Supplementary Material : Occlusion-Net: 2D/3D Occluded Keypoint Localization Using Graph Networks

N Dinesh Reddy Minh Vo Srinivasa G. Narasimhan
Carnegie Mellon University
{dnarapur,mpvo,srinivas}@cs.cmu.edu

a: Intersection-1 b: Intersection-2 c: Intersection-3 d: Intersection-4 e: Intersection-5

Figure 1: The top column consists of Google map view of the intersection used to capture the data. While the bottom row consists of snapshots from the cameras captured in the intersection.

1. Extended CarFusion Dataset

We show the Google map view of the intersections, which were used for capturing the data in the Figure 1 to evaluate the algorithm. The training for the dataset has been computed from the images captured in the Intersection-{1-4}, while the testing for the dataset has been computed from the images sampled from the Intersection-5. Some sample image from each intersection is shown in the bottom of the google map view. We plan on releasing the dataset for further research in the direction of Multi-View data for different tasks like keypoint detection, segmentation etc.

2. Multi-Car Qualitative Results

We extend the teaser images to different intersections and are plotted in the Figure 2. The images show challenging occlusion with different occlusion scenarios across the extended carfusion dataset. We observe that the 2D-KGNN performs well in occlusion situations compared to image based methods. Different colors in this figure depict different cars.

3. Ablation Analysis

We supplement the comparative qualitative results proposed in the paper for each category of occlusion. Figure 3 depicts the comparison for the self-occulsion category. Similarly, Figure 4, 5 and 6 correspond to the vehicle occluding car, other objects occluding car and truncated cars respectively. We observe that with increasing occlusions in the object, the graph based methods are out-performing the image based Maskrcnn method. We further show a conanical-3D reconstruction of each of cars. Observe the variance in occlusion and shape of the 3D reconstucted points.

4. Multi-View Detections

Figure 7 shows the results on different cameras over different time instances. We show 4 cameras capturing the data from different corners of the intersection with minimal field-of-view intersecting.

5. Failure Analysis and Future Work

In Figure 8, we observe that the detected bounding box plays a major role in the accuracy of the overall detection of the algorithm. Bounding box covering only partial cars generally lead to erroneous keypoint localization of the occluded points. Another major issue is the overlapping cars creating a duplicate hypothesis of the visible keypoints leading to distorted shape reconstruction.

