NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2026 · Sydney

Weather-Robust Cross-View Geo-Localization via Prototype-Based Semantic Part Discovery

Chi-Nguyen Tran[1,2*] Duy-Minh Dao-Sy[1,2*] Trung-Kiet Huynh[1,2*] Phu-Hoa Pham[1,2] Phu-Quy Nguyen-Lam[1,2] Long Tran-Thanh[3]

[1]University of Science, Ho Chi Minh City [2]Vietnam National University, Ho Chi Minh City [3]University of Warwick *Equal contribution

Paper

1 Motivation

Localizing a drone without GNSS

Delivery, inspection and search-and-rescue drones rely on GNSS, which can be jammed, spoofed or lost. **Cross-view geo-localization** recovers the position by matching the drone's oblique view to a geo-referenced satellite tile, on the drone's own limited compute.

Why it is hard

The drone sees the scene obliquely from 150 to 300 m, the satellite straight down, and bad weather (fog, rain, snow, darkness) can wash out most of the texture. What survives is the **layout**: where roads, roofs and fields sit relative to each other.

Three gaps in current methods

- **Layout and texture mixed:** a single global vector compresses the patch grid without separating layout, which both views share, from texture.
- **Altitude leak:** altitude-driven scale stays in the embedding.
- **Hand-tuned losses:** fixed scalars weight losses that live on incompatible gradient scales.

2 Contributions

A light head on a self-supervised ViT that separates scene layout from texture.



Prototype parts: what is there
Learnable prototypes compete for patch tokens, so the model discovers the recurring components of a scene without part labels.



Altitude-free embedding
FiLM, a per-altitude scale and shift, absorbs flight height during training, so retrieval needs no altitude.

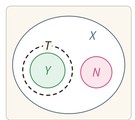


Layout graph: where it is
A graph attention network (GAT) encodes the spatial organization of the parts, which both views share.



Self-weighted losses
Each loss group learns its own weight from its uncertainty, so all four stay balanced without tuning.

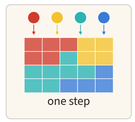
3 Theory



INFORMATION BOTTLENECK

Layout is the part worth keeping.
From a view X, the embedding T keeps location Y and drops nuisances N: texture and altitude change, layout does not.

$$\min I(T;X) - \beta I(T;Y),\ \ I(T;N) \to 0$$



SLOT ATTENTION

One competition step suffices.
Prototypes compete for patch tokens; the cosine assignment is the fixed point of one slot-attention step.

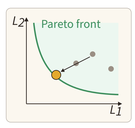
$$A_{i,k} = \mathrm{softmax}_k\left(\cos(p_k, z_i)\,/\,\tau\right)$$



FEATURE-WISE MODULATION

Mean FiLM drops altitude.
One FiLM per altitude bin in training; their mean at test, so no altitude input.

$$\gamma_\alpha \cdot z + \beta_\alpha \ \rightarrow\ \bar{\gamma}\cdot z + \bar{\beta}$$

