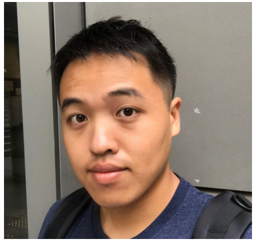
USC Viterbi School of Engineering

ICASSP 2021 46th IEEE International Conference on Acoustics, Speech, & Signal Processing, June 6 - 11, 2021, Toronto

Scene Completeness-Aware Lidar Depth Completion for Driving Scenario



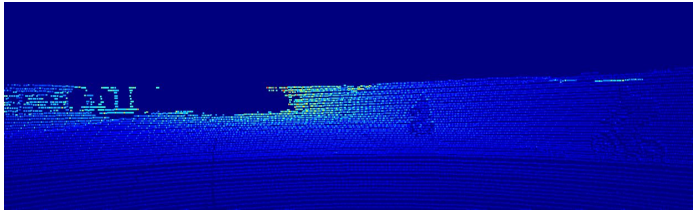

Cho-Ying Wu   Ulrich Neumann

CGIT Lab, Viterbi School of Engineering
University of Southern California

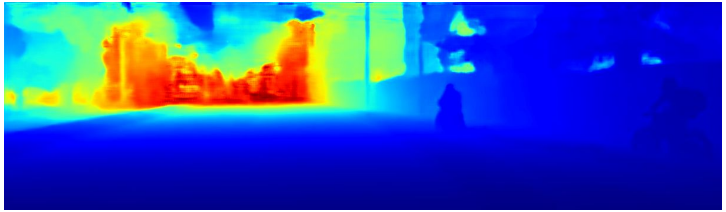
Background

Lidar Depth Completion

Image



Raw sparse depth

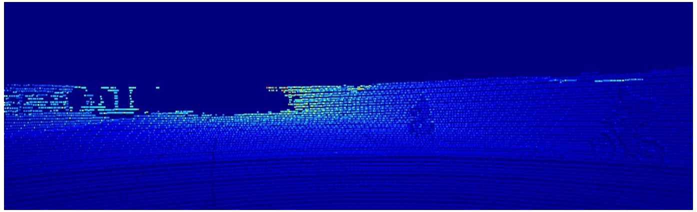


Dense depth map

Background

Lidar Depth Completion

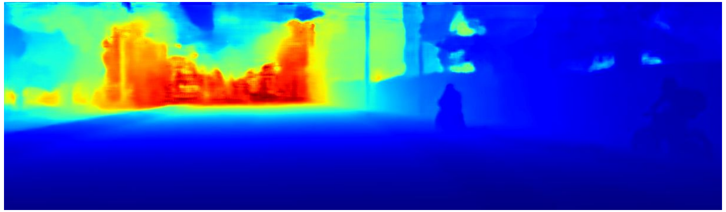
Image



Raw sparse depth

Advantage of dense depth maps:

- Benefit outdoor RGB-D methods, such as semantic or instance segmentation.



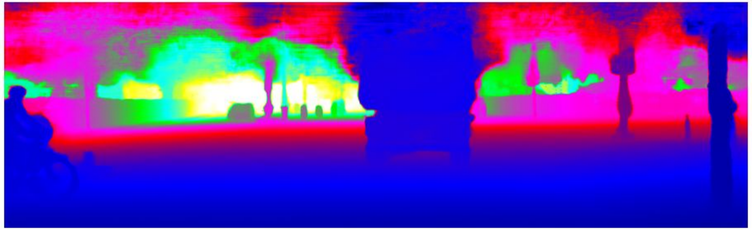
Dense depth map

Motivation

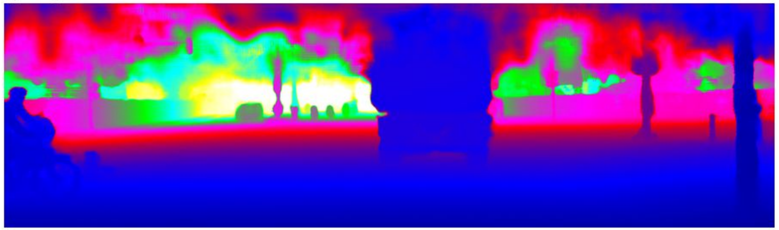
State-of-the-art methods

Scene

Completed Depth map



FCFR-Net (AAAI 21)



PENet (ICRA 21)

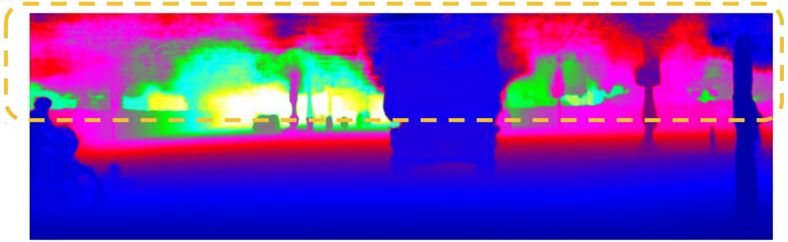
Motivation

Messy and unstructured upper scenes!

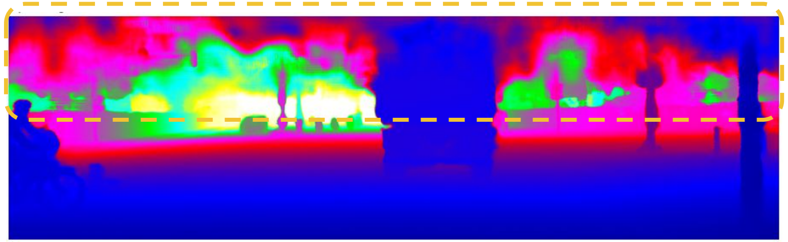
State-of-the-art methods

Scene

Completed Depth map



FCFR-Net (AAAI 21)

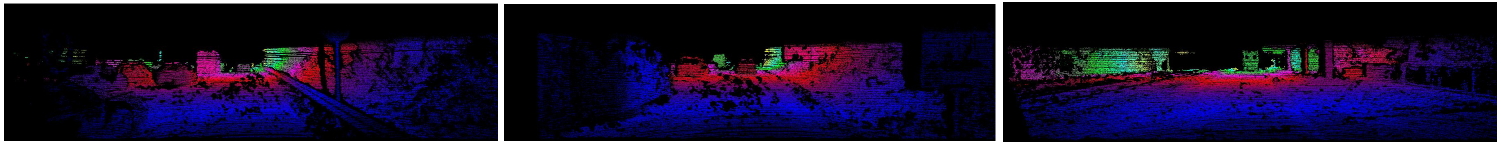


PENet (ICRA 21)

