



Core Contributions & Rationale



Minimalism Suffices

ARCHITECTURE

A single plain ViT (DINOv2) + minimal Depth-Ray targets are sufficient for any-view geometry. No complex multi-task bundles or bespoke 3D modules needed.



Implicit Pose via Rays

FORMULATION

Predicting dense rays avoids difficult orthogonality constraints of rotation matrices (SO3). Pose emerges naturally from ray convergence.



Empirical Dominance

RESULTS

+35.7% Pose AUC vs. VGGT. Validates that scale-invariant depth + ray maps is the optimal minimal set for foundation geometry.

DepthAnything3 proves that **generalist scaling beats specialized engineering** for 3D visual geometry.