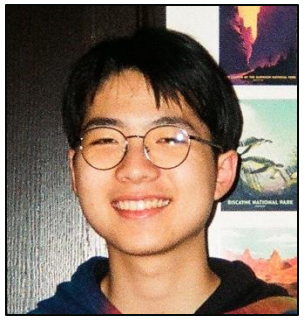
Infinigen for Zero-shot Stereo Matching

David Yan

PRINCETON
VISION & LEARNING LAB


David Yan


Alex Raistrick


Jia Deng

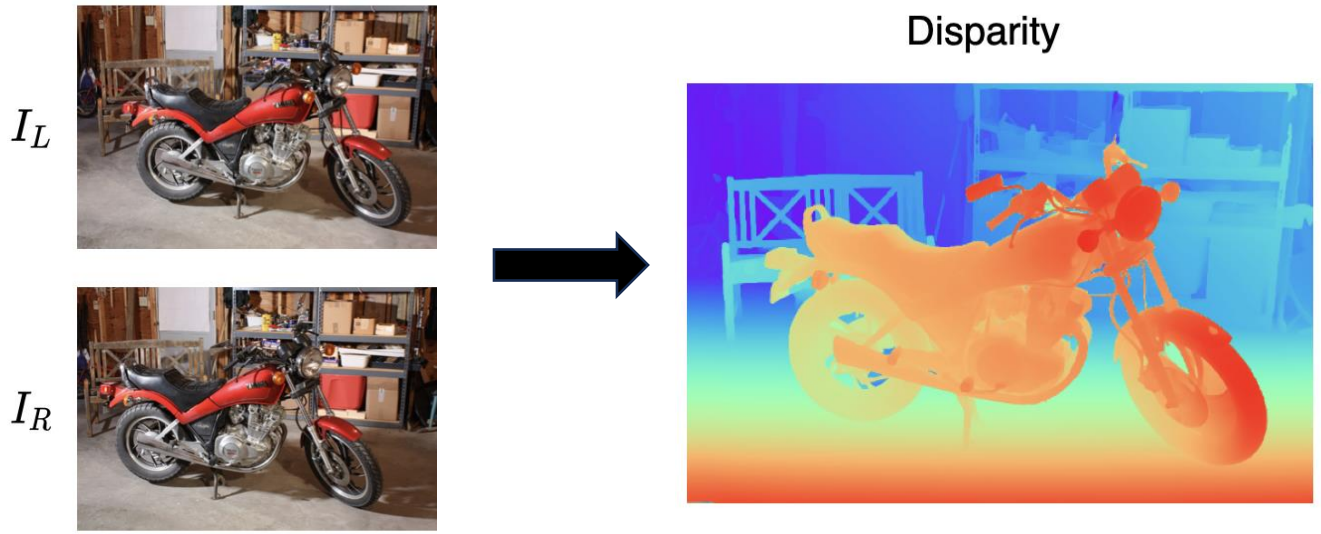
Stereo Depth Estimation Recap

Classic task

- Recover dense 3D from a rectified stereo image pair
- Equivalent to pixel correspondence (disparity)

