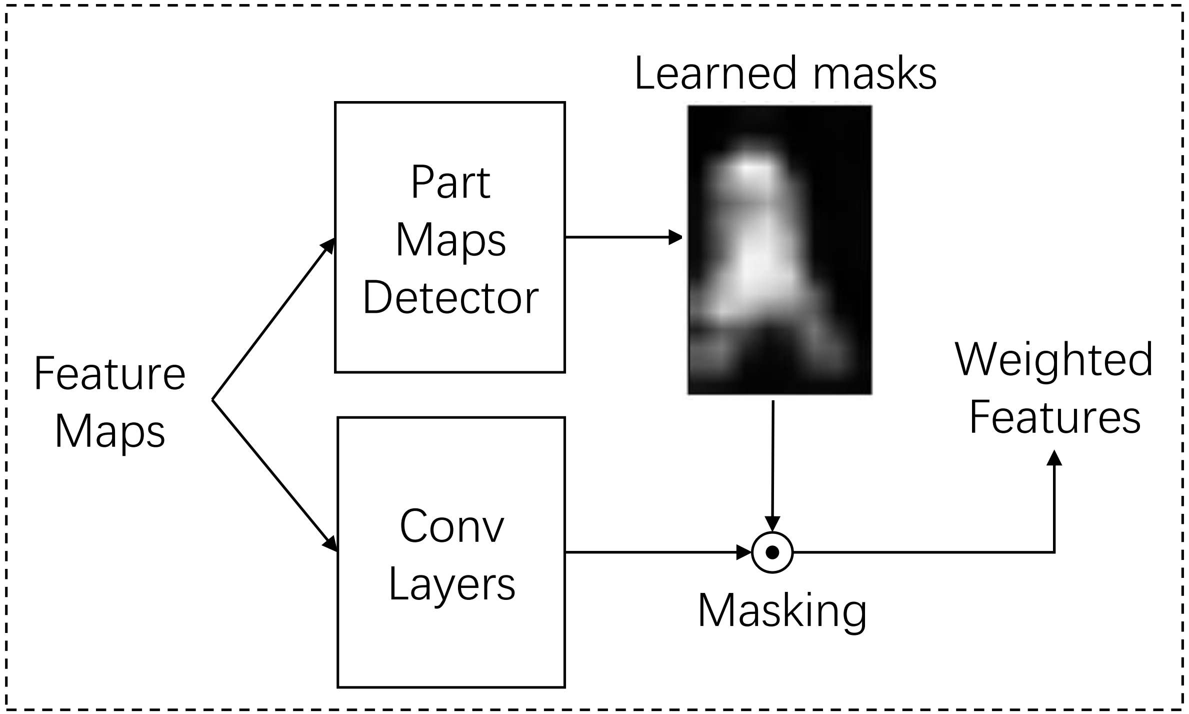
Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang.
Deeply-**L**earned **P**art-**A**ligned **R**epresentations for Person Re-Identification. ICCV, 2017.

Part Regions Learning: Part-Aligned (2017)

- An end-to-end solution to jointly
  - Learn the reid-sensitive maps for person matching
  - Learn the part-aligned deep representation



