

DepthAnything3

Recovering the Visual Space from Any Views

AUTHORS

Haotong Lin*, Sili Chen*, Jun Hao Liew*, Donny Y. Chen*, Zhenyu Li, Guang Shi, Jiashi Feng, Bingyi Kang[†]

ByteDance Seed

★ Core Contributions

- **Minimal Modeling:** Single ViT + per-pixel depth-ray targets.
- **SOTA Performance:** **+35.7%** pose AUC, **+23.6%** geometric accuracy vs VGGT.
- **Universal:** Works with monocular, multi-view, or video frame sequences.



Figure 1: DA3 reconstructs visual space from any number of images

Dual-DPT

NOVEL HEAD

DINOv2

BACKBONE