DeCoTR: Enhancing Depth Completion with 2D and 3D Attentions

Yunxiao Shi, Manish Kumar Singh, Hong (Herbert) Cai and Fatih Porikli

Qualcomm AI Research*

Qualcomm AI research

CVPR JUNE 17-21, 2024 SEATTLE, WA

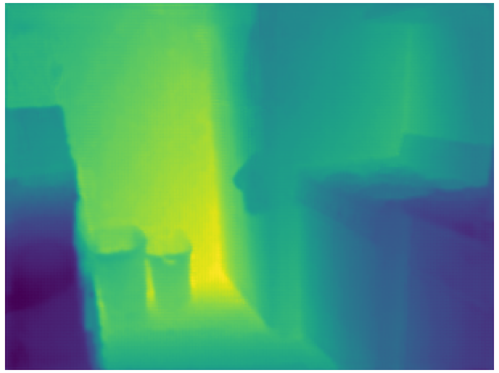
*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Image-guided depth completion

Sparse depth measurements + aligned image -> dense depth map

Aligned RGB image

Sparse depth measurements
(captured by Kinect)



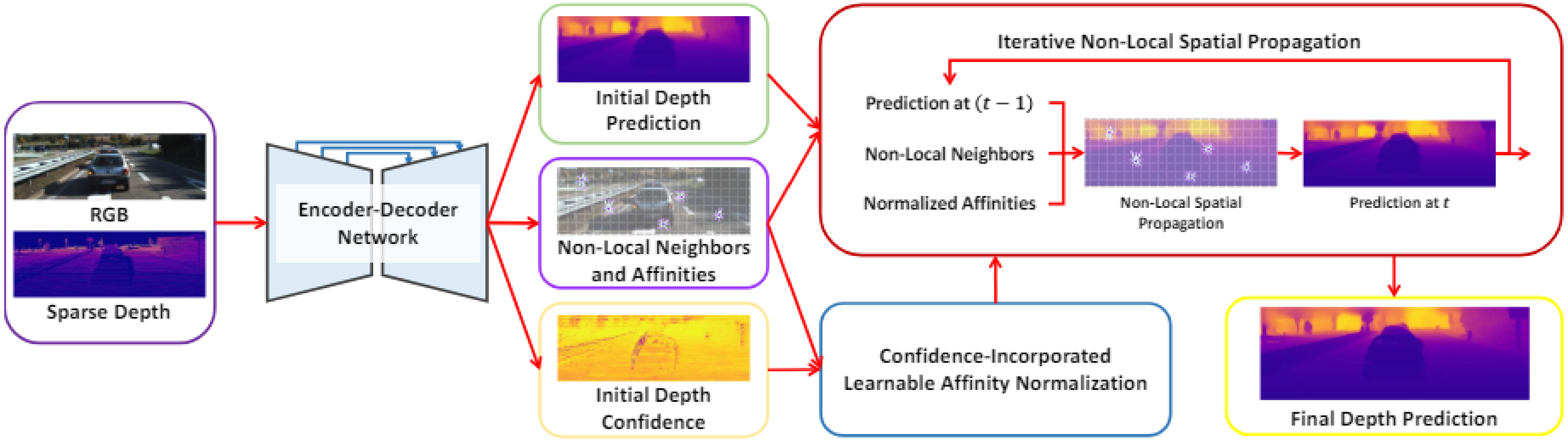
Completed dense depth

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Image-guided depth completion

Existing works rely on iterative spatial propagations as refinement

No 3D Geometry!
Edge Unfriendly!

