FINS: Fast Image-to-Neural Surface

Efficient Construction of Implicit Surface Models From a Single Image for Motion Generation

Wei-Teng Chu[1]* · Tianyi Zhang[2] · Matthew Johnson-Roberson[5] · Weiming Zhi[345]
[1]Stanford EE [2]Aurora Innovation [34]Univ. of Sydney / ACFR [5]Vanderbilt *waynechu@stanford.edu

Motivation

The Problem

Autonomous robots need reliable geometric representations of their environment. Signed Distance Fields (SDFs) are widely used for obstacle avoidance, path planning, and surface interaction.

Limitations of Existing Methods

① NeuS [1] requires 49+ multi-view images
② Training takes minutes to hours — impractical in real-time settings
③ Sparse-view [2] - [3] methods still need several images and remain slow

FINS at a Glance

1	~10s	1.8×	49×
Input Image	Training Time	Faster than NeuS2 [4]	Fewer Imgs vs NeuS [1]

Method: FINS Framework

① 3D Foundation Model

DUSt3R [5] / VGGT [6] lifts a single RGB image into a colored 3D point cloud. Low-confidence points are filtered before SDF training.

② Multi-Resolution Hash Grid Encoding [7]

Encodes coarse-to-fine geometry using compact hash features for fast convergence.

③ Geometry & Color Heads

Separate lightweight heads predict SDF and RGB for stable geometry–appearance learning.

④ Staged Hybrid Optimizer

Warm-up with Lion [8], then refine heads with K-FAC [9] for fast curvature-aware convergence.

Training Objectives

$\mathscr{L}_{\text{SDF}}$ — SDF Supervision

Supervises predicted SDF values against ground-truth distances at noisy point samples, with per-point confidence weighting.

$\mathscr{L}_{\text{eik}}$ — Eikonal Constraint

Enforces $\|\nabla d(x)\| = 1$ both near the surface and globally, ensuring the network learns a valid SDF rather than an arbitrary scalar field.

$\mathscr{L}_{\text{off-surf}}$ — Off-Surface Regularization

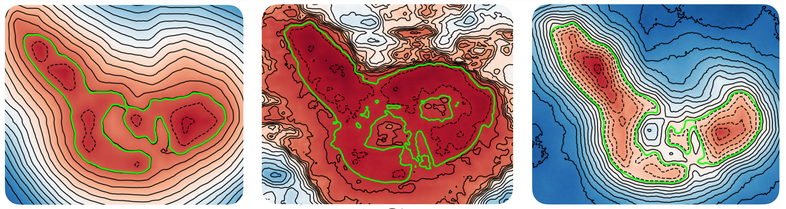
Supervises SDF values at off-surface samples, encouraging valid signed-distance structure beyond the observed surface.

Quantitative Comparison

Method	Input images↓	Training time (s)↓	DTU [10] avg. CD↓	BMVS [11] avg. CD↓
NeuS [1]	49	247	9.95	n/a
NeuS2 [4]	5	18	7.16 best	0.0540
SparseNeuS [2]	2	127	9.95	0.0758
SparseCraft [3]	3	85	655.86	2.35
Ours (FINS)	**1**	**10**	**7.96** +11%	**0.0286** best

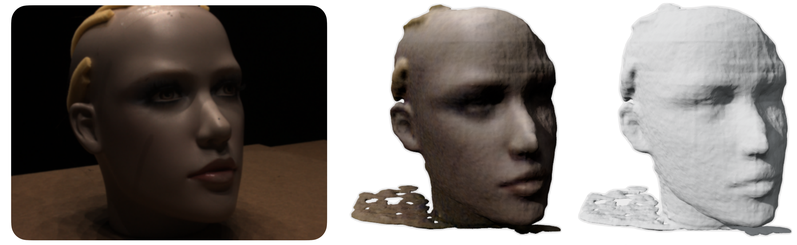
KEY TAKEAWAY
Ours uses **1 image** and trains in **~10 seconds** —
yet achieves CD within **11% of the best baseline on DTU [10] and best overall on BlendedMVS [11].**

