
Mamshad N Rizve, Ugur Demir, Praveen Tirupattur, Aayush J Rana, Kevin Duarte, Ishan Dave, Yogesh S Rawat, Mubarak Shah
Introduction

Detect activities in untrimmed security videos

- Human and Vehicles
- Activity types
  - Single actors
  - Interaction between actors
  - Actor-object interactions
Challenges

• Untrimmed nature
• Multiple activities
• Varying length of activities
• Multiple actors
Challenges

• Untrimmed nature
• Multiple activities
• Varying length of activities
• Multiple actors
• Multiple scales
Motivations

• Region proposal based approach [1, 2]
  • Scaling issue with videos
  • Multiple actors
    • How to pair?

• Object detection [3]
  • Time consuming
  • Multiple actors
    • How to pair?

Approach

• A two-stage process
  • Detect activity tubelets from long untrimmed videos
  • Recognize activities in the detected tubelets

• Encoder-decoder architecture
  • No region proposal

• Video level detection
  • No object detection
Our Approach
Our Approach
Our Approach
Our Approach
Our Approach
Our Approach
Our Approach
Our Approach
Our Approach
Datasets

• VIRAT [1]
  • 64 (2.47 hours) videos for training
  • 54 videos (1.93 hours) for validation
  • 40 activities

• MEVA [2]
  • 1056 videos (88 hours)
  • 37 activities

[2] Kitware inc, the multiview extended video with activities (meva) dataset.
Quantitative Results (VIRAT Dataset)

<table>
<thead>
<tr>
<th>Team</th>
<th>$P_{\text{miss@0.15}}^{T_{FA}}$</th>
<th>$P_{\text{miss@0.15}}^{R_{FA}}$</th>
<th>AUDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraunhofer</td>
<td>0.7747</td>
<td>0.8474</td>
<td>0.8270</td>
</tr>
<tr>
<td>vireoJD-MM</td>
<td>0.5482</td>
<td>0.7284</td>
<td>0.6012</td>
</tr>
<tr>
<td>NTT_CQUPT</td>
<td>0.5112</td>
<td>0.8725</td>
<td>0.6005</td>
</tr>
<tr>
<td>Hitachi</td>
<td>0.5099</td>
<td>0.8240</td>
<td>0.5988</td>
</tr>
<tr>
<td>BUPT-MCPRL</td>
<td>0.4328</td>
<td>0.7491</td>
<td>0.5240</td>
</tr>
<tr>
<td>MUDSML [20]</td>
<td>0.3915</td>
<td>0.7979</td>
<td><strong>0.4840</strong></td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>0.3858</strong></td>
<td><strong>0.7022</strong></td>
<td>0.4909</td>
</tr>
</tbody>
</table>
Quantitative Results (MEVA Dataset)

<table>
<thead>
<tr>
<th>Team</th>
<th>AUDC</th>
<th>$P_{<a href="mailto:miss@0.15TFA">miss@0.15TFA</a>}$</th>
<th>Processing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team-Vision</td>
<td>0.717</td>
<td>0.776</td>
<td>0.793</td>
</tr>
<tr>
<td>IBM-MIT-Purdue</td>
<td>0.641</td>
<td>0.733</td>
<td>0.272</td>
</tr>
<tr>
<td>Edge-Intelligence</td>
<td>0.628</td>
<td>0.754</td>
<td>0.939</td>
</tr>
<tr>
<td>INF</td>
<td>0.489</td>
<td>0.559</td>
<td>0.646</td>
</tr>
<tr>
<td>UMD [9]</td>
<td>0.475</td>
<td>0.544</td>
<td>0.725</td>
</tr>
<tr>
<td>Ours</td>
<td>0.438</td>
<td>0.523</td>
<td>0.362</td>
</tr>
</tbody>
</table>
Qualitative Results (Localization)
Thank You

Project Page:
https://tinyurl.com/y6gv8dpl