Smoothness, Synthesis, and Sampling: Re-thinking Unsupervised Multi-View Stereo with DIV Loss

Alex Rich

Noah Stier

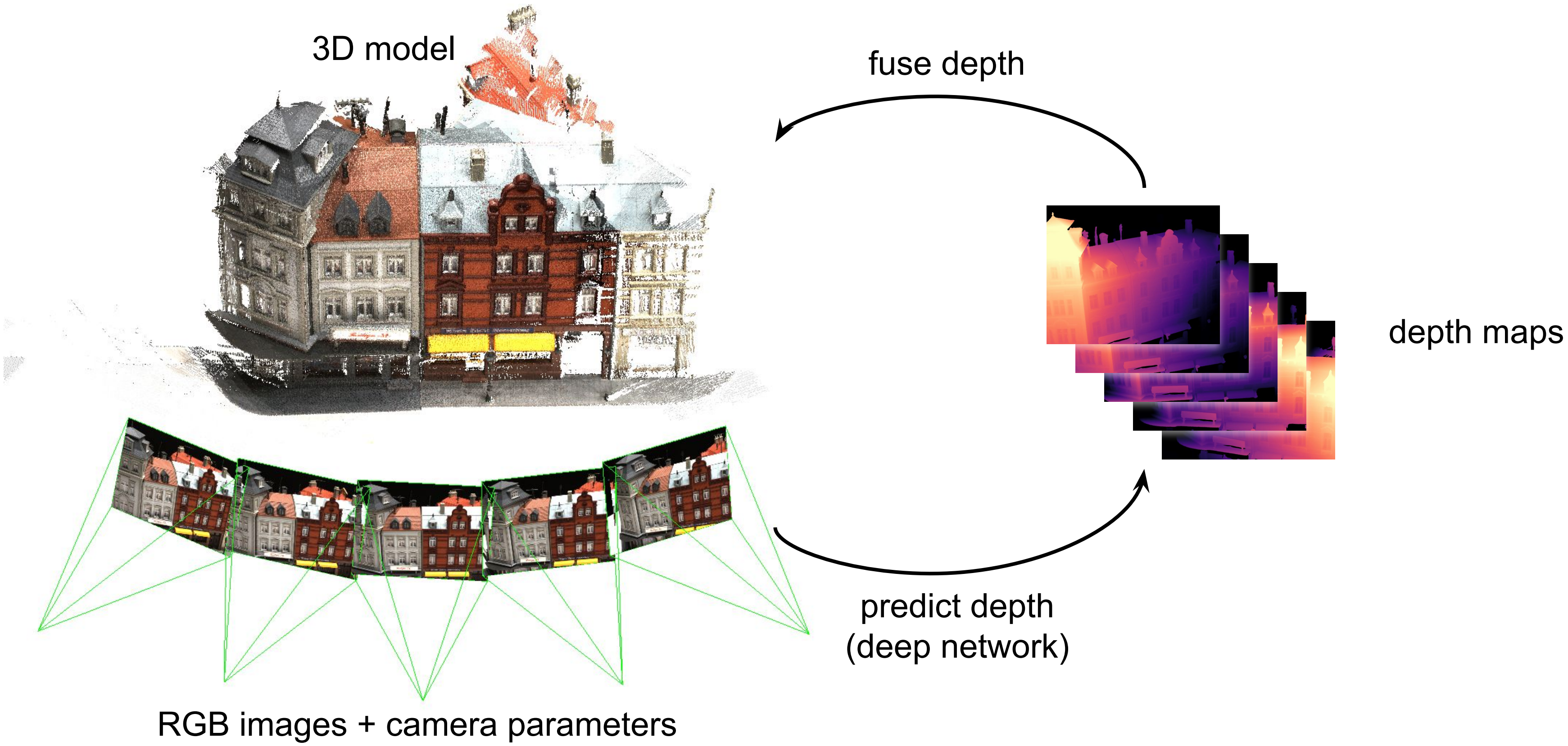
