1. Introduction

Goal
- Generate a video of human motion from a different view-point

Approach
- Motion learning through texture-less representation

Applications
- data augmentation

2. VA-LSTM

- Target-view learning through Conv-LSTM
- Incorporates the texture-less representation

3. View Adaptive Network (VA-Net)

- VA-LSTM for target-view feature approximation
- Two-stage pipeline for synthesis and refinement

Motion network:
- Separate foreground and background synthesis
- Foreground feature estimation using the teacher-student approach (see Sec. 3)

Refinement network:
- UNet-like network to retain spatial information
- Use explicit edge-loss (Sobel-filter) penalization to encourage high-frequency details:
  \[ L_e = \| C_x \cdot \hat{x}^i - C_x \cdot \hat{x}^j \| + \| C_y \cdot \hat{x}^i - C_y \cdot \hat{x}^j \| \]

4. Results

Comparison with state-of-the-art methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>SSIM</th>
<th>M-SSIM</th>
<th>PSNR</th>
<th>M-PSNR</th>
<th>L_2</th>
<th>PCK</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>View-LSTM</td>
<td>d^j, d^i</td>
<td>.821</td>
<td>.972</td>
<td>23.18</td>
<td>29.70</td>
<td>.14</td>
<td>.14</td>
<td>88.1</td>
<td>12.8</td>
</tr>
<tr>
<td>VA-LSTM</td>
<td>d^j, d^i</td>
<td>.830</td>
<td>.976</td>
<td>22.78</td>
<td>30.89</td>
<td>.16</td>
<td>.16</td>
<td>92.5</td>
<td>20.6</td>
</tr>
</tbody>
</table>

References