Toward Practical Monocular Indoor Depth Estimation

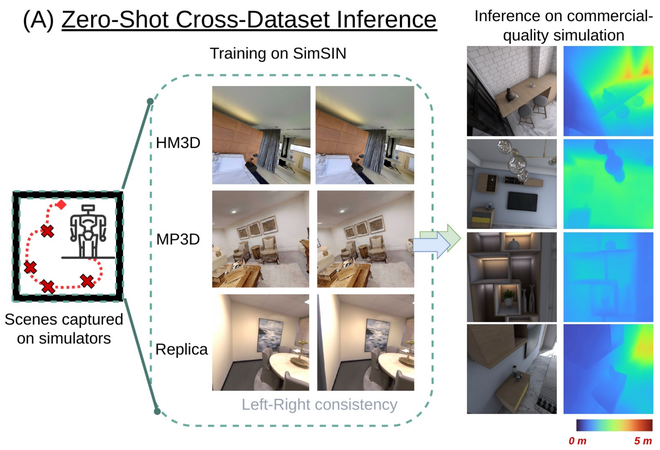
Cho-Ying Wu[1], Jialiang Wang[2], Michael Hall[2], Ulrich Neumann[1], Shuochen Su[2]

[1]University of Southern California

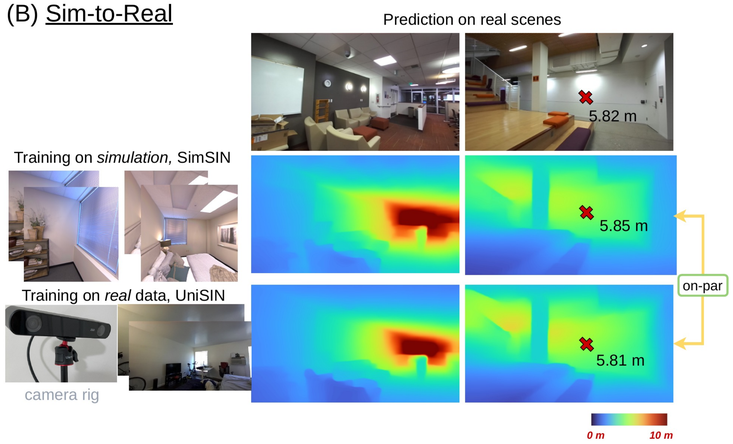
[2]Meta Reality Labs,

Advantages

(A) Zero-Shot Cross-Dataset Inference



(B) Sim-to-Real



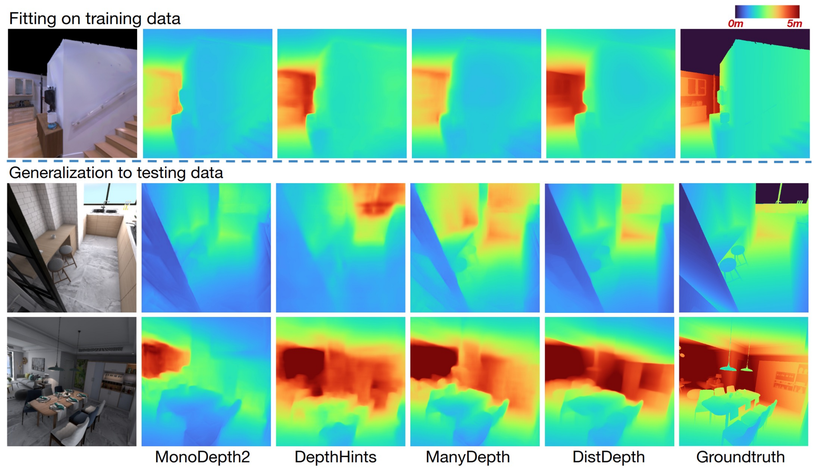
Practical Depth Estimation

- Learning without curated depth groundtruth
- Efficient and effective data collection by simulator
- High generalizability and accurate and real-time performance

Motivation

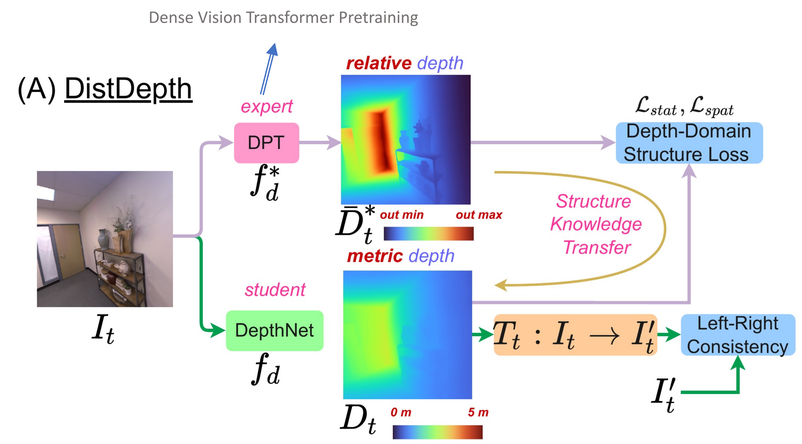
- Most prior works for self-supervised depth focus on outdoor driving scenes. However, the applicability is yet to be investigated.
- Either cross-data evaluation and sim-to-real evaluation for indoor scenes has not explored much.