Pradeep Sen

Tobias Höllerer

UC SANTA BARBARA

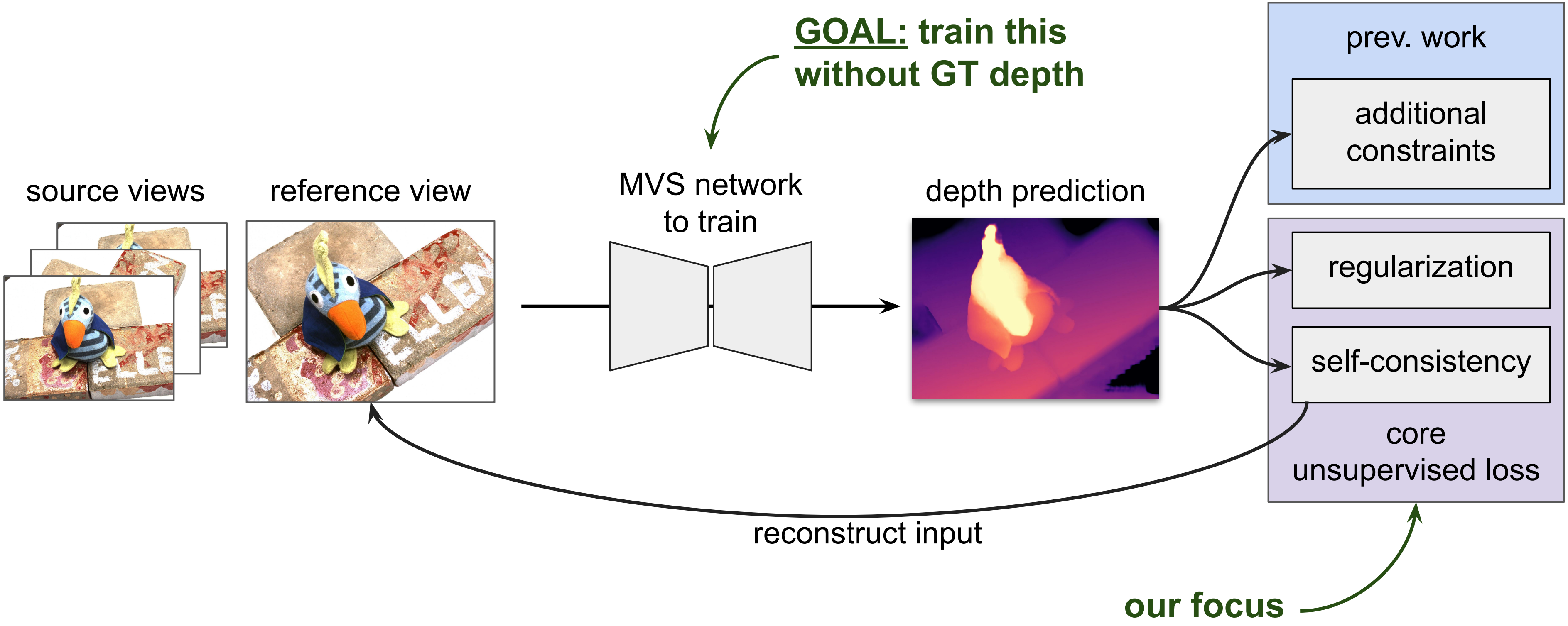
https://github.com/alexrich021/div-loss/