Framework



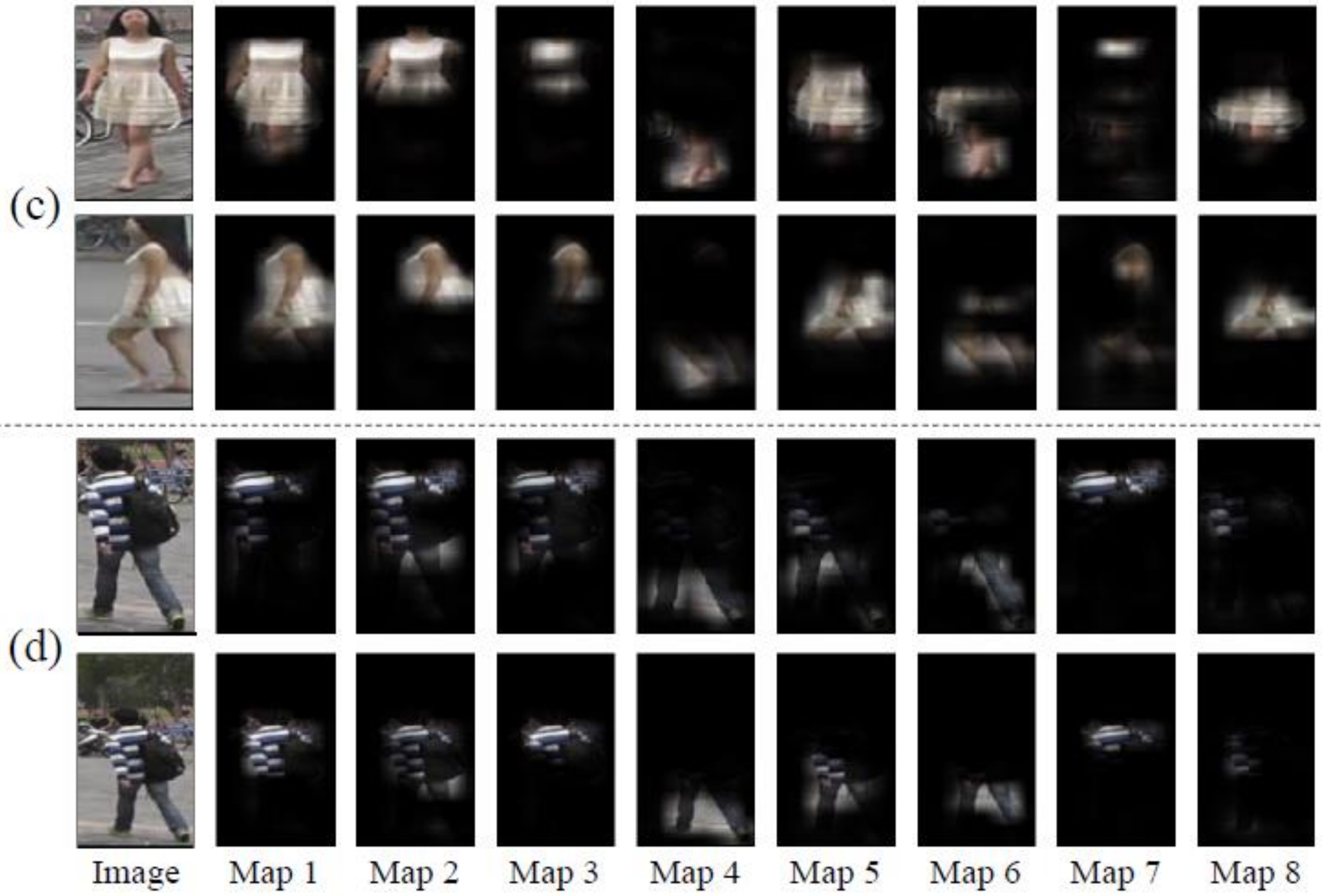
Proposed module

Part Regions Learning: Part-Aligned (2017)

Learn the maps without extra annotations.

ReID-sensitive regions

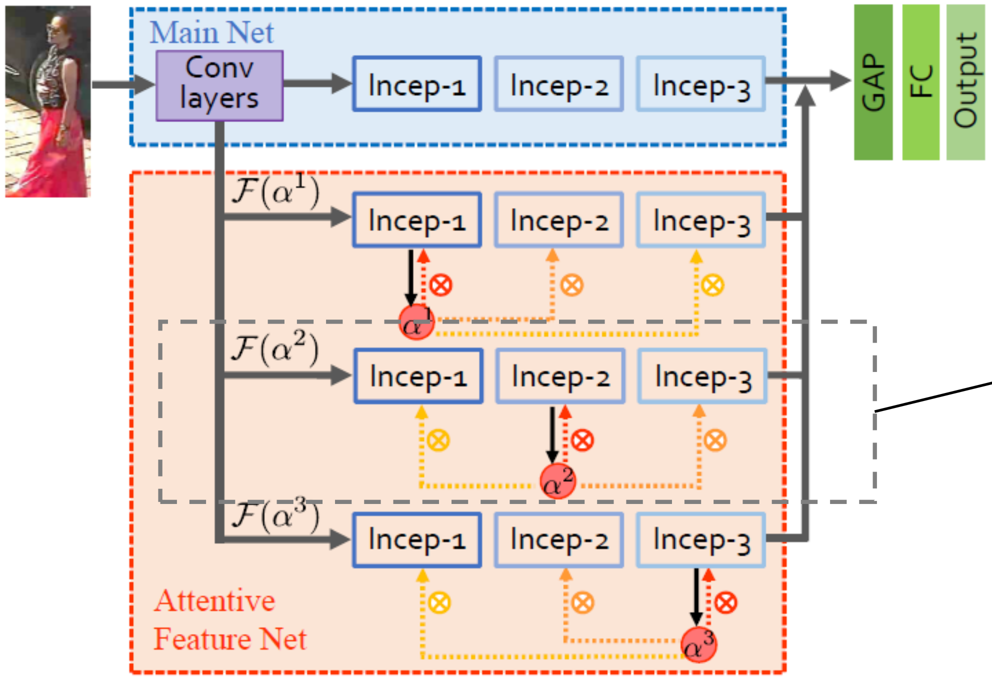
Different with traditional part segmentation