Motivation

No groundtruth annotation for the upper scenes in KITTI

Scenes

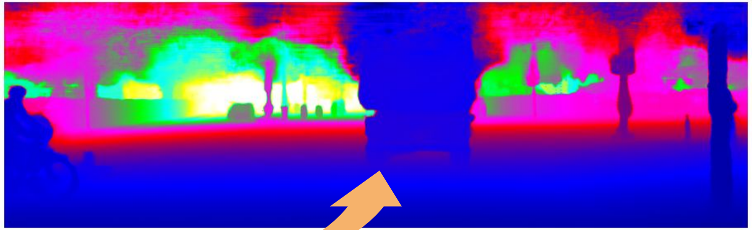


Semi-dense depth as groundtruth

Motivation

Scene

Completed Depth map

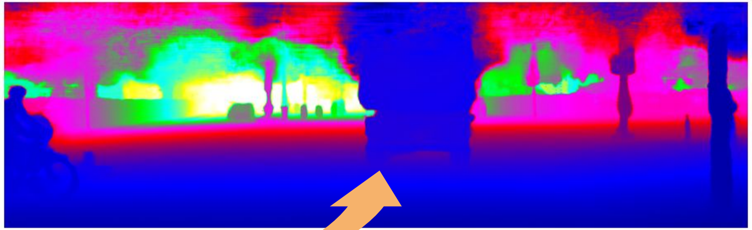


Object-level correspondence is not satisfied for outdoor RGB-D method

Motivation

Scene

Completed Depth map



Omitted issue of *scene completeness*

1. Depth Completion is treated as a standalone task.
2. Outdoor RGB-D methods are hard due to the obstacle of outdoor range-sensing.
3. Upper scenes are ignored since in most cases, upper scenes are sky and trees.

Motivation

Omitted issue of *scene completeness*
1. Depth Completion is treated as a *standalone* task.
2. Outdoor RGB-D methods are hard due to the obstacle of outdoor range-sensing.
3. Upper scenes are ignored since in most cases, upper scenes are sky and trees.

Improvements!

Our method for remedies on *scene completeness*
1. We validate our scene completeness-aware depth on semantic segmentation
2. We improve over previous SOTA work on outdoor semantic segmentation using our depth.
3. We raise counter examples that upper scenes are important. For example, traffic poles or signs extend to the upper, or in a case when there is a large truck in front.

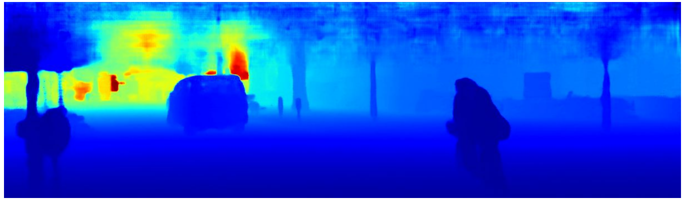
Motivation

Stereo matching

Stereo pair

Estimation from stereo pair

v.s.



Lidar Completion

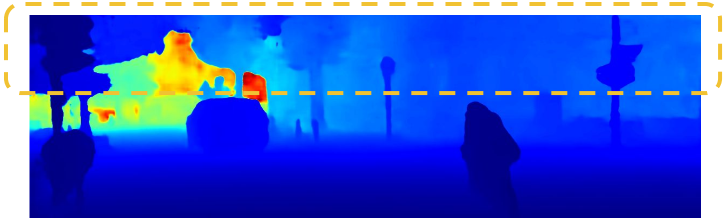
Motivation

Stereo matching

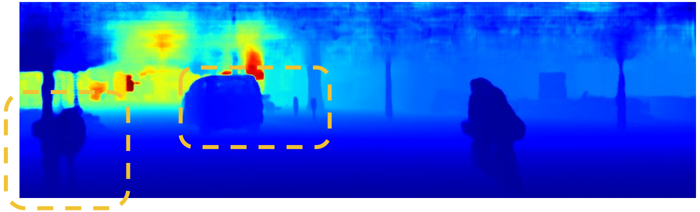
Stereo pair

Estimation from stereo pair

Advantage: Structured upper scenes

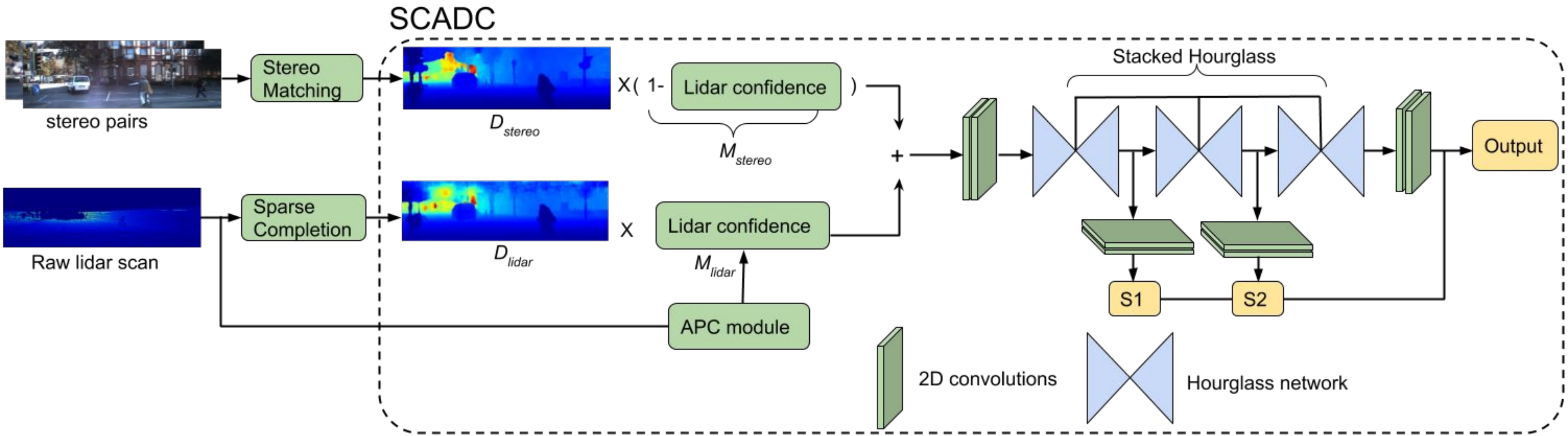


v.s.