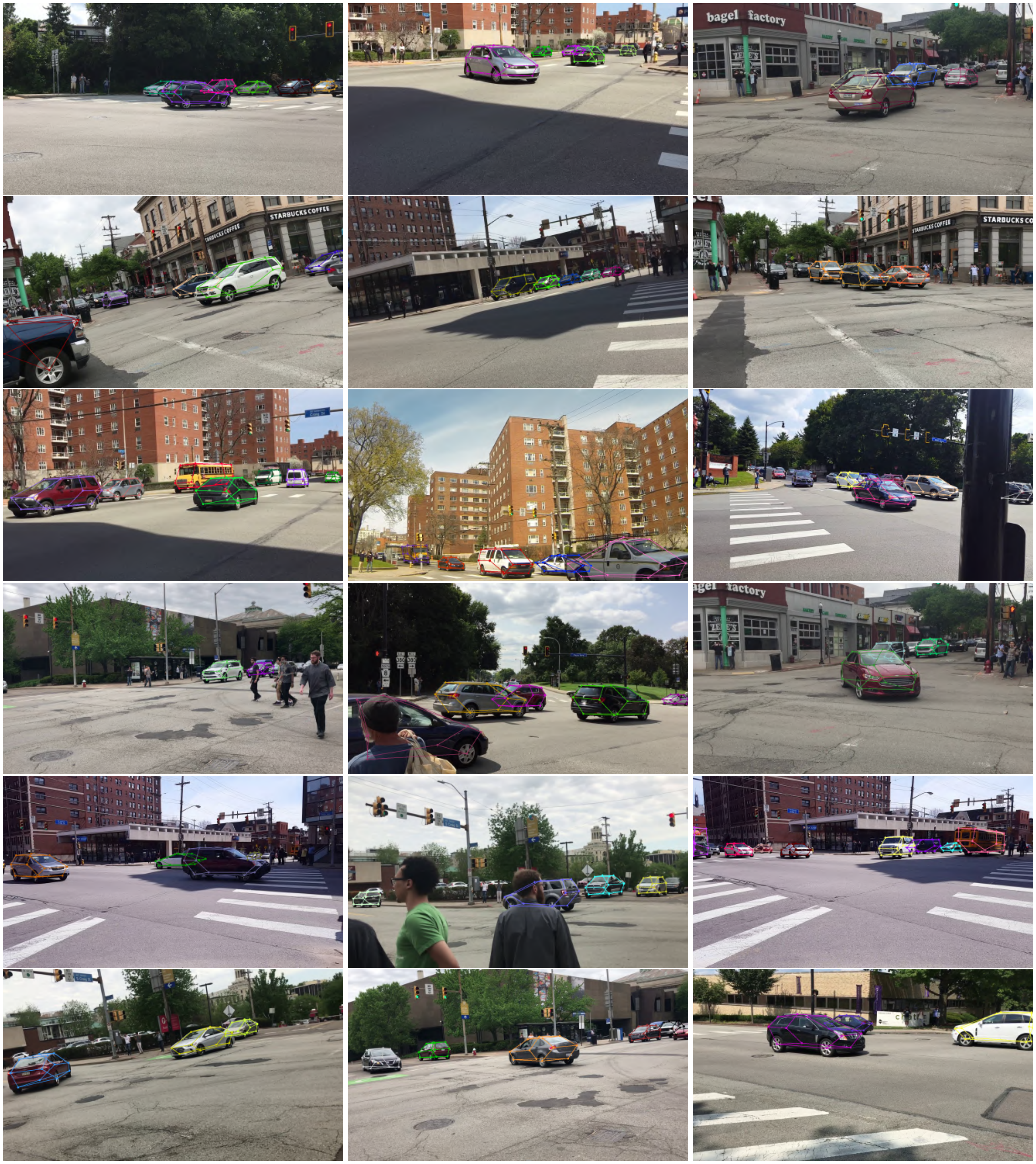