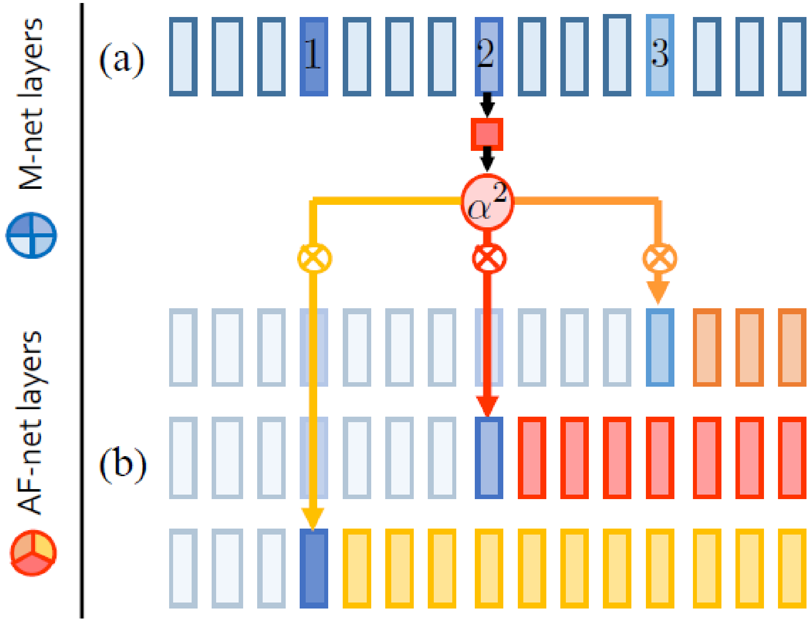
Attention Regions Learning: HydraPlus-Net (2017)

- Learn features by fusing three attention module with Softmax loss
- Learn attention map from different scale for each module
- Apply the attention map on different layers of the network

	DLPAR	HPNet
CUHK03	85.4%	91.8%
Market	81.0%	76.9%



Framework



Attention module F(α^2)

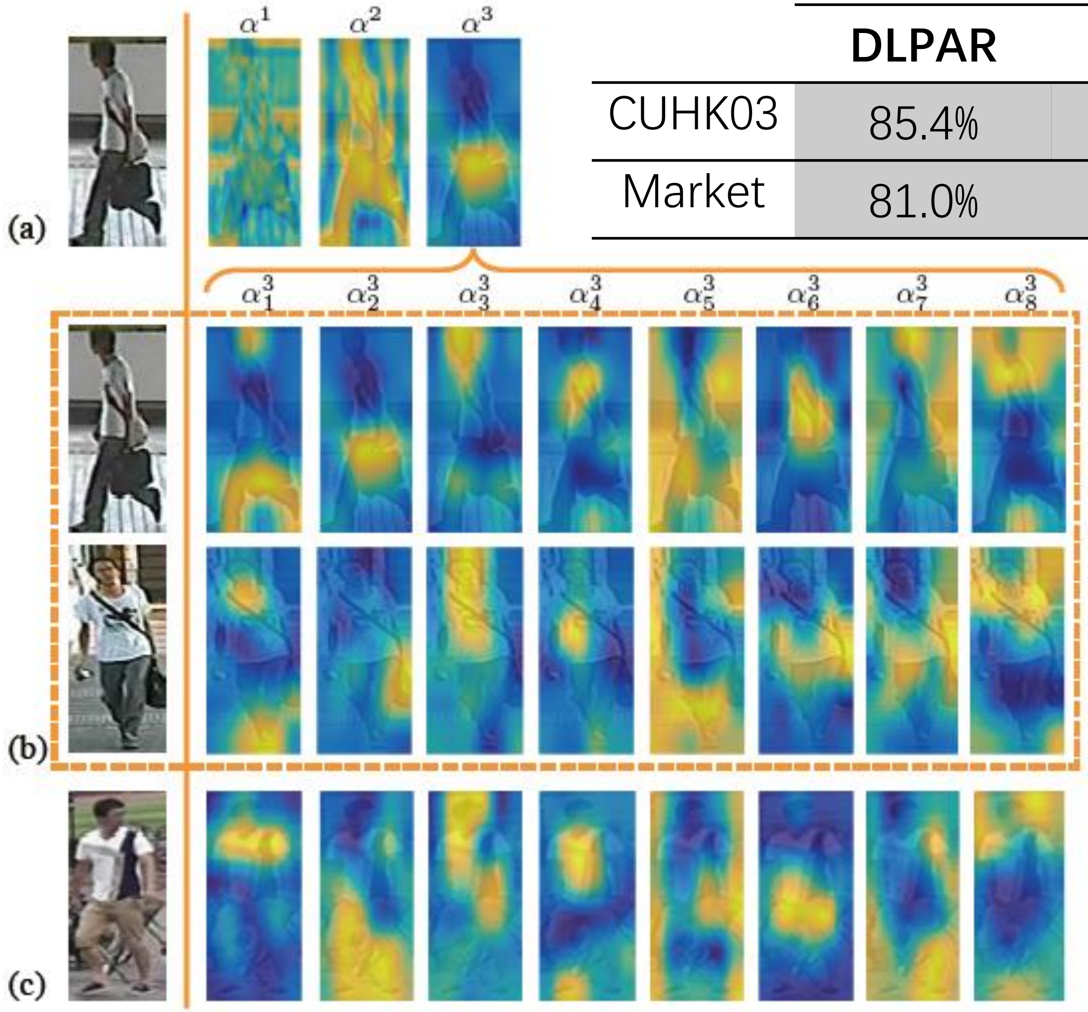
Xihui Liu, Haiyu Zhao, Maoqing Tian, Lu Sheng, Jing Shao, Shuai Yi, Junjie Yan, Xiaogang Wang: HydraPlus-Net: Attentive Deep Features for Pedestrian Analysis. ICCV 2017

