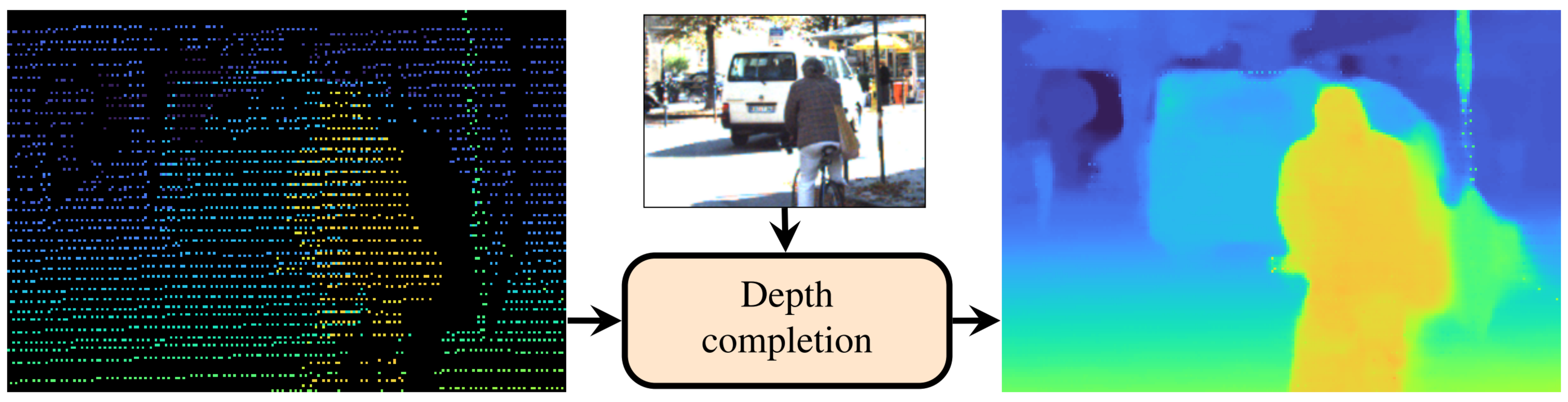
Temporal Lidar Depth Completion

Pietari Kaskela Philipp Fischer Timo Roman

NVIDIA Applied Deep Learning Research

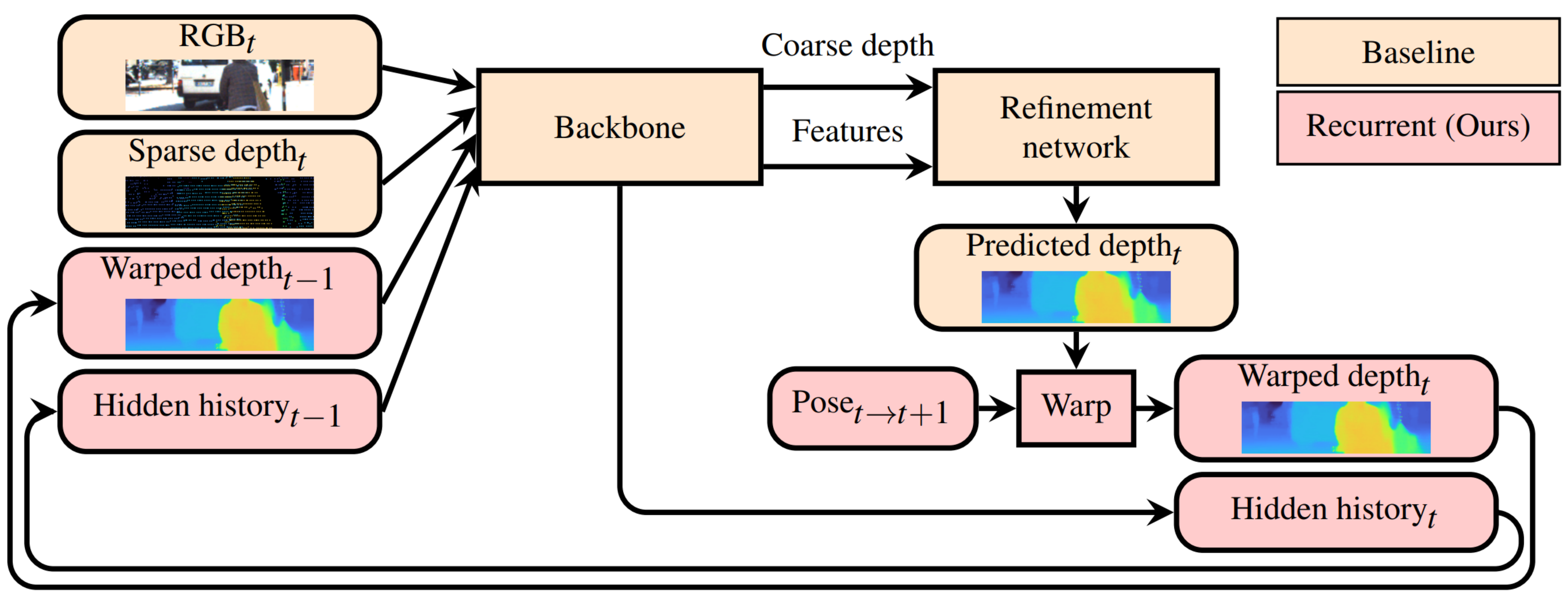
Introduction

- Depth completion: **Infill and interpolate a sparse depth image to a dense depth image,** using an RGB image as a guide.



- Most SOTA approaches are **non-temporal** and use a U-Net-style backbone followed by a spatial propagation refinement network.
  - PENet [1], SemAttNet [3], DySPN [2]
- **We propose a recurrent depth completion architecture**, which is able to effectively combine information from multiple timesteps of input.