Ablation Studies

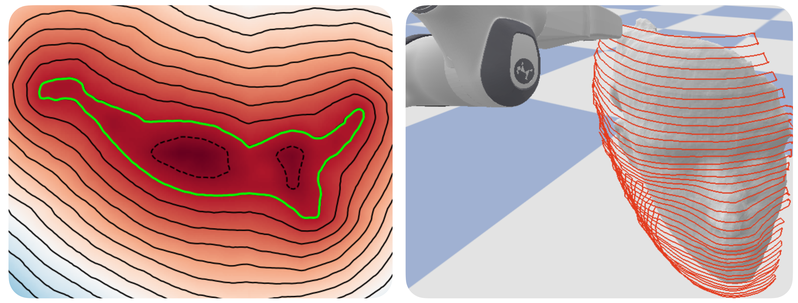


Full Model | w/o $\mathscr{L}_{\text{eik}}$ | w/o $\mathscr{L}_{\text{off-surf}}$

Qualitative Results



Input Image | Output Mesh



Signed Distance Fields | Surface Following

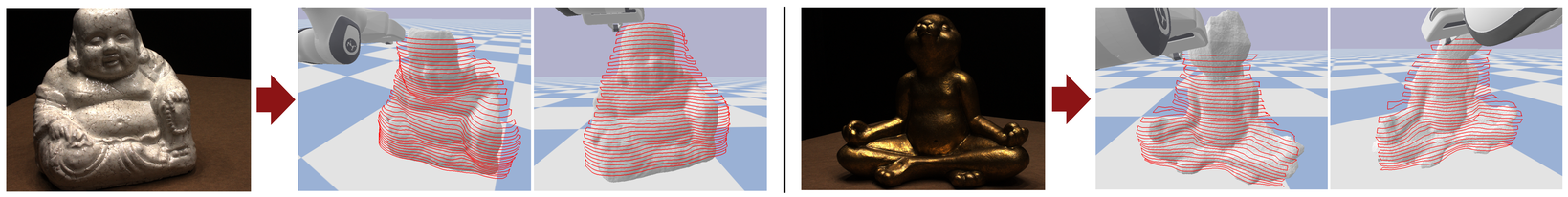
Contributions

① End-to-end single-image SDF reconstruction in ~10 seconds on consumer hardware.
② Foundation-model-guided SDF learning from DUSt3R [5] / VGGT [6] point-cloud priors.
③ Fast geometry optimization with hash encoding [7] and K-FAC [9] refinement.

Conclusions

FINS shows that high-fidelity SDF reconstruction from a single RGB image can be achieved in ~10 seconds on consumer hardware. By combining 3D foundation model priors, multi-resolution hash-grid encoding [7], and K-FAC [9] optimization, FINS provides an accurate and practical solution for real-time robotics deployment.

More Results



References

[1] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-viewreconstruction," arXiv preprint arXiv:2106.10689, 2021.
[2] X. Long, C. Lin, P. Wang, T. Komura, and W. Wang, "Sparseneus: Fast generalizable neural surface reconstruction from sparse views," in European Conference on Computer Vision. Springer, 2022, pp. 210–227.
[3] M. Younes, A. Ouasfi, and A. Boukhayma, "Sparsecraft: Few-shot neural reconstruction through stereopsis guided geometric linearization," arXiv preprint, 2024.
[4] Y. Wang, Q. Han, M. Habermann, K. Daniilidis, C. Theobalt, and L. Liu, "Neus2: Fast learning of neural implicit surfaces for multiview reconstruction," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
[5] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "Dust3r: Geometric 3d vision made easy," in CVPR, 2024.
[6] J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, "Vggt: Visual geometry grounded transformer," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025.
[7] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," ACM Trans. Graph., 2022.
[8] X. Chen, C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, Y. Lu, and Q. V. Le, "Symbolic discovery of optimization algorithms," 2023.
[9] J. Martens and R. Grosse, "Optimizing neural networks with kronecker-factored approximate curvature," in Proceedings of the 32nd International Conference on Machine Learning, 2015.
[10] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, "Large scale multi-view stereopsis evaluation," in IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[11] Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, and L. Quan, "Blendedmvs: A large-scale dataset for generalized multi-view stereo networks," Computer Vision and Pattern Recognition (CVPR), 2020.