

Research ownership and collaboration

Project	Documented contribution
FourieRF / MILO / Spokes	Coauthorship, mentoring of Diego Gomez, MILO code contributions and LIX/GRAPHDECO collaboration
ZeroKey	Conceived, implemented and evaluated the first-author project
PatchAlign3D	Joint research with Souhail Hadgi and collaborators
Is It a Good Match?	Scientific lead since April 2026, experiments and manuscript
3DHarnessBench	Experiment guidance, project coordination, TaskSolver and restricted Blender interface

Mesh complexity and surface accuracy

Surface sampling evaluates geometry independently of vertex density.

Method	Mesh vertices ↓ Mesh-rendering evaluation	Uniform F1 ↑	Virtual-scan F1 ↑
GOF	11.63 M	0.31	0.38
RaDe-GS	10.06 M	0.35	0.39
MILo	4.36 M	0.34	0.42
Spokes	5.80 M	0.48	0.53

Tanks & Temples. Vertices: mesh-rendering evaluation (Table 2). F1: six-scene surface evaluation (Table 1); Spokes F1 uses two pivots, without PAM.

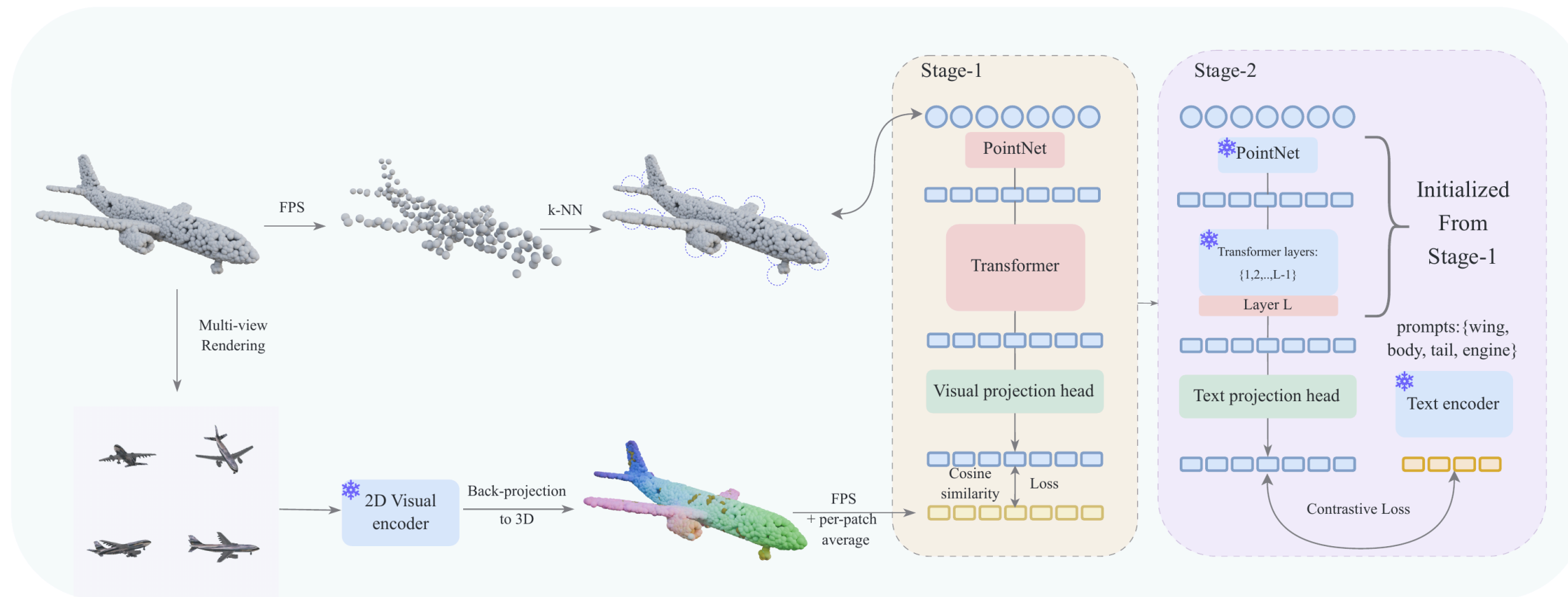
MILo has the fewest reported vertices; Spokes has the highest F1 among these methods.