Attention Regions Learning: HydraPlus-Net (2017)

(a) Attention maps in three different levels or scales

(b,c) Each level maps contain 8 channels

(b,c) One channel learn one part or things (bags)



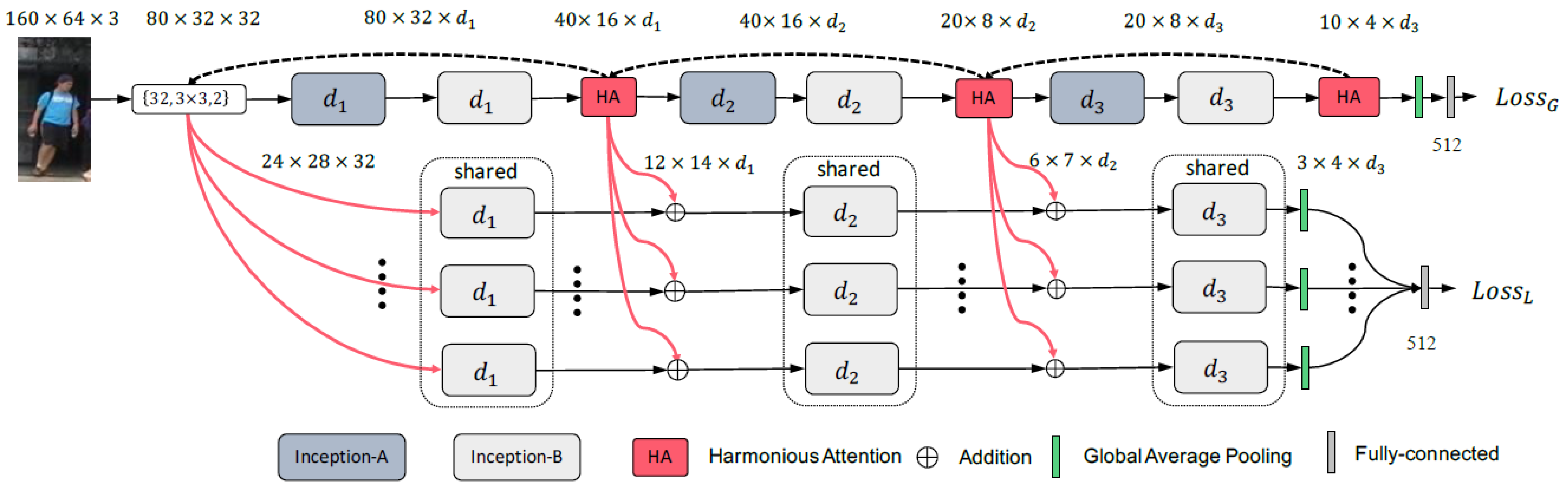
	DLPAR	HPNet
CUHK03	85.4%	91.8%
Market	81.0%	76.9%

Xihui Liu, Haiyu Zhao, Maoqing Tian, Lu Sheng, Jing Shao, Shuai Yi, Junjie Yan, Xiaogang Wang: HydraPlus-Net: Attentive Deep Features for Pedestrian Analysis. ICCV 2017

Attention Regions Learning: HA-CNN (2018)