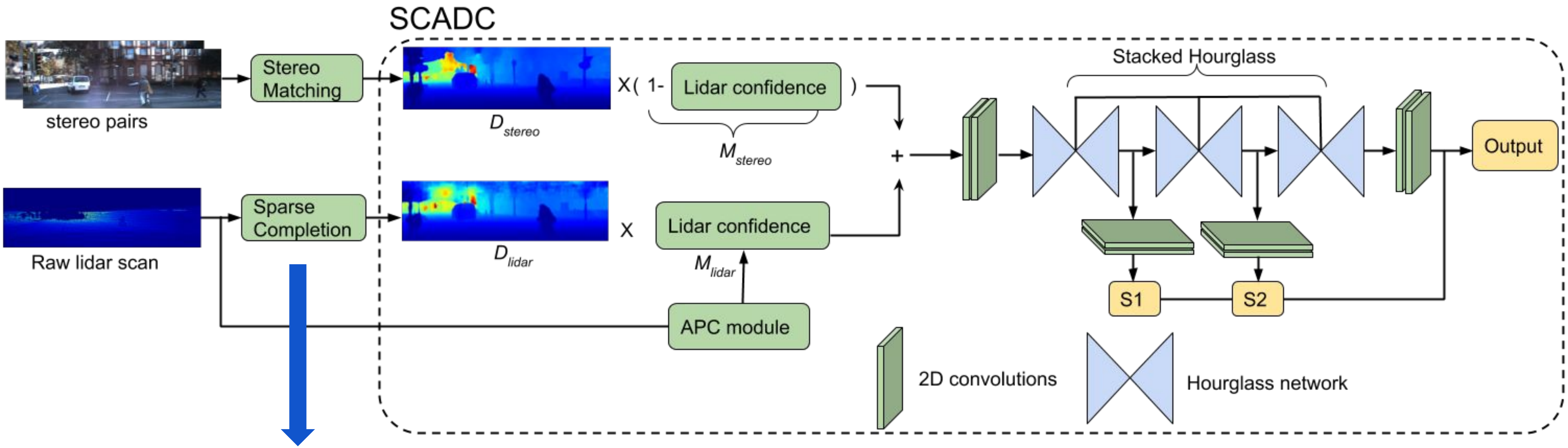


Advantage: Accuracy on lower scenes

Scene Completeness-Aware Depth Completion (SCADC)

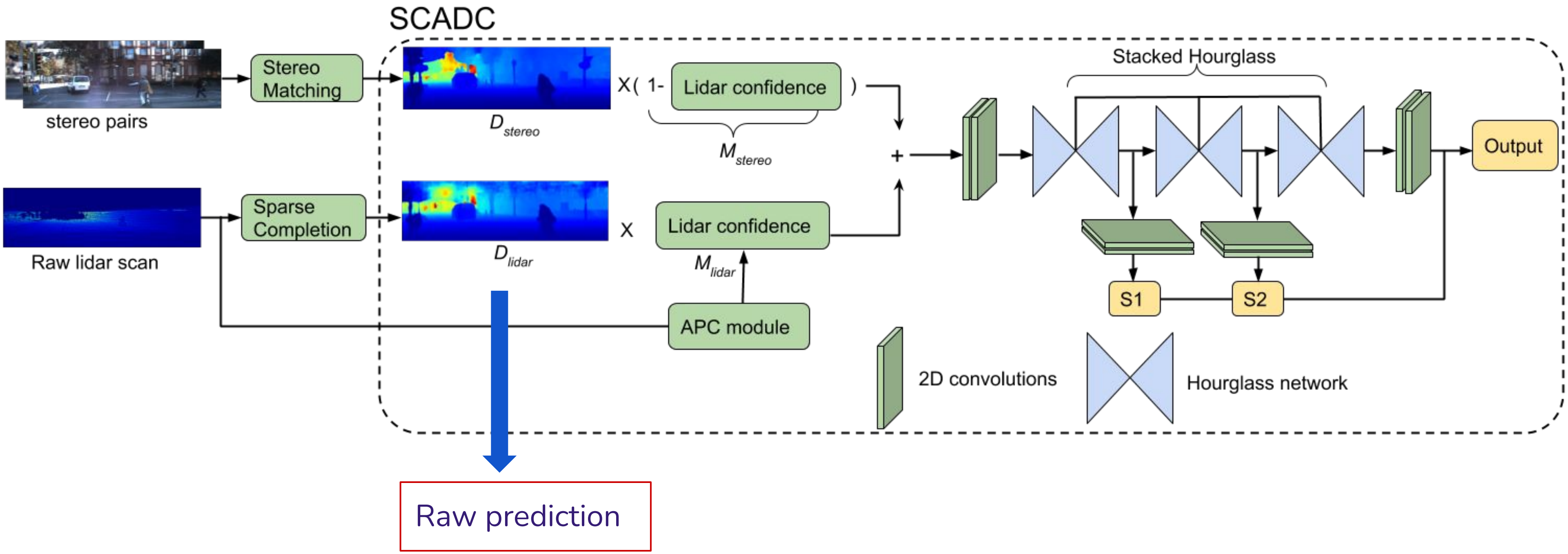


Scene Completeness-Aware Depth Completion (SCADC)