Fitting on Training Scenes

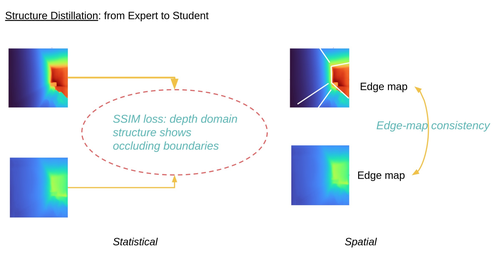


Prior left-right consistency methods can fit well on training scenes, but they do generalize to testing scenes with complex structures.

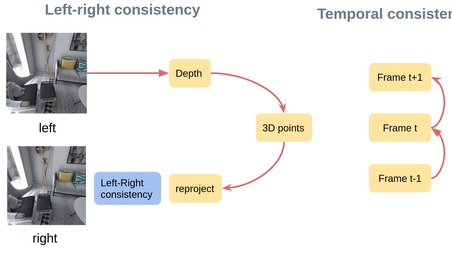
Method



(1) Depth-Domain Structure Loss:

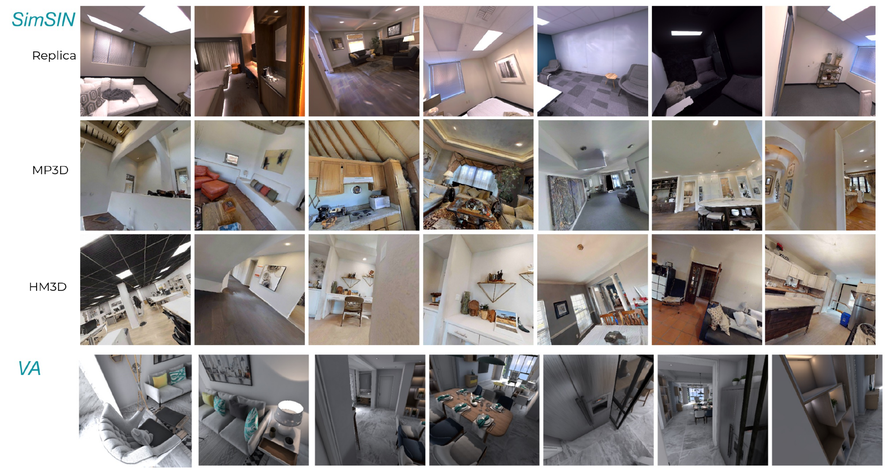


(2) Left-Right Consistency:

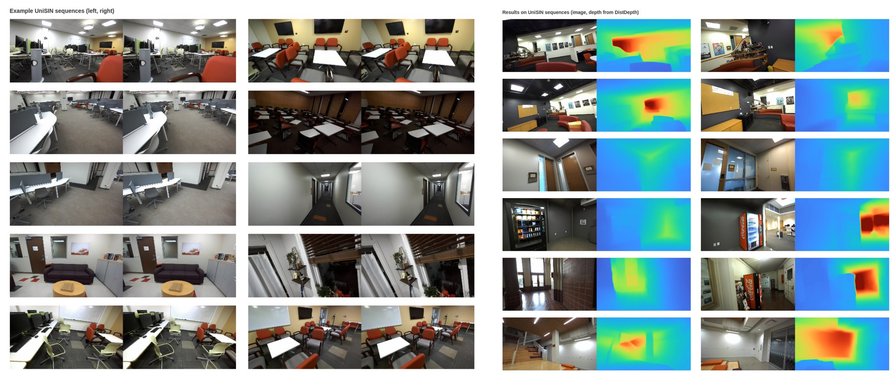


Combining metrics from left-right pairs and structures from relative depth pretraining

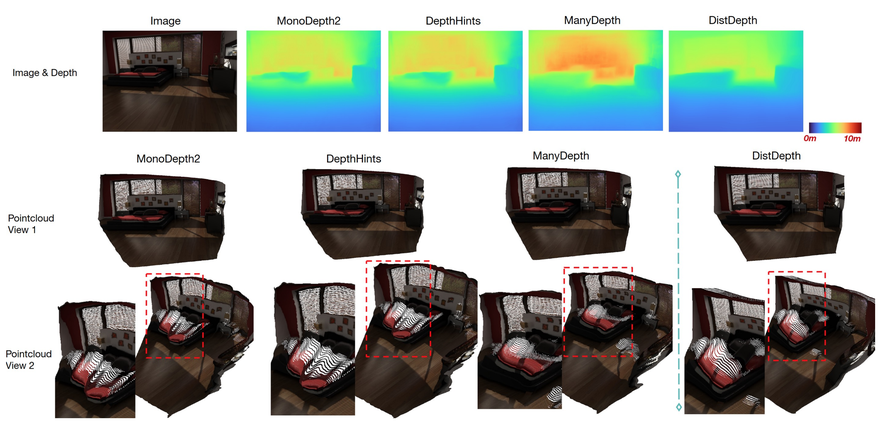
Collected Datasets and Results



UniSIN



Results



Take a photo to learn more:

SCAN ME