Motion Supervised Co-Part Segmentation

Aliaksandr Siarohin            Subhankar Roy            Stéphane Lathuilière
Sergey Tulyakov            Elisa Ricci            Nicu Sebe

International Conference on Pattern Recognition, 2020
Co-part Segmentation

- Given an image of an object, the goal is to identify segments corresponding to different object parts within the single object.

- For example, in a human body the relevant object parts correspond to hands, legs, torso and head.
Co-part Segmentation

Prior works:

• Collins et. al. (ECCV’18) used Non-negative Matrix Factorization on ConvNet features. Requires costly optimisation at inference time.

• Xu et. al. (ICLR’19) exploited hierarchical structure and dynamical structures from unlabelled videos. Assumes externally pre-computed motion information available during training.

• Hung et.al. (CVPR’19) proposed a self-supervised approach enforcing geometric, equivariance and semantic consistency constraints, using static images. Intricate interplay of several losses.

Our work:

• Exploit large collection of unlabelled videos using self-supervised learning objective.

• Learning to transfer motion from the unlabelled videos with the purpose of obtaining segments.
Motion Supervised Co-Part Segmentation

Segmentation Module

Encoder

Affine Motion Parameters

Segment Motion

Reconstruction Module

Background Visibility

Reconstructed Target Frame

Source Frame

Target Frame
Predicted Segmentation Masks on VoxCeleb

Input | DFF (ECCV’ 18) | SCOPS (CVPR’ 19) | Ours

--- | --- | --- | ---

![Input Image](image1.png) | ![DFF Image](image2.png) | ![SCOPS Image](image3.png) | ![Ours Image](image4.png)

![Input Image](image5.png) | ![DFF Image](image6.png) | ![SCOPS Image](image7.png) | ![Ours Image](image8.png)

![Input Image](image9.png) | ![DFF Image](image10.png) | ![SCOPS Image](image11.png) | ![Ours Image](image12.png)
<table>
<thead>
<tr>
<th>Input</th>
<th>DFF (ECCV’ 18)</th>
<th>SCOPS (CVPR’ 19)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Input" /></td>
<td><img src="image2.png" alt="DFF" /></td>
<td><img src="image3.png" alt="SCOPS" /></td>
<td><img src="image4.png" alt="Ours" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Input" /></td>
<td><img src="image6.png" alt="DFF" /></td>
<td><img src="image7.png" alt="SCOPS" /></td>
<td><img src="image8.png" alt="Ours" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Input" /></td>
<td><img src="image10.png" alt="DFF" /></td>
<td><img src="image11.png" alt="SCOPS" /></td>
<td><img src="image12.png" alt="Ours" /></td>
</tr>
</tbody>
</table>
State-of-the-art Comparison: Quantitative Results

Evaluation Protocol:

• Landmark regression MAE (lower the better)

• Foreground segmentation IoU (higher the better)

<table>
<thead>
<tr>
<th>Method</th>
<th>Tai-Chi-HD</th>
<th></th>
<th>VoxCeleb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE ↓</td>
<td>IoU ↑</td>
<td>MAE ↓</td>
</tr>
<tr>
<td>DFF</td>
<td>494.48</td>
<td>-</td>
<td>1254.25</td>
</tr>
<tr>
<td>SCOPS</td>
<td>411.38</td>
<td>0.5485</td>
<td>663.04</td>
</tr>
<tr>
<td>Ours</td>
<td>389.78</td>
<td>0.7686</td>
<td>424.96</td>
</tr>
</tbody>
</table>
Application: Video Part Swapping

Hair Swap

Beard Swap

Lips Swap

Eyes Swap
Summary

Training:
Self-supervised training on video frames

Testing:
Inference on a single image
Motion Supervised Co-Part Segmentation

International Conference on Pattern Recognition, 2020

Github