Deep Video Frame Interpolation using Cyclic Frame Generation
Yu-Lun Liu\textsuperscript{1,2,3}, Yi-Tung Liao\textsuperscript{1,2}, Yen-Yu Lin\textsuperscript{1}, Yung-Yu Chuang\textsuperscript{1,2}
\textsuperscript{1}Academia Sinica, \textsuperscript{2}National Taiwan University, \textsuperscript{3}MediaTek
Code Available at: https://github.com/alex04072000/CyclicGen

Introduction

• **Goal:** Predict the intermediate frame between two consecutive frames

• **Challenges:**
  - Conventional methods → computationally expensive
  - CNN-based methods → artifacts and over-smoothed results

A two-stage training procedure

- **Stage 1:**
  - Pre-train the baseline model
  - Reconstruction loss
  \[ L_c = \sum_{i=1}^{N} \left( \| f(t_{i+1}) - t_{i} \| _{1} \right) \]

- **Stage 2:**
  - Include the cycle consistency loss
  \[ L = L_c + \sum_{i=1}^{N} \| f(t_{i+1}) - t_{i} \| _{1} \]
  - Motion linearity loss
  \[ L_m = \sum_{i=1}^{N} \left( \| f(t_{i+1} - t_{i}) - 2 \cdot f(t_{i+1} - t_{i}) \| _{1} \right) \]
  - Edge-guided training

Visual comparisons

Ablation studies on UCF dataset

- The introduced components help video interpolation

<table>
<thead>
<tr>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Ours)</td>
<td>35.98</td>
</tr>
<tr>
<td>+ Cycle</td>
<td>36.37</td>
</tr>
<tr>
<td>+ Cycle + Motion</td>
<td>36.38</td>
</tr>
<tr>
<td>+ Cycle + Edge</td>
<td>36.36</td>
</tr>
<tr>
<td>full model</td>
<td>36.36</td>
</tr>
</tbody>
</table>

- Cycle consistency loss improves the robustness to few training data

Comparison with SoTAs

- On UCF-101 dataset and a high-quality video *See You Again*

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVF</td>
<td>35.89</td>
<td>0.945</td>
</tr>
<tr>
<td>SepCom</td>
<td>36.49</td>
<td>0.950</td>
</tr>
<tr>
<td>Ours</td>
<td>36.96</td>
<td>0.953</td>
</tr>
</tbody>
</table>

- On Middlebury dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVF</td>
<td>35.89</td>
<td>0.945</td>
</tr>
<tr>
<td>SepCom</td>
<td>36.49</td>
<td>0.950</td>
</tr>
<tr>
<td>Ours</td>
<td>36.96</td>
<td>0.953</td>
</tr>
</tbody>
</table>

Conclusion

• We present a novel loss, the cycle consistency loss, which
  - can be integrated with existing video frame interpolation methods and trained end-to-end
  - synthesizes more plausible frames possessing similar characteristics with the original frames

• We propose two extensions, motion linearity loss and edge guided training, that
  - regularize the training procedure
  - further improve model performance

• The proposed approach better utilizes the training data, not only enhancing the interpolation results, but also reaching better performance with less training data.

References: