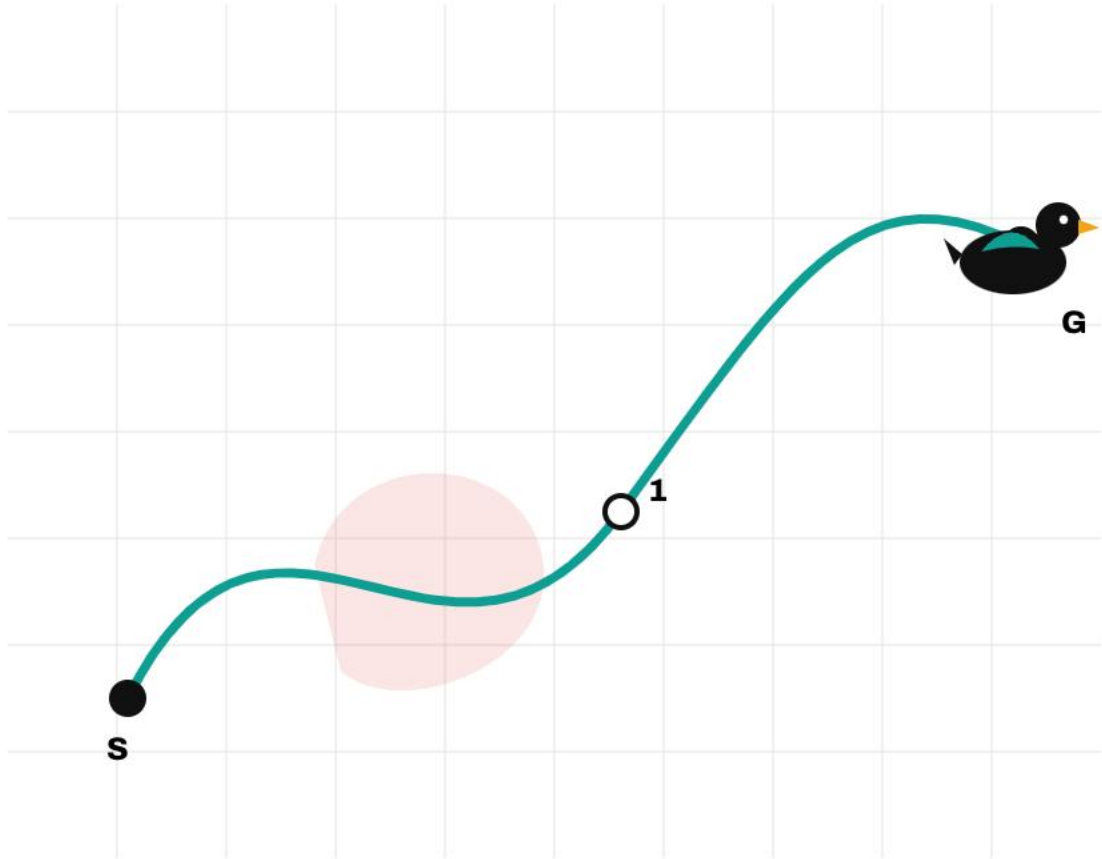
TAAI 2026 · DOMESTIC ORAL SESSION 4

Mapping Intent

Agentic UAV route planning with semantic cost maps



Jeng Wen Joshua Lean[1], Ting-Yu Yen[1], Kang-Chiao Lin[2], Xu-Peng He[2], Shih-Hsuan Hung[1], Hung-Kuo Chu[1]

[1]National Tsing Hua University · [2]Industrial Technology Research Institute

Paper #208

21 Nov 2026

“Walk.”