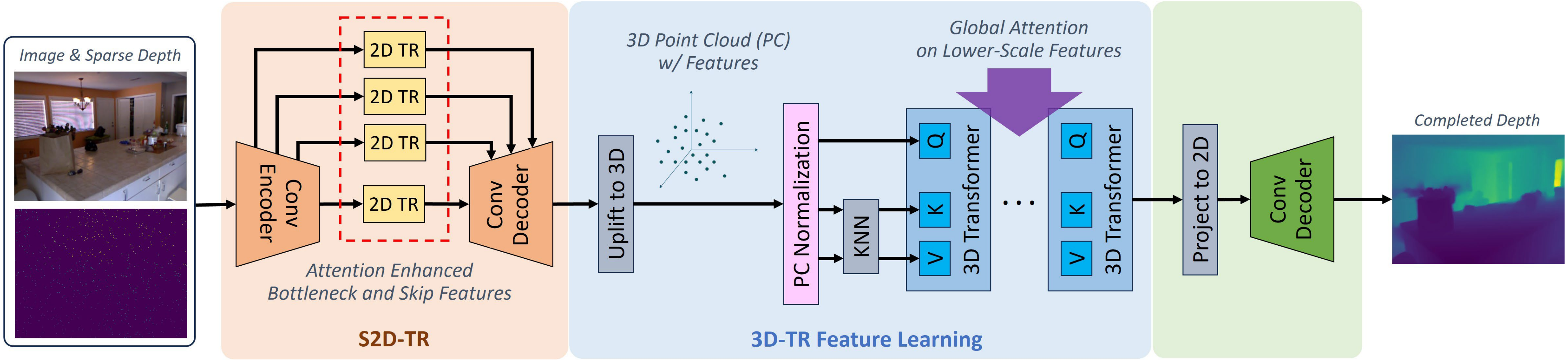


Park, Jinsun, et al. "Non-local spatial propagation network for depth completion.", *ECCV 2020.*

Our Method: DeCoTR

Fully transformer-based architecture

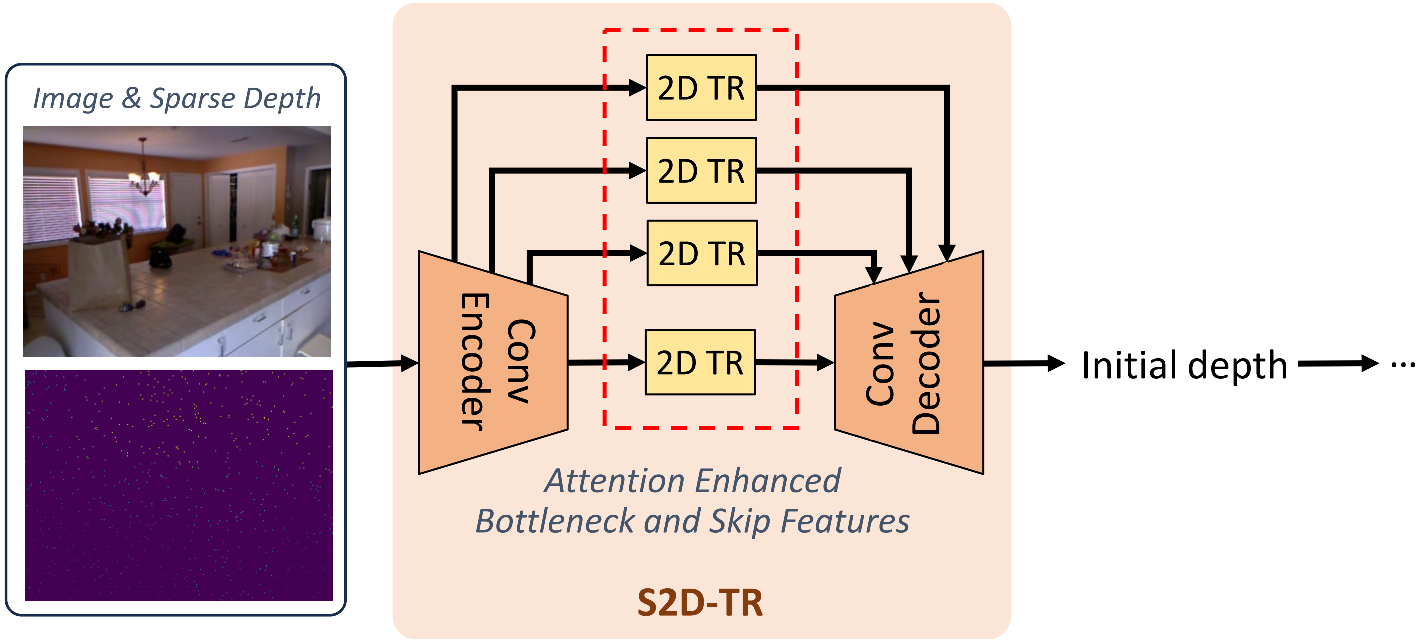
- No spatial propagation
- Efficient 2D attention to improve initial completed depth
- Powerful 3D point transformer as refinement baking in 3D Geometry



