Person Re-identification with Deep Similarity-Guided Graph Neural Network

Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen, Xiaogang Wang
CUHK-SenseTime Joint Lab, SenseTime Research
ECCV 2018
Problem

• **Weakness of the Existing Person Re-ID Models:**
  • Ignore the relationship information between different probe-gallery pairs.
  • Hard samples are difficult to get proper similarity scores.

• **Main Idea:** Update s2 by s1 and s3.
Comparison

Conventional Approach

SGGNN Approach
Graph Formulation

- An undirected **complete graph** $G(V, E)$.
- Each node $v_i$ represents a pair of probe-gallery images.
- Node features are processed difference features.
Naïve Node **Loss Function I**

- $L = - \sum_{i=1}^{N} y_i \log(f(d_i)) + (1 - y_i) \log(1 - f(d_i))$
- $f()$ is a linear classifier followed by a sigmoid function.
Similarity-Guided Graph Neural Network

- **Intuition**: Using gallery-gallery similarity scores to guide the refinement of the probe-gallery relation features.
- **Updating Node Feature**: Original Feature + Fusion Feature.
- \[ d_i^{(t+1)} = (1 - \alpha) d_i^{(t)} + \alpha \sum_{j=1}^{N} W_{ij} t_j^{(t)} \] for \( i = 1, 2, \ldots, N \)
Similarity-Guided Graph Neural Network

- **Updating**: \( d^{(t+1)}_i = (1 - \alpha)d^{(t)}_i + \alpha \sum_{j=1}^{N} W_{ij}t^{(t)}_j \) for \( i = 1, 2, \ldots, N \)

- \( W_{ij} \) is a scalar edge weight, represents the relation importance between node \( i \) and node \( j \).

- \( W_{ij} = \begin{cases} \frac{\exp(S(g_i,g_j))}{\sum_j \exp(S(g_i,g_j))}, & i \neq j \\ 0, & i = j \end{cases} \)

- \( S() \) is a pairwise similarity function.

- Set \( t = 1 \) in both training and testing.

Avoid Self-Enhancing
Update Node Loss Function II

- $L = - \sum_{i=1}^{N} y_i \log(s_i) + (1 - y_i) \log(1 - s_i)$
- Similarity estimator is a linear classifier followed by a sigmoid function.
Datasets, Metrics, Experiments

• **Datasets:** CUHK03, Market-1501, DukeMTMC
• **Metrics:** mAP and CMC top-1, top-5, top-10 accuracies.