- One global feature extraction branch
- Several local feature extraction branches (3 branches illustrated in figure)
- Harmonious Attention: learn a set of complementary attention maps
  - hard (regional) attention for the local branch
  - soft (pixel-level and scale-level) attention for the global branch

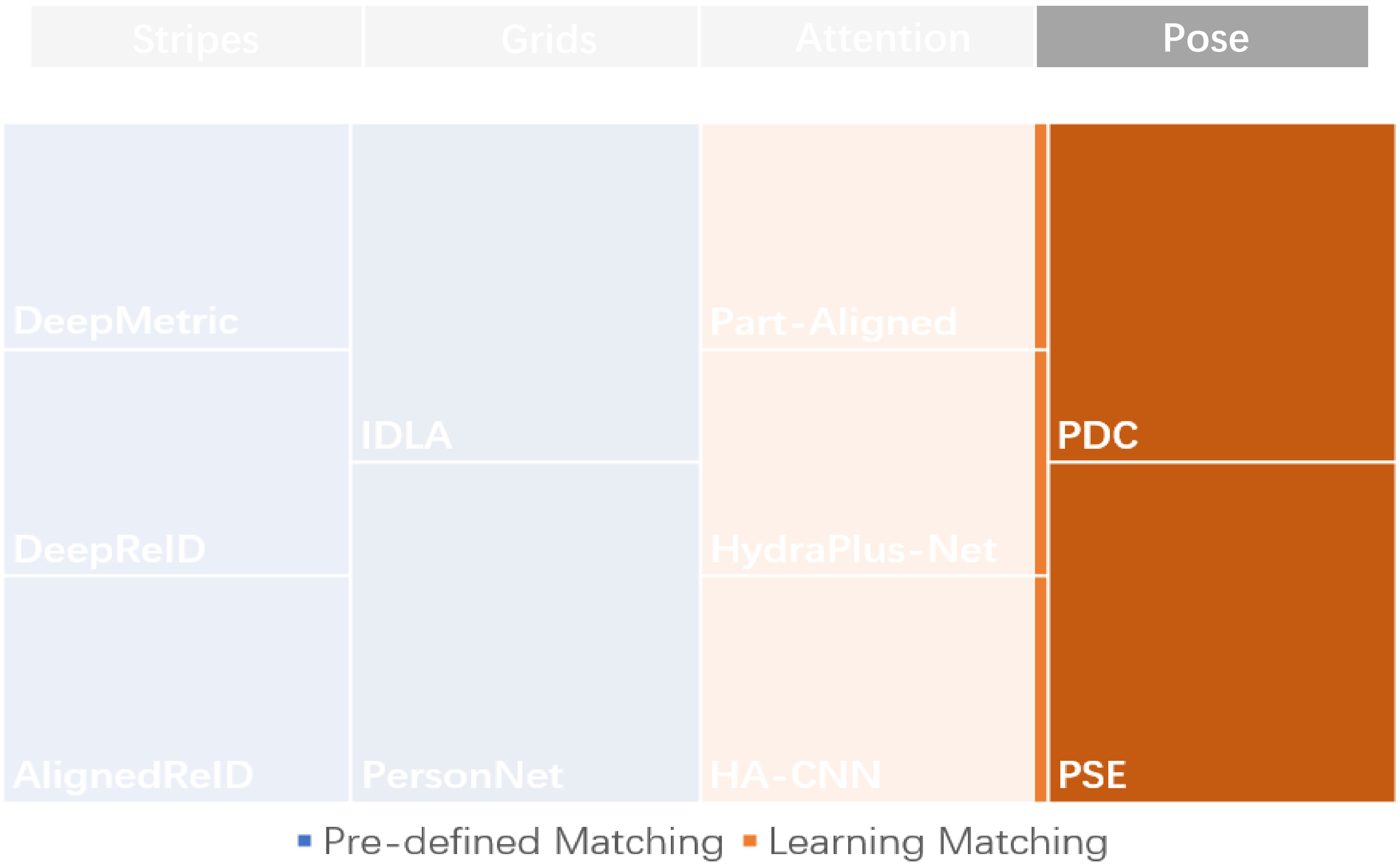
	HPNet	HA-CNN
CUHK03	91.8%	-
Market	76.9%	91.2%



Wei Li, Xiatian Zhu, and Shaogang Gong:
Harmonious Attention Network for Person Re-Identification. CVPR 2018

