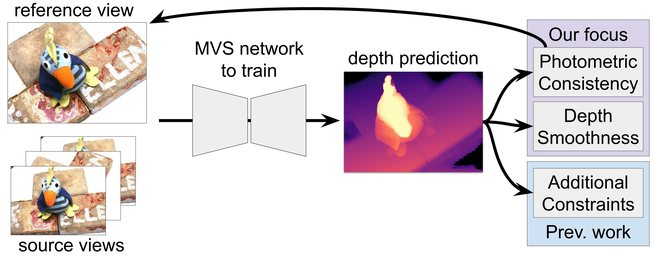
Smoothness, Synthesis, and Sampling: Re-thinking Unsupervised Multi-View Stereo with DIV Loss

Alex Rich ◇ Noah Stier ◇ Pradeep Sen ◇ Tobias Höllerer
University of California, Santa Barbara

FourEyesLab UCSB — ECCV MILANO

PROBLEM SETUP & CONTRIBUTION

Unsupervised MVS: training depth-prediction networks without access to ground-truth depth

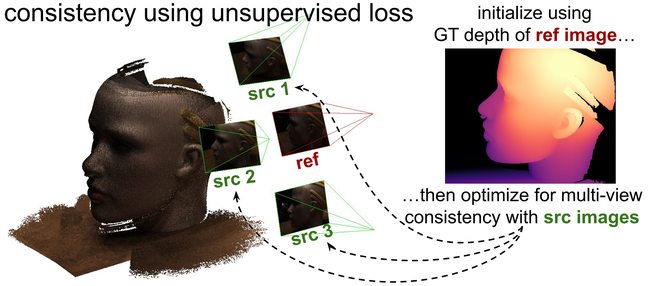


DIV Loss: A novel core unsupervised loss formulation

- **D**epth smoothness + **I**mage synthesis + **V**iew sampling
- Easily drops into existing unsupervised MVS pipelines
- Improves network performance for minimal additional training cost

MOTIVATION

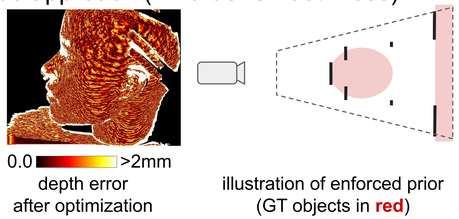
Experiment: optimized ground-truth depth for multi-view consistency using unsupervised loss



Results: previous approach (1st-order smoothness)

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

❌ stair stepping
❌ boundary blurring

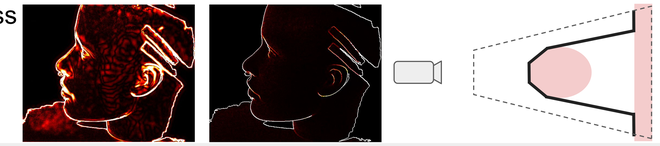


RELAXED 2ND-ORDER DEPTH SMOOTHNESS

Results: 2nd-order smoothness

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

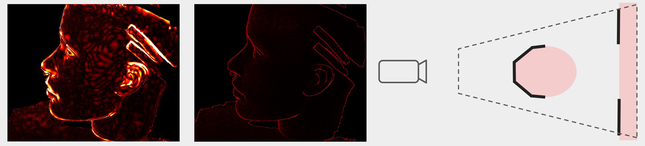
✅ smooth surfaces
❌ boundary blurring



Results: ours (**relaxed** 2nd-order smoothness)

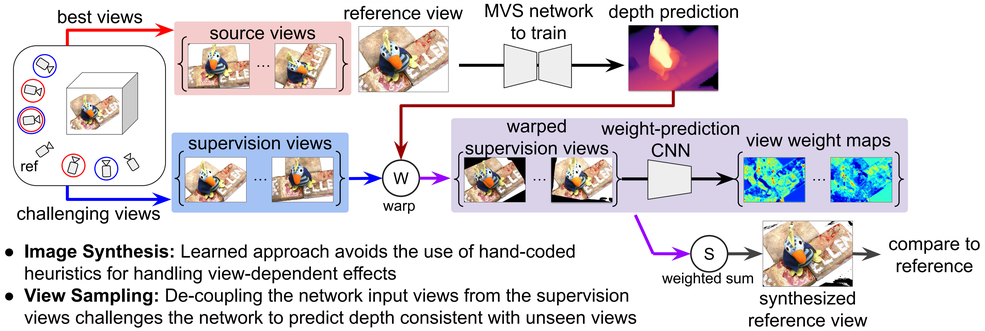
$$\mathcal{L}_2^{ours} = \begin{cases} \mathcal{L}_2 & \mathcal{L}_2 < \alpha \\ \alpha & \text{otherwise} \end{cases}$$

