**Introduction**

*Task:* Synthesize person images in arbitrary poses, based on an image of that person and a novel pose.

*Motivation:* Provide users more control over the generation process.

*Key idea:* Guide the generation process explicitly by an appropriate representation of that intention.

**Dataset**

- i) DeepFashion[1]: in-shop clothes retrieval dataset with 256x256 resolution.
- ii) Market-1501[2]: person re-identification dataset with 128x64 resolution.

**Contributions**

- i) A novel task of conditioning image generation on a reference image and an intended pose.
- ii) A novel mask loss is proposed to encourage the model to focus on transferring the human body appearance instead of background information.
- iii) Divide the problem into two stages, with stage-I focusing on global structure and stage-II on filling in appearance details.

**Generation results**

**Qualitative results**

- On Market-1501 dataset.

**Quantitative results**

Table 1: Quantitative evaluation. For all measures, higher is better.

<table>
<thead>
<tr>
<th>Model</th>
<th>SSIM</th>
<th>IS</th>
<th>SSIM</th>
<th>IS</th>
<th>mask-SSIM</th>
<th>mask-IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepFashion</td>
<td></td>
<td></td>
<td>Market-1501</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1-CE-L1</td>
<td>0.694</td>
<td>2.395</td>
<td>0.210</td>
<td>2.568</td>
<td>0.771</td>
<td>2.455</td>
</tr>
<tr>
<td>G1-HEM-L1</td>
<td>0.735</td>
<td>2.427</td>
<td>0.294</td>
<td>3.171</td>
<td>0.802</td>
<td>2.506</td>
</tr>
<tr>
<td>G1-L1</td>
<td>0.735</td>
<td>2.427</td>
<td>0.304</td>
<td>3.006</td>
<td>0.809</td>
<td>2.455</td>
</tr>
<tr>
<td>G1-poseMaskLoss</td>
<td>0.779</td>
<td>2.668</td>
<td>0.340</td>
<td>3.326</td>
<td>0.817</td>
<td>2.682</td>
</tr>
<tr>
<td>G1-D</td>
<td>0.761</td>
<td>3.091</td>
<td>0.283</td>
<td>3.499</td>
<td>0.803</td>
<td>3.310</td>
</tr>
<tr>
<td>G1+G2-D</td>
<td>0.762</td>
<td>3.090</td>
<td>0.253</td>
<td>3.460</td>
<td>0.792</td>
<td>3.435</td>
</tr>
</tbody>
</table>

**Further analysis**

- Comparison examples with [3]
- Our failure cases on DeepFashion

---

