DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification

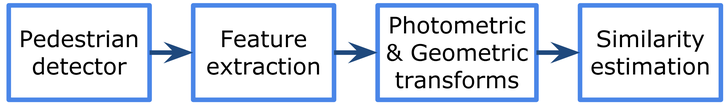
Wei Li Rui Zhao Tong Xiao Xiaogang Wang

香港中文大學
The Chinese University of Hong Kong

IEEE 2014 Conference on Computer Vision and Pattern Recognition
CVPR 2014

Motivation:

- Current algorithms are developed and evaluated on manually cropped dataset. In practice, images come from a detector which introduces larger inter-camera variations.
- Current algorithms optimize components of the re-identification pipeline individually.

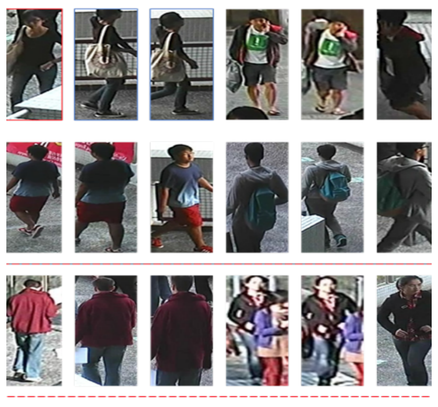


Contribution:

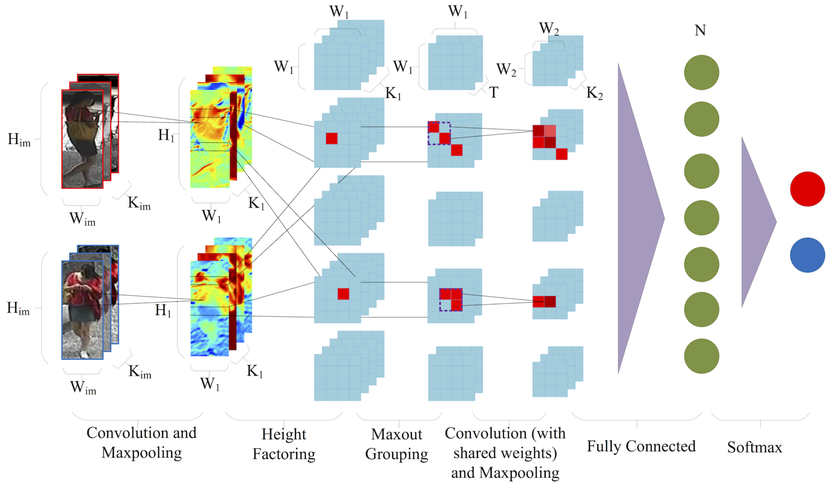
- A new dataset providing bounding boxes from a pedestrian detector[1] *and* manually labeling.
- A new deep model for jointly optimizing all components in the re-identification pipeline. It is designed to be robust to imperfect detection.

A New Dataset:

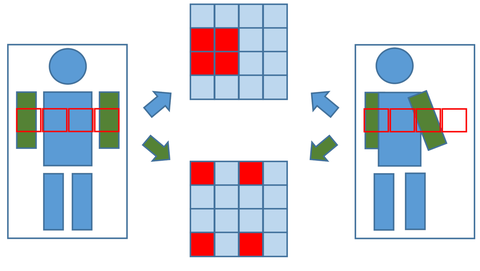
- We build a new dataset in more realistic settings.
  - Pedestrians images are automatically detected by detectors
  - 5 pairs of camera settings in uncontrolled environment
  - 1360 identities, each identity has ~4.8 images per view
  - Manually labeled bounding boxes are also provided



Deep Neural Network for ReID:

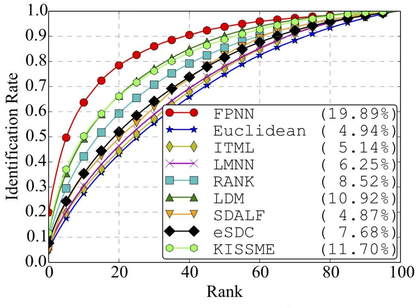


- Jointly optimizing key components of re-identification pipeline: *feature extraction*, *photometric transformations*, *geometric transformations*.
- **Feature Extraction** is modeled by two parallel convolutional layers and multiple filter pairs.
- **Photometric Transformation** is modeled by height factoring and maxout grouping layers, and the difference detected by paired filters.

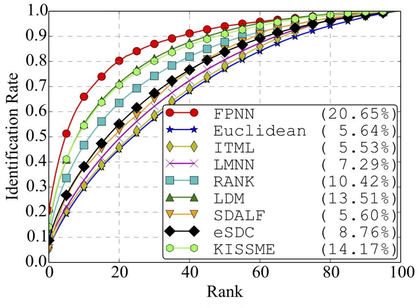


- **Geometric Transformation** is modeled by convolution on displacement matrices.
- A **Mixture of Transformations** are modeled by fully connected layer.

Experiments:

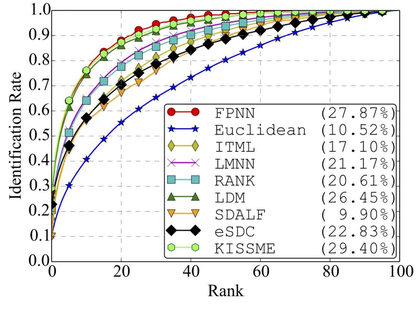


Detected



Labeled

- Our model works best on both detected and labeled samples. Our model drops 0.76%(relatively **-3.7%**), while the second best (KISSME) drops 2.47%(**-17.4%**) when automatically detected bounding boxes are used.



- Our model performs reasonably well on a much smaller dataset:
  - 971 identities
  - 2 images per identity per view

Reference:

[1] *P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan,* Object Detection with Discriminatively Trained Part Based Models, PAMI 2010