

Cho-Ying Wu Ulrich Neumann

University of Southern California, Viterbi School of Engineering

Code and Data



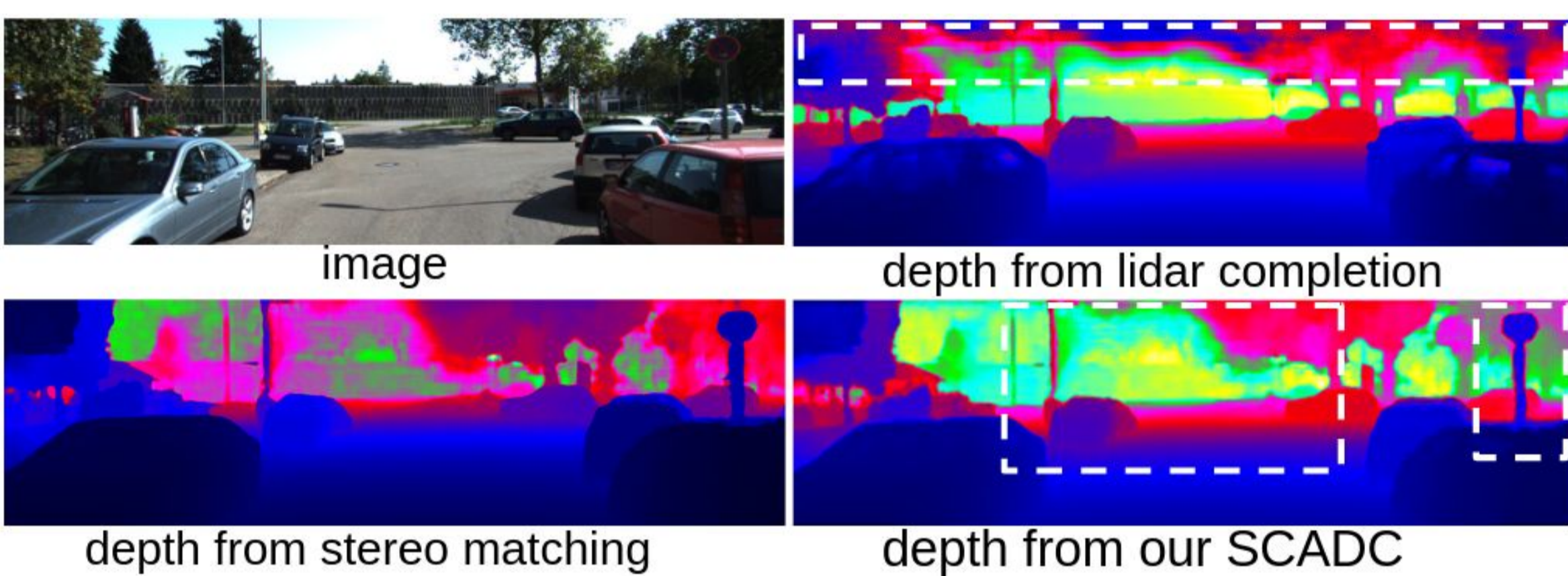
[click for code and data](#)

[click for demo](#)

