Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion

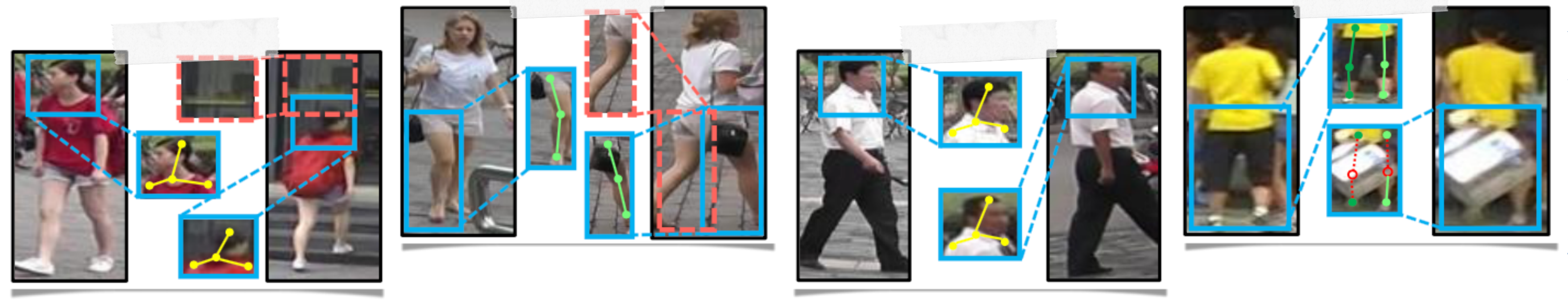
Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, Xiaoou Tang
SensetTime Group Limited *The Chinese University of Hong Kong*

Our code and dataset

IEEE 2017 Conference on Computer Vision and Pattern Recognition

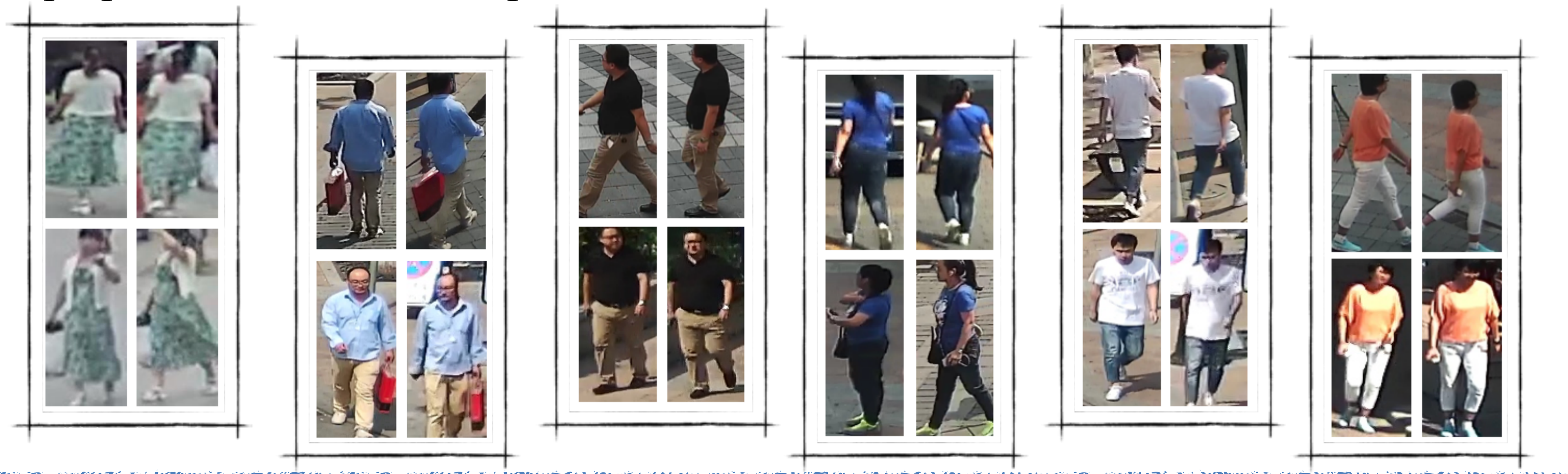
Person Re-identification

- Person re-identification (ReID) aims at associating person images across cameras and temporal periods. Given one query image of one person, a person ReID system is expected to provide all the images of the same person from large gallery database.
- It is of great security interest and can be used for various surveillance applications.
- Although person ReID has been studied for years, it is still quite challenging.
  - Human body regions cannot be well aligned across images.
  - The general appearances of the two persons are quite similar.
  - Some body parts may be occluded which makes the association process more difficult.

