Introduction

- Dystopian censorship as portrayed in *White Christmas*
- Current video object tracking plus segmentation algorithms are very slow\(^{[13]}\)
- Multiple object trackers are fast, operating at over 30 FPS\(^{[3]}\)

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM 1 pixel FOV</td>
<td>54.17%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SVM 5 pixel FOV</td>
<td>56.05% (3 image overlap, 89.3%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SVM 15 pixel FOV</td>
<td>51.20%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-Layer Neural Network</td>
<td>61.38%</td>
<td>0.6012</td>
<td>0.6315</td>
<td>106 FPS</td>
</tr>
<tr>
<td>FCC Network w/DenseNet</td>
<td>81.10%</td>
<td>0.7008</td>
<td>0.8341</td>
<td>35 FPS</td>
</tr>
</tbody>
</table>

Test Set Performance

Fig 1: Real-time censorship from *Black Mirror* episode *White Christmas* (Netflix, 2014)

Fig 2: Left – Image from MS-COCO (upper) with annotated GT (below). Right – Image from MOT2016 (upper) with annotated GT (below)

Fig 3: 3 image overfitted example on 5 pixel FOV SVM with 89% accuracy. Pixel-by-pixel evaluation.

Fig 4: Left two images - average example of input/output pair for wide image. Right two - average input/output for narrow. Demonstrates ‘blobbingness’ of prediction.

Fig 5: Left 4 images – FCC mask of man from sequence 5 frames apart. Right 4 images – FCC mask of woman from sequence 5 frames apart

Fig 6: Full pipeline example. Top is the object tracking. Below shows numbered mask examples from the frames from the FCC output.

Fig 7: Another full pipeline example. Left is shopping mall single frame, while right is two selected instances from the frame.

Discussion

- SVM completely ineffective, likely not a good domain for application.
- Neural network had weak performance, with no person-like characteristics. However, clearly did learn general location.
- FCC Network had strong performance, with 81.1% binary accuracy.
- Failure cases of FCC still mostly robust to application space. This can be seen in high recall value of 0.834.
- High speed can be observed in all models.

Conclusion

- Overall, a strong pipeline for video object tracking plus segmentation, especially if precision unneeded.
- Huge speed improvements allow for real-time application.
- Real-time camera input as well as post-mask blurring needed to recreate *White Christmas.*