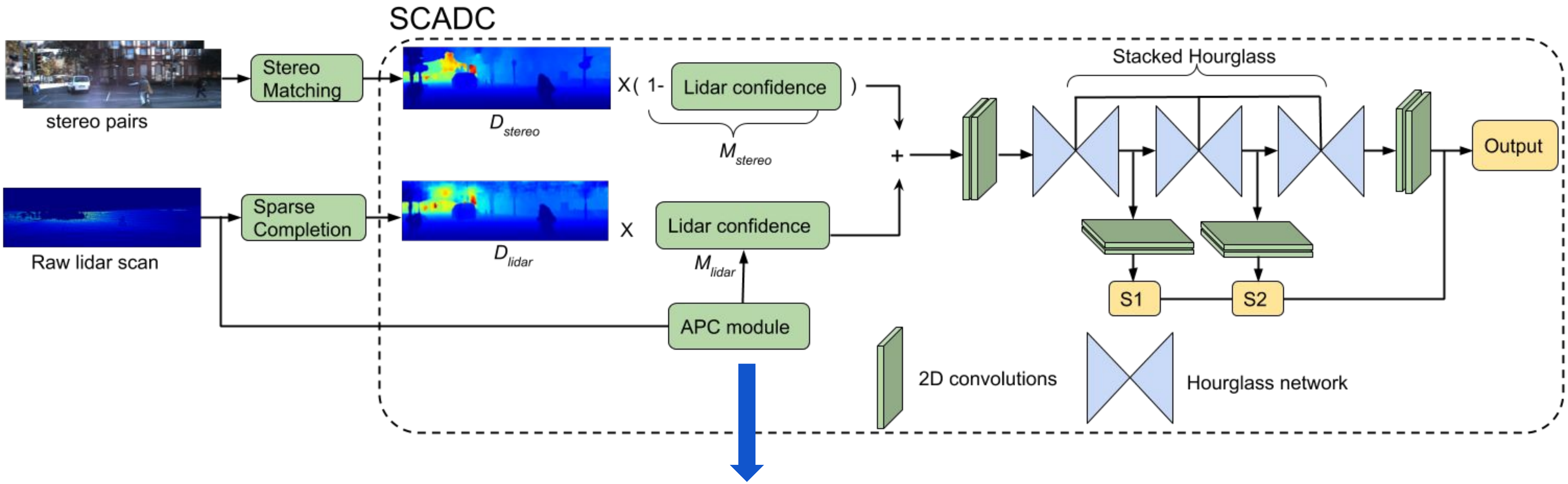


base units for stereo matching and lidar completion

Scene Completeness-Aware Depth Completion (SCADC)

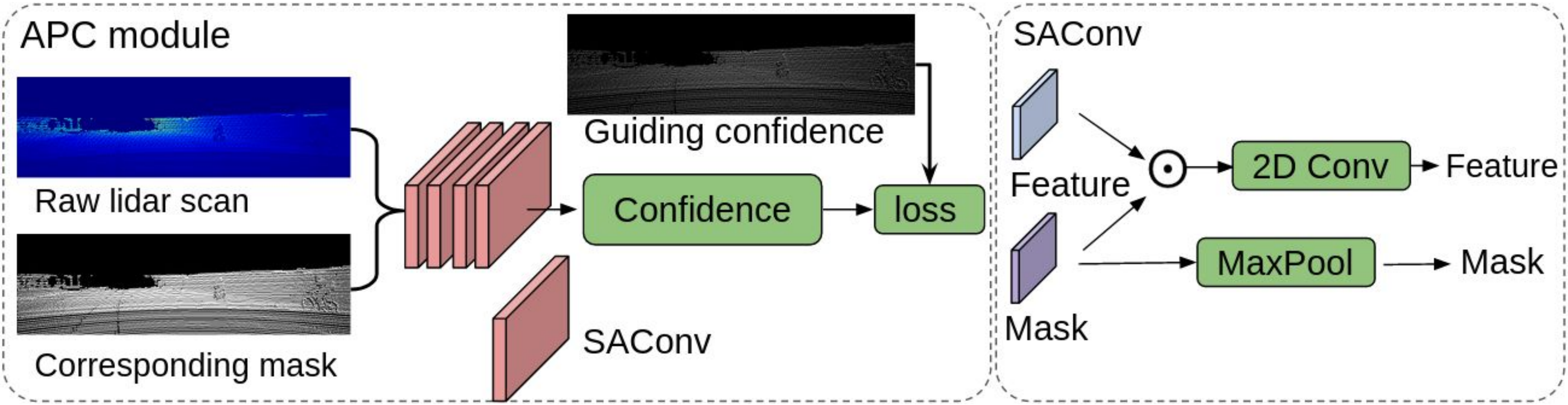


Scene Completeness-Aware Depth Completion (SCADC)



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

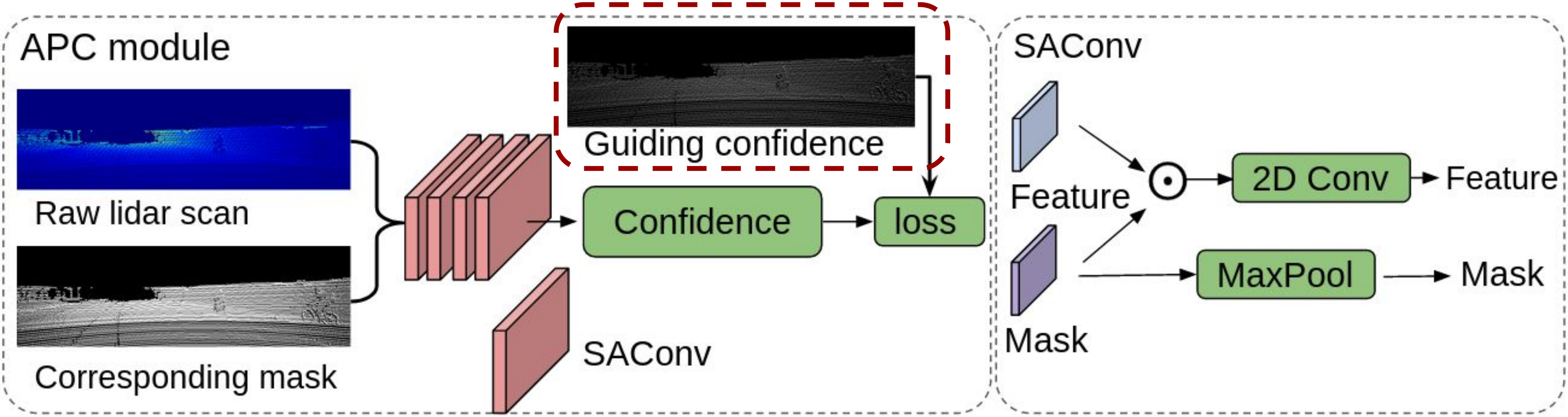
Scene Completeness-Aware Depth Completion (SCADC)



Estimate pixel-wise confidence with Sparse Convolution (CFCNet, NeurIPS 2019)