Classic recipe

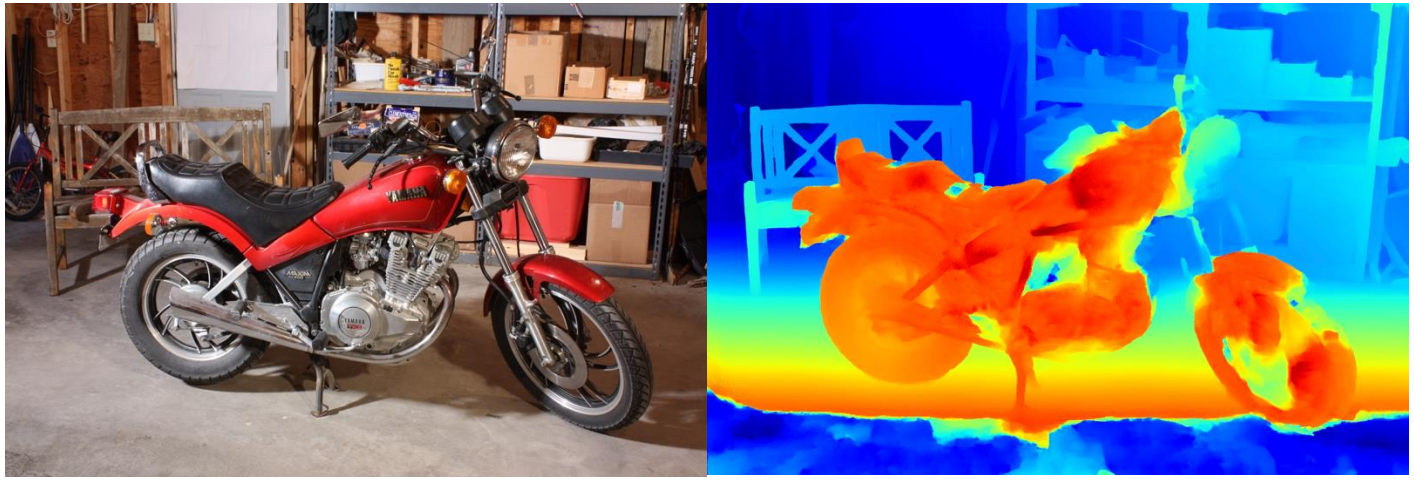
- Deep neural networks
- Labeled data



RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching [3DV 2021]

Robust stereo remains challenging

- We still train models for each benchmark
  - Including the benchmark train set!
- Zero-shot transfer is a better proxy for in-the-wild performance



Failure case of cross-dataset transfer from KITTI to Middlebury

Why study data for zero-shot stereo?

A lesson from mono-depth (MiDaS):

- A well-curated dataset is key to robust performance

A typical stereo network will be trained on

- >600k synthetic stereo pairs
- < 3k real-world stereo pairs

Let’s take a hard look at our synthetic datasets!

Synthetic data has a massive design space

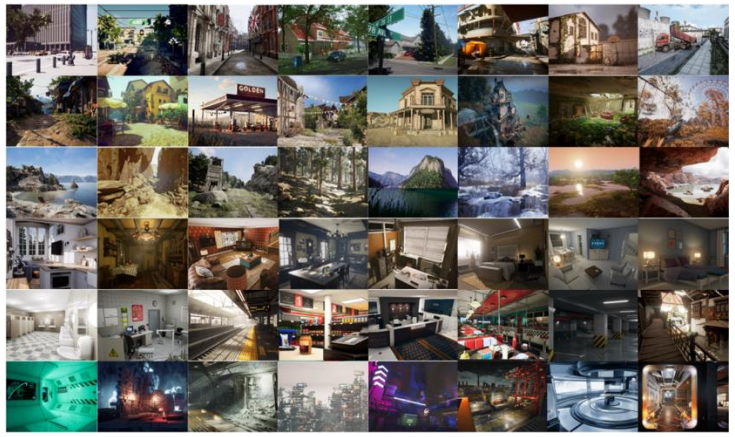
High-level Design Choices:

- Random object placement
- Artist-designed scenes
- Domain-specific simulators

Low Level Design Choices:

- How many floating objects?
- What materials?
- What lighting?
- …

FlyingThings3D (Sceneflow)



