

Person Re-Identification

Hongzhi Liu

May 3, 2018

1 Existing Methods and Research Background

We often see this scene in action movies: The agents identify the different locations of the target personnel by retrieving the monitoring images in surveillance camera. After reading Professor Li's thesis, I realize it is a technique about person re-identification.

Nowadays, existing person re-identification methods either assume the availability of well-aligned person bounding box images as model input or rely on constrained attention selection mechanisms to calibrate misaligned images. Different from these, Professor Li presents a *Harmonious Attention Convolutional Neural Network* (HA-CNN) for joint learning of person re-identification attention selection and feature representations in an end-to-end fashion [1]. In contrast to most existing re-id methods that either ignore the matching misalignment problem or exploit stringent attention learning algorithms, the proposed model is capable of extracting multiple complementary attention and maximising their latent complementary effect for person re-id in a unified lightweight CNN architecture.

2 Harmonious Attention Network

The team aims to learn a deep feature representation model optimal for person re-id matching under significant viewing condition variations. To this end, Professor Li formulates a *Harmonious Attention Convolutional Neural Network* that aims to con-

currently learn a set of harmonious attention, global and local feature representations for maximising their complementary benefit and compatibility in terms of both discrimination power and architecture simplicity. The overall design of the HA-CNN architecture is shown in Figure. 1.

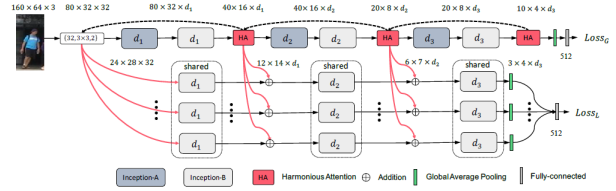


Figure 1: The Harmonious Attention Convolutional Neural Network. The symbol d_l ($l \in 1, 2, 3$) denotes the number of convolutional filter in the corresponding Inception unit at the l -th block.

This HA-CNN model contains two branches: (1) One local branch: Each stream aims to learn the most discriminative visual features for local image regions of a person bounding box image. (2) One global branch: This aims to learn the optimal global level features from entire person images. For both branches, the team select the Inception-A/B [2] units as the basic building blocks.

3 Datasets and Evaluation Protocol

For evaluation, Professor Li selects three large-scale person re-id benchmarks, Market-1501, DukeMTMC-ReID and CUHK03. They adopt the

standard person re-id setting including the training split and test protocol as Table 1 shows. For performance measure, they use the cumulative matching characteristic and mean Average Precision metrics. Existing deep re-id methods typically benefit significantly from these operations at the price of not only much higher computational cost but also notoriously difficult and time-consuming model tuning.

Table 1: Re-id evaluation protocol. TS: Test Setting; SS: Single-Shot. SQ: Single-Query; MQ: Multi-Query

Dataset	#ID	#Train	#Test	#Image	Test Setting
CUHK03	1,467	767	700	14,097	SS
Market-1501	1,501	751	750	32,668	SQ/MQ
DukeMCMT-ReID	1,402	702	702	36,411	SQ

References

- [1] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. 2018.
- [2] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. pages 1249–1258, 2016.