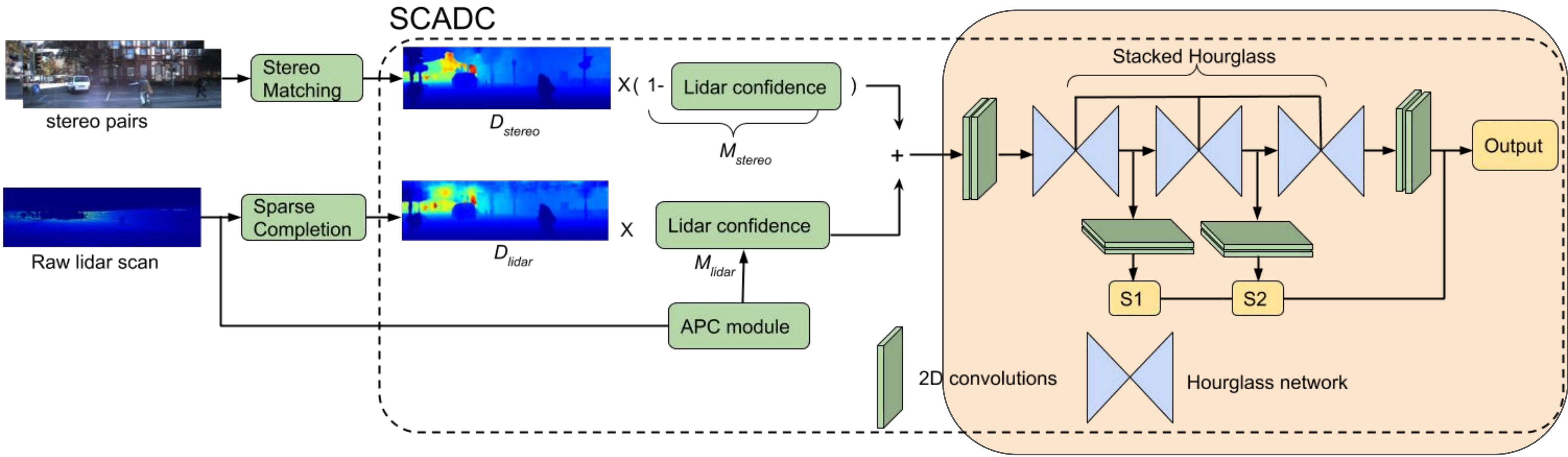
Deep RGB-D Canonical Correlation Analysis for Sparse Depth Completion, *Cho Ying Wu*, Yiqi Zhong*, Suya You, Ulrich Neumann (*Equal Contribution),* Neural Information Processing System (NeurIPS), 2019

Scene Completeness-Aware Depth Completion (SCADC)

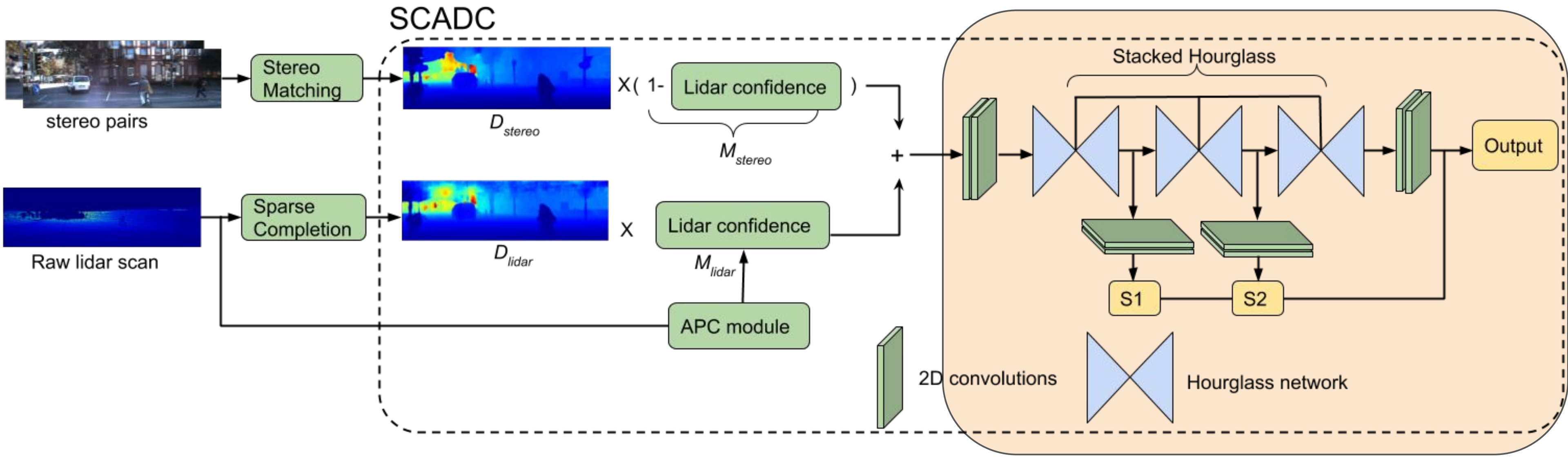


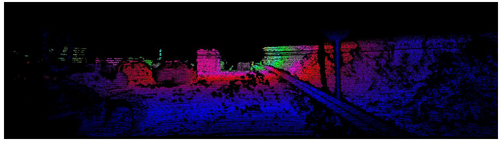
- Density along a scanline for KITTI is 44.6% at the center and 30.6% near the left/right side.
- Create the guide from the corresponding mask. Confidence = 1.0 for the raw point position.
- Dilate with a 3x3 kernel and choose a variance that drops values to half with 1-pixel distance from the center.

