25th International Conference on Pattern Recognition, Milan, Italy, Jan.10-15, 2021

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]

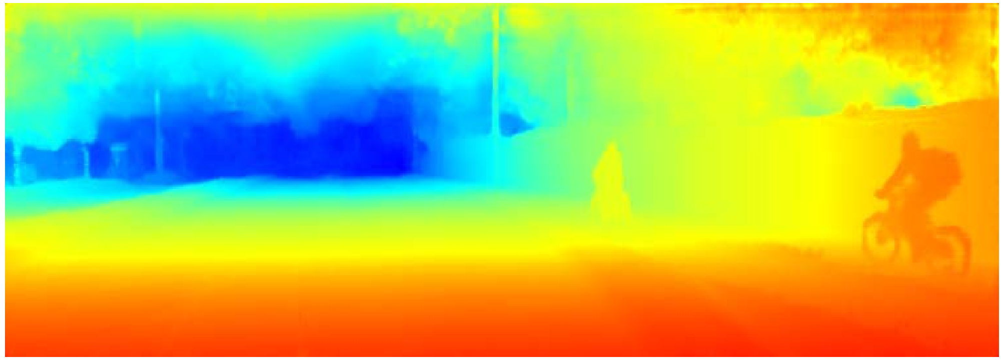
[1]Xi'an Jiaotong University, China

[2]Tencent Robotics X, China

Depth Completion

sparse depth map
~ 4% valid pixels

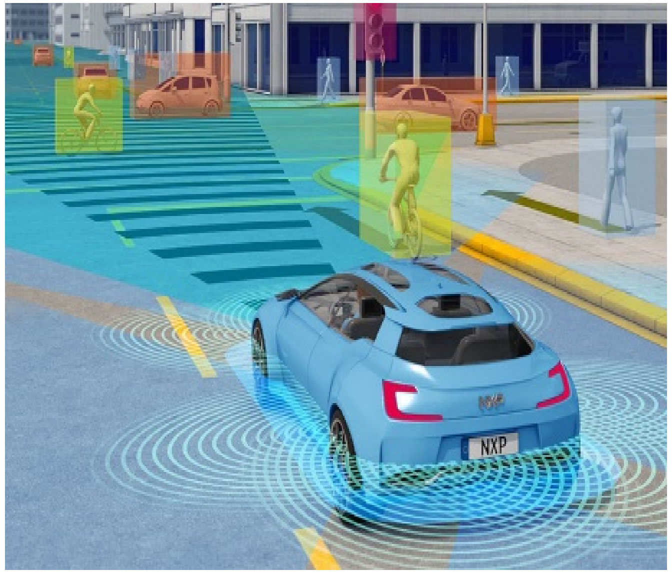
camera image



dense depth map

Applications


AR/VR



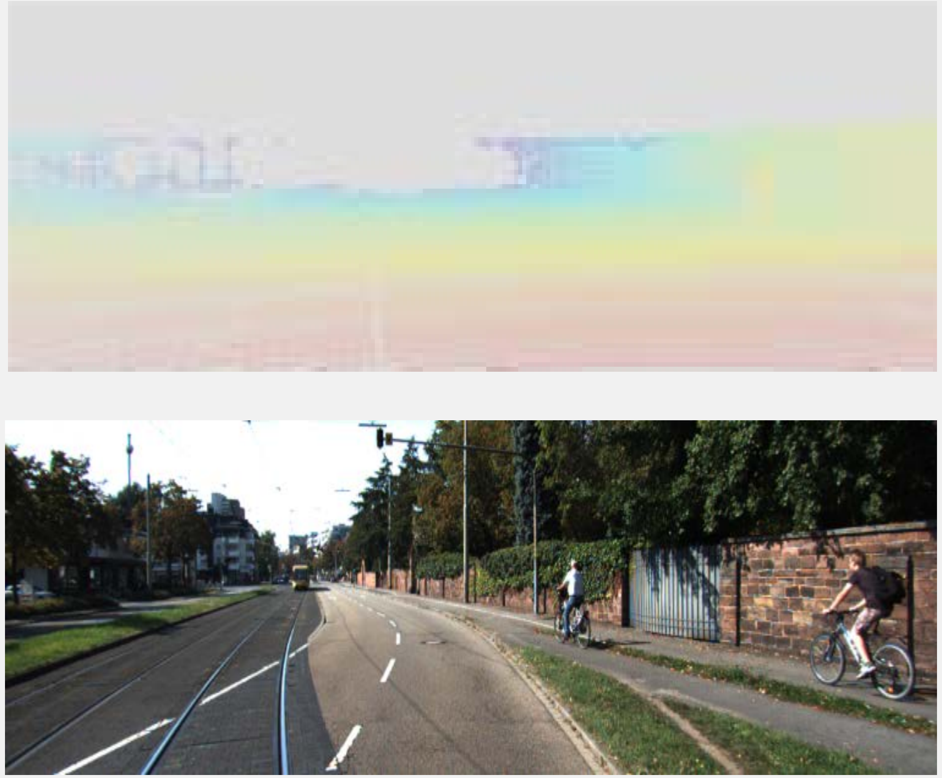
Autonomous Driving


Robot

A good trade-off between accuracy and speed is an increasing demand for a depth completion method.