### CUHK03 vs Market-1501 vs DukeMTMC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mAP</td>
<td>top-1</td>
<td>top-5</td>
<td>top-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>top-1</td>
<td>top-5</td>
<td>top-10</td>
</tr>
<tr>
<td>Quadruplet Loss [9]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>75.5</td>
<td>95.2</td>
</tr>
<tr>
<td>OIM Loss [65]</td>
<td>CVPR 2017</td>
<td>72.5</td>
<td>77.5</td>
<td>-</td>
</tr>
<tr>
<td>SpindleNet [73]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>88.5</td>
<td>97.8</td>
</tr>
<tr>
<td>MSCAN [26]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>74.2</td>
<td>94.3</td>
</tr>
<tr>
<td>SSD [2]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>76.6</td>
<td>94.6</td>
</tr>
<tr>
<td>k-reciprocal [78]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>67.6</td>
<td>61.6</td>
</tr>
<tr>
<td>VI+LSRO [77]</td>
<td>ICCV 2017</td>
<td>87.4</td>
<td>84.6</td>
<td>97.6</td>
</tr>
<tr>
<td>SVDNet [61]</td>
<td>ICCV 2017</td>
<td>84.8</td>
<td>81.8</td>
<td>95.2</td>
</tr>
<tr>
<td>OLMANS [80]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>61.7</td>
<td>88.4</td>
</tr>
<tr>
<td>Pose Driven [60]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>88.7</td>
<td>98.6</td>
</tr>
<tr>
<td>Part Aligned [74]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>85.4</td>
<td>97.6</td>
</tr>
<tr>
<td>HydraPlus-Net [39]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>91.8</td>
<td>98.4</td>
</tr>
<tr>
<td>MuDeep [49]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>76.3</td>
<td>96.0</td>
</tr>
<tr>
<td>JLML [29]</td>
<td>IJCAI 2017</td>
<td>-</td>
<td>83.2</td>
<td>98.0</td>
</tr>
<tr>
<td>MC-PPIM [43]</td>
<td>AAAI 2018</td>
<td>-</td>
<td>86.4</td>
<td>98.5</td>
</tr>
<tr>
<td><strong>Proposed SGGNN</strong></td>
<td></td>
<td><strong>94.3</strong></td>
<td><strong>95.3</strong></td>
<td><strong>99.1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Reference</th>
<th>Market-1501 [70]</th>
<th>DukeMTMC [52]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mAP</td>
<td>top-1</td>
<td>top-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>top-1</td>
<td>top-5</td>
</tr>
<tr>
<td>OIM Loss [65]</td>
<td>CVPR 2017</td>
<td>60.9</td>
<td>82.1</td>
</tr>
<tr>
<td>SpindleNet [73]</td>
<td>CVPR 2017</td>
<td>-</td>
<td>76.9</td>
</tr>
<tr>
<td>MSCAN [26]</td>
<td>CVPR 2017</td>
<td>53.1</td>
<td>76.3</td>
</tr>
<tr>
<td>SSD [2]</td>
<td>CVPR 2017</td>
<td>68.8</td>
<td>82.2</td>
</tr>
<tr>
<td>k-reciprocal [78]</td>
<td>CVPR 2017</td>
<td>63.6</td>
<td>77.1</td>
</tr>
<tr>
<td>Point 2 Set [81]</td>
<td>CVPR 2017</td>
<td>44.3</td>
<td>70.7</td>
</tr>
<tr>
<td>CADL [35]</td>
<td>CVPR 2017</td>
<td>47.1</td>
<td>73.8</td>
</tr>
<tr>
<td>VI+LSRO [77]</td>
<td>ICCV 2017</td>
<td>66.1</td>
<td>84.0</td>
</tr>
<tr>
<td>SVDNet [61]</td>
<td>ICCV 2017</td>
<td>62.1</td>
<td>82.3</td>
</tr>
<tr>
<td>OLMANS [80]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>60.7</td>
</tr>
<tr>
<td>Pose Driven [60]</td>
<td>ICCV 2017</td>
<td>63.4</td>
<td>84.1</td>
</tr>
<tr>
<td>Part Aligned [74]</td>
<td>ICCV 2017</td>
<td>63.4</td>
<td>81.0</td>
</tr>
<tr>
<td>HydraPlus-Net [39]</td>
<td>ICCV 2017</td>
<td>-</td>
<td>76.9</td>
</tr>
<tr>
<td>JLML [29]</td>
<td>IJCAI 2017</td>
<td>65.5</td>
<td>85.1</td>
</tr>
<tr>
<td>HA-CNN [30]</td>
<td>CVPR 2018</td>
<td>75.7</td>
<td>91.2</td>
</tr>
<tr>
<td><strong>Proposed SGGNN</strong></td>
<td></td>
<td><strong>82.8</strong></td>
<td><strong>92.3</strong></td>
</tr>
</tbody>
</table>
Ablation Study

• **Base Model**: Only use the naïve node loss function.

• **SGGNN w/o SG**: 
\[ d_i^{(t+1)} = (1 - \alpha)d_i^{(t)} + \alpha \sum_{j=1}^{N} h(d_i, d_j)t_j^{(t)}, \]
where \( h(\ast, \ast) \) is an inner product function.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Market-1501</th>
<th>CUHK03</th>
<th>DukeMTMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mAP</td>
<td>top-1</td>
<td>mAP</td>
</tr>
<tr>
<td>Base Model</td>
<td>76.4</td>
<td>91.2</td>
<td>88.9</td>
</tr>
<tr>
<td>Base Model + SGGNN w/o SG</td>
<td>81.2</td>
<td>90.6</td>
<td>92.7</td>
</tr>
<tr>
<td>Base Model + SGGNN</td>
<td>82.8</td>
<td>92.3</td>
<td>94.3</td>
</tr>
</tbody>
</table>

Normal Ablation Study.

To show SGGNN also learns better visual features. Evaluate the performance by directly calculating the l2 distance between probe and gallery image features from ResNet-50 model.

2018/10/09 Xu Gao, Peking University
Conclusion

• Present SGGNN to incorporate the rich gallery-gallery similarity information into training process.
• + Consider the relationship between each probe-gallery pair.
• + Add directly label supervision for guidance.
• - A complete graph might be slow when the number of nodes increasing.