Overview

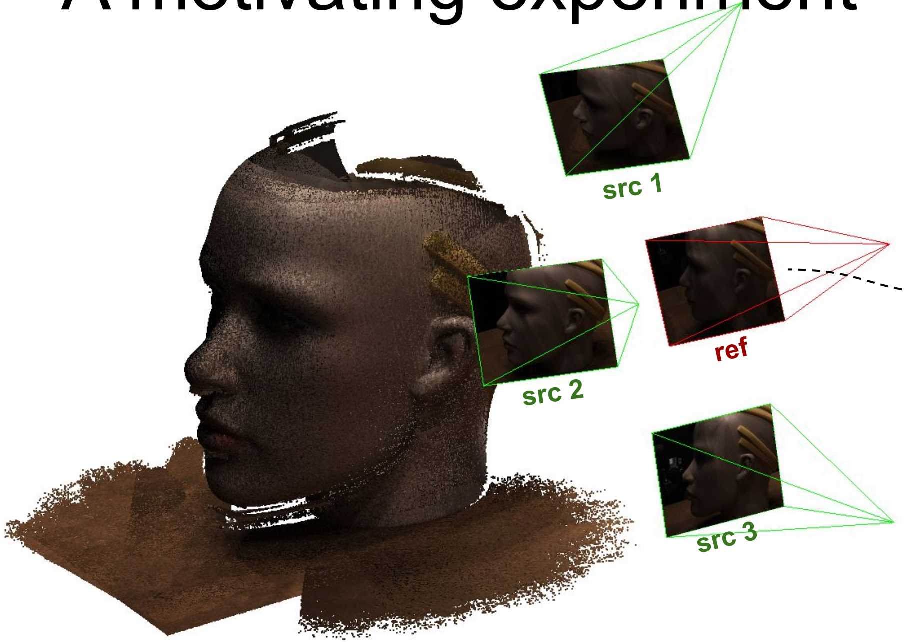

3D model
fuse depth
depth maps
predict depth
(deep network)
RGB images + camera parameters

Overview

Unsupervised MVS allows us access to large amounts of data



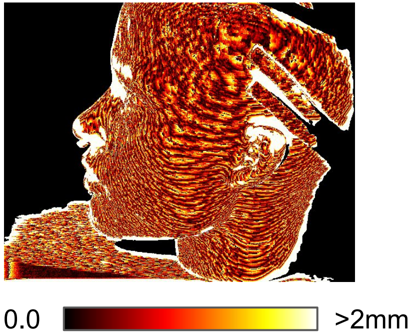
A motivating experiment



initialize using
GT depth of **ref image**…

…then update to minimize
core unsupervised loss

Results: standard loss



depth error
after optimization

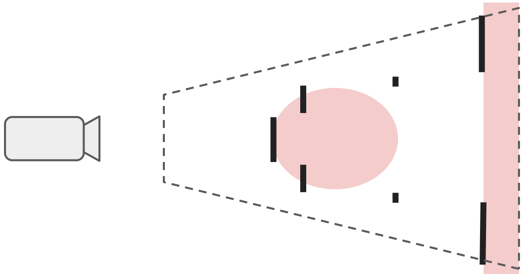
The standard unsupervised loss produces **artifacts**

this is our training objective

DTU dataset Jensen et al., 2016

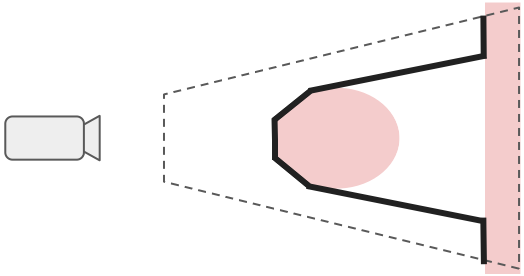
Regularization

$$\mathcal{L}_1 = |\nabla D(p)|$$



❌ stair stepping

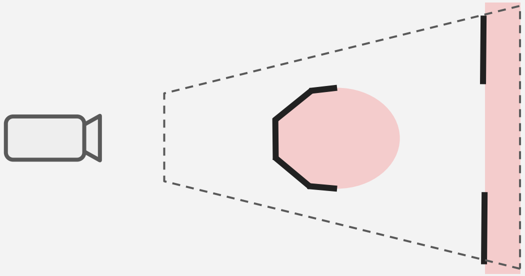
$$\mathcal{L}_2 = |\nabla^2 D(p)|$$



❌ boundary blurring

Ours

$$\mathcal{L}_2^{ours} = \begin{cases} \mathcal{L}_2 & \mathcal{L}_2 < \alpha \\ \alpha & \text{otherwise} \end{cases}$$