Scene Completeness-Aware Depth Completion (SCADC)



Scene Completeness-Aware Depth Completion (SCADC)





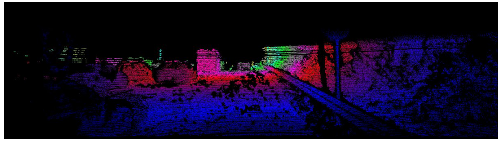
Semi-dense depth map is used for supervision

Scene Completeness-Aware Depth Completion (SCADC)

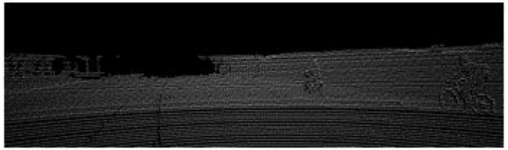
Total Loss

$$L_1 + L_2 + L_3 + L_c$$

Supervision on stage outputs for Stacked Hourglass

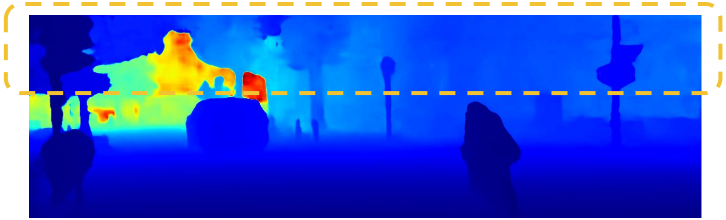


Supervision on confidence



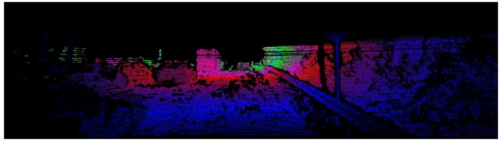
Scene Completeness-Aware Depth Completion (SCADC)

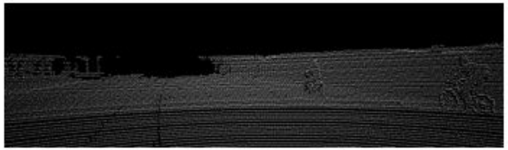
Where does upper scene structures come from? Prior from stereo matching



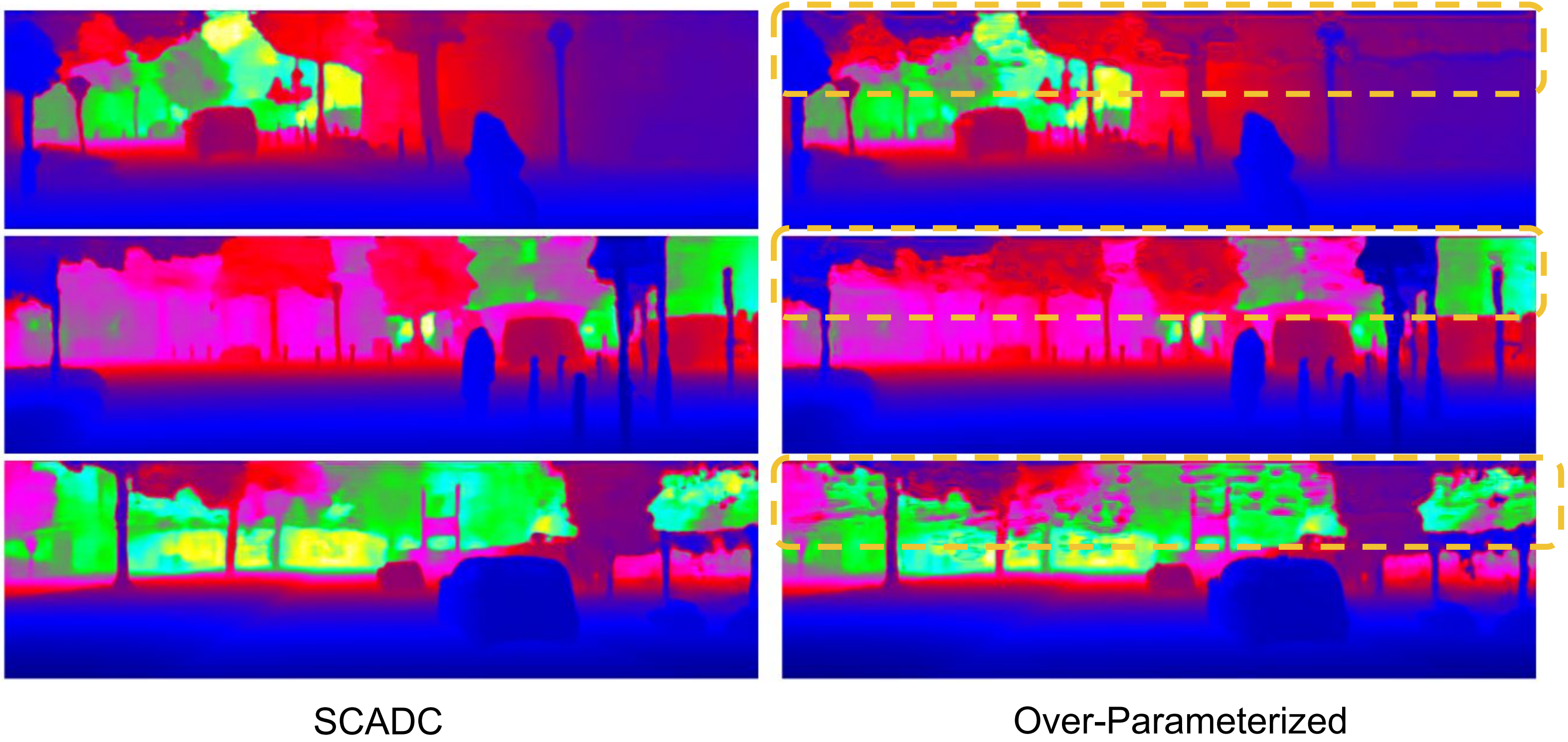
$$D_f = D_{stereo} \times M_{stereo} + D_{lidar} \times M_{lidar}.$$

The stacked hourglass network learns a depth mapping from **coarse estimation to finer depth**. However, only lower scenes have available annotations. The network should not be over-parameterized that overfits the lower scene!





Scene Completeness-Aware Depth Completion (SCADC)



Scene Completeness-Aware Depth Completion (SCADC)

Advantage? Combining prior information of structured upper scenes from stereo matching and accurate depth estimation by lidar completion.

Obtain a both scene completeness-aware and accurate scene depth!