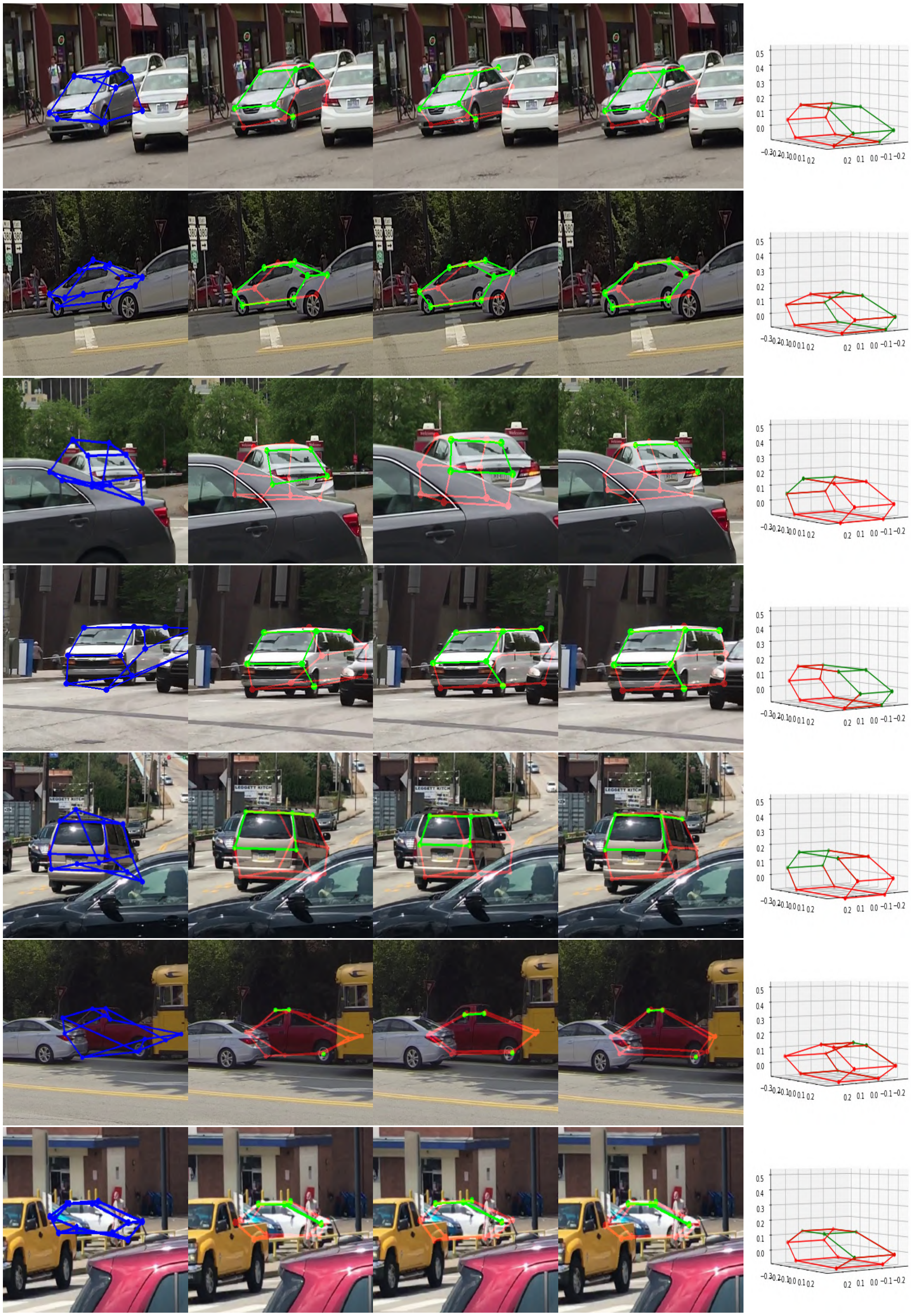