Method

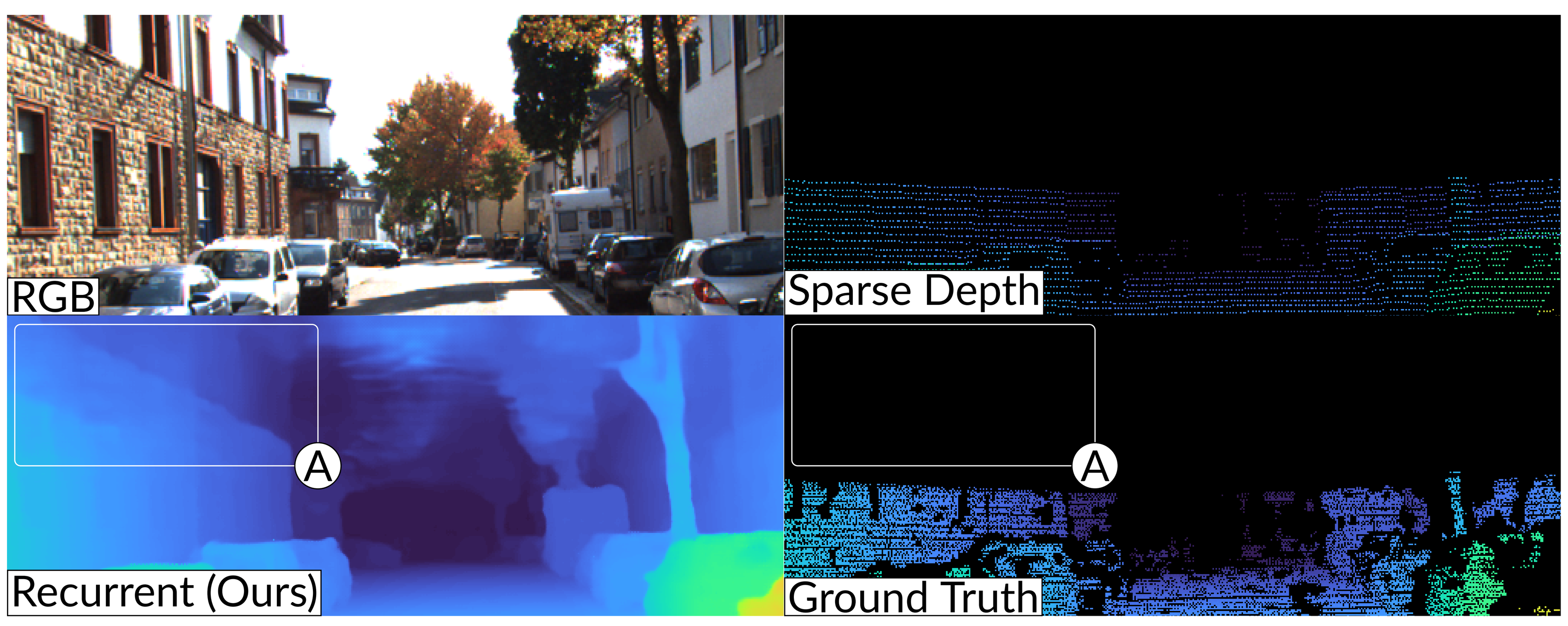


- We build on the open-source PENet [1] which consists of a U-Net-like backbone followed by a spatial propagation refinement network.
- We introduce recurrency with **warped previous depth** and **hidden history** as input to the network from the previous timestep.
  - The warping/reprojection is performed using the corresponding pose matrix between the timesteps.
  - Hidden history is a single output channel from the last convolution of the U-Net backbone.
- Temporally-aware training is performed using truncated backpropagation through time (TBPTT).