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Efficient 2D attention

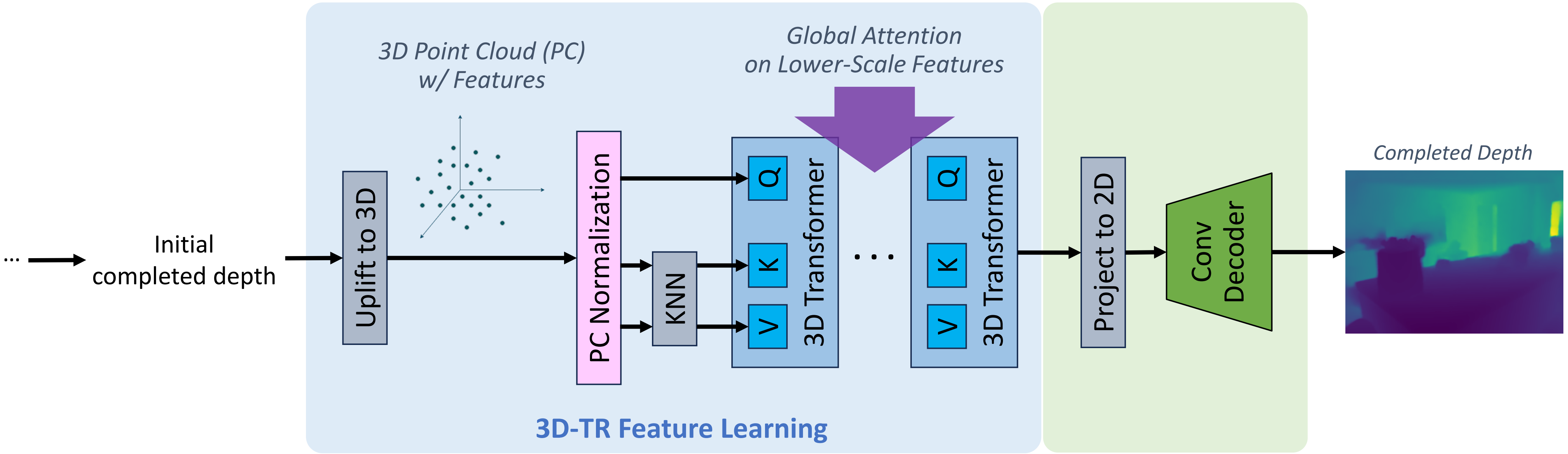
- Self-attention on down-sampled (via DS Conv) image features
- **Makes computing self-attention tractable -> better initial depth**



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Feature Cross-Attention in 3D

- Local vector cross-attention on uplifted point features
- Global dot-product attention on features of last-encoding stage
- **Effectively captures 3D geometry in learning**



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Evaluation

State-of-the-Art results on NYU Depth v2 and KITTI

Method	RMSE ↓	Abs Rel ↓	$\delta < 1.25$ ↑	$\delta < 1.25^2$ ↑	$\delta < 1.25^3$ ↑
S2D [27]	0.204	0.043	97.8	99.6	99.9
DeepLiDAR [31]	0.115	0.022	99.3	99.9	100.0
CSPN [4]	0.117	0.016	99.2	99.9	100.0
DepthNormal [41]	0.112	0.018	99.5	99.9	100.0
ACMNet [47]	0.105	0.015	99.4	99.9	100.0
GuideNet [37]	0.101	0.015	99.5	99.9	100.0
TWISE [19]	0.097	0.013	99.6	99.9	100.0
NLSPN [29]	0.092	**0.012**	99.6	99.9	100.0
RigNet [42]	0.090	0.013	99.6	99.9	100.0
DySPN [24]	0.090	**0.012**	99.6	99.9	100.0
CompletionFormer [45]	0.090	**0.012**	-	-	-
PRNet [23]	0.104	0.014	99.4	99.9	100.0
CostDCNet [20]	0.096	0.013	99.5	99.9	100.0
PointFusion [18]	0.090	0.014	99.6	99.9	100.0
GraphCSPN [26]	0.090	**0.012**	99.6	99.9	100.0
PointDC [44]	0.089	**0.012**	99.6	99.9	100.0
DeCoTR (ours)	0.087	**0.012**	99.6	99.9	100.0
DeCoTR w/ GA (ours)	**0.086**	**0.012**	99.6	99.9	100.0

