A. Additional Details on Surfel Scene Reconstruction

For vehicle reconstruction, we aggregate LiDAR points across frames into a single model using bounding box annotation and local registration method (ICP). We also exploit the symmetry of man-made objects to complete their geometry. We reconstruct a total of 46,786 vehicle models. Some examples are shown in Figure 7. We also include some surfel scene reconstruction results in Figure 8. These surfel maps are built based on camera-LiDAR sequence from Waymo Open Dataset training split.

B. Additional Qualitative Results of the Synthesized Images

We provide additional qualitative results: Figure 9 contains synthesized output when gradually perturbing current view point. Figure 10 contains synthesized output when we also perturb the other objects in the scene. Figure 13 provide a visual comparison between different model variants. Results of generated surfel image and semantic map of SurfelGAN-SAC can be found in Figure 14. More visual results of SurfelGAN-SAC can be find in Figure 11 and Figure 12.

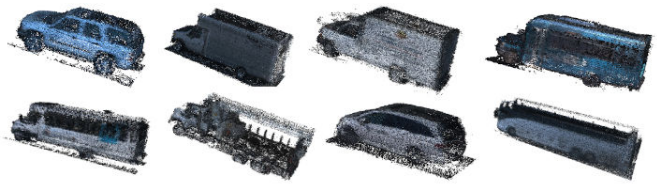

Figure 7. Sample reconstructed surfel vehicles.

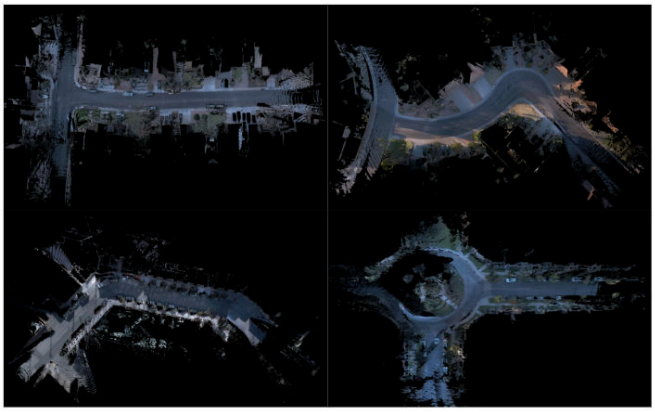

Figure 8. Reconstructed surfel scene maps from four different scenes.

C. Analysis on Surfel Image Coverage

We examine the detector's metric on different surfel image coverage ratio. Specifically, we define the surfel image coverage ratio as the percentage of area in the surfel image that is non-empty. The results can be found in Tab. 5. We found that our SurfelGAN model performs similarly when the observation ratio is above 30%, but the performance declines significantly when the input surfel map becomes more and more incomplete. It can happen when the SDV is looking through a direction where the surfel scene reconstruction is incomplete; for example, it stands at the boundary of surfel map and looking outward. This observation suggests building more complete surfel scene maps over multiple runs.

Perturbation	AP@50	AP@75	AP
$r \leq 0.3$	0.490	0.155	0.218
$0.3 < r \leq 0.5$	0.577	0.235	0.279
$0.5 < r$	0.566	0.172	0.253

Table 5. Detector metric break down at different region of observation ratio (r) on WOD-EVAL-NV.

Figure 9. Synthesized images when gradually perturbing the camera view point. Each column contains one example.

Figure 10. Synthesized images by perturbing objects in the scene. Top row: original surfel rendering. Center row: surfel rendering after perturbing the objects. Third row: synthesized images.