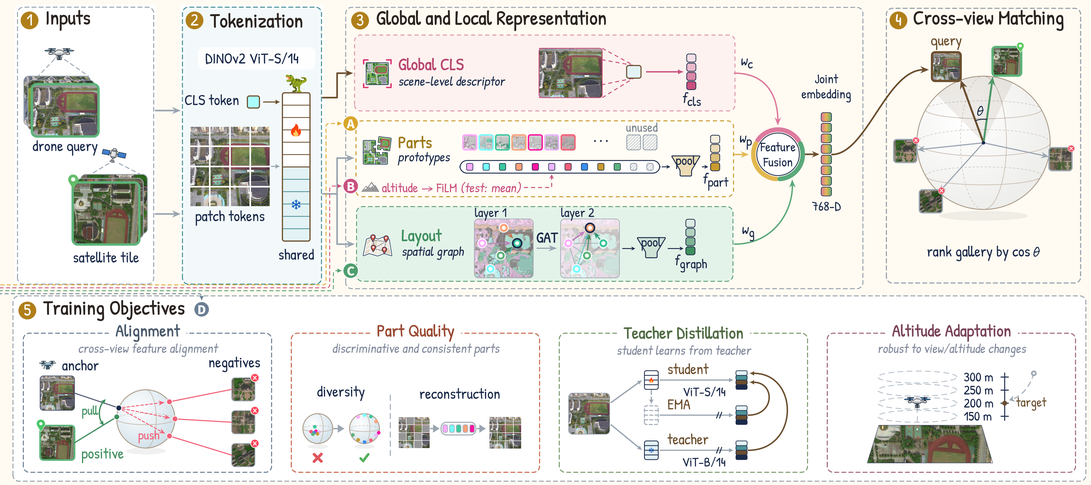


MULTI-TASK LEARNING

Self-weighting is Pareto-safe.
Positive weights make a stationary point Pareto-stationary for all four loss groups.

$$L = \sum_g \left[\, e^{-s_g} L_g + s_g \,\right]$$

4 Method



One shared **DINOv2 ViT-S/14** encodes both views. The CLS token (the ViT's global image token), the parts (*what*) and the layout graph (*where*) are fused into one 768-D vector, and retrieval is a single cosine ranking. The training objectives (bottom row) are dropped at test time.

5 Main results

Method	Params	GFLOPs	SUES-200 D→S	University-1652 D→S	University-1652 S→D	DenseUAV D→S
LPN	62.4M	36.78	73.57	77.71	90.30	–
Sample4Geo	87.6M	90.24	96.86	92.65	95.14	49.38
MCCG	56.6M	51.04	90.12	89.28	94.29	83.14†
CCR	156.6M	160.6	93.22	92.54	95.15	–
CAMP	91.4M	90.24	97.60	94.46	96.15	88.72
MEAN	36.5M	26.18	98.09	93.55	96.01	90.18
SkyPart	**26.95M** −26%	**22.14** −15%	**98.725** +0.64	**96.47** +2.01	**98.43** +2.28	**91.85** +1.67

SkyPart leads every column and is the smallest, cheapest model in this table.

R@1 (%), D→S: drone query vs. satellite gallery; one 448 px pass, no re-ranking or TTA. Three-seed mean ± SD: 98.74 ± 0.11, 96.47 ± 0.25, 91.85 ± 0.26. Underlined: best baseline in this table; green: gain over it; † 28 M MCCG variant.

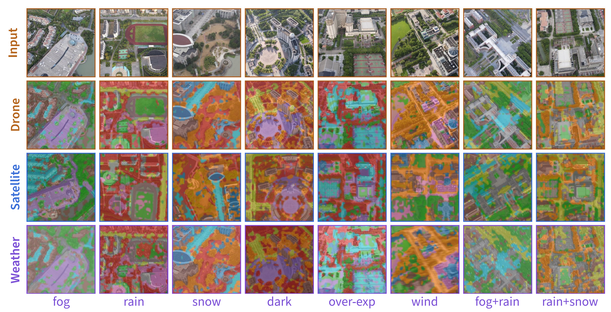
6 Ablation

Variant	Clean	Weather	Weather lost
Full SkyPart	**98.69**	**96.03**	–
Loss group removed			
− alignment	54.49	27.67	**−68.4**
− distillation	97.63	94.65	−1.4
− part quality	97.28	94.92	−1.1
− altitude	97.47	95.22	−0.8
Readout changed			
CLS readout only	93.56	85.64	**−10.4**
Part branch only	97.51	94.09	−1.9
No graph branch	97.56	94.86	−1.2
Prototype gate removed			
All prototypes active	97.79	93.72	−2.3

Without the alignment loss the model collapses; the part and graph readouts bring most of the weather robustness.

SUES-200 D→S R@1 (%); 448 px ablation checkpoint (panel 8: a separate 392 px weather-trained model).