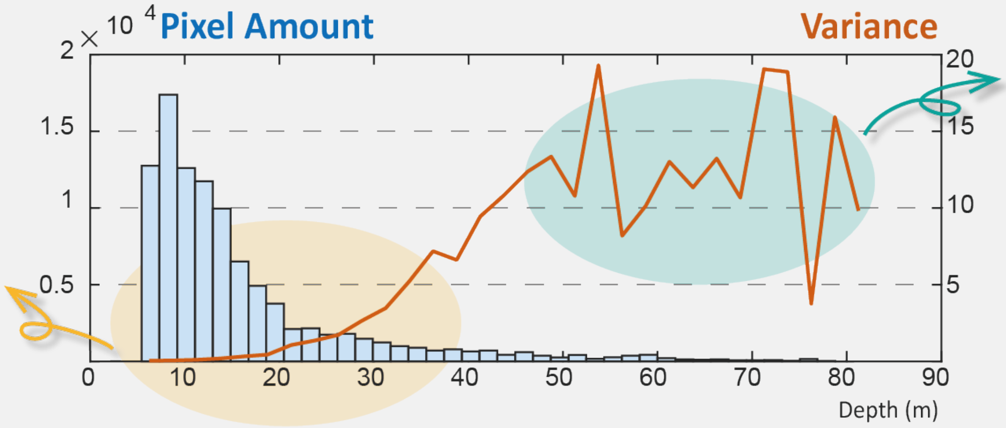
Challenges

① Multimodality data



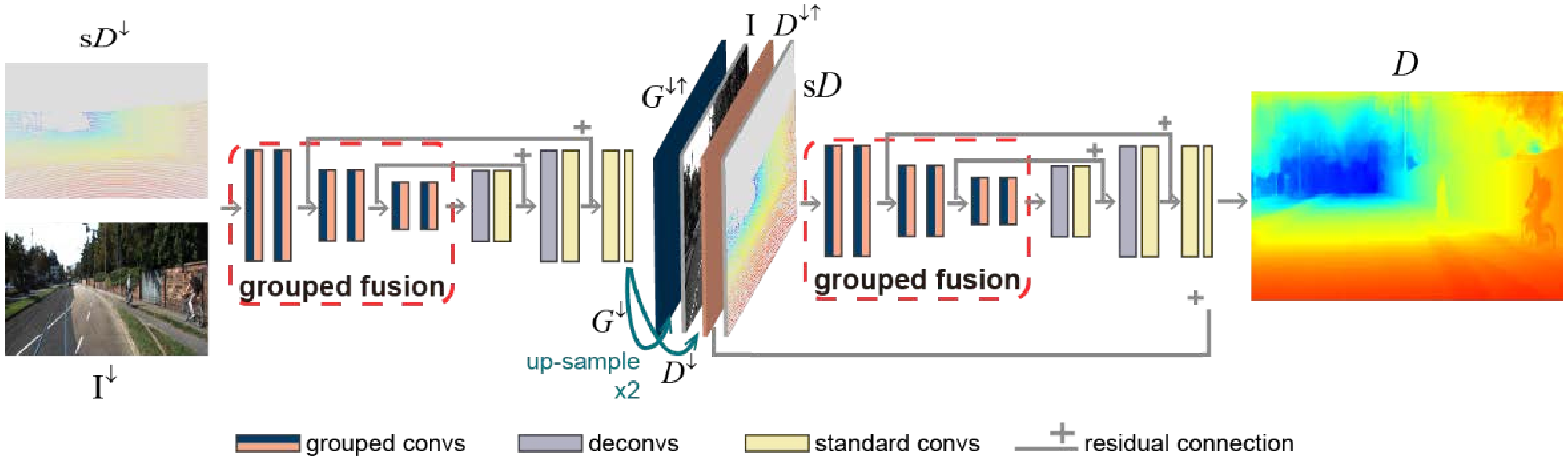
② Diverse structures