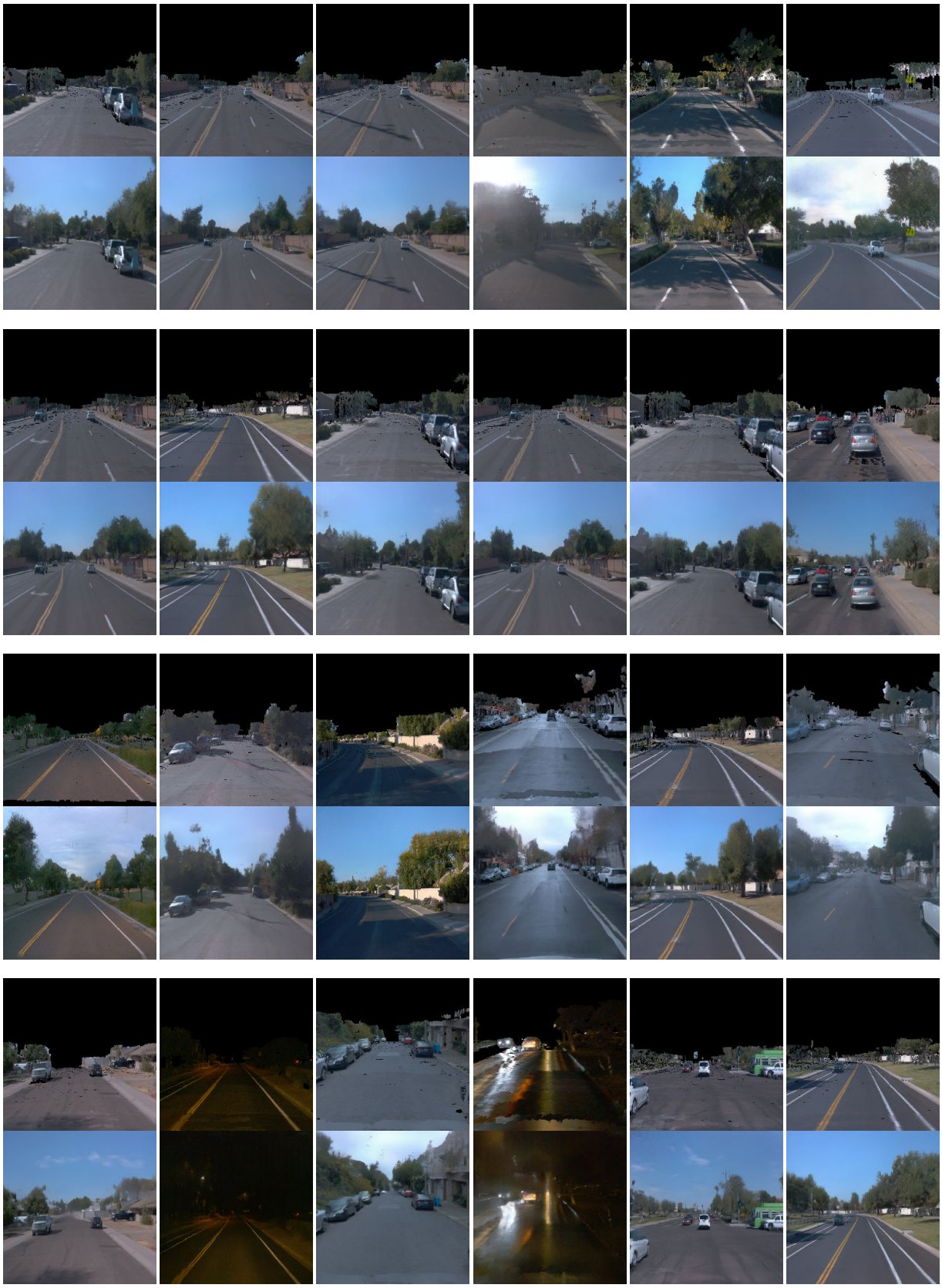

Figure 11. Qualitative results of SurfelGAN-SAC (1/2). We show pairs of surfel rendering and synthesized image.