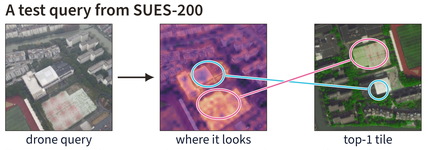
7 Learned parts



Each colour is one learned prototype, found without part labels.

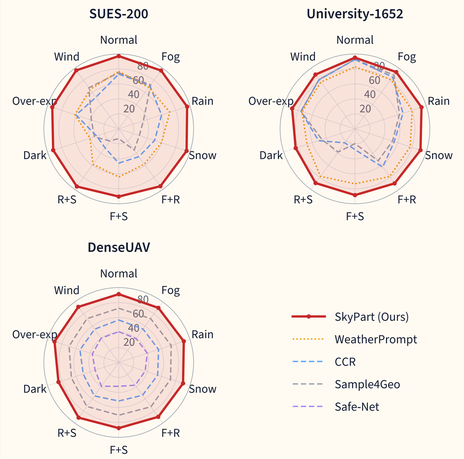
- **Across views:** one prototype marks the same roofs and roads in both views.
- **Under weather:** the parts stay fixed in all eight conditions.
- **Kept apart:** without the diversity loss they collapse in about 15 epochs.

A test query from SUES-200



The query's building and courts are found again in the true tile; the links cross because the drone faces the other way.

8 Weather robustness



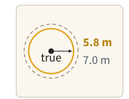
SkyPart is best on every condition of all three datasets; on SUES-200 Fog+Snow, **90.44** against 64.36 for the best prior method (WeatherPrompt).

The ten conditions



D→S R@1 per condition; F = fog, R = rain, S = snow. SkyPart is retrained at 392 px with these conditions; baselines from WeatherPrompt (384 px), on DenseUAV retrained by us.

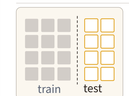
9 Generalisation



DENSEUAV

Error in metres
Mean distance of the top-1 tile to the true position.

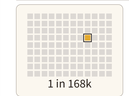




UAV-VISLOC

Unseen sites
Geographically disjoint test sites; R@5 shown, R@1 is 0.07 lower.





UNIVERSITY160K-WX

Large gallery
168k satellite images, queries under weather; R@1.



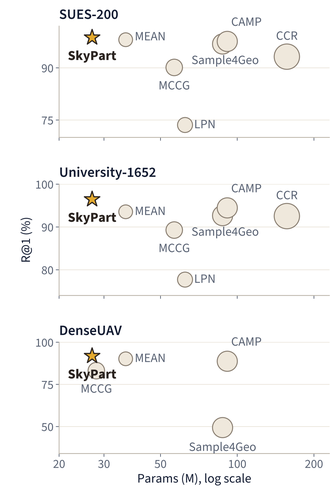


SUES-200 → UNIVERSITY-1652

No retraining
Trained on SUES-200 only, zero-shot on University-1652; R@1.



10 Efficiency



SkyPart is the top-left point on all three: best R@1 at the fewest parameters.

Same numbers as panel 5; bubble area grows with GFLOPs.

11 Limitations

✗ TRIED, DID NOT HELP

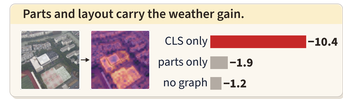
~~polar warping~~ ~~ArcFace margins~~ ~~Sinkhorn OT matching~~ ~~rotating both views~~

? NOT COVERED YET

- Returns the matching tile; a metric pose needs a separate registration step.
- Train and test share a region; seasons, sensors and map vintages are untested.
- Prototypes have no fixed names; the bank size is fixed at 12.
- Dual use: could aid aerial surveillance.

12 Key findings

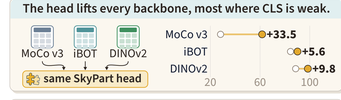
Parts and layout carry the weather gain.



Altitude helps weather more than clean.



The head lifts every backbone, most where CLS is weak.



Results are stable across three seeds.