Fine structures in the distant areas with less samples



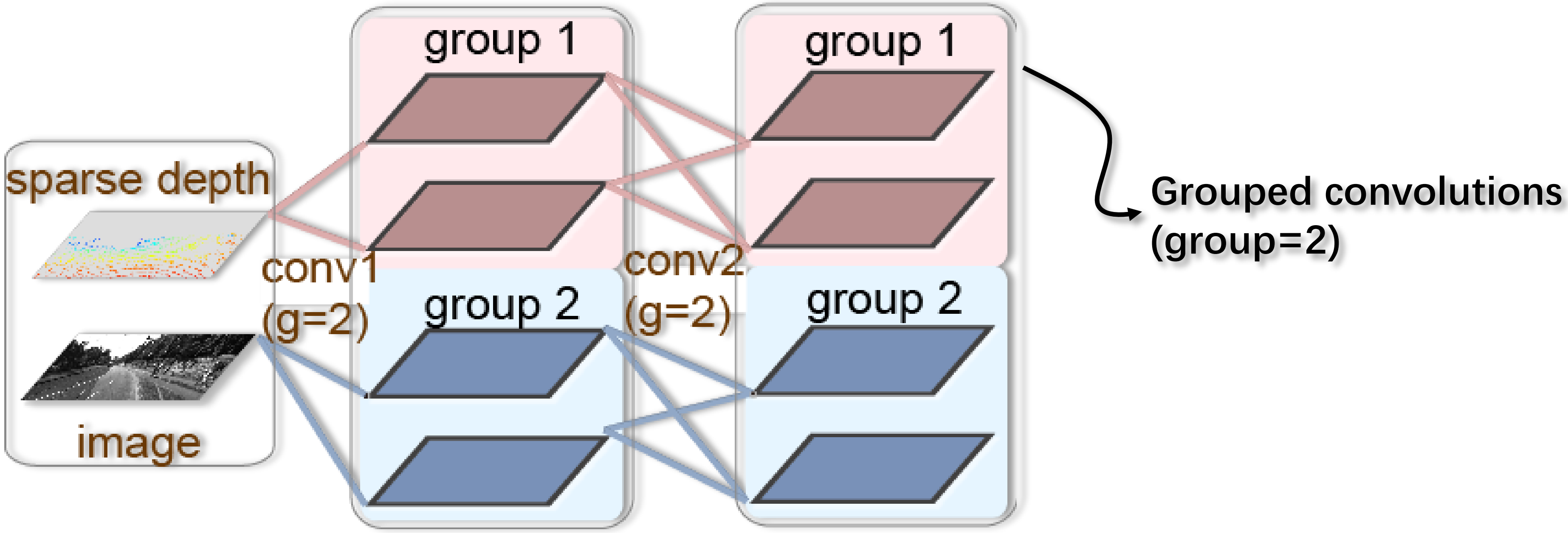
Large structures in the near areas with more supporting data