ZeroKey: localization thresholds and comparison scope

Method	0.01	0.05	0.10
B2-3D · few-shot, 62 views	20.29	57.72	70.57
B2-3D · few-shot, 26 views	14.39	44.11	56.90
GPT-4o localizer	0.48	6.04	20.73
CLIP-DINOiser	1.41	9.80	25.56
StablePoints	5.80	19.91	38.22
Binary back-projection	5.84	29.81	54.64
ZeroKey · 26 views	13.16	56.60	79.43

KeypointNet: airplane, chair and table. IoU (%) at geodesic thresholds.
StablePoints receives the ground-truth keypoint count; benchmark text descriptions are manually assigned.

Visual distillation, then local language alignment



2,048 XYZ points · 128 overlapping patches of 32 points · 12 transformer blocks

Stage 1 · visual structure

Match 3D patch features to frozen DINOv2 targets with a cosine loss.

Stage 2 · language alignment

Freeze the first 11 blocks; train the last block and projection into OpenCLIP's text space.

Patch features can support ZeroKey’s point proposals

Semantic support for voting

Use patch–text similarity to modulate ZeroKey’s soft-voting weights.

Local features for clustering

Add learned feature cues when grouping the 3D point proposals.

Reported KeypointNet comparison · IoU (%) at geodesic distance thresholds

Method	0.01	0.05	0.10
ZeroKey	13.16	56.60	79.43
PatchAlign3D + ZeroKey	10.42	58.87	81.27

IoU improves at 0.05 and 0.10, but falls at 0.01.

The hybrid still uses ZeroKey’s multiview pipeline.

The harness changes the observation and interaction protocol

Setting	Target access	Target geometry	Refinement
Single view	1 fixed image	Hidden	3 rounds
Multi-view	4 fixed images	Hidden	3 rounds
Active Visual	Camera / viewport tools	Blocked by allowlist	Agent-directed
Full 3D	Unconstrained Blender tools	Available to query	Agent-directed

100 sampled objects from 3DCodeBench.
Original procedural programs are withheld.

Active Visual separates target and reconstruction processes.
Full 3D prohibits copying the target.

Evaluation conventions determine what a score can mean

Geometric alignment

Sample 8,192 surface points per mesh.
Center each cloud and scale to the unit sphere.

$$CD = \min_{R \in R_{24}} d_{\text{sym}}(RX, Y)$$

d_{sym} : mean squared nearest-neighbor distances in both directions.
 R_{24} : the 24 right-handed axis-aligned rotations.

Translation, scale and these rotations are discounted.

Complementary checks

Appearance: four 512^2 renders; match views by the best one-to-one assignment.

$$e_{\text{topo}} = \sum_{i=0..2} |\beta_i^{\text{GT}} - \beta_i^{\text{rec}}| / \sum_{i=0..2} \beta_i^{\text{GT}}$$

Topology: median across objects.

Other quality metrics: mean across objects.

No loadable mesh: zero similarity; Chamfer penalty of $1.5 \times$ the worst valid value in that run.