  - TBPTT(k_1,k_2): k_1 = weight update interval, k_2 = backpropagation length

Configuration	RMSE ↓ (mm)	MAE ↓ (mm)
Baseline	773.9±3.2	218.0±0.8
Prev. Depth, TBPTT(1,1)	762.4 (-11.5)	215.1
Prev. Depth, TBPTT(1,2), Hidden	758.5 (-15.4)	214.2
Warped Prev. Depth, TBPTT(1,1)	728.7 (-45.2)	204.9
Warped Prev. Depth, TBPTT(1,2), Hidden	**720.8 (-53.1)**	203.5

Table 1. Ablation metrics for the full KITTI depth completion validation set.

Dataset



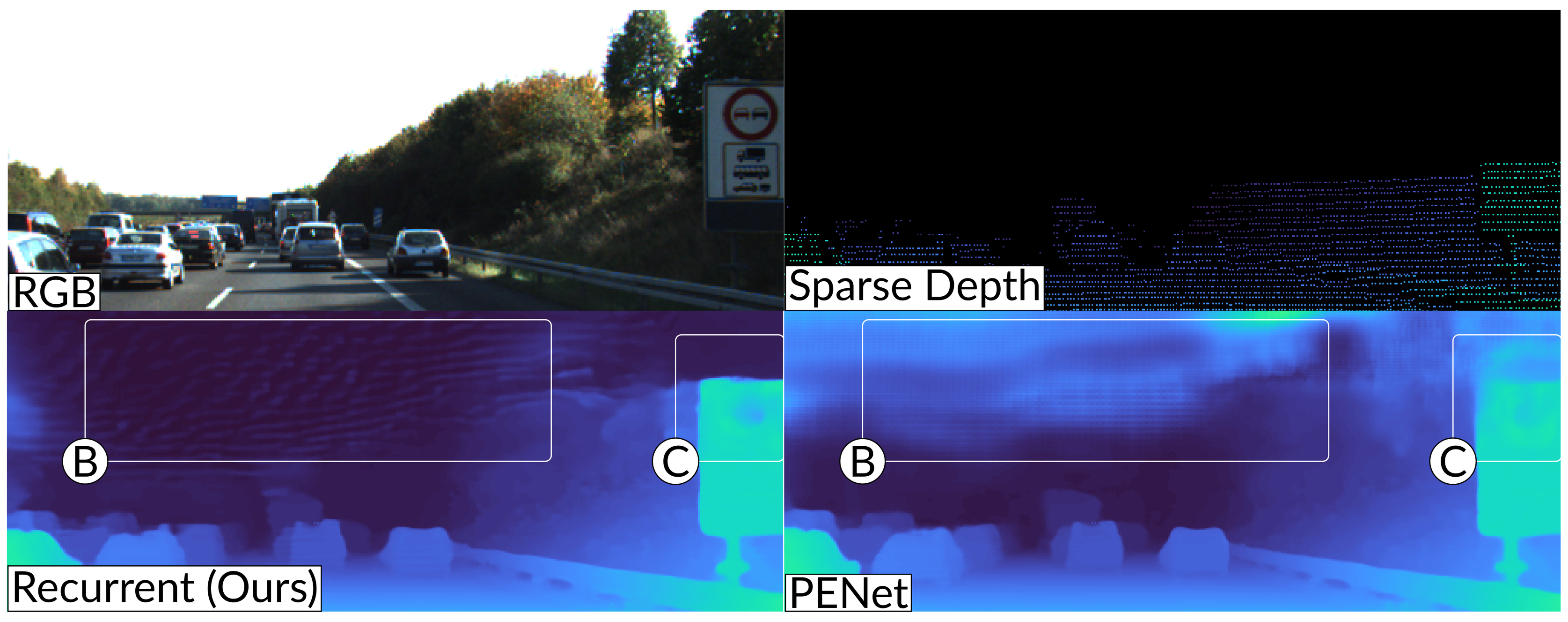
- The most popular benchmark is **KITTI depth completion** with ∼94k images. Sparse depth input contains 6% valid depth values and ground truth contains 16% valid depth values.