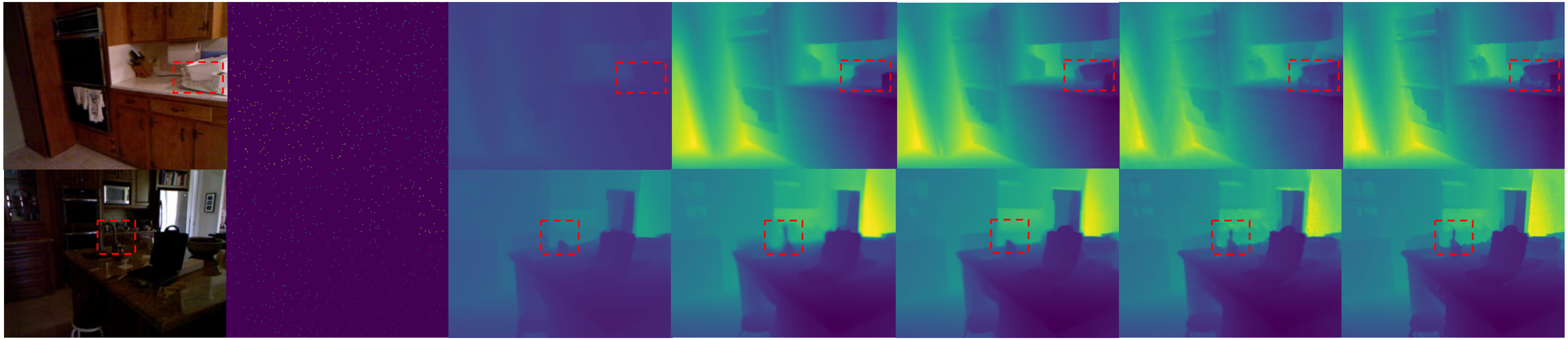
On NYUDv2

Method	RMSE ↓	MAE ↓	iRMSE ↓	iMAE ↓
CSPN [4]	1019.64	279.46	2.93	1.15
TWISE [19]	840.20	195.58	2.08	**0.82**
ACMNet [47]	744.91	206.09	2.08	0.90
GuideNet [37]	736.24	218.83	2.25	0.99
NLSPN [29]	741.68	199.59	1.99	0.84
PENet [17]	730.08	210.55	2.17	0.94
GuideFormer [32]	721.48	207.76	2.14	0.97
RigNet [42]	712.66	203.25	2.08	0.90
DySPN [24]	709.12	**192.71**	**1.88**	**0.82**
CompletionFormer [45]	**708.87**	203.45	2.01	0.88
PRNet [23]	867.12	204.68	2.17	0.85
FuseNet [3]	752.88	221.19	2.34	1.14
PointFusion [18]	741.9	201.10	1.97	0.85
GraphCSPN [26]	738.41	199.31	1.96	0.84
PointDC [44]	736.07	201.87	1.97	0.87
DeCoTR (ours)	717.07	195.30	1.92	0.84