A measured speed–quality tradeoff

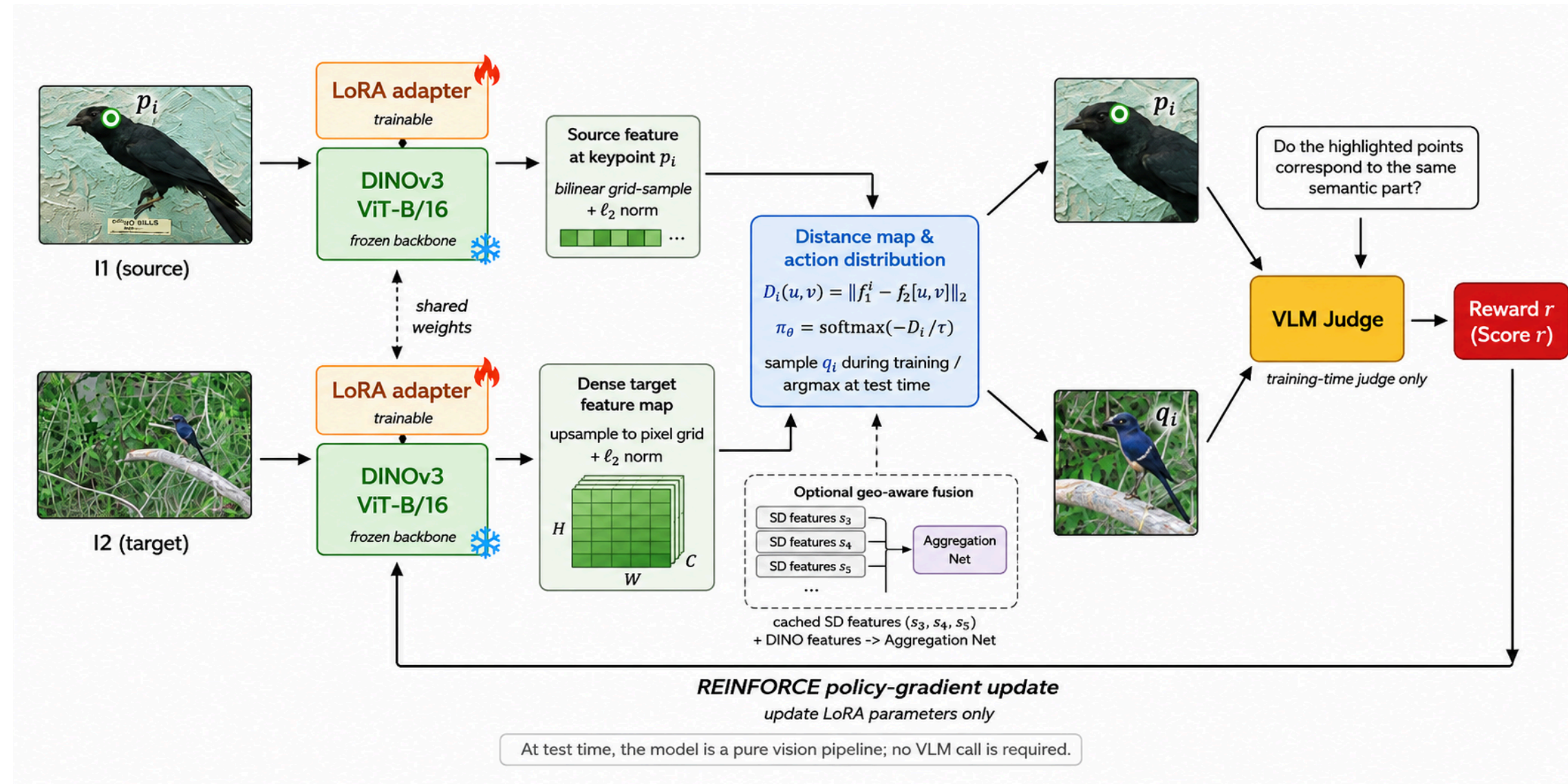
5.57× faster training than ZeroRF,
with 0.212 dB lower PSNR.

Method	Relative training time ↓	4-view Blender PSNR ↑
FourieRF-VM	0.930	21.728 dB
ZeroRF	5.181	21.940 dB

Reported Blender setup: 10,000 iterations, RTX 4090. Training time normalized to TensoRF = 1.

Frequency control regularizes the field. Geometric use also needs an explicit surface.