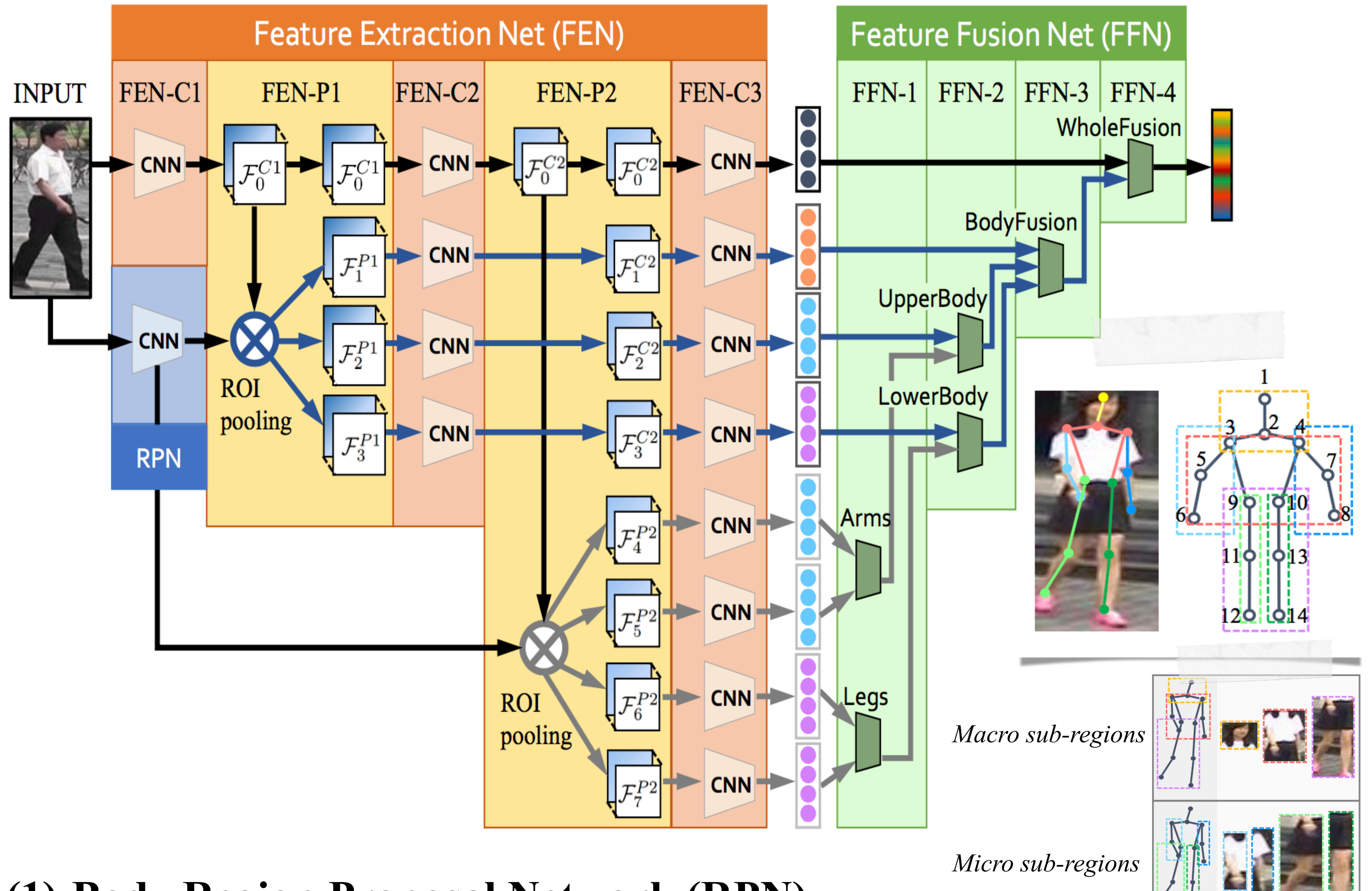


Our Contributions

- It is the first time human body structure information is considered in a ReID pipeline. It can help align body region features across images and local detail information can be better described.
- The Spindle Net is designed for the ReID task. Features of different body regions are first extracted by a multi-stage ROI pooling framework, and features of different semantic levels are pooled out separately at different stages. Then the regions features of different semantic levels are merged by a tree-structured fusion network with a competitive strategy.
- A real surveillance ReID dataset (SenseReID) is proposed for performance evaluation purpose only. Our proposed method can achieve state of the art performance on the proposed dataset and multiple standard datasets.



