**Part-Aligned Bilinear Representations for Person Re-identification**

Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, Kyoung Mu Lee

Seoul National University, Microsoft Research Asia, Max Planck Institute for Intelligent Systems, University of Tübingen *JD AI Research

**PROPOSED METHOD**

**Network architecture**

- Training:
  - Pre-train the pose map extractor using the pose estimation dataset
  - Train the whole network to minimize the re-identification loss

**Image-to-image similarity**

Image similarity is the average of local feature similarities over every position pair:

\[
sim_{m}(I, I') = \frac{1}{25} \sum_{x,y} \sum_{x',y'} \sim_{p}(x,y,x',y') \sim_{a}(x,y,x',y')
\]

Local similarity is a product of local appearance and part similarity:

\[
sim_{l}(I, I') = \text{vec} \!(\text{vec}(\tilde{f}_{I}, \tilde{f}_{I'})) \approx < \tilde{a}_{xy}, \tilde{a}_{xy'} > < \tilde{p}_{xy}, \tilde{p}_{xy'} >
\]

From (1) and (2), image similarity becomes the average local appearance similarities at the corresponding body parts:

\[
sim_{m}(I, I') = \frac{1}{25} \sum_{x,y} \sum_{x',y'} s_{m}(x,y,x',y') s_{m}(x,y,x',y')
\]

**Part-aligned feature by bilinear pooling of part and appearance feature maps**

- Decomposes part & appearance using a two stream network
- Attention based
  - Learn attention with weak supervision
  - No explicit pose estimation

**Related works**

- Bounding-box based:
  - Crop bounding boxes around the parts of interests discovered from pose estimation
  - Hard to make detailed part discrimination

**Introduction & Background**

**Person Re-identification Problem**

- How can we identify the same person across different cameras reliably?

**Challenge & Motivation**

- Body part misalignment is a key challenge
- Part-aligned representation is necessary

**Proposal/Main Idea**

- New part-aligned feature (representation):
  - Decomposes part & appearance using a two stream network
  - Part-aligned feature by bilinear pooling of part and appearance feature maps

- Part-aware matching using part-aligned features:
  - Image-to-image similarity (matching) becomes the Average local app similarities at the corresponding body part

**Visualization of the image-to-image similarity**

- Comparison with baselines while varying dataset/architecture/pooling
- Comparison with the State-of-the-Arts

**Experiments**

- Our network successfully separates appearance and part information
- Code is available at: github.com/yuminsuh/part_bilinear_reid

**INTRO & BACKGROUND**

**Person Re-Identification Problem**

- How can we identify the same person across different cameras reliably?

**Challenge & Motivation**

- Body part misalignment is a key challenge
- Part-aligned representation is necessary

**Related works**

- Bounding-box based:
  - Crop bounding boxes around the parts of interests discovered from pose estimation
  - Hard to make detailed part discrimination

**Proposal/Main Idea**

- New part-aligned feature (representation):
  - Decomposes part & appearance using a two stream network
  - Part-aligned feature by bilinear pooling of part and appearance feature maps

- Part-aware matching using part-aligned features:
  - Image-to-image similarity (matching) becomes the Average local app similarities at the corresponding body part

**Visualization of the image-to-image similarity**

- Comparison with baselines while varying dataset/architecture/pooling
- Comparison with the State-of-the-Arts

**Experiments**

- Our network successfully separates appearance and part information
- Code is available at: github.com/yuminsuh/part_bilinear_reid