Fastcompletion: Multiscale Cascaded Networks



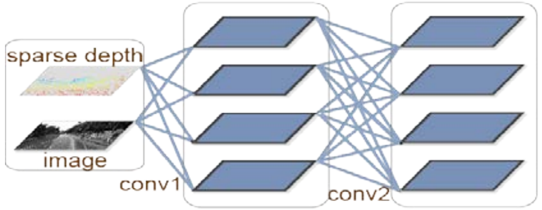
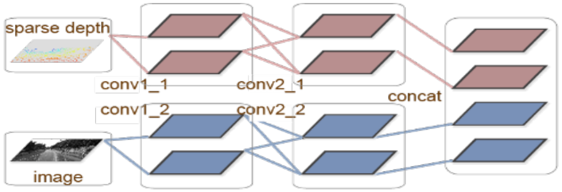
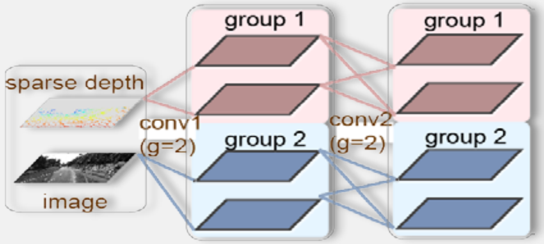
Each subnetwork is specialized for certain structures and has a lightweight architecture.

Fastcompletion: Grouped Fusion



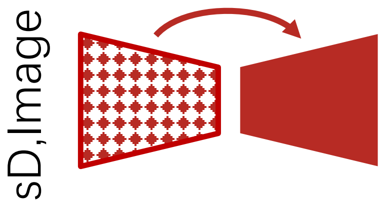
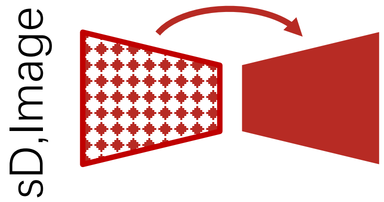
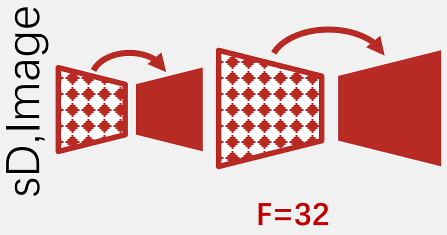
Grouped convolutions: extract depth and guidance features in parallel and fusing them naturally in the feature spaces.

Fastcompletion: Grouped Fusion

	Early fusion	**Late fusion**	**Grouped fusion**
Methods			
Complexity	$MHWK^2N$	$1/2MHWK2N$	$1/2MHWK^2N$
# Params	MK^2N	$1/2MK^2N$	$1/2MK^2N$
# Conv. layers	$1C$	$2C$	$1C$

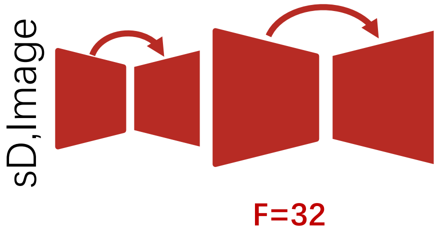
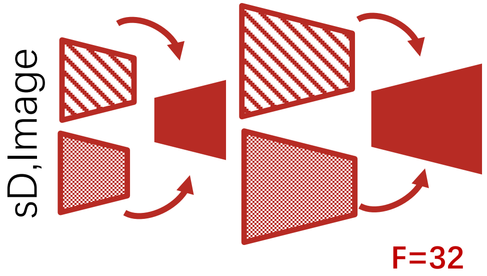
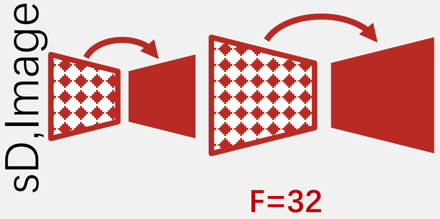
low computational complexity & high degree of parallelism.