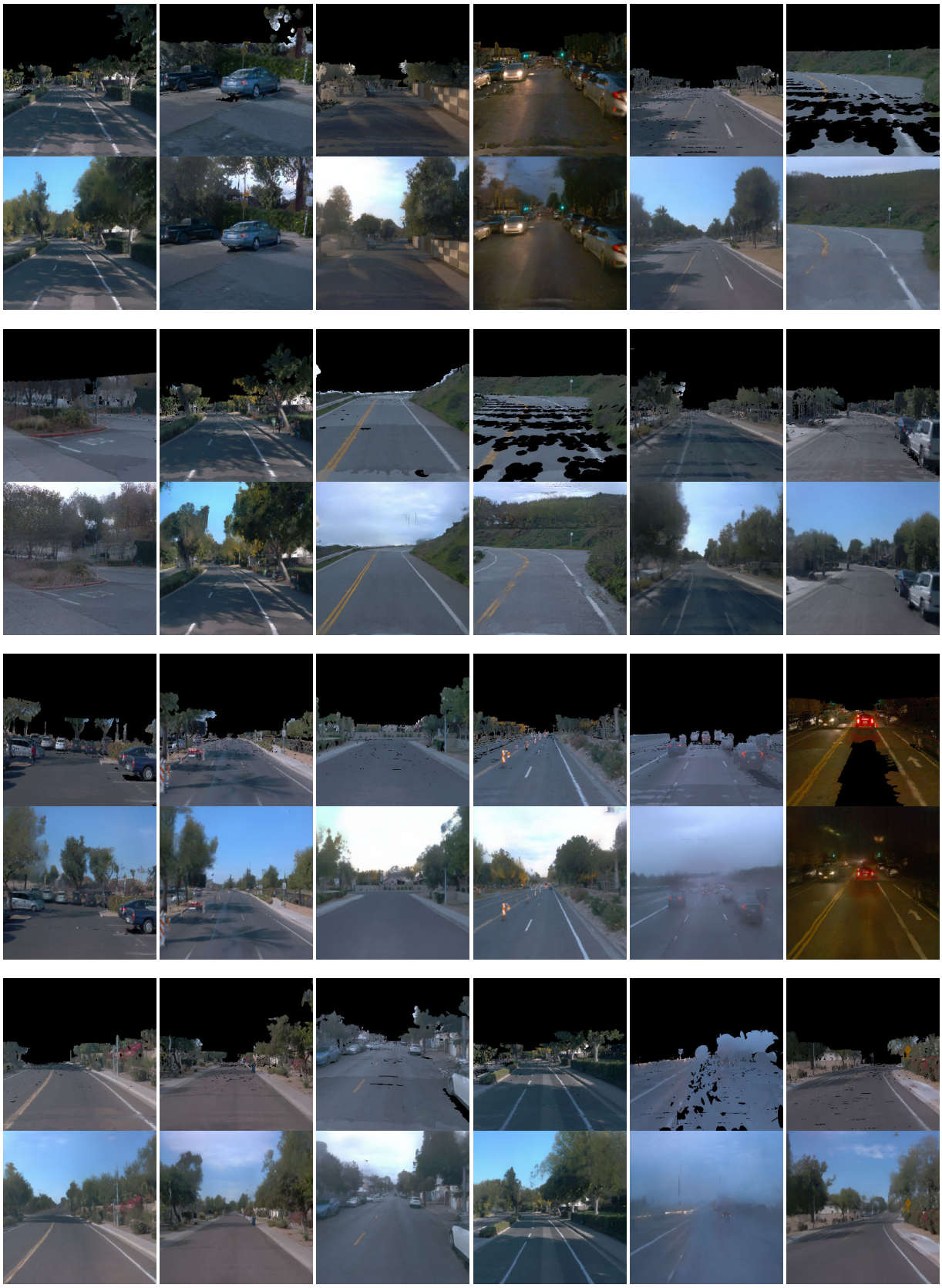

Figure 12. Qualitative results of SurfelGAN-SAC (2/2). We show pairs of surfel rendering and synthesized image.