Deep Learning Based Methods

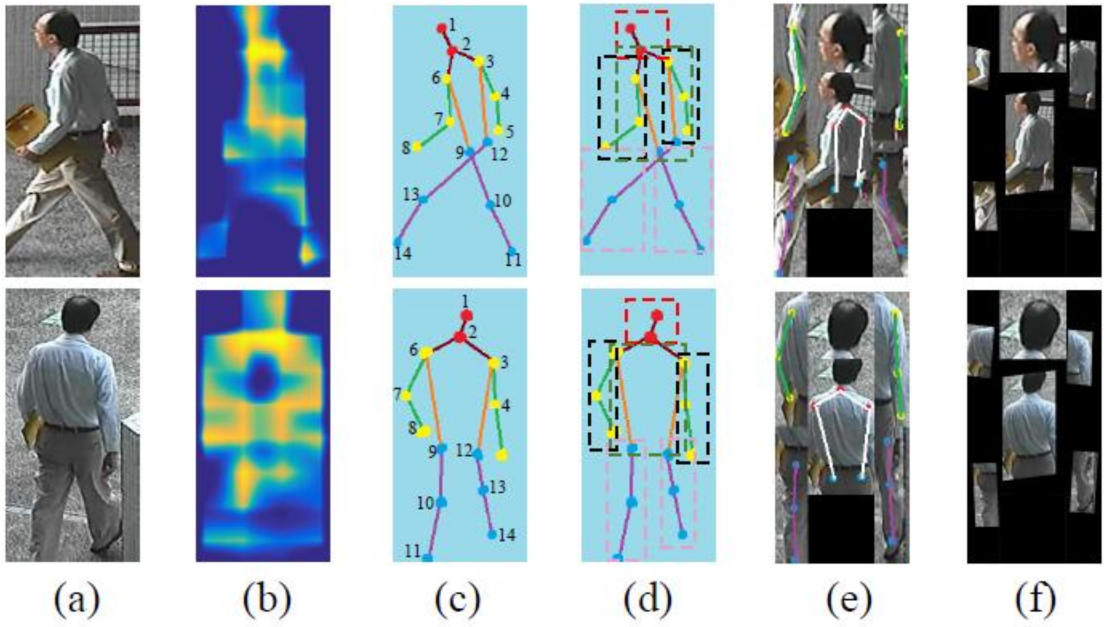
different matching or partitioning strategies



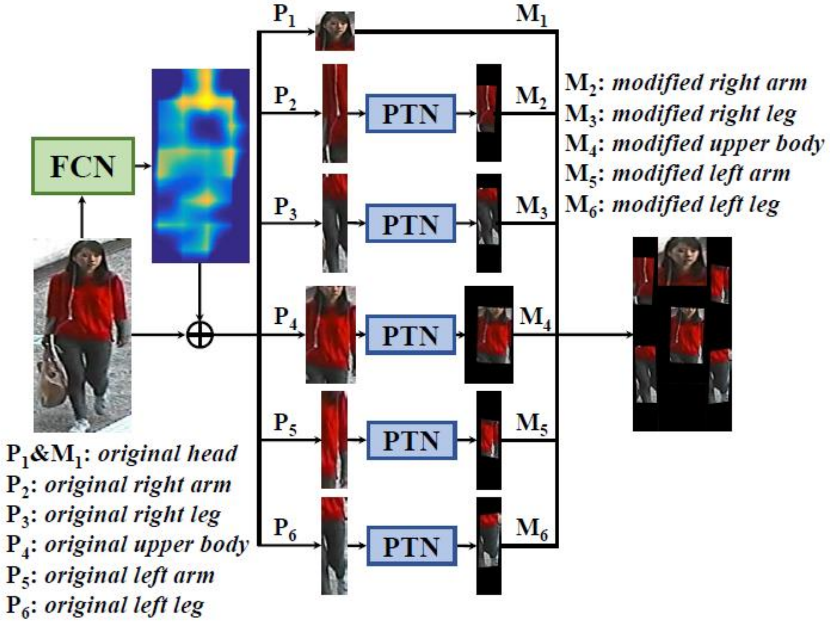
Pose-driven Embedding: PDC (2017)

- To better alleviate the challenges from pose variations
- Propose a FEN to learn and normalize human parts
  - Estimate 14 joints using a separated pose estimator (FCN).
  - Merge 14 joints to 6 parts and normalize to pre-defined locations
  - Generate a transformed and modified part image

	HPNet	PDC
CUHK03	91.8%	88.7%
Market	76.9%	84.1%



Feature Embedding sub-Net (FEN)