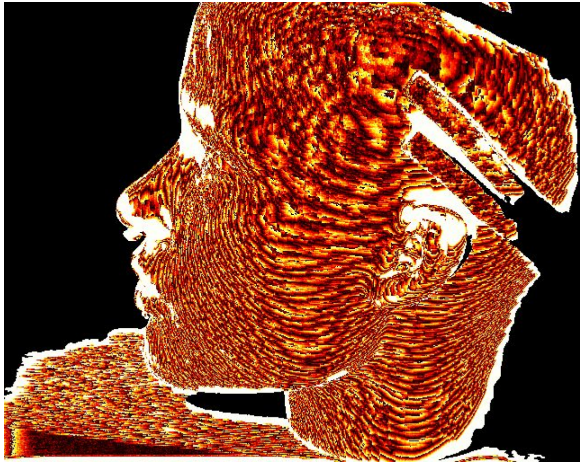
✅ smooth surfaces

✅ object boundaries

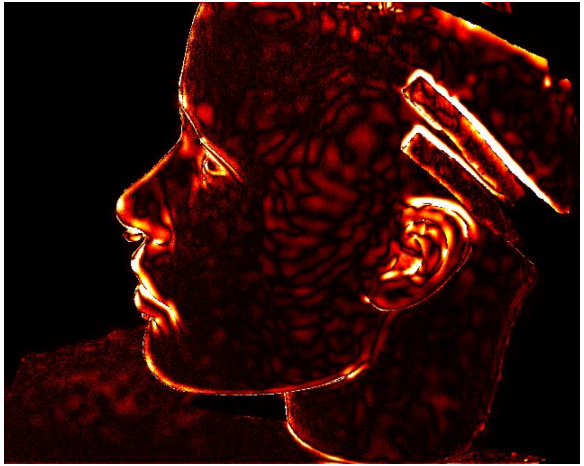
(GT objects in **red**)

Regularization

Results: standard loss (1^{st}-order)



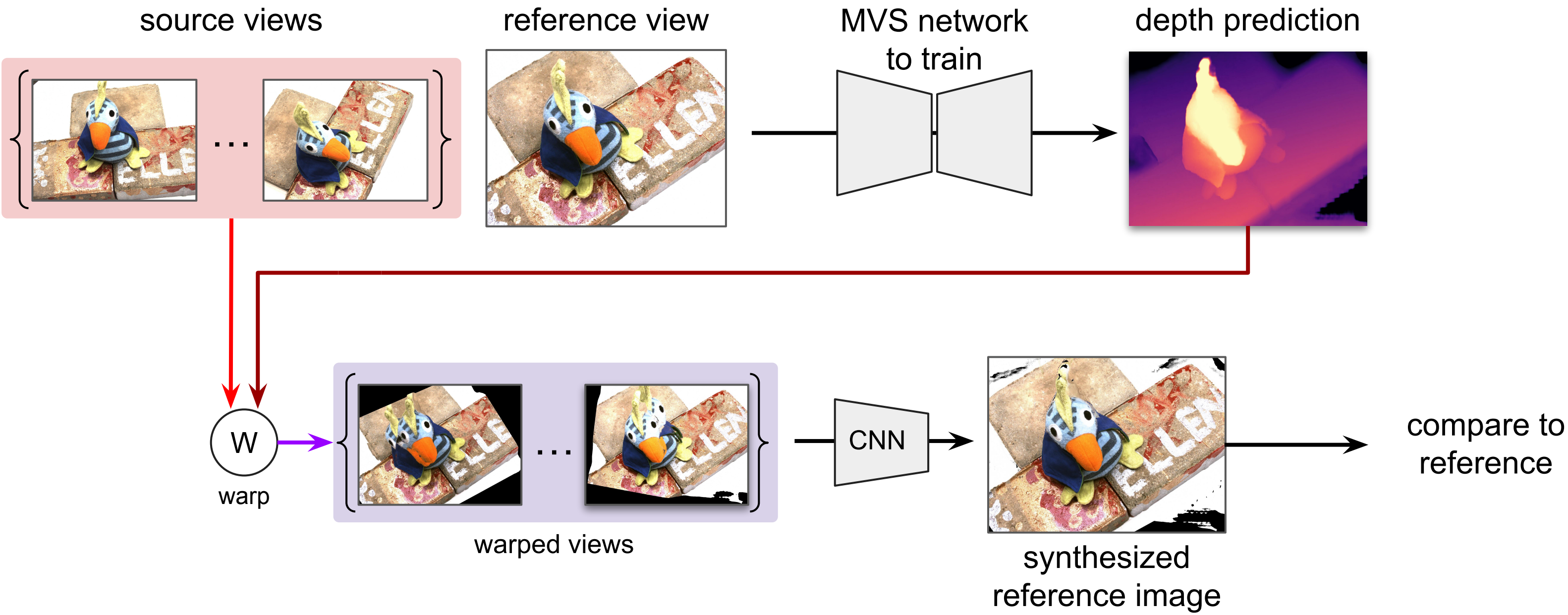
Results: ours (relaxed 2^{nd}-order)



