Video Frame Synthesis using Deep Voxel Flow

Ziwei Liu\textsuperscript{1} \quad Raymond A. Yeh\textsuperscript{2} \quad Xiaoou Tang\textsuperscript{1} \\
Yiming Liu\textsuperscript{3} \quad Aseem Agarwala\textsuperscript{4}

1. The Chinese University of Hong Kong \\
2. University of Illinois at Urbana-Champaign \\
3. Pony AI \\
4. Google Inc.
Problem

Video interpolation/extrapolation
Challenges

1. Complex motion (camera motion & scene motion)
2. High-res images (1280 * 720)
Previous Attempts

Hallucination-based Methods
[Ranzato et al.], [Srivastava et al.], [Mathieu et al.]

Flow-based Methods
Voxel Flow

symmetric bi-directional flows
Voxel Flow

selection mask between frames
Voxel Flow

differentiable bilinear sampling
Deep Voxel Flow

Input Video → Convolutional Encoder-Decoder → Voxel Flow → Synthesized Frame
Multi-scale Deep Voxel Flow

Convolutional Encoder-Decoder (scale 2)

Input Video (scale 2)

Convolutional Encoder-Decoder (scale 0)

Input Video (scale 0)

Convolutional Encoder-Decoder (scale 1)

Input Video (scale 1)

- Convolution
- Max Pooling
- Deconvolution
- Skip Connection
Multi-scale Deep Voxel Flow

(a) Voxel Flow
(b) Multi-scale Voxel Flow
Comparisons

- UCF-101
Comparisons

- UCF-101
Comparisons

• UCF-101
Comparisons

- KITTI
Comparisons

- KITTI
Comparisons

- KITTI
## Self-supervised Learning

<table>
<thead>
<tr>
<th>Method</th>
<th>EPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD Flow [3]</td>
<td>12.4</td>
</tr>
<tr>
<td>EpicFlow [22]</td>
<td>3.8</td>
</tr>
<tr>
<td>Ours (w/o ft.)</td>
<td>14.6</td>
</tr>
<tr>
<td>Ours</td>
<td>9.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>39.1</td>
</tr>
<tr>
<td>Unsup. Video [30]</td>
<td>43.8</td>
</tr>
<tr>
<td>ImageNet [14]</td>
<td>63.3</td>
</tr>
<tr>
<td>Ours (w/o ft.)</td>
<td>48.7</td>
</tr>
<tr>
<td>Ours</td>
<td>52.4</td>
</tr>
</tbody>
</table>

Flow estimation  
Action Recognition
Spatio-temporal Coherence
User Study

Diagonal-split Comparison

Method 1 \ Method 2
User Study

Preference Percentage

EpicFlow  Ground Truth  Ours

Booth  Dog  Kids  Park  Throw  Street  Sky  Baseball  Subway  Balloon
Real-life Applications

Video Frame Synthesis using Deep Voxel Flow

Ziwei Liu¹, Raymond Yeh², Xiaou Tang¹, Yiming Liu³, Aseem Agarwala³

¹The Chinese University of Hong Kong
²University of Illinois at Urbana-Champaign
³Google
Conclusions

• Unified motion representation

• End-to-end learning system

• Long-range dependencies in the future
Thanks!

Project Page: https://liuziwei7.github.io/projects/VoxelFlow