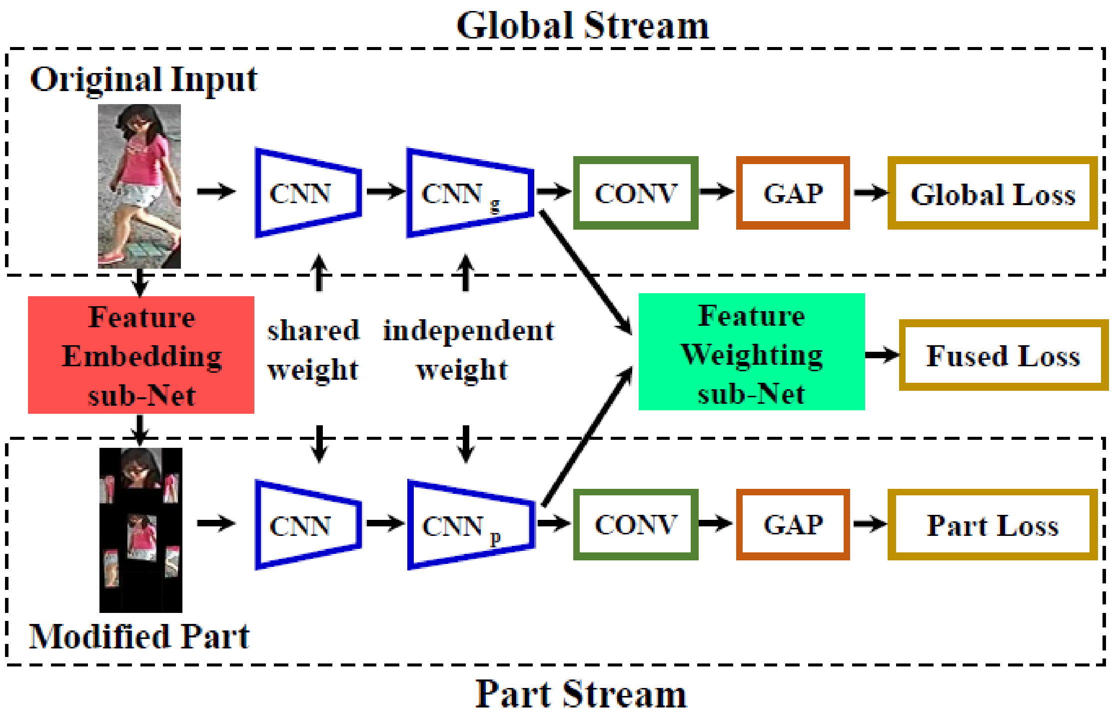
Feature Embedding sub-Net (FEN)

Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian:
Pose-driven Deep Convolutional Model for Person Re-identification. ICCV 2017

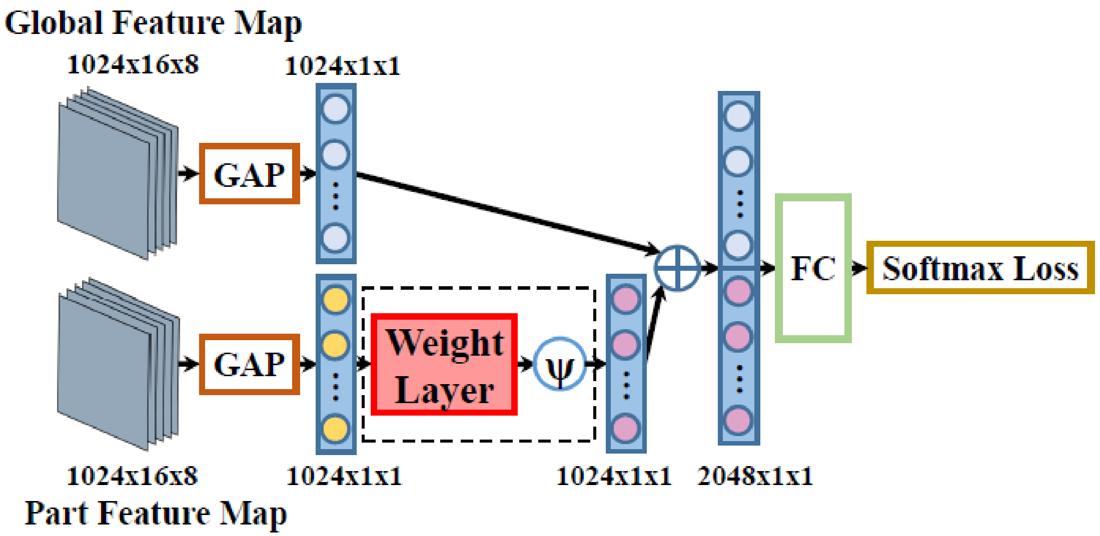
Pose-driven Embedding: PDC (2017)