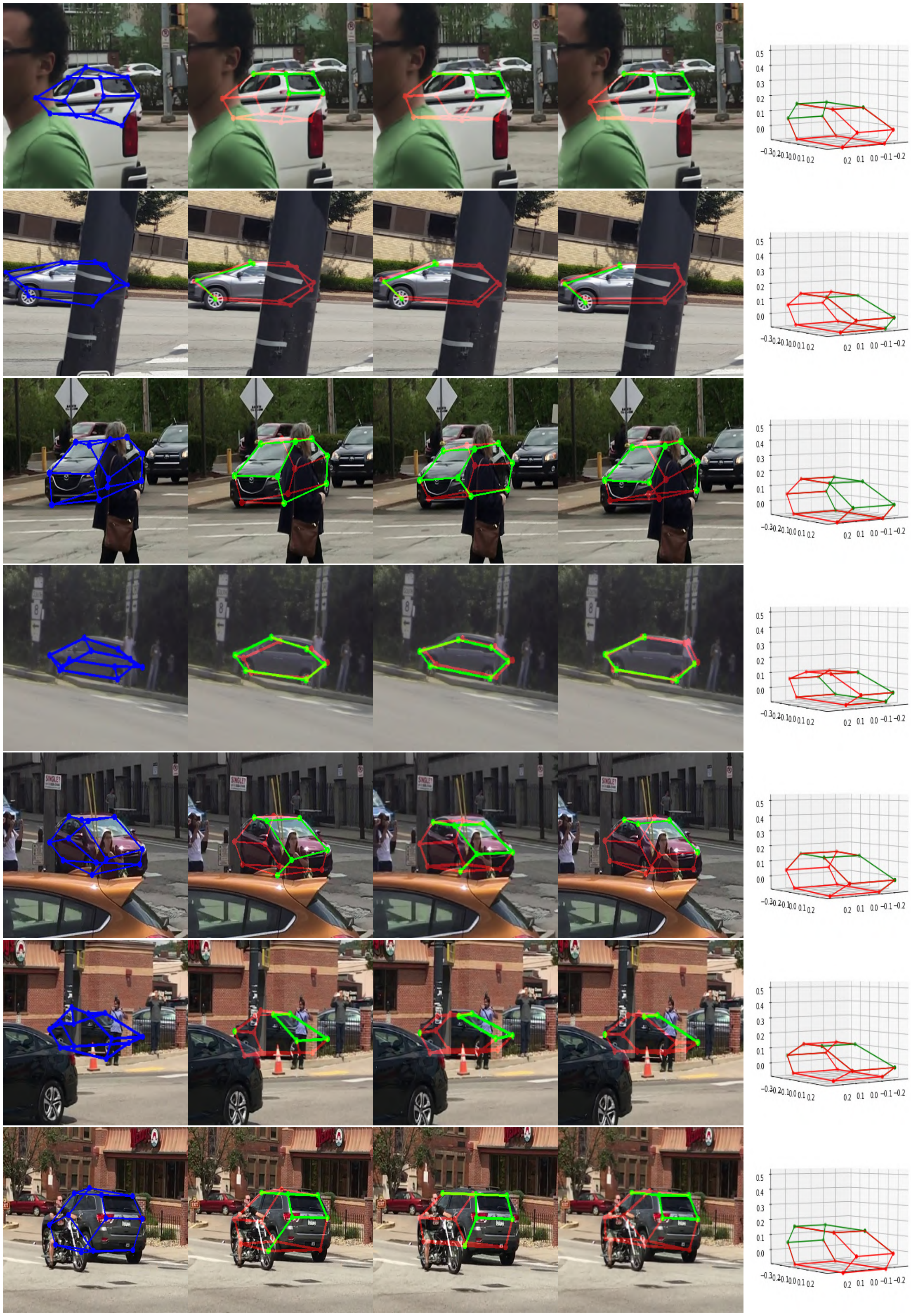
Figure 2: Extended Multi-car qualitative results on the extended carfusion dataset. Images have been randomly sampled from the extended carfusion dataset and passed through our multi-car network. Different colors depict different objects in the scene.