#SensorFusion #DepthCompletion #Practicability

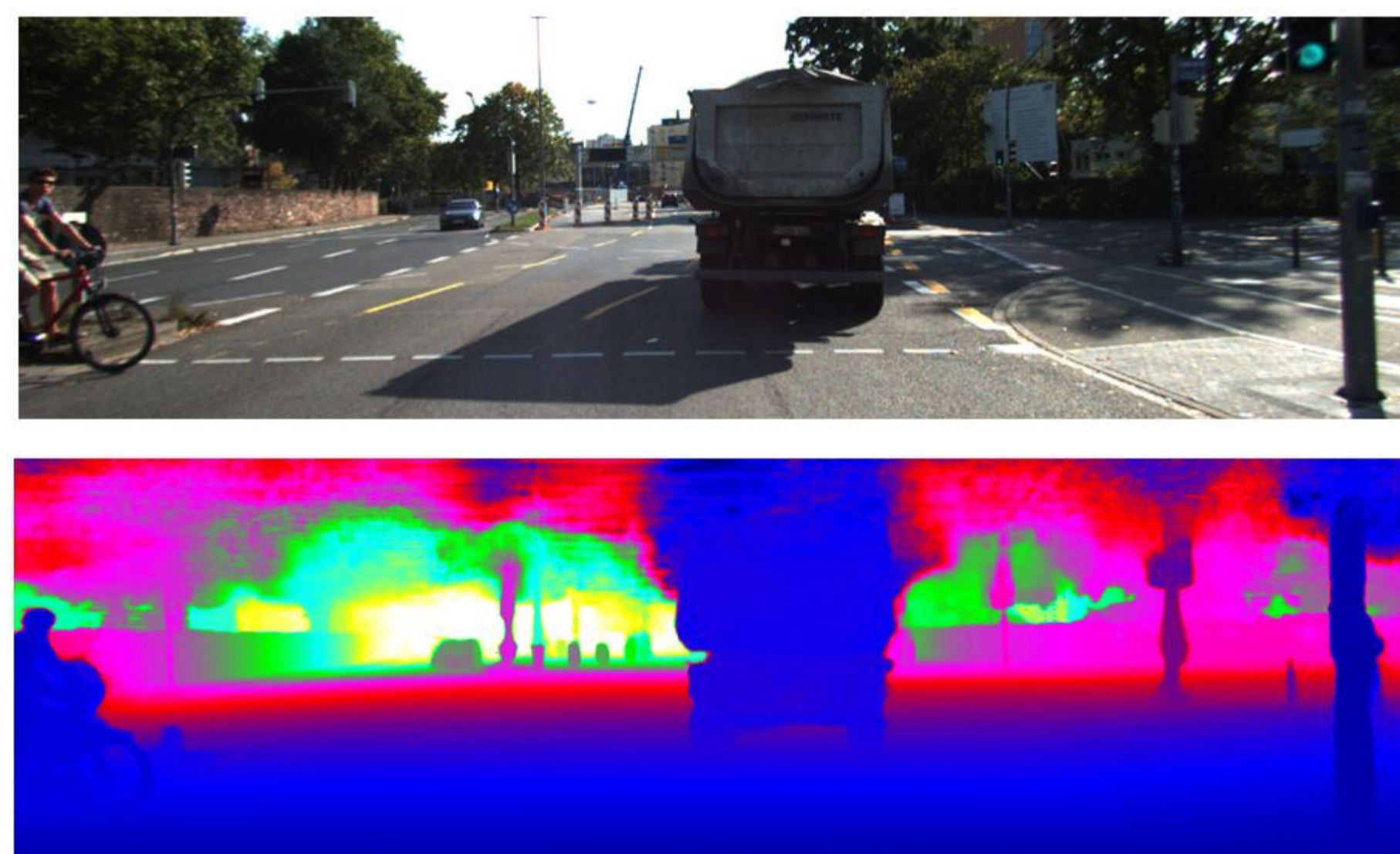
## Motivation

- Depth Completion from lidar produces irregular upper scenes due to no groundtruth guidance
- Stereo Matching based methods suffer from edge bleeding and the accuracy is generally lower than lidar



- Scene Completeness is omitted by prior arts. In real-world scenario, upper scenes are also important.

For example, depth completion from the SOTA method (FCFRNet, AAAI 2021)

