SCAM! Transferring humans between images with Semantic Cross Attention
Nicolas Dufour  David Picard  Vicky Kalogeiton

Introduction
Goal: Easily swap subjects in an image.

Approach: Separate the latent space with a segmentation mask.

Prior work: Single style code per image [ParkCVPR19], Single style code per segmentation region [ZhuCVPR20].

Our method: Multiple style codes per semantic region.

Method

Semantic Cross Attention (SCA)

- Semantically constrained attention on pixels and latents.
- Pixel or latent can only attend same semantic label latents.

\[
\text{SCA}(\mathbf{x}, \mathbf{y}) = \left( \mathbf{R}(\mathbf{x} + \mathbf{y} - \mathbf{v}) \right)_y
\]

Semantic Cross Attention Modulation (SCAM)

- Multiple latents per semantic region.
- More representative power for coarse semantic regions.
- Discover unsupervised semantic structures inside the labels.

Training setup
- Hinge GAN loss + L1 loss + Perceptual Loss.
- PatchGAN discriminator.

Semantic Attention Transformer (SAT)

SAT Operation: transformer-like operation with SCA.

SAT encode/decode data in a semantically constrained fashion.

Unsupervised Semantic Knowledge

Pixels and group of latents have an assigned semantic label.
- Pixels or latents can only attend same semantic label latents.

SAT encode/decode data in a semantically constrained fashion.

Discover unsupervised semantic structures inside the labels.

Subject Transfer Results

Quantitative Results

<table>
<thead>
<tr>
<th>Method</th>
<th>iDesigner</th>
<th>CelebAMask-HQ</th>
<th>ADE20K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR ↑</td>
<td>R-FID ↓</td>
<td>S-FID ↓</td>
</tr>
<tr>
<td>SPADE [ParkCVPR19]</td>
<td>10.4</td>
<td>66.7</td>
<td>67.5</td>
</tr>
<tr>
<td>CLAIDE [TanTPAMI21]</td>
<td>11.1</td>
<td>45.4</td>
<td>46.1</td>
</tr>
<tr>
<td>SEAN-CLAIDE [TanTPAMI21]</td>
<td>15.3</td>
<td>48.4</td>
<td>56.1</td>
</tr>
<tr>
<td>INADE [TanCVPR21]</td>
<td>12.0</td>
<td>33.8</td>
<td>33.9</td>
</tr>
<tr>
<td>SEAN [ZhuCVPR20]</td>
<td>14.8</td>
<td>53.5</td>
<td>56.7</td>
</tr>
<tr>
<td>SCAM (Ours)</td>
<td>21.4</td>
<td>13.2</td>
<td>26.9</td>
</tr>
</tbody>
</table>