Supplementary Material for Video Propagation Networks

Varun Jampani\textsuperscript{1}, Raghudeep Gadde\textsuperscript{1,2} and Peter V. Gehler\textsuperscript{1,2}
\textsuperscript{1}Max Planck Institute for Intelligent Systems, Tübingen, Germany
\textsuperscript{2}Bernstein Center for Computational Neuroscience, Tübingen, Germany
\{varun.jampani,raghudeep.gadde,peter.gehler\}@tuebingen.mpg.de

1. Parameters and Additional Results

In this supplementary, we present experiment protocols and additional qualitative results for experiments on video object segmentation, semantic video segmentation and video color propagation. Table 1 shows the feature scales and other parameters used in different experiments. Figures 1, 2 show some qualitative results on video object segmentation with some failure cases in Fig. 3. Figure 4 shows some qualitative results on semantic video segmentation and Fig. 5 shows results on video color propagation.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Feature Type</th>
<th>Feature Scale-1, (\Lambda_a)</th>
<th>Feature Scale-2, (\Lambda_b)</th>
<th>(\alpha)</th>
<th>Input Frames</th>
<th>Loss Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Object Segmentation</td>
<td>((x, y, Y, Cb, Cr, t))</td>
<td>(0.02,0.02,0.07,0.4,0.01)</td>
<td>(0.03,0.03,0.09,0.5,0.2)</td>
<td>0.5</td>
<td>9</td>
<td>Logistic</td>
</tr>
<tr>
<td>Semantic Video Segmentation</td>
<td>((x, y, R, G, B, t))</td>
<td>(0.08,0.08,0.2,0.0.2,0.04)</td>
<td>(0.11,0.11,0.2,0.2,0.2)</td>
<td>0.5</td>
<td>3</td>
<td>Logistic</td>
</tr>
<tr>
<td>with CNN1 \cite{5}-NoFlow</td>
<td>((x+u_x, y+u_y, R, G, B, t))</td>
<td>(0.11,0.11,0.14,0.14,0.14)</td>
<td>(0.08,0.08,0.12,0.12,0.01)</td>
<td>0.65</td>
<td>3</td>
<td>Logistic</td>
</tr>
<tr>
<td>with CNN1 \cite{5}-Flow</td>
<td>((x+u_x, y+u_y, R, G, B, t))</td>
<td>(0.08,0.08,0.2,0.2,0.2)</td>
<td>(0.09,0.09,0.25,0.25,0.25)</td>
<td>0.5</td>
<td>4</td>
<td>Logistic</td>
</tr>
<tr>
<td>Video Color Propagation</td>
<td>((x, y, I, t))</td>
<td>(0.04,0.04,0.2,0.04)</td>
<td>No second kernel</td>
<td>1</td>
<td>4</td>
<td>MSE</td>
</tr>
</tbody>
</table>

Table 1. Experiment Protocols. Experiment protocols for the different experiments presented in this work. **Feature Types**: Feature spaces used for the bilateral convolutions, with position \((x, y)\) and color \((R, G, B\) or \(Y, Cb, Cr\)) features \(\in [0, 255]\). \(u_x, u_y\) denotes optical flow with respect to the present frame and \(I\) denotes grayscale intensity. **Feature Scales** \((\Lambda_a, \Lambda_b)\): Validated scales for the features used. \(\alpha\): Exponential time decay for the input frames. **Input Frames**: Number of input frames for VPN. **Loss Type**: Type of loss used for back-propagation. “MSE” corresponds to Euclidean mean squared error loss and “Logistic” corresponds to multinomial logistic loss.
Figure 1. **Video Object Segmentation.** Shown are the different frames in example videos with the corresponding ground truth (GT) masks, predictions from BVS [2], OFL [4], VPN (VPN-Stage2) and VPN-DLab (VPN-DeepLab) models.
Figure 2. **Video Object Segmentation.** Shown are the different frames in example videos with the corresponding ground truth (GT) masks, predictions from BVS [2], OFL [4], VPN (VPN-Stage2) and VPN-DLab (VPN-DeepLab) models.
Figure 3. **Failure Cases for Video Object Segmentation.** Shown are the different frames in example videos with the corresponding ground truth (GT) masks, predictions from BVS [2], OFL [4], VPN (VPN-Stage2) and VPN-DLab (VPN-DeepLab) models.
Figure 4. **Semantic Video Segmentation.** Input video frames and the corresponding ground truth (GT) segmentation together with the predictions of CNN \[5\] and with VPN-Flow.
Figure 5. Video Color Propagation. Input grayscale video frames and corresponding ground-truth (GT) color images together with color predictions of Levin et al. [1] and VPN-Stage1 models.
References


