UBnormal: New Benchmark for
Supervised Open-Set Video Anomaly Detection – Supplementary

Andra Acsintoaie\textsuperscript{1,*}, Andrei Florescu\textsuperscript{1,*}, Mariana-Iuliana Georgescu\textsuperscript{1,2,3,*}, Tudor Mara\textsuperscript{3,*}, Paul Sumedrea\textsuperscript{1,*}, Radu Tudor Ionescu\textsuperscript{1,3,⋄}, Fahad Shahbaz Khan\textsuperscript{2,4}, Mubarak Shah\textsuperscript{5}

\textsuperscript{1}University of Bucharest, Romania, \textsuperscript{2}MBZ University of Artificial Intelligence, UAE
\textsuperscript{3}SecurifAI, Romania, \textsuperscript{4}Linköping University, Sweden, \textsuperscript{5}University of Central Florida, US

\textsuperscript{*}equal contribution, \textsuperscript{⋄}corresponding author

1. Discussion

In Figure 1, we present the scenes of our data set. We have 22 outdoor scenes and 7 indoor scenes. There are various illumination conditions, corresponding to different weather conditions (\textit{e.g.} sunny, cloudy, foggy) and different day times (\textit{e.g.} day, sunset, night). The last image (at the bottom right) illustrates one of the existing scenes with the fog effect.

We present the animated characters from the UBnormal data set in Figure 2. There are 19 unique characters, but in order to increase the variety of the data set, we augment the characters by changing the color of their hair and clothes.

In Figure 3, we show the 5 object categories (excluding people) from UBnormal. We variate the colors of objects to increase diversity, as for the animated characters.

In Figure 4, we illustrate an example for each abnormal action category from the UBnormal benchmark. There
2. Statistics

In Table 1, we report several statistics about the UBnormal data set. The number of videos in the data set is 543, with 268 videos for training, 64 for validation and the remaining 211 for testing. There are a total of 660 anomalies divided into 195, 76 and 389 for training, validation and test, respectively. The total number of abnormal frames is 89,015, with the largest fraction of 49,850 frames belonging to the test set. There are 142,107 abnormal regions in the data set, which accounts for an average of 1.60 normal regions per abnormal frame. The average number of objects per frame is 4.44. The overall length of our data set is 132 minutes.

- **Figure 2.** Various characters animating the scenes in the UBnormal benchmark. There are 19 unique characters that serve as seeds for generating people with different hair color and clothes. Best viewed in color.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>#anomalies</td>
<td>195</td>
<td>76</td>
<td>389</td>
<td>660</td>
</tr>
<tr>
<td>#abnormal frames</td>
<td>25,227</td>
<td>13,938</td>
<td>49,850</td>
<td>89,015</td>
</tr>
<tr>
<td>#normal frames</td>
<td>90,860</td>
<td>14,237</td>
<td>42,790</td>
<td>147,887</td>
</tr>
<tr>
<td>#abnormal minutes</td>
<td>14.02</td>
<td>7.74</td>
<td>27.69</td>
<td>49.45</td>
</tr>
<tr>
<td>#normal minutes</td>
<td>50.48</td>
<td>7.91</td>
<td>23.77</td>
<td>82.16</td>
</tr>
<tr>
<td>#videos</td>
<td>268</td>
<td>64</td>
<td>211</td>
<td>543</td>
</tr>
<tr>
<td>#abnormal regions</td>
<td>38,048</td>
<td>21,321</td>
<td>82,738</td>
<td>142,107</td>
</tr>
<tr>
<td>#unique objects</td>
<td>1,443</td>
<td>351</td>
<td>1,114</td>
<td>2,908</td>
</tr>
<tr>
<td>avg. objects / frame</td>
<td>4.58</td>
<td>4.53</td>
<td>4.24</td>
<td>4.44</td>
</tr>
</tbody>
</table>

Table 1. Detailed statistics for the UBnormal data set. Our videos are generated at 30 FPS.

are 20 anomaly types related to objects (e.g. people, cars) and two non-object anomaly types (fire, smoke). The figure shows the diversity of our anomalies.
Figure 3. Various objects animating the scenes in the UBnormal benchmark. There are 5 object categories besides people. To increase variation, we apply different colors to the objects. Best viewed in color.
Figure 4. The abnormal action categories from UBnormal. The abnormal objects are emphasized through a red contour. To improve readability, we apply a magnifying effect to smaller objects. Best viewed in color.