Figure 3: Qualitative Comparisons for self occluding category from the extended carfusion dataset. We observe a improvement in the keypoint localization of occluded points



(a) MaskRCNN (b) 2D-KGNN (c) 3D-KGNN (d) Ground-Truth (e) Canonical-3D

Figure 4: Qualitative Comparisons for Vehicle-occluding car category from the extended carfusion dataset. We observe a improvement in the keypoint localization of occluded points

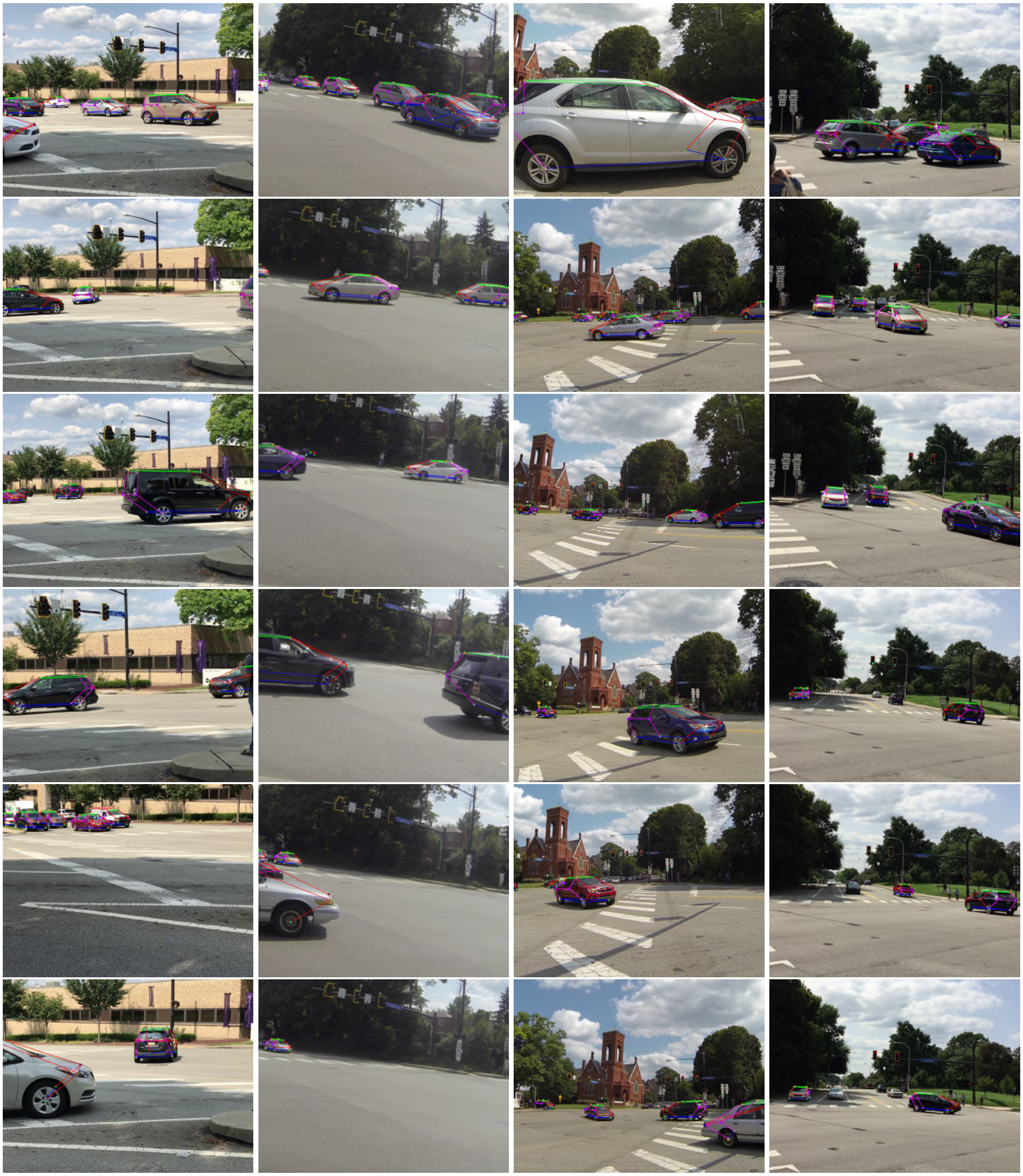


(a) MaskRCNN (b) 2D-KGNN (c) 3D-KGNN (d) Ground-Truth (e) Canonical-3D

Figure 5: Qualitative Comparisons for other-occluding car category from the extended carfusion dataset. This category has the most number of occluded keypoints compared to the rest of the categories and we observe the graph networks out performing the MaskRCNN baseline.

(a) MaskRCNN (b) 2D-KGNN (c) 3D-KGNN (d) Ground-Truth (e) Canonical-3D

Figure 6: Qualitative Comparisons for Truncation car category from the extended carfusion dataset. This category contains cars partially visible because of lying outside the viewing range of the camera.



(a) Camera-1 (b) Camera-2 (c) Camera-3 (d) Camera-4

Figure 7: Detector results on different cameras at over time on the intersection-5. We observe the same car detected from different views and the occluded keypoints localized accurately.

Figure 8: Failure situations of the algorithm are shown in this figure. Our method uses the bounding-box obtained from the previous method (MaskRCNN) to segment the object for computing the occluded keypoints.