Unsupervised Video Summarization with Adversarial LSTM Networks
Behrooz Mahasseni, Michael Lam and Sinisa Todorovic

Problem: Video Summarization
- Goal: Select a sparse set of key frames

Key Idea: Adversarial Training
- The summary should have the same content as the original → GAN
- Sparse selection of frames should represent well the original → VAE
- Contribution:
  - Unsupervised Summarization with VAE + GAN

Adversarial Video Summarizer

Regularized Training

Accuracy in [%] on SumMe & TVSum

<table>
<thead>
<tr>
<th>Method</th>
<th>SumMe</th>
<th>TVSum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avila et al. '11</td>
<td>33.7</td>
<td>-</td>
</tr>
<tr>
<td>Li et al. '10</td>
<td>26.6</td>
<td>-</td>
</tr>
<tr>
<td>Lhosla et al. '13</td>
<td>-</td>
<td>36.0</td>
</tr>
<tr>
<td>Song et al. '15</td>
<td>26.6</td>
<td>50</td>
</tr>
<tr>
<td>Zhao et al. '14</td>
<td>-</td>
<td>46.0</td>
</tr>
<tr>
<td>Ours</td>
<td>39.1</td>
<td>51.7</td>
</tr>
</tbody>
</table>

Method
- Gygli et al. '15: 39.7
- Zhang et al. '16: 40.9
- Gygli et al. '14: 39.5
- Zhang et al. '16: 38.6
- Ours: 39.1

(F-score for variations of σ)

Acknowledgments: DARPA XAI & NSF RI1302700.