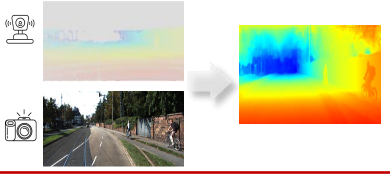
FastCompletion: A Cascade Network with Multiscale Group-Fused Inputs for Real-Time Depth Completion

Ang Li[1] Zejian Yuan[1]
Yonggen Ling[2] Wanchao Chi[2] Shenghao Zhang[2] Chong Zhang[2]
[1]Xi'an Jiaotong University, China [2]Tencent Robotics X, China
Xi'an Jiaotong University
Xi'an, P.R.China

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Task

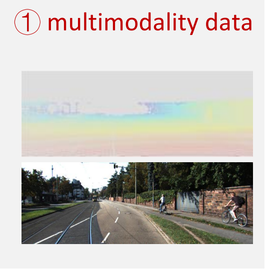
- **Depth completion**
  To estimate a dense depth map from sparse depth points in combination with an aligned high-resolution camera image.

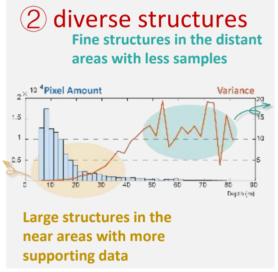


Challenges

- **To have a low computational complexity**

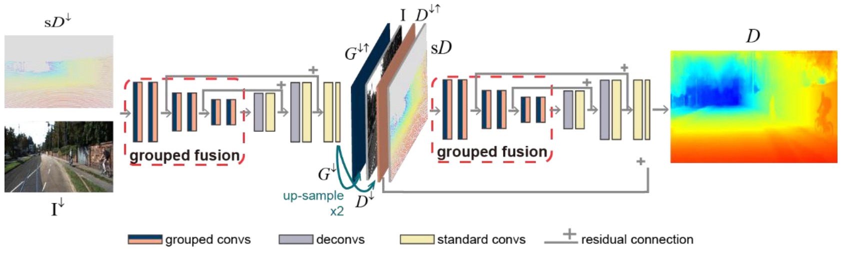
1. Dealing with multimodality data introduces extra computations.
2. Capturing diverse structures requires large models.





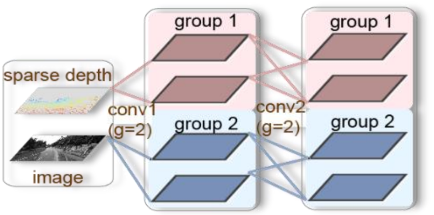
Cascaded Networks with Multiscale Inputs

- 2 cascaded hourglass networks are employed as the backbone.
- The cascaded networks receive input maps at different resolutions.
- Each subnetwork is specialized for certain structures and has a lightweight architecture.



Grouped Fusion

- The sparse depth and the gray image are fed into convolutional layers with 2 filter groups.
- Low computational complexity
- High degree of parallelism



Ablation Study

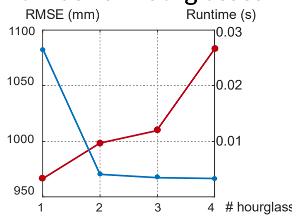
- Grouped fusion saves inference time without losing accuracy

	MAE	RMSE	♯Params	MACs	FPS
early	271.74	1006.34	187.91K	6.29	261.63
late cat	274.51	980.66	104.96K	5.05	203.79
late add	275.15	971.44	280.45K	8.63	189.24
early (H)	294.69	1019.28	1.76M	12.77	190.42
late cat (H)	285.93	982.05	1.18M	9.59	182.30
late add (H)	281.03	981.47	2.93M	19.08	160.85
group (H)	283.58	1001.63	1.18M	9.59	210.88
this work	273.35	980.61	140.96K	5.06	270.43

- Cascaded networks result in high accuracy & low runtime.

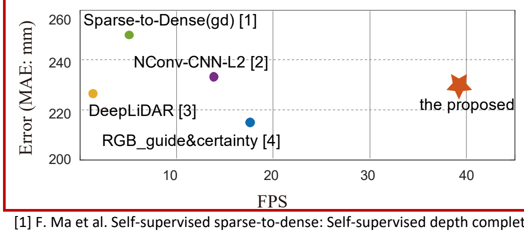
	MAE	RMSE	♯Params	MACs	FPS
holistic-32	298.40	1024.49	106.78K	5.19	332.49
holistic-256	283.58	1001.63	1.18M	9.59	210.88
holistic-512	263.54	967.41	4.72M	38.03	95.58
holistic-512-dw	280.33	1007.64	2.65M	19.66	104.86
this work	273.35	980.61	140.96K	5.06	270.43

- The influence of the number of hourglasses.

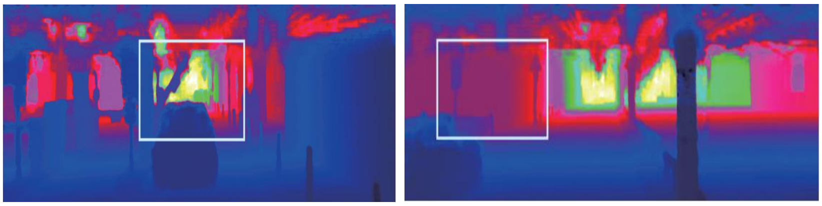


Results

- Comparison with state-of-the-art methods on NVIDIA Jetson AGX Xavier.



- The proposed model runs at more than 39 frames per second (FPS).
- Example results



[1] F. Ma et al. Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In *ICRA*, 2019

[2] A. Eldesokey, M. Felsberg, and F. S. Khan, Confidence propagation through cnns for guided sparse depth regression, *TPAMI*, 2019.

[3] J. Qiu et al. Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In *CVPR*, 2019.

[4] W. V. Gansbeke et al. Sparse and noisy lidar completion with RGB guidance and uncertainty. In *MVA*, 2019.