Results

Methods	**Holistic (32 ch.)** 	**Holistic (512 ch.)** 	**Cascade (32 ch.)** 
MAE (mm)	298.40	263.54	**273.35**
RMSE(mm)	1024.49	967.41	**980.61**
#Params	106.78k	4.72M	**140.96k**
MACs	5.19	38.03	**5.05**
FPS	332.49	95.58	**270.43**

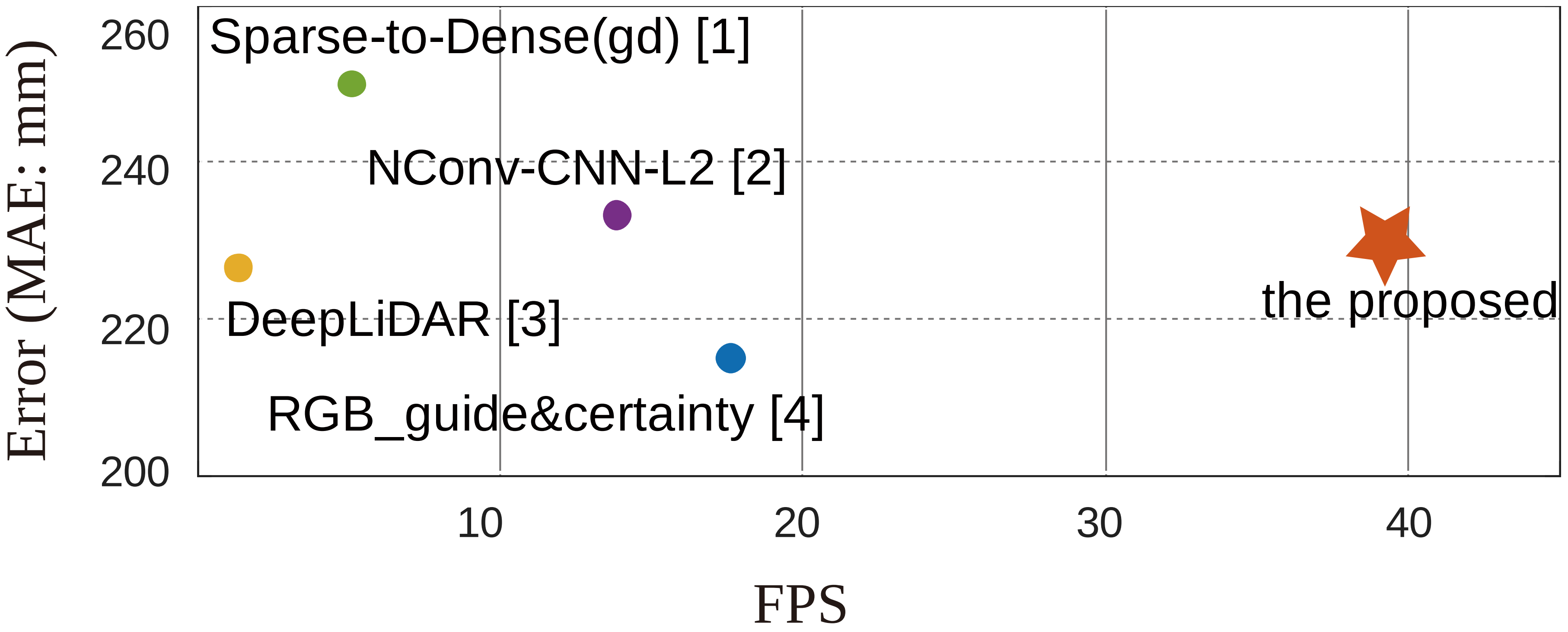
Cascaded networks result in high accuracy & low runtime.

Results

Methods	Early fusion	Late fusion	Grouped fusion
			
MAE (mm)	281.74	274.51	**273.35**
RMSE(mm)	1006.34	980.66	**980.61**
#Params	187.91k	**140.96k**	140.96k
MACs	6.29	**5.05**	5.05
FPS	261.63	203.79	**270.43**

Grouped fusion saves inference time without losing accuracy.

Results



A good trade-off between accuracy and speed .

[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.

Results

Over 39 FPS on an embedded GPU (NVIDIA Jetson AGX Xavier) .

Conclusion

Lightweight cascaded hourglass networks for diverse structures

Grouped fusion for efficiently extracting and fusing depth and guidance features

Feasibile for applications in realworld scenarios

For more details, pls refer to:

Ang Li, Zejian Yuan, Yonggeng Ling, Wanchao Chi, Shenghao Zhang, and Chong Zhang, ***FastCompletion****: A Cascade Network with Grouped Fusion Inputs for Real-time Depth Completion, 25th International Conference on Pattern Recognition (****ICPR****), Milan, Jan.10-15,* ***2021****.*