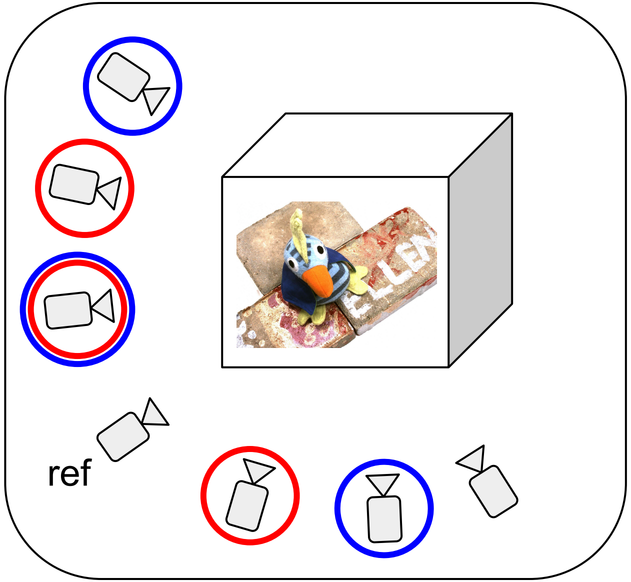
depth error

DTU dataset Jensen et al., 2016

Self-Consistency



Self-Consistency



network input views

≠ (de-coupled)

network supervision views

Our method: DIV Loss

Depth smoothness + **I**mage synthesis + **V**iew sampling

A novel supervision strategy for unsupervised multi-view stereo

- Easily drops into existing pipelines
- Improves results quantitatively and qualitatively
- Requires minimal additional GPU memory and time during training

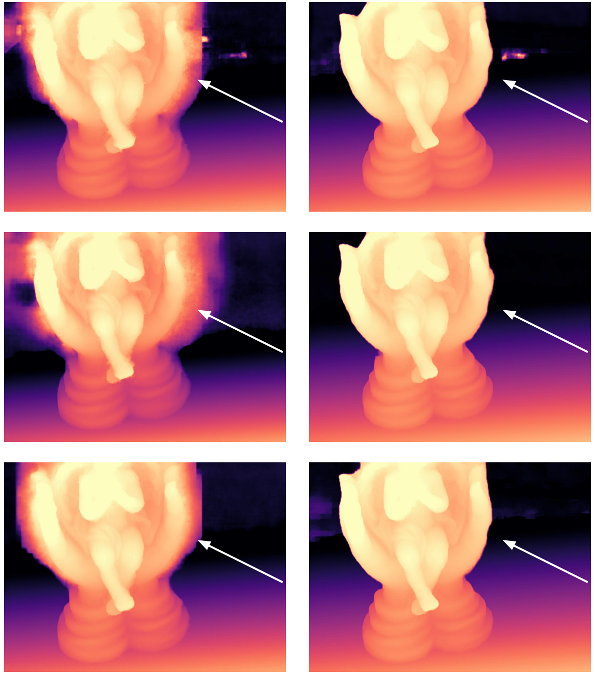
Results

DTU Results

DTU dataset Jensen et al., 2016

Method	Ovr. ↓	Diff
Baseline	0.361	
+ DIV loss (Ours)	0.330	-0.031
RC-MVSNet	0.345	
+ DIV loss (Ours)	0.333	-0.017
CL-MVSNet	0.330	
+ DIV loss (Ours)	$\boxed{0.321}$	-0.009