TartanAir

VirtualKITTI

Dataset design is not well understood

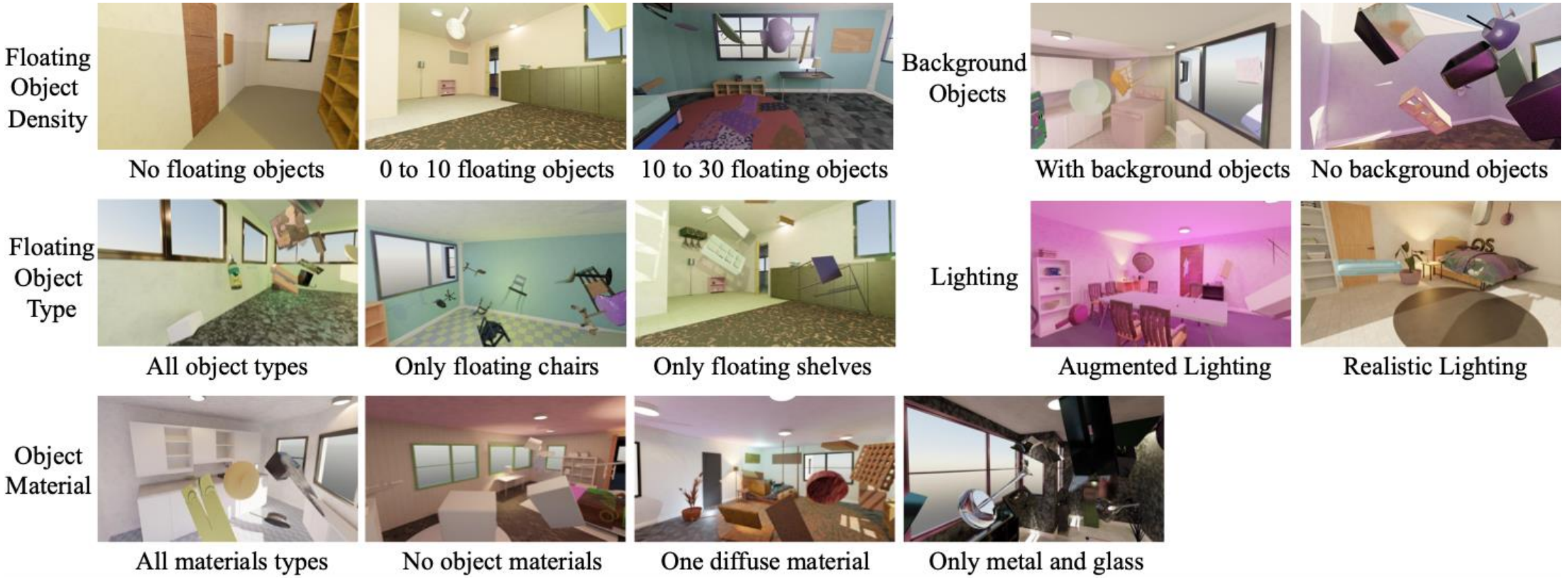
Lots of new datasets in the last ~10 years…

But almost no ablation studies on dataset design!

What makes a good synthetic stereo dataset?

Dataset Ablation

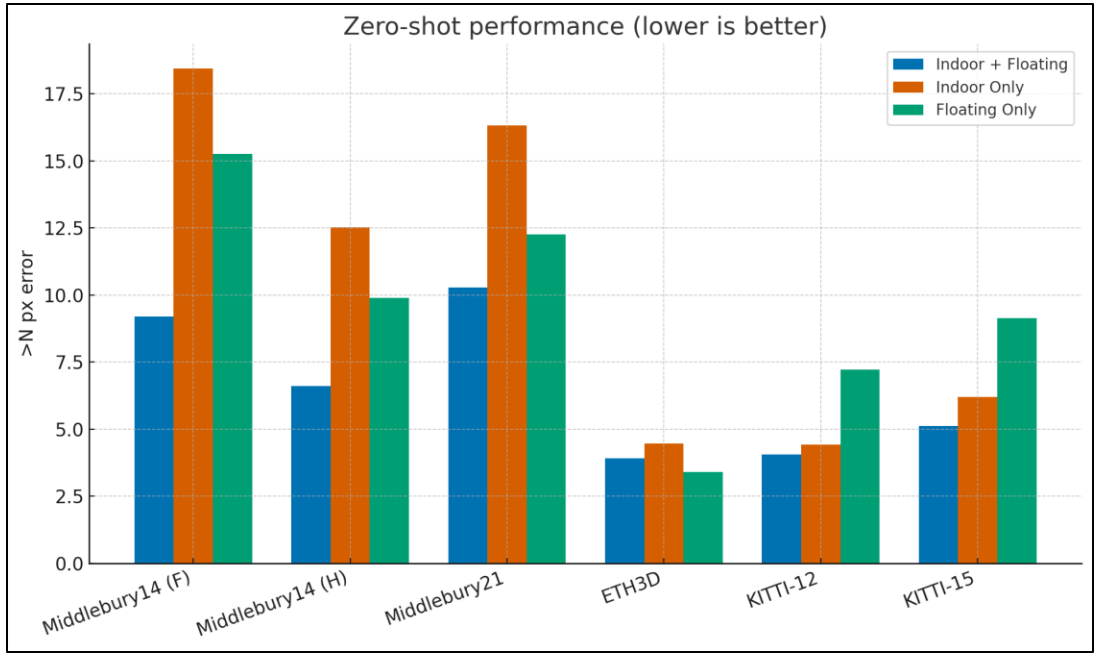
1. Render small datasets using different Infinigen settings
2. Train stereo networks and examine zero-shot performance



We discover better arrangements

Combine floating objects with realistic scenes!

