Controllable Garment Image Synthesis Integrated with Frequency Domain Features
–Supplemental Material–

Xinru Liang, Haoran Mo, and Chengying Gao †
Sun Yat-sen University

1. Comparison with Existing Approaches
Figures 1 and 2 show results of our framework compared with FashionGAN [CLGS18], TextureGAN [XSA∗18], MUNIT [HLBK18], ReferenceGAN [LKL∗20], SSSIM [LZSE21] and DiSS [CCC∗23].

2. Ablation Studies

2.1. Effectiveness of FFT-based Generator
We show more comparison results between a FFT-based generator (FcF generator [JZYS23]) and a generator without FFT (co-modulated StyleGAN2 [ZCS∗21]) in Figure 3.

2.2. Frequency Perceptual Loss
We show more comparison results between models with and without the frequency perceptual loss in Figure 4.

2.3. High Receptive Field Perceptual Loss
We validate the effectiveness of high receptive field perceptual loss $L_{hp}$. As shown in Figure 5, it improves the structural details of the synthetic garment images (e.g., details and realism of collar). The quantitative results are shown in Table 1.

2.4. Feature Fusion Module
We also study the effectiveness of different feature fusion modules. Our framework uses attention-based SCFT [LKL∗20] to fuse the sketch feature vector and the texture feature vector from the dual-branch encoder. We also evaluate the performance of directly concatenating the two feature vectors in the channel dimension. Table 1 shows the quantitative comparisons.

3. Diversity of Texture Patterns
Figure 6 shows the diversity results with various textures.

Table 1: Ablation study.

<table>
<thead>
<tr>
<th></th>
<th>FID↓</th>
<th>LPIPS↓</th>
<th>c-FID↓</th>
<th>c-LPIPS↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o $L_{hp}$</td>
<td>31.143</td>
<td>0.264</td>
<td>44.128</td>
<td>0.529</td>
</tr>
<tr>
<td>Direct concatenation</td>
<td>17.389</td>
<td>0.127</td>
<td>33.197</td>
<td>0.373</td>
</tr>
<tr>
<td>Ours</td>
<td>17.146</td>
<td>0.126</td>
<td>31.796</td>
<td>0.371</td>
</tr>
</tbody>
</table>

References

Figure 1: Qualitative comparison with baseline methods. The texture patch of different sizes and the garment sketch in the first and the second columns are used as the inputs.
Figure 2: Qualitative comparison with baseline methods. The texture patch of different sizes and the garment sketch in the first and the second columns are used as the inputs.
Figure 3: Comparisons between methods with and without FFT-based generator.
Figure 4: Comparisons between methods with and without our proposed frequency perceptual loss ($L_{fp}$).
Figure 5: Comparisons between methods with and without high receptive field perceptual loss perceptual loss ($L_{hp}$).
Figure 6: Diversity results with various textures per garment sketch.