**Motivation & Contribution**
- The IPT dataset aims to facilitate the application of modern computer vision methods to privacy-critical applications such as:
  - Ambient Assisted Living (AAL)
  - Security
- Privacy is protected using a depth sensor as an anonymizing imaging method. In many cases behavior analysis does not need to know who is acting but rather what they are doing.
- We present:
  - A new public dataset for identity preserved human detection and tracking in 2d or 3d.
  - Baseline results for person detection and tracking.

**Dataset Overview**
- 10 sequences featuring:
  - Indoor environments
  - Both scripted and natural behavior
  - Frequent person-person and person-object occlusions
- 72k frames, static depth sensor at 640×480 resolution, 30 fps.
- Split by sequence into training (70%), validation (20%) and test sets (10%).

```
0 1 2 3 4 5 6 7 8
```

Representative sample frames for each recorded sequence

**Labeling**
- 3d location of actors
  - Consistent actor ID across all sequences
- 2d bounding boxes
  - Inferred from 3d locations

**Results**
- 2d detection baseline using a YOLOv3 model.
- Improved model performance using an efficient background model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Average Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Validation</td>
</tr>
<tr>
<td>YOLOv3</td>
<td>77.2%</td>
</tr>
<tr>
<td>YOLOv3 + BG Model</td>
<td>85.2%</td>
</tr>
<tr>
<td>YOLOv3 Tiny</td>
<td>71.9%</td>
</tr>
</tbody>
</table>

Baseline person detection scores on validation and test set

- 2d tracking baseline using the SORT and DeepSORT tracking algorithm.
- Evaluation using full MOT Challenge metrics

<table>
<thead>
<tr>
<th>Mode</th>
<th>MOT Accuracy</th>
<th>Identity F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORT</td>
<td>76.6%</td>
<td>14.5%</td>
</tr>
<tr>
<td>DeepSORT</td>
<td>77.2%</td>
<td>17.5%</td>
</tr>
<tr>
<td>DeepSORT Short Tracks</td>
<td>77.3%</td>
<td>58.8%</td>
</tr>
</tbody>
</table>

Baseline tracking results using MOT Challenge metrics

**Conclusion**
- We presented the IPT dataset and demonstrated baseline human detection and tracking results.
- The depth modality can serve as an alternate imaging mode in privacy-sensitive applications.

**Future work**
- Future versions of IPT will increase in size, improve instance labeling and add annotations for human interactions.
- Multimodal behavior analysis based on depth/thermal fusion.