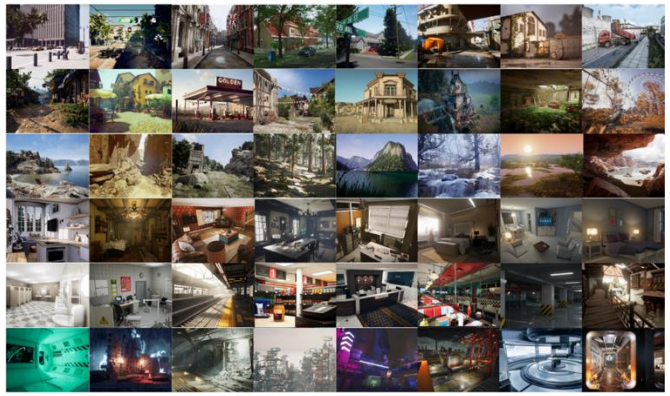
- More effective than realistic scenes or floating objects alone



We discover better arrangements

Hypothesis:

- Flying objects are extremely sample efficient, but have large domain gap
- Realistic scenes reduce domain gap, but are sample inefficient
- Combining the two gets the best of both worlds

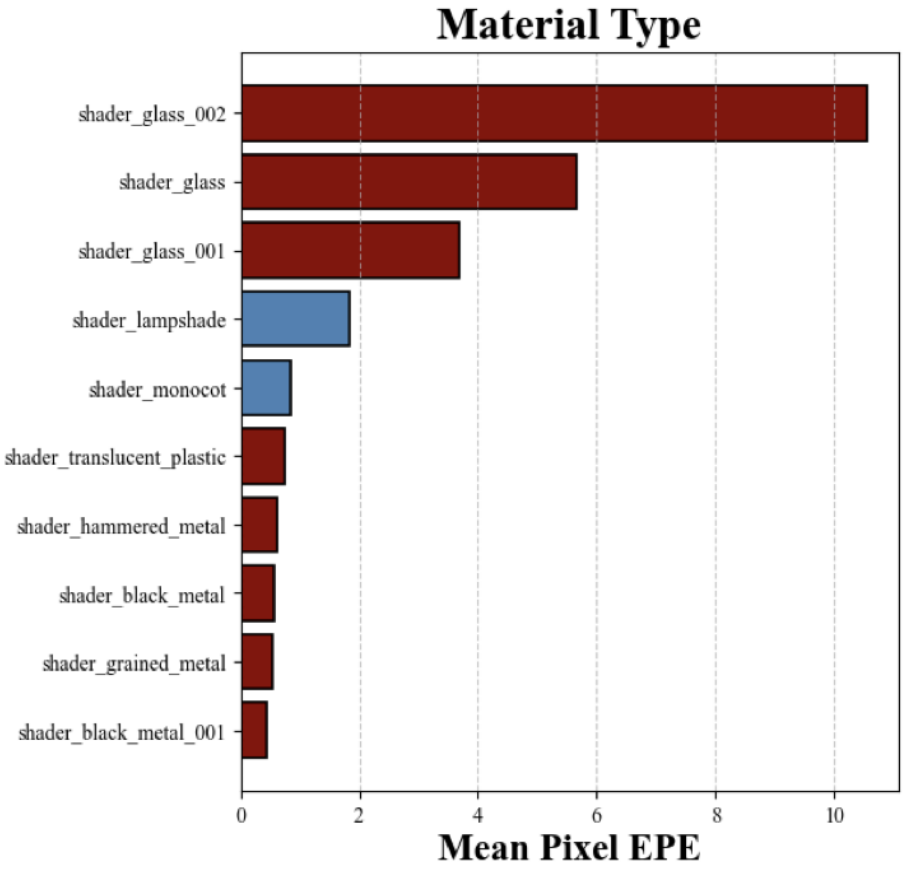




Current dataset designs can hurt performance

Training on very reflective and transparent objects hurts performance on diffuse regions

Hypothesis:

- Most models have insufficient inductive bias for these ill-posed surfaces
- Recent work on mono-depth integration might help mitigate this trade-off



Novel Dataset

Construct full-scale (~150k) dataset

- Use parameters found in ablation study
- Performance optimizations to make generation cost tractable