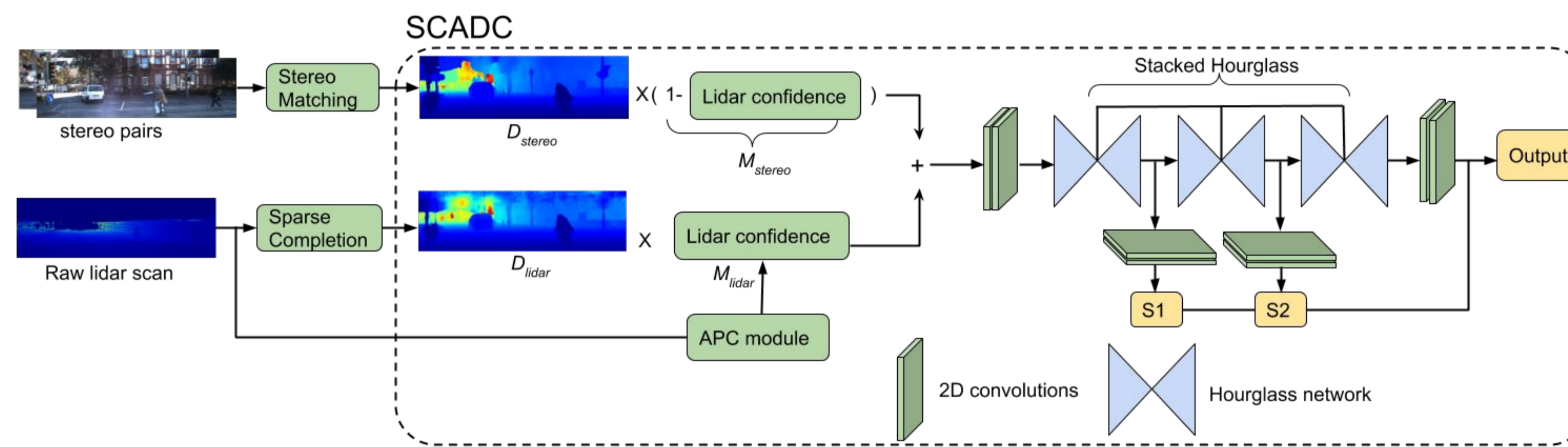


Irregular structures of traffic signs and large truck in front!

Practicability of completed depth maps is an issue.

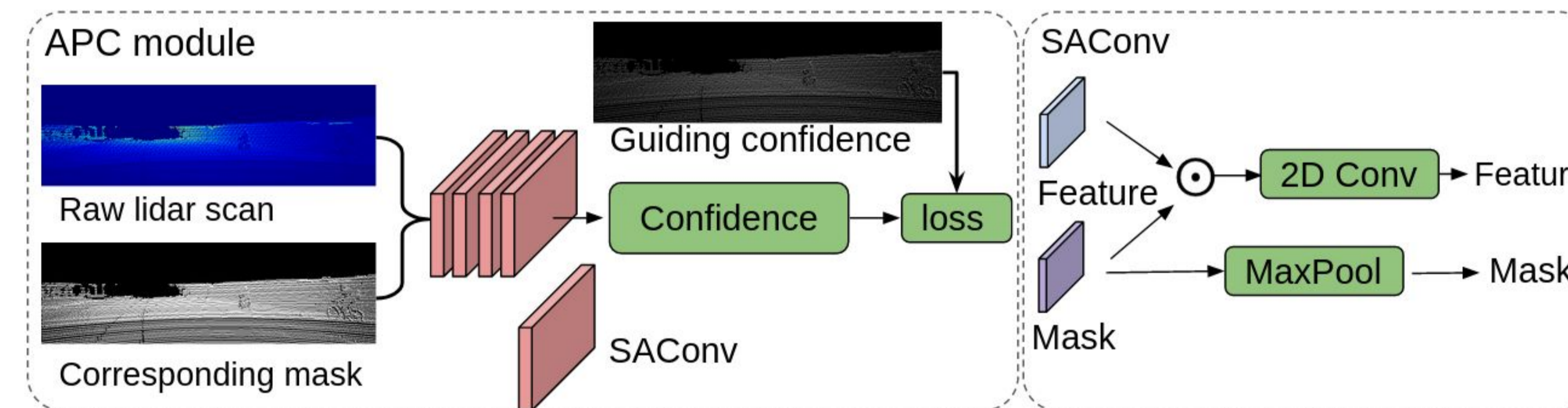
## Method & Contribution

- Sensor Fusion** of lidar and stereo camera is desirable: fusing upper scene information from *stereo pairs* and utilizing the high accuracy from *lidars*
- We propose **Scene Completeness-Aware Depth Completion (SCADC)** to fuse stereo disparity and maps and depth from lidar



$$D_f = D_{stereo} \times M_{stereo} + D_{lidar} \times M_{lidar}$$

- Attentional Point Confidence (APC)** estimates confidence map for lidar with a pseudo-guide, using SAConv (CFCNet, NeurIPS 2019)



Total loss  $L_1 + L_2 + L_3 + L_c$

Supervision on stage outputs from Stacked Hourglass      Supervision on confidence