A one-step policy over target pixels



$$W' = W + (\alpha / r_L) BA \cdot Q/K/V \text{ LoRA: } r_L = 16, \alpha = 8 \cdot \tau = 0.1$$

Main path: train LoRA. Optional fusion also trains the aggregation network.

A useful reward can still miss geometric precision

SPair-71k cat test · Matched evaluation · PCK@bbox 0.10 (%)

Method	PCK ↑
Frozen SD + DINOv3	74.87
Geometry-aware LoRA · Kimi reward	77.74 ± 0.64
Ground-truth reward oracle	82.95

+2.87 points under the matched protocol; reward quality still limits accuracy.

Calibration limit	Negative 3D transfer
Judge scores plateau below about 20 pixels.	CP2P: 41.33 → 31.52 after Kimi-reward adaptation.

Local v4 manuscript · preliminary. Source keypoints are given; target-coordinate labels do not enter the VLM reward.

Teaching evidence and roles

Experience	Documented role
HKU, 2016–2021	Teaching assistant for eight courses: tutorials, laboratories, assignment support and project supervision
Pedagogical training	Certificate in Teaching and Learning in Higher Education (CTLHE), completed during the Ph.D.
MVA, autumn 2026	Appointed teaching assistant for Geometry Processing and Geometric Deep Learning, Master 2
Research mentoring	Six co-mentees across doctoral, master's and research-intern roles, with eleven co-authored papers and manuscripts

HKU courses span graphics and games, AI, programming, discrete mathematics, computer organisation and data engineering.

Questions that bound the programme

Starting point	Question to test
GoDe + MILO-base	Does geometric feedback preserve surfaces at matched bytes, beyond retraining alone?
FINER + Good Match	Does better factuality also improve fine-coordinate ranking and then matcher PCK?
ZeroKey + SMACC	Does surface evidence improve rig attachment beyond joint names and a click?
Reconstruction + relaxation	With supplied mechanics, does recovery justify its shape, motion and energy distortion?
Longer-term video feedback	When does Proprio’s learned score agree with independent physical checks?

A useful negative result can define when a component should not be used.

Structured signals and compact avatars

Jhony H. Giraldo · COSIMO

Refine local semantics on a mesh

Project PatchAlign3D features to vertices;
diffuse vertex, edge and face signals.

A stability test for COSIMO

MILO reconstructions introduce measured
feature noise and missing surfaces.

First test

One fixed mesh: vertex diffusion vs COSIMO.
Measure part-boundary and keypoint errors.

Limit: published mesh-class results do not
establish part segmentation or remeshing stability.

Xiaoran Jiang · MISTA

Compress a canonical avatar

Spatial ordering + tensor-train factors;
keep the supplied SMPL animation model.

What remains geometrically stable?

Adapt our surface and normal checks
to measure drift during deformation.

First test

One supplied-SMPL sequence, equal budgets:
render quality plus surface drift over motion.

Limit: human-body prior; compact parameters
do not establish real-time or generic animation.

Motion scores and observation robustness

Kaouther Messaoud · Proprio

Calibrate a generator's self-score

Compare its frozen latent flow residual with VLM feedback and explicit checks.

Isolate the cause of a score change

Our executable-scene workflow can vary one declared contact or motion constraint.

First test

10 matched motion pairs: contact violations, appearance and motion magnitude controlled.

Limit: a learned score cannot certify physics.
Measure prompt fidelity separately.

3DHarnessBench currently evaluates static objects. These motion tests are proposed extensions.

Kaouther Messaoud · OmniTraj

Stress the observation interface

Reuse explicit frame-rate conditioning and missing-cue masks from OmniTraj.

Preserve motion; change visibility

Controlled rendering supplies paired inputs for the same human trajectory.

First test

One motion sequence: vary FPS and occlusion; export pose/box cues and measure ADE/FDE.

Limit: needs a valid human-motion source.
Object keypoints do not supply human pose.