3D-MAN: 3D Multi-frame Attention Network for Object Detection

Supplementary Material

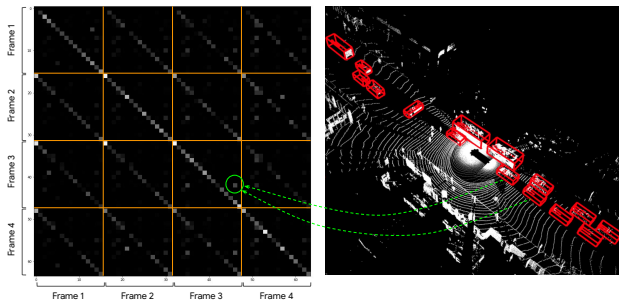## 1. Analysis

1.1. Attention weights

Figure 1. Left: Visualization of the attention matrix computed during the alignment stage. Right: Visualization of the corresponding scene and detections from our model. Each row and column corresponds to a detection proposal for a given frame. The rows and columns of the matrix have been reordered so as to maintain the same ordering of objects for each frame. The matrix has also been truncated to remove proposals with low confidence scores. The diagonal bands suggests that attention matrix of the alignment stage is able to relate the same objects across frames. The off-diagonal elements capture interactions: one of these interactions is visualized by the dotted line between the two diagrams.

Attention networks are used in the multi-view alignment module that relate all the previous frames to the new frames. Since we use all stored proposals from all past frames, the multi-view alignment module performs attention between all pairs of past frames (Figure 1). We find that the attention network generally focuses on the same objects across different frames, suggesting that it is able to relate objects across frames.