Multi-Task Curriculum Transfer Deep Learning of Clothing Attributes

Qi Dong
q.dong@qmul.ac.uk
Shaogang Gong
s.gong@qmul.ac.uk
Xiattian Zhu
xiatian.zhu@qmul.ac.uk

1 Introduction

Problem
Domain transfer learning for recognizing fine-grained multi-label clothing attributes in the street (wild) given limited training data.

Limitation of Existing Methods
• Handcrafted features.
• Single task deep learning for multi-label recognition.
• Lack of end-to-end cross domain transfer learning.

Contributions
• Novel Multi-Task Curriculum Transfer (MTCT) deep learning strategy.
• Effective Multi-Task Network (MTN) for learning from sparse target data.

2 Overview of method

Clothing detection
Faster R-CNN[4] for clothing detection

Stage 1: Shop domain (clean)
Pretrain MTN on ImageNet and train on shop images.

Stage 2: Street domain (wild)
Initialize 3MTN by shop domain images trained model and then fine-tune FC layers using cross-domain triplet information for transfer learning

3 Multi-task deep learning

Stage 1: Multi-Task (MT)

Stage 2: Curriculum Transfer (Ct)

4 Experiments

Comparison to the State-of-The-Arts

<table>
<thead>
<tr>
<th>Methods</th>
<th>Category</th>
<th>Button</th>
<th>Colour</th>
<th>Length</th>
<th>Pattern</th>
<th>Shape</th>
<th>Collar</th>
<th>Slev-Len</th>
<th>Slv-shp</th>
<th>mAP@5</th>
<th>mAP@10</th>
<th>mAP@15</th>
<th>mAP@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN1[1]</td>
<td>12.56</td>
<td>24.13</td>
<td>20.72</td>
<td>35.91</td>
<td>61.67</td>
<td>47.14</td>
<td>31.17</td>
<td>80.63</td>
<td>73.96</td>
<td>43.10</td>
<td>45.41</td>
<td>52.20</td>
<td></td>
</tr>
<tr>
<td>DNN2[2]</td>
<td>52.55</td>
<td>37.48</td>
<td>58.24</td>
<td>51.49</td>
<td>67.53</td>
<td>47.70</td>
<td>47.77</td>
<td>82.04</td>
<td>73.72</td>
<td>57.61</td>
<td>57.79</td>
<td>67.29</td>
<td></td>
</tr>
<tr>
<td>FashionNet[3]</td>
<td>55.85</td>
<td>39.52</td>
<td>60.33</td>
<td>53.08</td>
<td>68.65</td>
<td>49.79</td>
<td>51.27</td>
<td>83.79</td>
<td>75.34</td>
<td>59.84</td>
<td>59.97</td>
<td>69.74</td>
<td></td>
</tr>
<tr>
<td>MTCT</td>
<td>65.96</td>
<td>43.57</td>
<td>66.86</td>
<td>58.27</td>
<td>70.55</td>
<td>51.40</td>
<td>58.79</td>
<td>86.05</td>
<td>77.54</td>
<td>64.35</td>
<td>64.97</td>
<td>75.66</td>
<td></td>
</tr>
</tbody>
</table>

MTN and Transfer learning

End-to-End and Curriculum Transfer learning

Different cross-domain loss functions

Model Robustness vs. target data size

4.2 Implementational Considerations

We built on Stage 1 to transfer learned localised attribute information, followed by optimising the vectorised feature maps of the conv layers during the CT Stage-2 learning.

Stage 1: MTN stream and all FC layers of both Source and Target training data ratio

Attribute order from top to bottom: Category, Button, Colour, Length, Pattern, Shape, Collar, Slev-Len, Slv-Shape

A qualitative evaluation of MTCT

IEEE 2017 Winter conference on Applications of Computer Vision