The Proposed Spindle Net



(1) Body Region Proposal Network (RPN)

- Given an input image, the RPN generates seven rectangle region proposals representing seven sub-regions of the person body in the image, including the head-shoulder region, the upper body region, the lower body region, two arm regions and two leg regions.
- The RPN contains two main steps, i.e. body joint localization (14 joints) and body region generation (7 regions).

(2) Feature Extraction Network (FEN)

- One 256-dimensional feature vector can be extracted from each of the eight regions, including a full body region, and seven sub-regions proposed by the RPN, corresponding to three macro sub-regions and four micro sub-regions.
- The FEN contains three convolution stages (FEN-C1, FEN-C2, FEN-C3) and two ROI pooling stages (FEN-P1, FEN-P2).

(3) Feature Fusion Network (FFN)

- In the FFN, the eight feature vectors are combined together to generate one compact 256 dimensional feature vector that can well represent the whole image.
- A tree-structured fusion strategy is proposed and features representing micro body sub-regions are merged in early stages and some macro features are merged in later stages.

Investigations on Spindle Net

(1) Investigations on FEN

- The combination of Full+FEN-C1/C2 achieves the best accuracies. FEN-C1/C2 means the macro features are pooled from FEN-C1 and the micro features are pooled from FEN-C2.
- Even without the FFN the Top-1 accuracy can be improved by 2.6% to 74.7% by introducing the macro and micro region features.

	Market-1501	Top-1	Top-5	Top-10	Top-20
(a)	Full only	72.1	88.9	92.9	95.6
(b)	Full+FEN-C1/C1	74.3	90.5	94.1	96.5
	Full+FEN-C2/C2	73.1	90.0	93.8	96.2
	Full+FEN-C3/C3	67.8	85.9	90.6	93.9
(c)	**Full+FEN-C1/C2**	**74.7**	**90.8**	**94.3**	**96.6**
	Full+FEN-C1/C3	73.7	90.0	93.7	96.2
	Full+FEN-C2/C3	72.5	89.3	93.2	95.8
(d)	Full+FEN-C2/C1	74.0	90.5	94.1	96.5
	Full+FEN-C3/C1	72.2	89.4	93.3	95.9
	Full+FEN-C3/C2	72.0	89.2	93.2	95.9

(2) Investigations on FFN

- There are two key factors of the proposed FFN, i.e. the tree fusion structure and the feature competition strategy.
- The proposed FFN (Tree+Max.) achieves the best performance. Global fine-tuning the whole Spindle Net can further improve the performance.

Market-1501	Top-1	Top-5	Top-10	Top-20
Linear + Concat.	72.8	89.1	93.0	95.6
Linear + Avg.	62.7	82.0	87.3	91.4
Linear + Max.	62.8	82.0	87.2	91.3
i-Tree + Concat.	66.5	86.4	91.3	94.7
i-Tree + Avg.	68.6	87.4	91.9	95.0
i-Tree + Max.	41.9	66.4	76.2	84.1
Tree + Concat.	67.1	84.7	88.9	92.1
Tree + Avg.	74.3	90.4	93.9	96.3
Tree + Max. (Ours)	**76.3**	**91.1**	**94.5**	**96.5**
Fine-tune (Ours)	**76.9**	**91.4**	**94.6**	**96.7**

Experiments and Results

(1) Quantitative Evaluation

- State of the art performance can be achieved on multiple datasets and the robustness of the proposed Spindle Net is demonstrated on a new dataset (SenseReID) without fine-tuning.

Market-1501	Top-1	Top-5	Top-10	Top-20
WARCA-L	45.2	68.2	-	-
NFST	61.0	-	-	-
PersonNet	37.2	-	-	-
S-CNN	**65.9**	-	-	-
BoW-best	42.6	-	-	-
Spindle (Ours)	**76.9**	91.5	94.6	96.7

CUHK03	Top-1	Top-5	Top-10	Top-20
WARCA-χ^2	**78.4**	94.6	-	-
NFST	62.6	90.1	94.8	98.1
PersonNet	64.8	89.4	94.9	98.2
S-CNN	61.8	80.9	88.3	-
JSTL	75.3	-	-	-
Spindle (Ours)	**88.5**	97.8	98.6	99.2

SenseReID dataset	Top-1	Top-5	Top-10	Top-20
JSTL	23.0	34.8	40.6	46.3
BoW-best	22.4	-	-	-
Spindle (Ours)	**34.6**	52.7	59.9	66.7

(2) Qualitative Evaluation

- For each of the probe image (shown in black box), the top-10 results of the JSTL model and the results of the proposed Spindle Net, are shown in Rows (a) and (b), respectively.
- Correct results are shown in green boxes and the incorrect ones are shown in red boxes.