Indoor Floating Objects

Dense Floating Objects

Nature

Zero-shot Benchmark Performance

Our data beats a combined suite of existing synthetic datasets

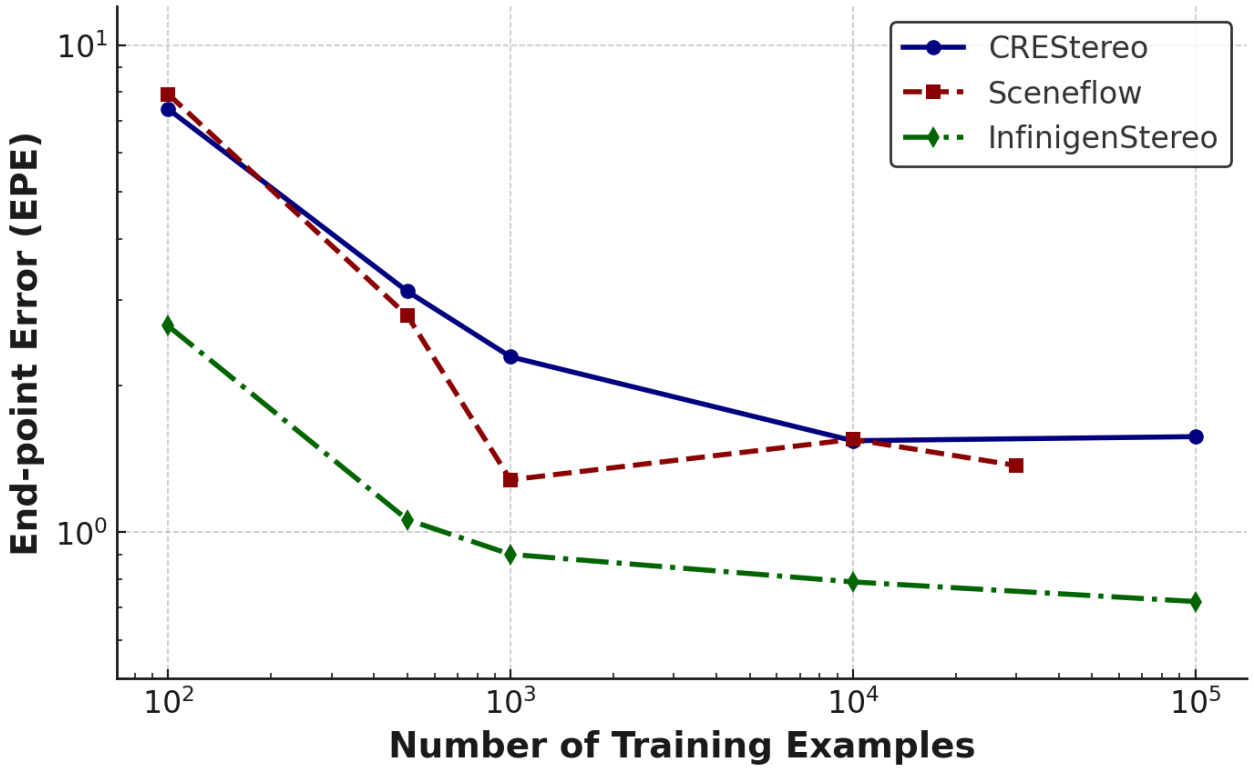
- Mixed Sceneflow+CREStereo+TartanAir+IRS (~640k)
- Ours (~150k)

Method	Middlebury 2014				Middlebury 2021		ETH3D		KITTI				Booster (Q)	
	H		F		-		-		2012		2015		-	
	2px	EPE	2px	EPE	2px	EPE	1px	EPE	3px	EPE	3px	EPE	2px	EPE
RAFT-Mixed	5.50	0.75	10.4	3.09	8.97	1.34	**2.58**	**0.26**	3.64	0.80	4.95	1.04	11.46	3.18
RAFT-Infinigen	**4.48**	**0.62**	**9.4**	**2.16**	**8.17**	**1.26**	2.93	**0.26**	**3.25**	**0.75**	**4.25**	**0.98**	**9.17**	**2.05**
DLNR-Mixed	5.21	**0.76**	9.31	2.15	9.30	1.36	**2.50**	0.25	3.68	0.87	4.95	1.08	12.17	2.75
DLNR-Infinigen	**3.85**	**0.76**	**6.16**	**2.06**	**7.23**	**1.08**	2.63	**0.23**	**3.33**	**0.77**	**4.60**	**1.01**	**8.75**	**1.83**
Selective-IGEV-Mixed	5.24	0.85	9.11	3.98	8.24	1.57	**2.37**	0.31	3.97	0.86	5.31	1.12	11.00	3.04
Selective-IGEV-Infinigen	**3.61**	**0.72**	**6.00**	**1.98**	**7.62**	**1.22**	2.47	**0.24**	**3.26**	**0.77**	**4.55**	**1.03**	**8.84**	**2.01**