Experiments

Dataset: KITTI Depth Completion, including 42K training paired data (stereo pairs and lidars) and 3.4K validation data

Depth recovery accuracy metrics: Root Mean Square Error (RMSE), Rel (Relative Error), and delta series

$$\delta_i = \frac{|\{\hat{d} : max(\frac{\hat{d}}{d}, \frac{d}{\hat{d}}) < 1.25^i\}|}{|\{d\}|},$$

Numerical Performance

Evaluation on KITTI Depth Completion Validation

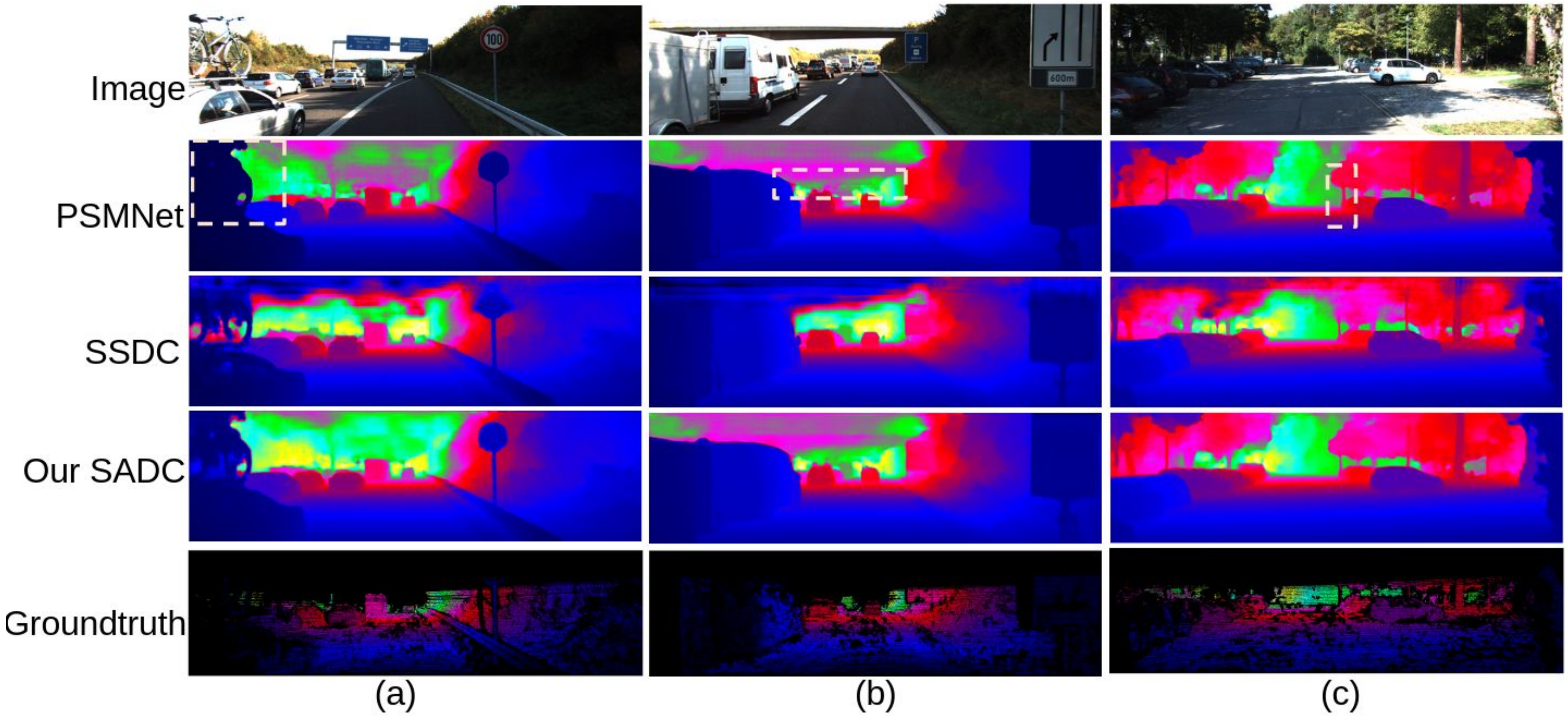
Methods	RMSE	Rel	$\delta 1$	$\delta 2$	$\delta 3$
PSMNet	2.4107	0.1296	98.6	99.8	99.9
SSDC	1.0438	**0.0191**	99.3	99.8	99.9
SCADC	**1.0096**	0.0226	**99.5**	**99.9**	**100.0**

PSMNet (CVPR 2018): Stereo-matching based method
SSDC (ICRA 2019): Lidar Completion based method
SCADC: Our method

Chang, Jia-Ren, and Yong-Sheng Chen. "Pyramid stereo matching network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

Ma, Fangchang, Guilherme Venturelli Cavalheiro, and Sertac Karaman. "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