✅ smooth surfaces
✅ object boundaries



Automatically relaxes 2nd-order gradient penalty at large depth boundaries, allowing sharp discontinuities where required

IMAGE SYNTHESIS AND VIEW SAMPLING



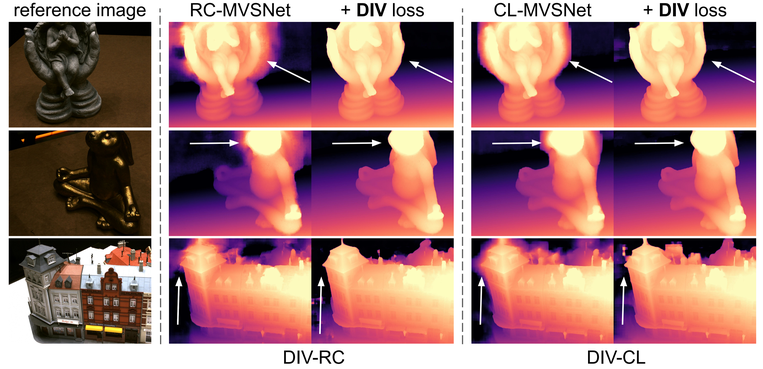
- **Image Synthesis:** Learned approach avoids the use of hand-coded heuristics for handling view-dependent effects
- **View Sampling:** De-coupling the network input views from the supervision views challenges the network to predict depth consistent with unseen views

QUANTITATIVE RESULTS (DTU)

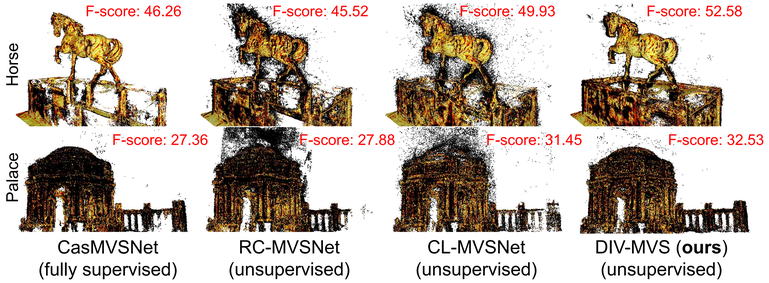
pipeline	DTU Ovr. ↓ without DIV	DTU Ovr. ↓ with DIV	DTU Ovr. ↓ diff	DTU Abs. Depth Error (mm) ↓ without DIV	DTU Abs. Depth Error (mm) ↓ with DIV	DTU Abs. Depth Error (mm) ↓ diff	Training Memory (GB) without DIV	Training Memory (GB) with DIV	Training Memory (GB) diff
DIV-MVS	0.361	0.330	-0.031	19.34	16.32	-3.02	10.50	10.52	+0.02
DIV-RC	0.350	0.333	-0.017	21.76	21.01	-0.75	12.24	12.26	+0.02
DIV-CL	0.330	0.321	-0.009	17.88	15.38	-2.50	11.64	11.70	+0.06

- **DIV Loss** improves network performance in all cases while requiring >0.1GB additional GPU memory during training
- **DIV-CL** achieves SOTA performance on DTU among unsupervised methods, **DIV-MVS**, **DIV-RC** rank highly

DEPTH RESULTS (DTU)



POINT CLOUD FUSION RESULTS (T&T)



SOTA on T&T among (1) unsupervised methods and (2) *all* methods which only train on DTU

ABLATION STUDY (DTU)