## Result and Comparison

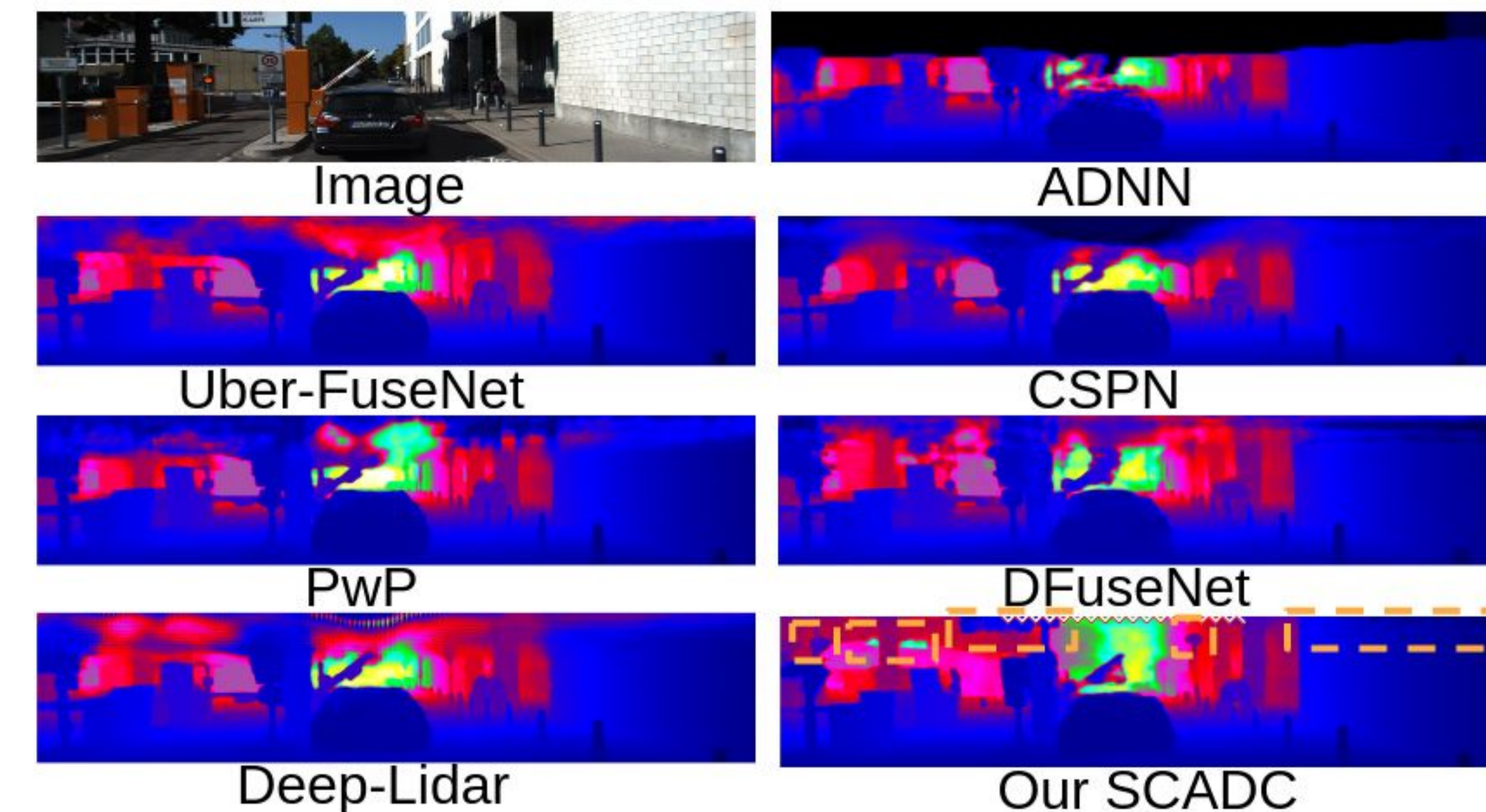
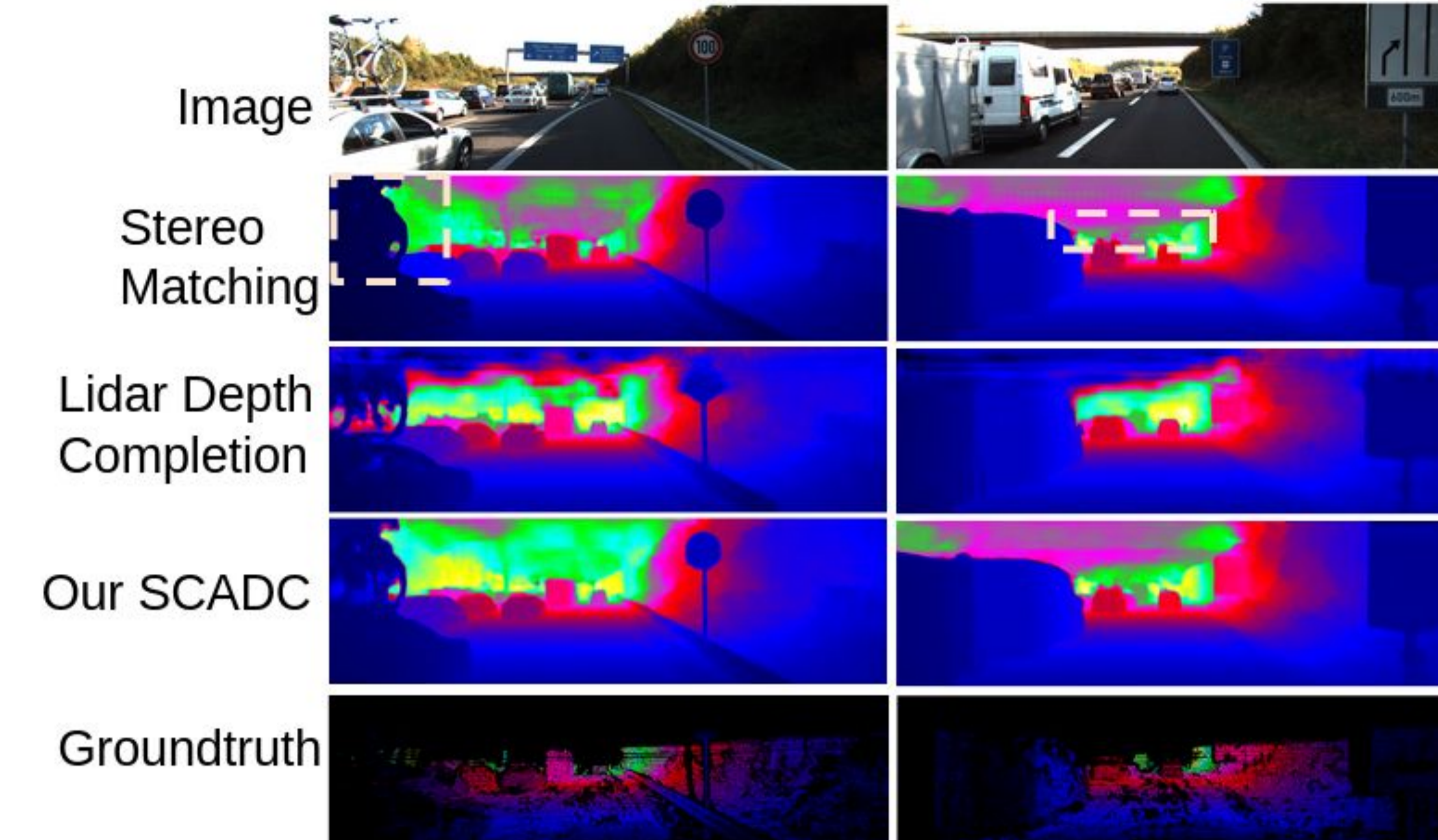
**Accuracy and scene structure:**

Dataset

- KITTI Completion: 42K frames of stereo pair images and lidar scans. 3.4K validation set.

Table 1. Evaluation on KITTI Depth Completion val set.

Methods	RMSE	Rel	$\delta_1$	$\delta_2$	$\delta_3$
PSMNet	2.4107	0.1296	98.6	99.8	99.9
SSDC	1.0438	<b>0.0191</b>	99.3	99.8	99.9
SCADC	<b>1.0096</b>	0.0226	<b>99.5</b>	<b>99.9</b>	<b>100.0</b>



Preserving the depth accuracy from lidar completion-based method (SSDC, ICRA 2019) and retain the structures from Stereo Matching (PSMNet, CVPR 2018).

**Practicability:** Outdoor RGBD Semantic Segmentation with our scene completeness-aware depth map