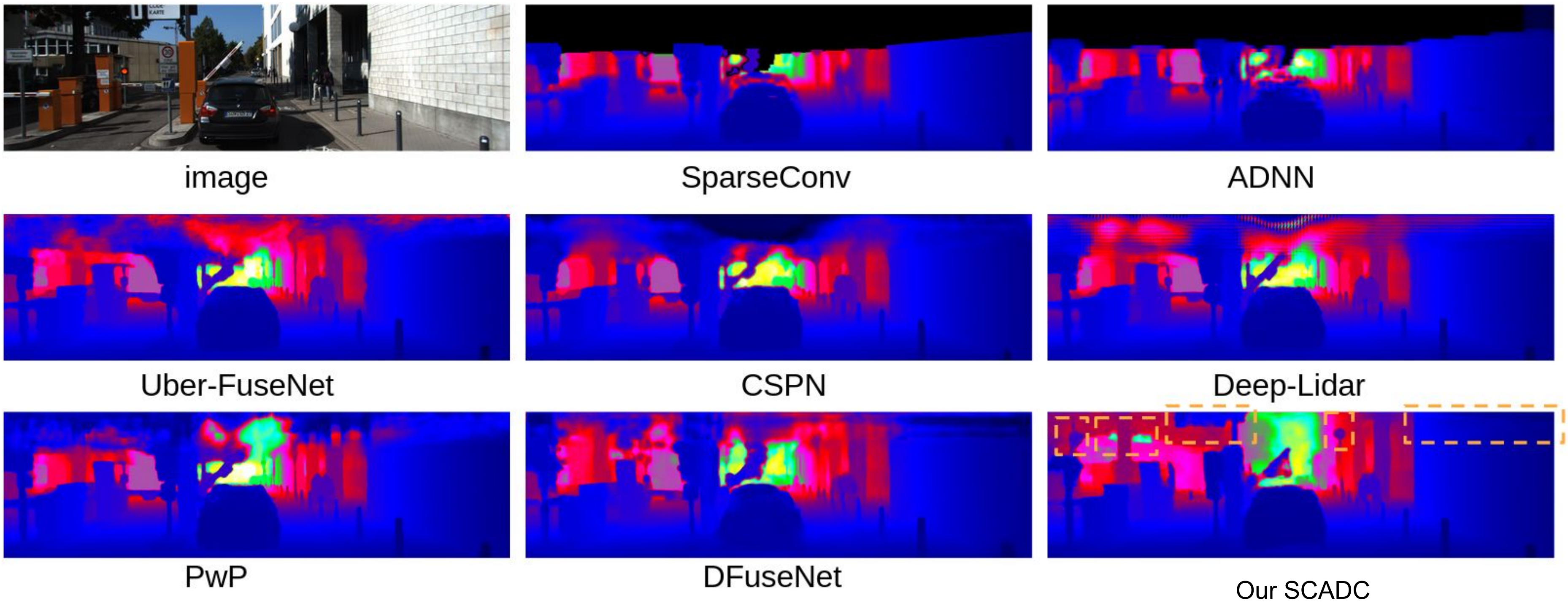
Upper Scene Recovery



Chang, Jia-Ren, and Yong-Sheng Chen. "Pyramid stereo matching network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

Ma, Fangchang, Guilherme Venturelli Cavalheiro, and Sertac Karaman. "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

Upper Scene Recovery - Benchmarking



Practicability - Application for completed depth maps

RGB-D outdoor semantic segmentation with our depth maps

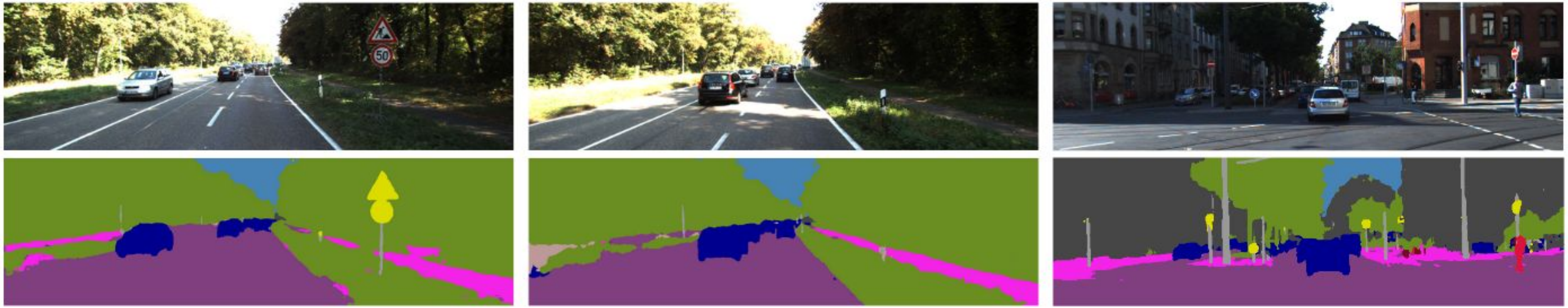
Dataset: KITTI Semantic Segmentation. Only 142/200 frames are associated with stereo pairs and lidar scans. We separate the available data into 121/21 for training and testing sets.

Metrics: mean Intersection over Union (mIoU)

Methods	mIoU
SDNet (GCPR 2019)	51.15
SGDepth (ECCV 2020)	53.04
SSMA (IJCV 2019)	54.76
SSMA + Our SCADC depth	**61.57**

Practicability

Visualizations: SSMA + our depth map



Demo

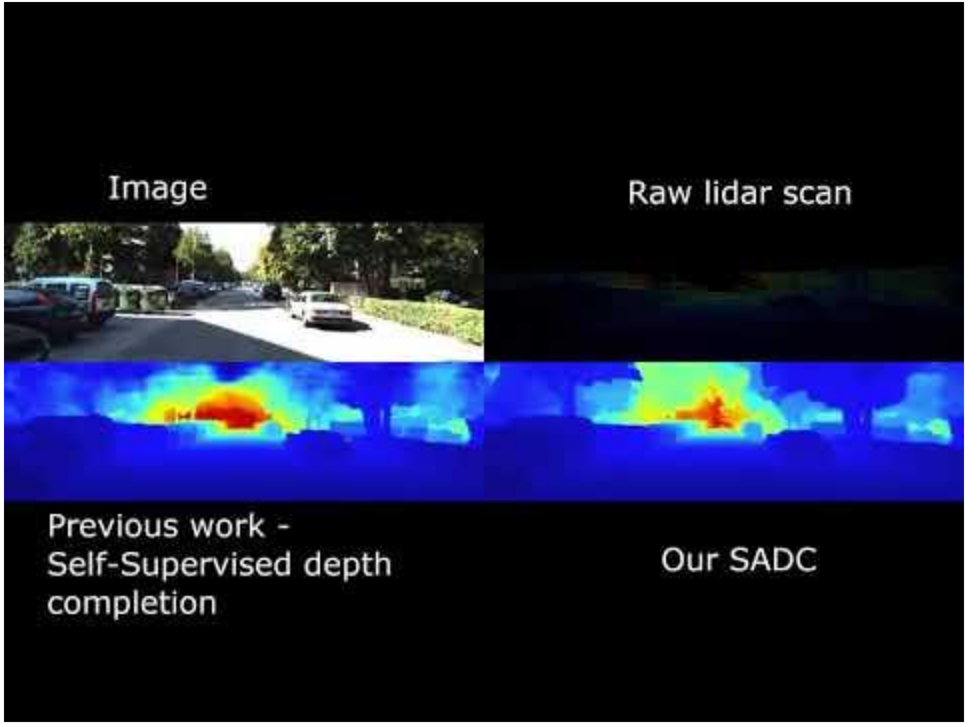


Image
Raw lidar scan
Previous work -
Self-Supervised depth
completion
Our SADC

Summary

- Sensor fusion for lidar and stereo cameras obtains both scene complete and accurate depth.
- Counter examples for the non-importance of upper scene depth are raised. Many examples show that upper scene structures are important for the driving scenario.
- We illustrate real-world applications for completed depth on outdoor RGB-D semantic segmentation, contrary to previous works that treat depth completion as a standalone task.

Summary

Thank you and please take a look at our poster #2152 for more illustrations, demos, and code and data link.