Figure 13. Qualitative comparison between different SurfelGAN variants on WOD-EVAL

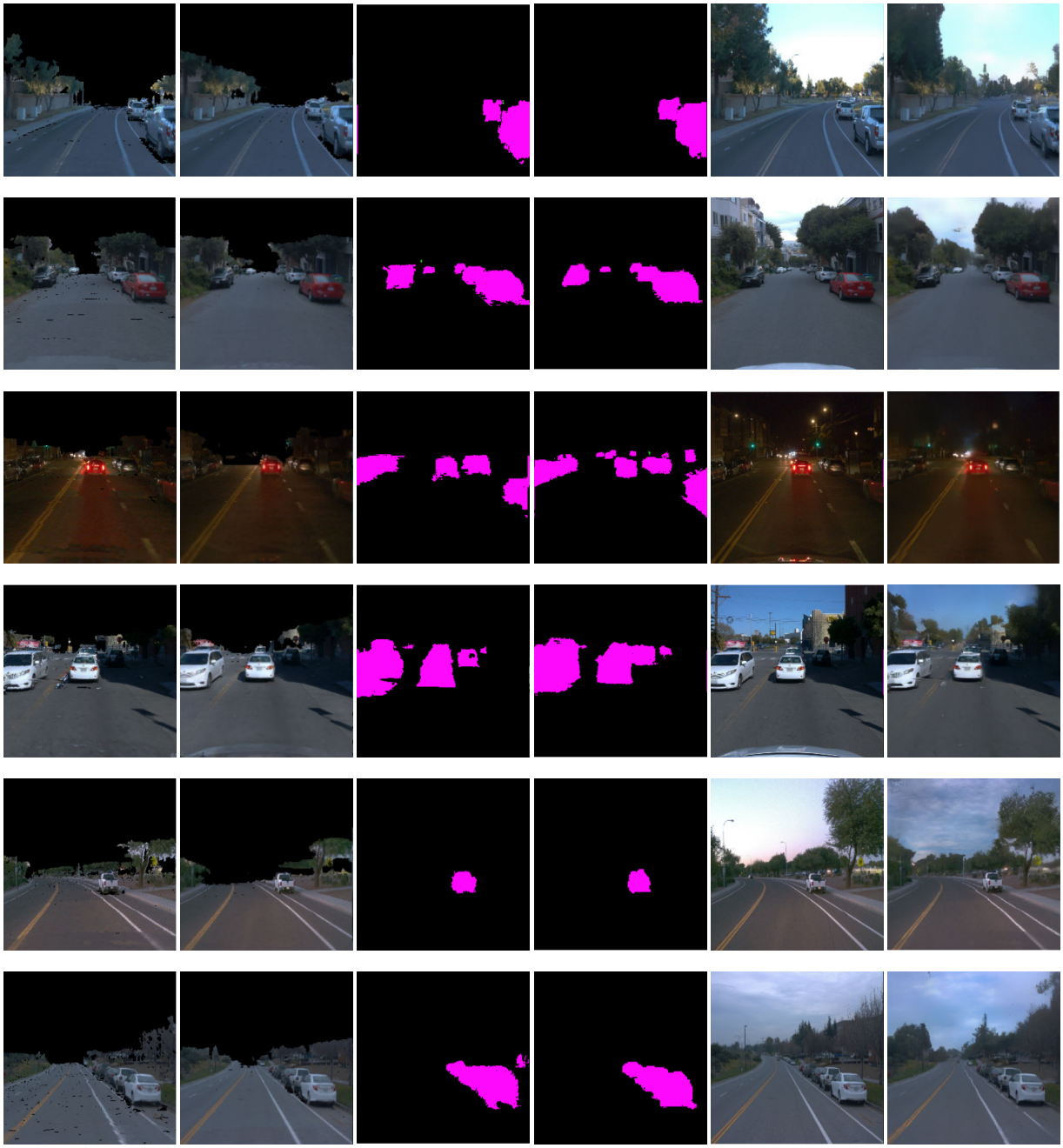

Figure 14. Visualization of different output head of SurfelGAN-SAC. From left to right: surfel rendering, generated surfel rendering, semantic map, generated semantic map, camera image, generated camera image.