On KITTI

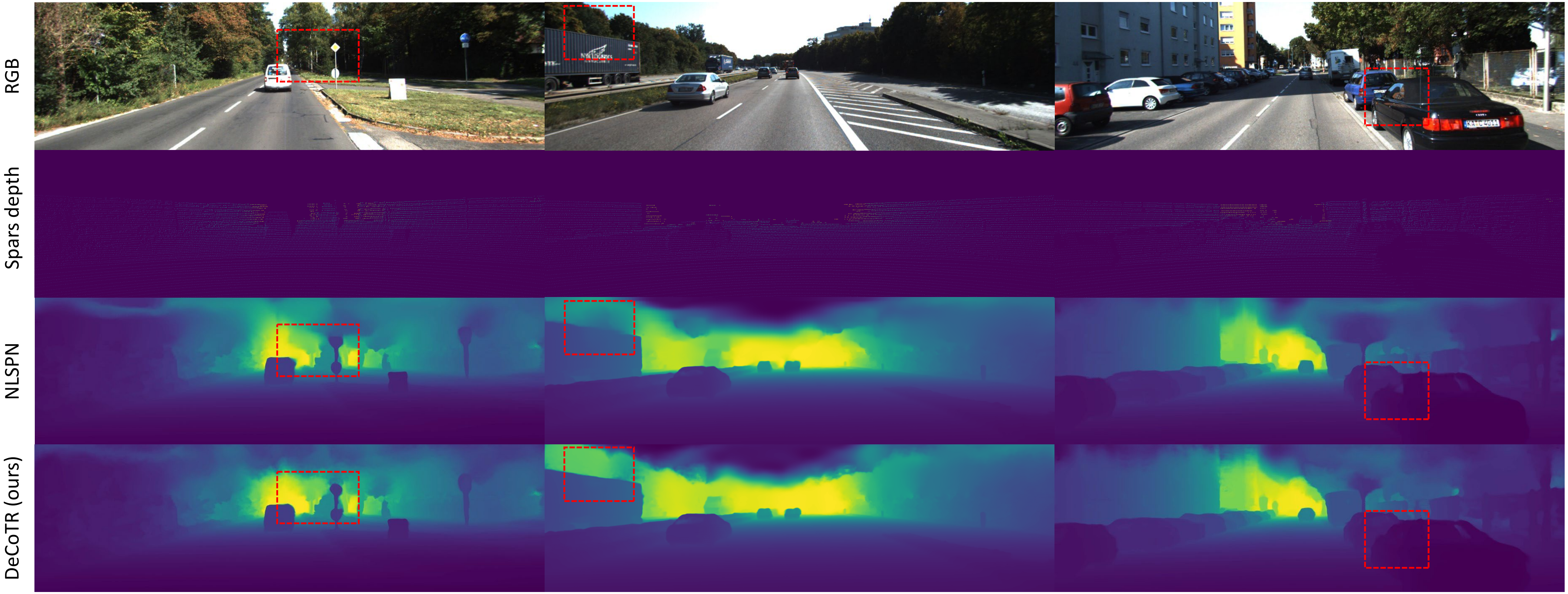
Visualization – NYUDv2

RGB | Sparse depth | NLSPN | GraphCSPN | CompletionFormer | DeCoTR (ours) | GT



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

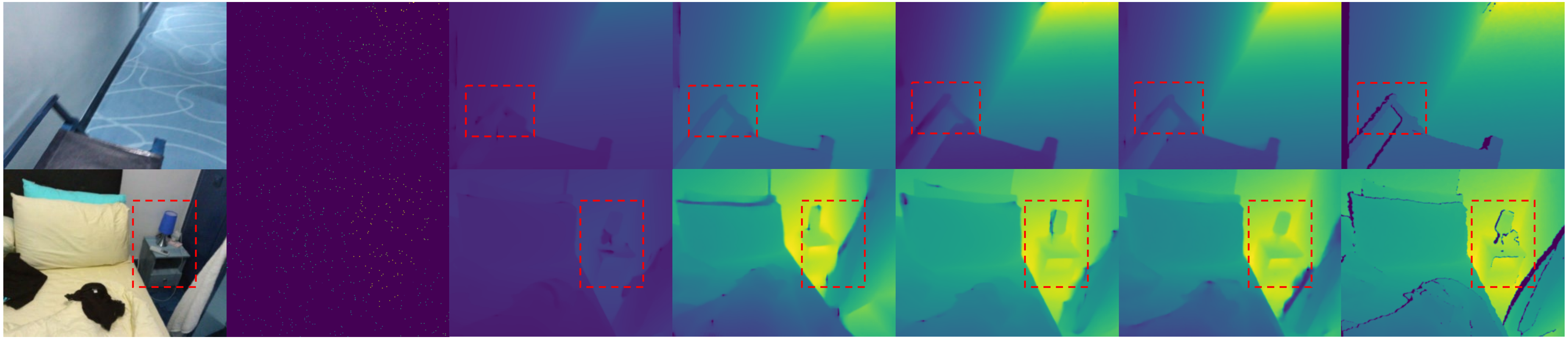
Visualization - KITTI



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Cross-domain zero-shot inference

State-of-the-Art zero-shot inference performance on ScanNet-v2 and DDAD



Visualization on ScanNet-v2

Method	RMSE ↓	$\delta < 1.25$ ↑
NLSPN [29]	0.198	97.3
GraphCSPN [26]	0.197	97.3
CompletionFormer [45]	0.194	97.3
DeCoTR (ours)	**0.188**	**97.6**

On ScanNet-v2

Method	RMSE ↓	MAE ↓
NLSPN [29]	701.9	309.6
CompletionFormer [45]	889.3	400.1
DeCoTR (ours)	399.2	263.1

On DDAD

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Conclusion

- Exploiting 3D geometry leveraging attention is an effective way to improve image-guided depth completion.
- Our method, DeCoTR, achieves SotA results in both indoor and outdoor scenes.
- Superior zero-shot inference performance on unseen datasets is also observed.