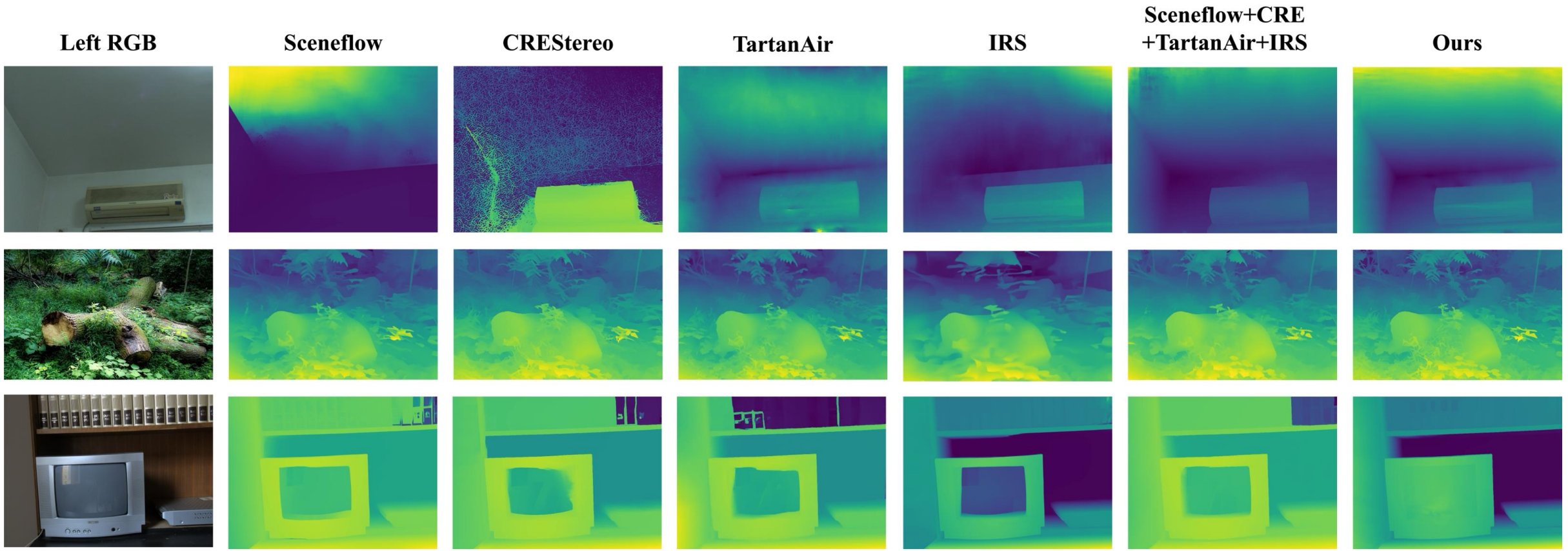
Scaling Results

Our data scales better when controlling for dataset size

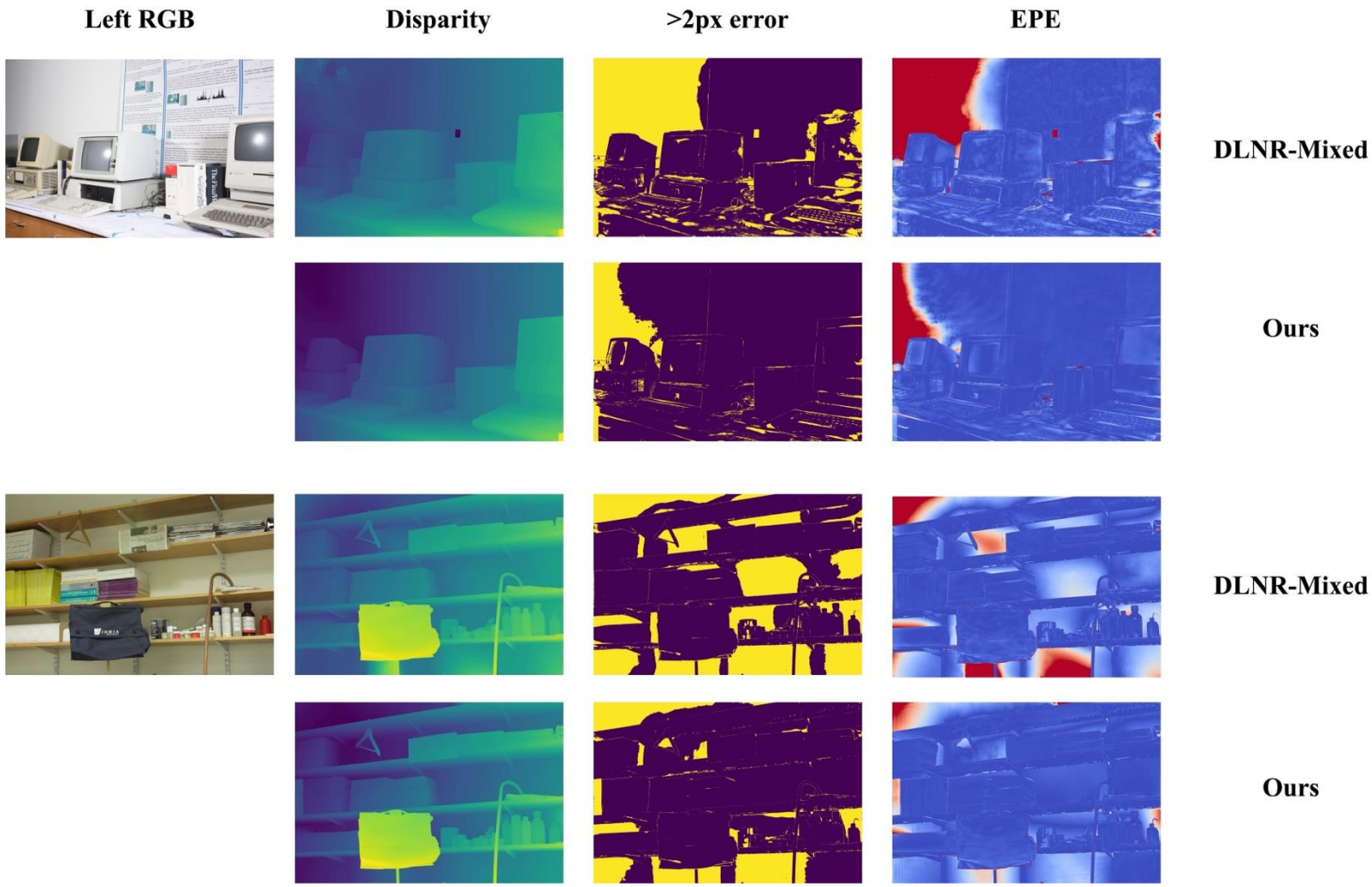
- 500 Infinigen > 100,000 CREStereo



Qualitative Results



Qualitative Results



What have we learned?

- Careful study of data can lead to important insights
- It is worth our time to ablate dataset design like we do for model design
- *Controllable* synthetic data enables us to perform such studies
  - Our stereo-tuned Infinigen checkpoint is open-source
  - We hope others will build on our work!

Infinigen as a tool to study data

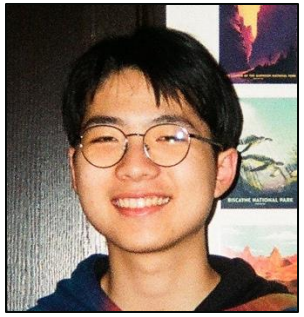
- Controllability
  - Easy ablation setup
- Fine-grained error analysis
  - Material and Object Segmentation Maps
- Scalable Realism
  - Over 1800 unique background scenes
- Open-source generation code
  - Purchasing artist 3D scenes typically does not give redistribution license

Procedural Dataset Generation for Zero-Shot Stereo Matching

David Yan
Princeton University
yan.david@princeton.edu

Alexander Raistrick
Princeton University
araistrick@princeton.edu

Jia Deng
Princeton University
jiadeng@princeton.edu



↑
applying to PhDs this fall!