EVAL: Explainable Video Anomaly Localization
Ashish Singh, Michael Jones, Erik G. Learned-Miller
University of Massachusetts Amherst, Mitsubishi Electric Research Laboratories (MERL)

Motivation
- We are interested in the problem of spatio-temporal localization of anomalous activities in videos of a given scene.
- Our aim is to design an interpretable, robust and accurate anomaly detection system.
- We are motivated by how humans are able to detect changes in a given scene after exposure to it by decomposing it into specific objects and their corresponding motion patterns.

Method
- Our method consists of three stages: Attribute Learning, Model Building, and Anomaly Detection.

- **Attribute Learning:** Attributes are human-understandable features of the objects present in a video volume. They give the system a high-level understanding of what is going on in a scene. In this work, the attributes we use are (a) Object Classes and (b) Motion directions, speeds and fraction of stationary pixels. We train neural networks to estimate these attributes given a video volume. The networks are only trained once (not trained for each different scene).

- **Model Building:** Given nominal video of a scene, learn a set of exemplars (attribute embeddings) for each different spatial region of the scene. Exemplar selection uses a simple greedy algorithm.

- **Anomaly Detection:** Given test video from the same scene, compute attribute embeddings for each video volume and find nearest neighbor exemplar from the nominal model. Distance to nearest neighbor is the anomaly score.

### Experimental Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>Avenue</th>
<th>Shanghai/Tech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ionescu (2019)</td>
<td>15.77</td>
<td>27.07</td>
</tr>
<tr>
<td>Ramachandra (2020)</td>
<td>30.50</td>
<td>80.90</td>
</tr>
<tr>
<td>Georgescu (2021)</td>
<td>57.00</td>
<td>58.30</td>
</tr>
<tr>
<td>Liu (2018)</td>
<td>19.59</td>
<td>56.01</td>
</tr>
<tr>
<td>Liu (2021)</td>
<td>41.05</td>
<td>86.18</td>
</tr>
<tr>
<td>Georgescu + Ristea (2021)</td>
<td>65.99</td>
<td>64.91</td>
</tr>
</tbody>
</table>

- **Our Method**

### Explanation Visualization

- Frame 12 of ShanghaiTech. The bounding box indicates spatiotemporal region containing a ground truth anomaly (cyclist on sidewalk).
- Frame 447 of Avenue. The bounding box indicates spatiotemporal region containing an anomaly (motorcycle on sidewalk).
- Frame 113 of UCSD Ped2 Test001. The bounding box indicates a particularly spatiotemporal region containing a ground truth anomaly (cyclist on sidewalk).

- **Stage 1: Attribute Learning**
- **Stage 2: Model Building**
- **Stage 3: Anomaly Detection**