Results

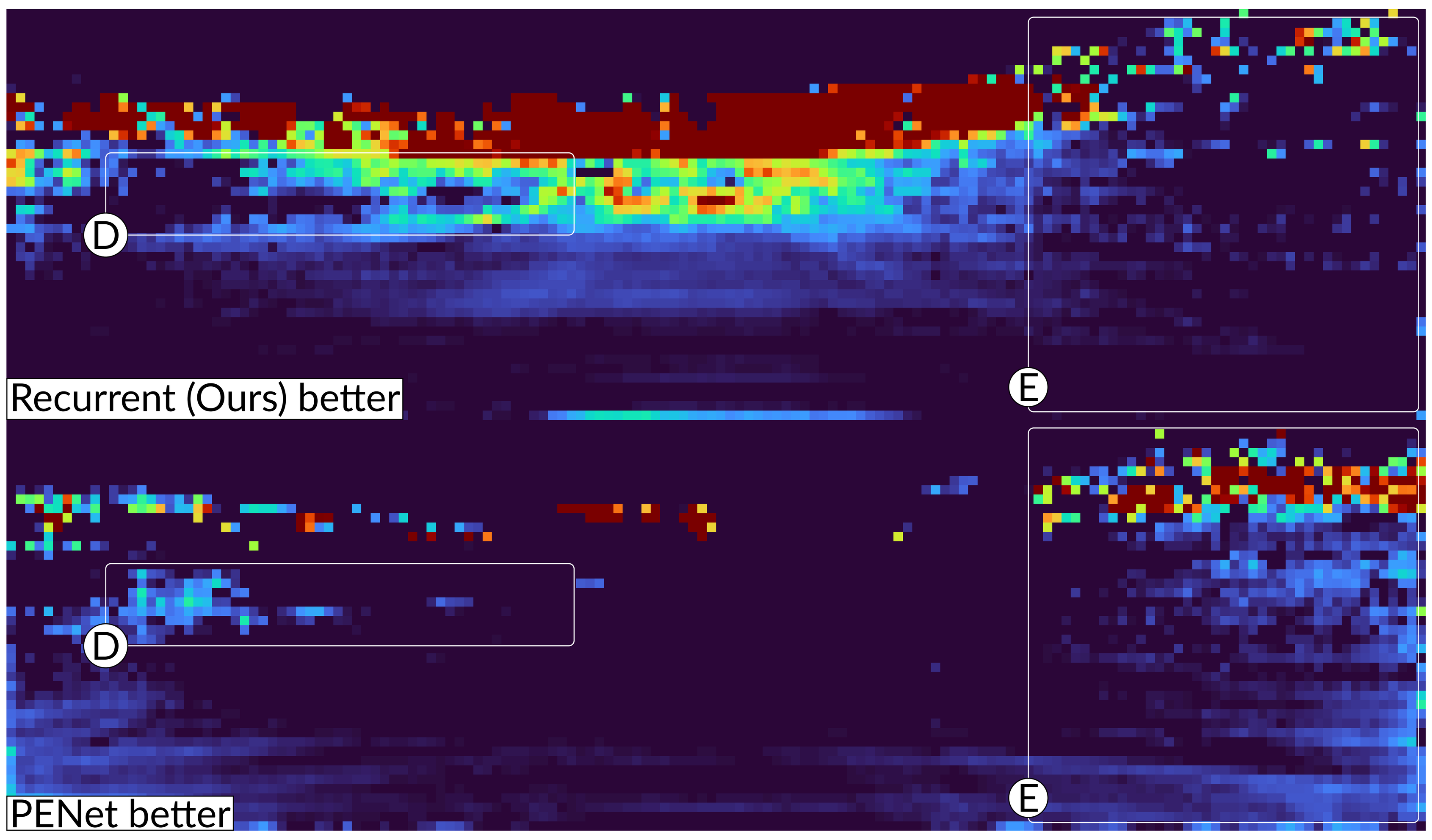
Model	RMSE ↓ (mm)	MAE ↓ (mm)
PENet [1]	757.2	209.0
DySPN [3]	739.4	**191.4**
SemAttNet [2]	738.1	204.5
Recurrent (Ours)	**722.2**	204.0

Table 2. Comparison to SOTA methods on the KITTI depth completion validation set.