- Global feature learnt from original image with Softmax loss
- Part feature learnt from modified part image with Softmax loss
- Fusing global and part features with a sub-Net

	HPNet	PDC
CUHK03	91.8%	88.7%
Market	76.9%	84.1%



Framework



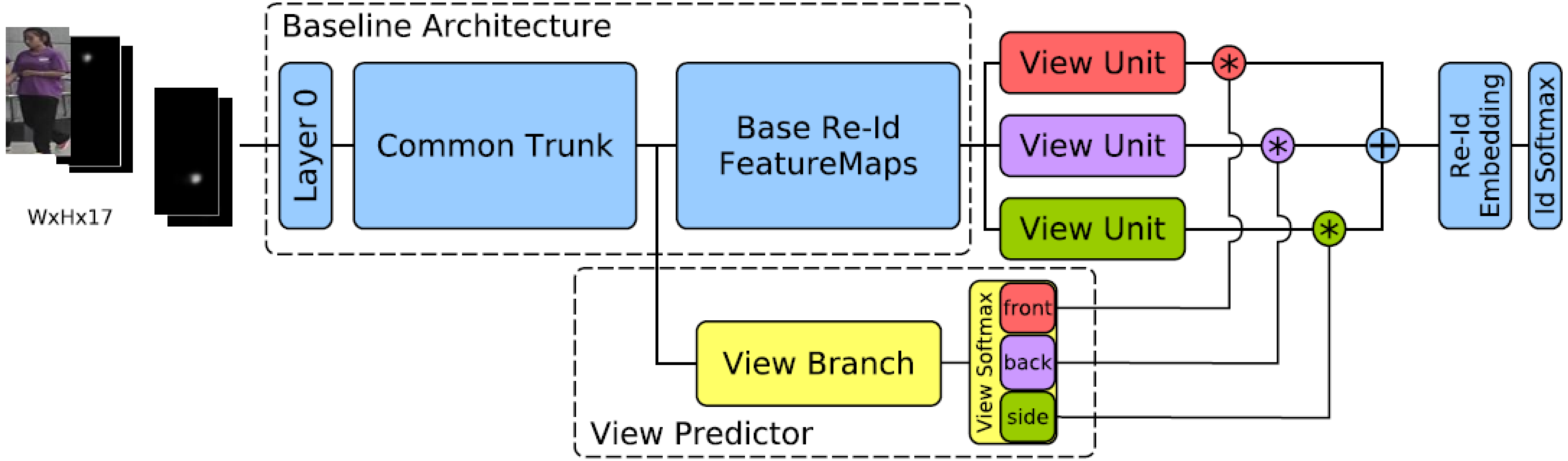
Feature Weighting sub-Net

Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian:
Pose-driven Deep Convolutional Model for Person Re-identification. ICCV 2017

Pose-driven Embedding: PSE (2018)

- Separately trained View Predictor and off-the-shelf pose estimator
- Combine pose maps and RGB images as input
- Using view predictions to select one of three CNN units
- Embedding the pose and view information by simply training the CNN

	PDC	PSE
CUHK03	88.7%	-
Market	84.1%	90.3%



M. Saquib Sarfraz, Arne Schumann, Andreas Eberle, Rainer Stiefelhagen:
A Pose-Sensitive Embedding for Person Re-Identification with Expanded Cross Neighborhood Re-Ranking. CVPR 2018

Performances

Methods	Publication	Types	CUHK03	Market-1501
DeepReID	CVPR 2014	Stripes + Matching	20.7	-
IDLA	CVPR 2015	Grid Patches + Matching	54.7	-
PersonNet	ArXiv 2016	Grid Patches + Matching	64.8	37.2
DCSL	IJCAI 2016	Structure Learning+ Matching	80.2	-
HydraPlus-Net	ICCV 2017	Attention + Embedding	91.8	76.9
Part-Aligned	ICCV 2017	Attention + Embedding	85.4	81.0
PDC	ICCV 2017	Pose + Embedding	88.7	84.1
PSE	CVPR 2018	Pose + Embedding	-	90.3
HA-CNN	CVPR 2018	Attention + Embedding	-	91.2
AlignedReID	ArXiv 2017	Stripes Association + Embedding	92.4	91.8

Conclusion

Person ReID with deep learning	
	Extracting feature maps using CNN
	Matching features by comparing on different locations

Matching on pre-defined spatial locations	Matching with learnt semantic regions
• Stripes • Grid • Patches	• Learn correspondence implicitly in the network • Learn key part regions for matching • Learn attention regions for feature embedding • Using off-the-shelf pose/view estimator for feature embedding

Acknowledgement

Dr. Liming Zhao

Dr. Yaqing Zhang