state-of-the-art among unsupervised methods

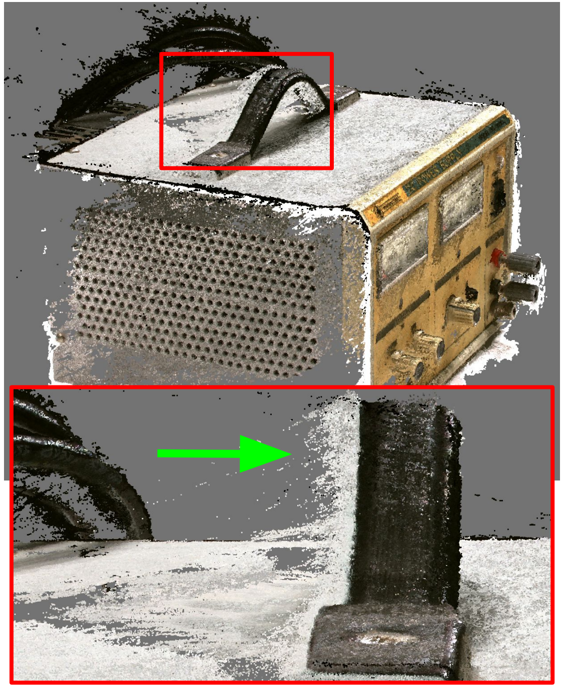


baseline + DIV loss

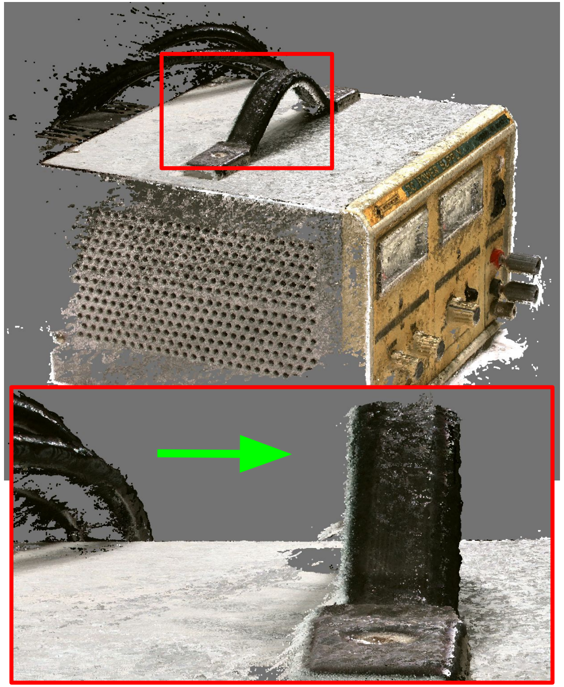
DTU Results

DTU dataset Jensen et al., 2016

baseline



+ DIV loss



Example Input Images

Additional Results

Tanks and Temples dataset, Knapitsch et al., 2017
ScanNet++ dataset, Yeshwanth et al., 2023

	Method	DTU only	T&T intermed. F-score ↑	T&T adv. F-score ↑	ScanNet++ F-score ↑
Supervised	CasMVSNet	✓	56.84	31.12	-
	CVP-MVSNet	✓	54.03	-	-
	AttMVS	✓	60.05	31.93	-
	PatchmatchNet	✓	53.15	32.31	-
	GeoMVSNet	✗	65.89	41.52	-
	MVSFormer-H	✗	**66.41**	**41.70**	-
Multi-Stage Self-Sup.	Self_sup CVP	✓	46.71	-	-
	U-MVS	✓	57.15	30.97	-
	KD-MVS	✗	**64.14**	**37.96**	-
E2E Unsup.	M^3VSNet	✓	37.67	-	-
	JDACS-MS	✓	45.48	-	-
	DS-MVSNet	✓	54.76	-	-
	ElasticMVS	✗	57.88	37.81	-
	RC-MVSNet	✓	55.04	30.82	37.42
	CL-MVSNet	✓	59.39	37.03	40.71
	DIV-MVS (Ours)	✓	**60.36**	**38.36**	**41.64**

(trained on DTU with no fine-tuning on additional data)

Conclusion

DIV loss: **D**epth smoothness + **I**mage synthesis + **V**iew sampling

A novel supervision strategy for unsupervised multi-view stereo

- Easily drops into existing pipelines
- Improves results quantitatively and qualitatively
- Requires minimal additional GPU memory and time during training

Poster #210

Tuesday, 4:30pm

UC SANTA BARBARA

https://github.com/alexrich021/div-loss/