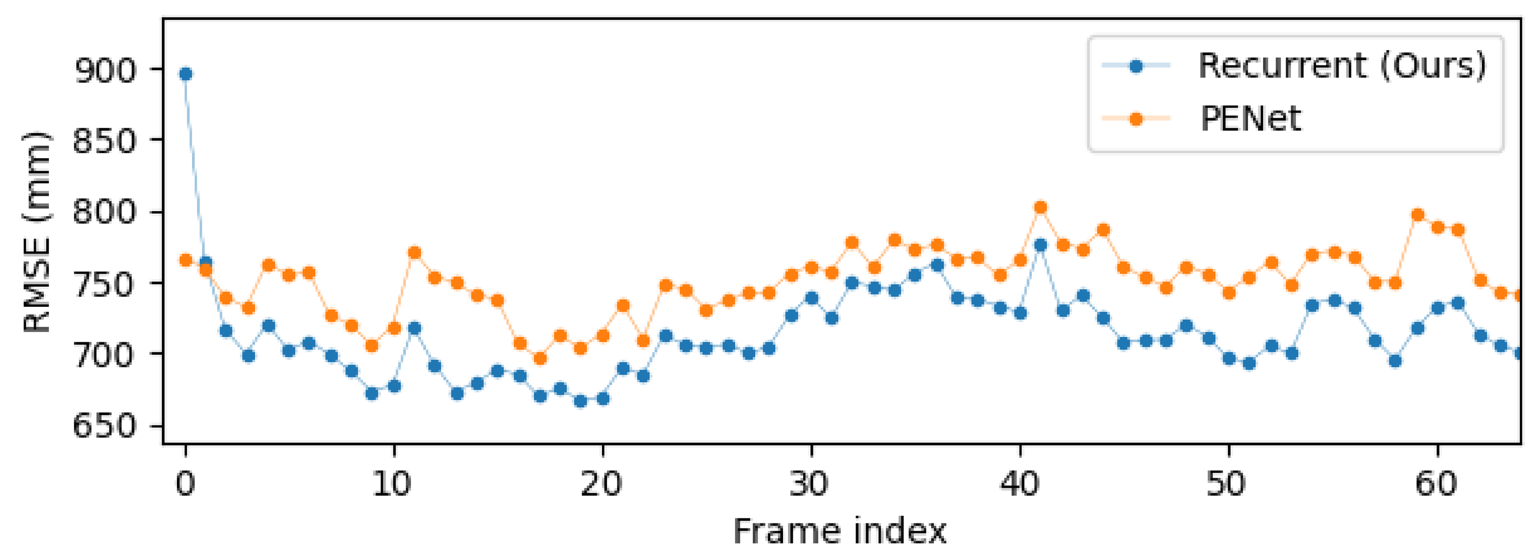
- Our method achieves a **new SOTA result on the KITTI depth completion validation set**.
  - Sequence data and pose information are required, which the test set does not contain. These are available in a real-world setting.
- We also observe **a large improvement in regions which do not contain ground truth or input depth** in the current timestep.



- Our method excels in regions with sparse sampling, but doesn't lead to much improvement in regions where warping is incorrect.
  - Box (D) highlights a region with cars often moving on the opposing lane.



- The recurrent method is worse on the first timestep when still uninitialized, but on average 50 RMSE better after the second timestep.



References

[1] Mu Hu et al. "Towards Precise and Efficient Image Guided Depth Completion". 2021.

[2] Yuankai Lin et al. *Dynamic Spatial Propagation Network for Depth Completion.* 2022.

[3] Danish Nazir et al. "SemAttNet: Towards Attention-based Semantic Aware Guided Depth Completion". 2022.

 pkaskela@nvidia.com