CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On

Matiur Rahman Minar¹, Thai Thanh Tuan¹, Heejune Ahn¹, Paul Rosin², Yu-Kun Lai²

¹SeoulTech, ROK, ²Cardiff University, UK

CVPRW 2020
3rd Workshop on Computer Vision for Fashion, Art and Design
Virtual try on

- Retain body shape and pose
- Reserve characteristics of target clothes
- Eliminate old clothes and replace with target clothes
- Retain non-relevant clothes
VITON

• Coarse-to-fine approach, using two-stage network

• Generator Stage
  • encoder-decoder generator
  • coarse synthesized image result $I'$

• Refinement Stage
  • generate warped image $c'$ using TPS
  • refine using $c'$ and $I'$
VITON

• Coarse-to-fine approach, using two-stage network
  
  • Generator Stage
    • encoder-decoder generator
    • coarse synthesized image result $I'$
  
  • Refinement Stage
    • generate warped image $c'$ using TPS
    • refine using $c'$ and $I'$

VITON: An Image-based Virtual Try-on Network, Xintong Han et al. CVPR 2018
VITON

• Coarse-to-fine approach, using two-stage network

• Problem
  • Warping is vulnerable to mask, blurry in rich details
CP-VTON

- Preserving the characteristics of clothes

- Geometric Matching Module (GMM)
  - estimating transformation parameters (TPS)
  - generate warped image \( \hat{c} \)

- Try-On Module (TOM)
  - A network to estimate \( M \) and coarse person image
  - generate final try on image \( I_r \)
  - fuse \( M \), \( I_r \) and \( \hat{c} \)

Toward Characteristic-Preserving Image-based Virtual Try-On Network, Bochao Wang et al., ECCV 2018
CPVTON+

Add skin label to VITON dataset

Using target cloth mask

Add un-upper cloth

Add warped cloth mask

Add regularization loss

Supervised ground truth

6/1/2020
Clothing Warping Stage: Adding skin label

- Neck and bare chest area → wrongly labeled as background
- Improvement:
  - Add new label ‘skin’

Reference image
Human parsing
From VITON dataset
Body shape
In CPVTON
Update human
 parsing
With skin label
Body shape
In CPVTON+
Clothing Warping Stage: using of cloth mask

• Colored cloth → Cloth mask

\[ \theta = f_\theta(f_H(H_t), f_C(C_i)) \]
Clothing Warping Stage: TPS parameters regularization

- Reveal that warped clothing is often severely distorted. → Add regularization on the TPS parameters.

\[
L_{GMM}^{CP, VTON^+} = \lambda_1 \cdot L1(C_{warped, Ict}) + \lambda_{reg} \cdot L_{reg}
\]

\[
L_{reg}(G_x, G_y) = \sum_{i=-1,1} \sum_{x} \sum_{y} |G_x(x + i, y) - G_x(x, y)| + \sum_{j=-1,1} \sum_{x} \sum_{y} |G_x(x, y + j) - G_x(x, y)|
\]
Blending Stage: Retain un-upper clothes area

- Reference image
- CPVTON head only input
- CPVTON result
- CPVTON+ up-upper clothes area input
- CPVTON+
Blending Stage: Supervised ground truth mask

\[
L_{\text{TOM}}^{CP,VTON} = \lambda_1 \cdot L1(I_0 - I_{GT}) + \lambda_{VGG} \cdot LVGG(I_0, I_{GT}) + \lambda_{\text{mask}} \cdot L1(1, M_0)
\]

\[
L_{\text{TOM}}^{CP,VTON+} = \lambda_1 \cdot L1(I_0 - I_{GT}) + \lambda_{VGG} \cdot LVGG(I_0, I_{GT}) + \lambda_{\text{mask}} \cdot L1(M_{GT}, M_0)
\]
Blending Stage: improve background color in inshop clothes

• TOM could not recognize the white cloth area.
### Experiments and Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Warped (IoU)</th>
<th>SSIM</th>
<th>LPIPS</th>
<th>IS (mean ± std.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP-VTON[4]</td>
<td>0.7898</td>
<td>0.7798</td>
<td>0.1397</td>
<td>2.7809 ± 0.0594</td>
</tr>
<tr>
<td>CP-VTON+ (w/o GMM regularization &amp; mask loss)</td>
<td>0.7602</td>
<td>0.8076</td>
<td>0.1263</td>
<td>3.0735 ± 0.0531</td>
</tr>
<tr>
<td>CP-VTON+ (w/o GMM mask loss)</td>
<td>0.7920</td>
<td>0.8077</td>
<td>0.1231</td>
<td>3.1312 ± 0.0837</td>
</tr>
<tr>
<td>CP-VTON+ (Ours)</td>
<td><strong>0.8425</strong></td>
<td><strong>0.8163</strong></td>
<td>0.1144</td>
<td><strong>3.1048 ± 0.1068</strong></td>
</tr>
</tbody>
</table>
Ablation Study

Reference image
Inshop cloth
CPVTON
Corrected human representation
With $L_{reg}$
Added mask
Add skin label in GMM
Discussions

- 2D transformation can not handle strong 3D deformations.
- Better human parsing is crucial for better try on results
Conclusion

• Proposed a refined image based VTON system, CPVTON+

• Solving issues in previous approaches:
  • Errors in human representation and dataset
  • Network design
  • Loose cost function

• Future work:
  • 3D reconstruction would be use for handle strongly clothing deformations
Project site

• [https://minar09.github.io/cpvtonplus/](https://minar09.github.io/cpvtonplus/)
• [https://github.com/minar09/cp-vton-plus](https://github.com/minar09/cp-vton-plus)
References