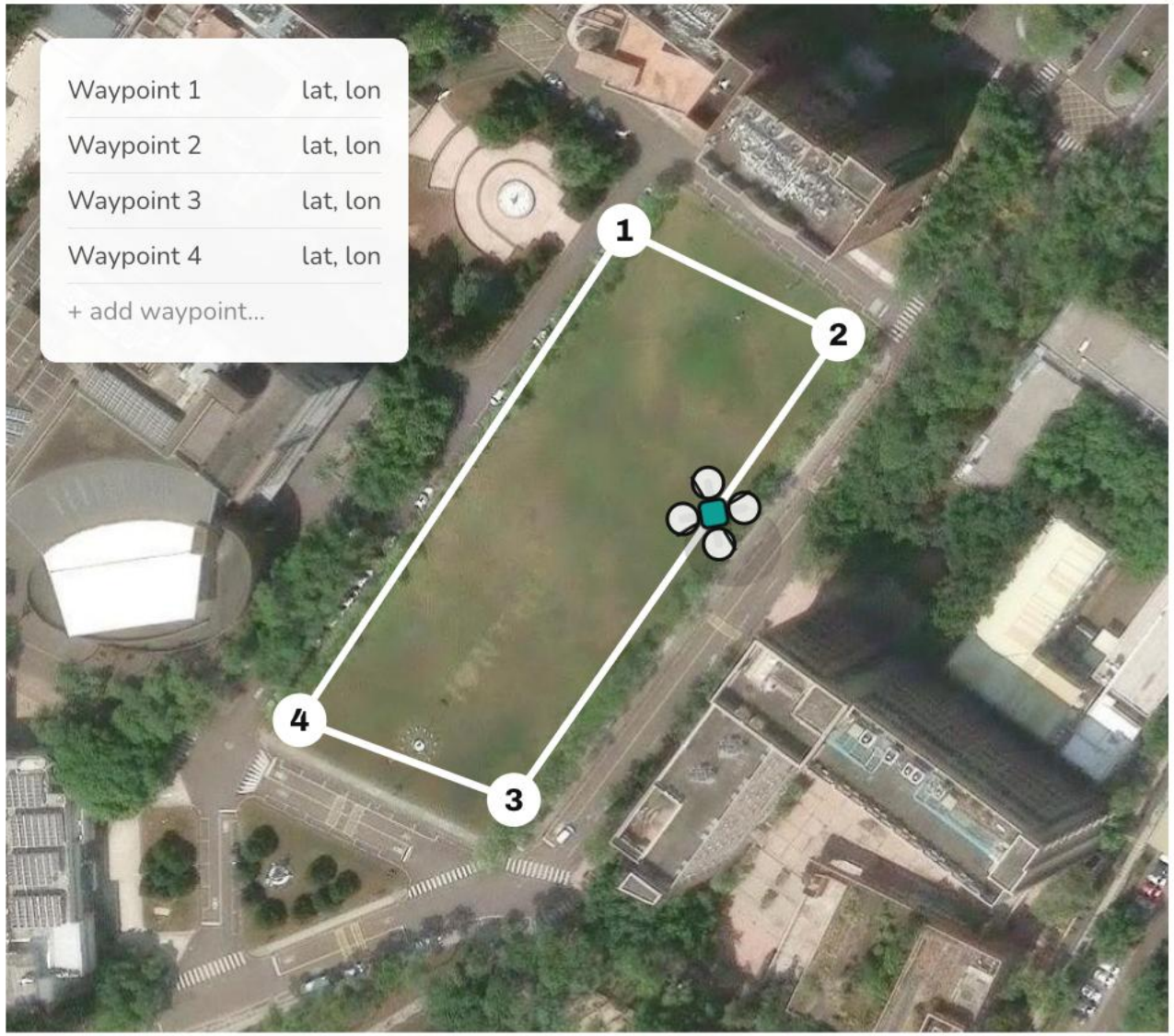
MICRODUCK · POLLEN ROBOTICS / HUGGING FACE

“Fly!”



SOMEDAY?

Video: Pollen Robotics (simulated Microduck walk policy)



Drones still speak coordinates.

Today an operator clicks waypoints, one by one.

What we'd rather do:

“Fly one loop around the grass field.”

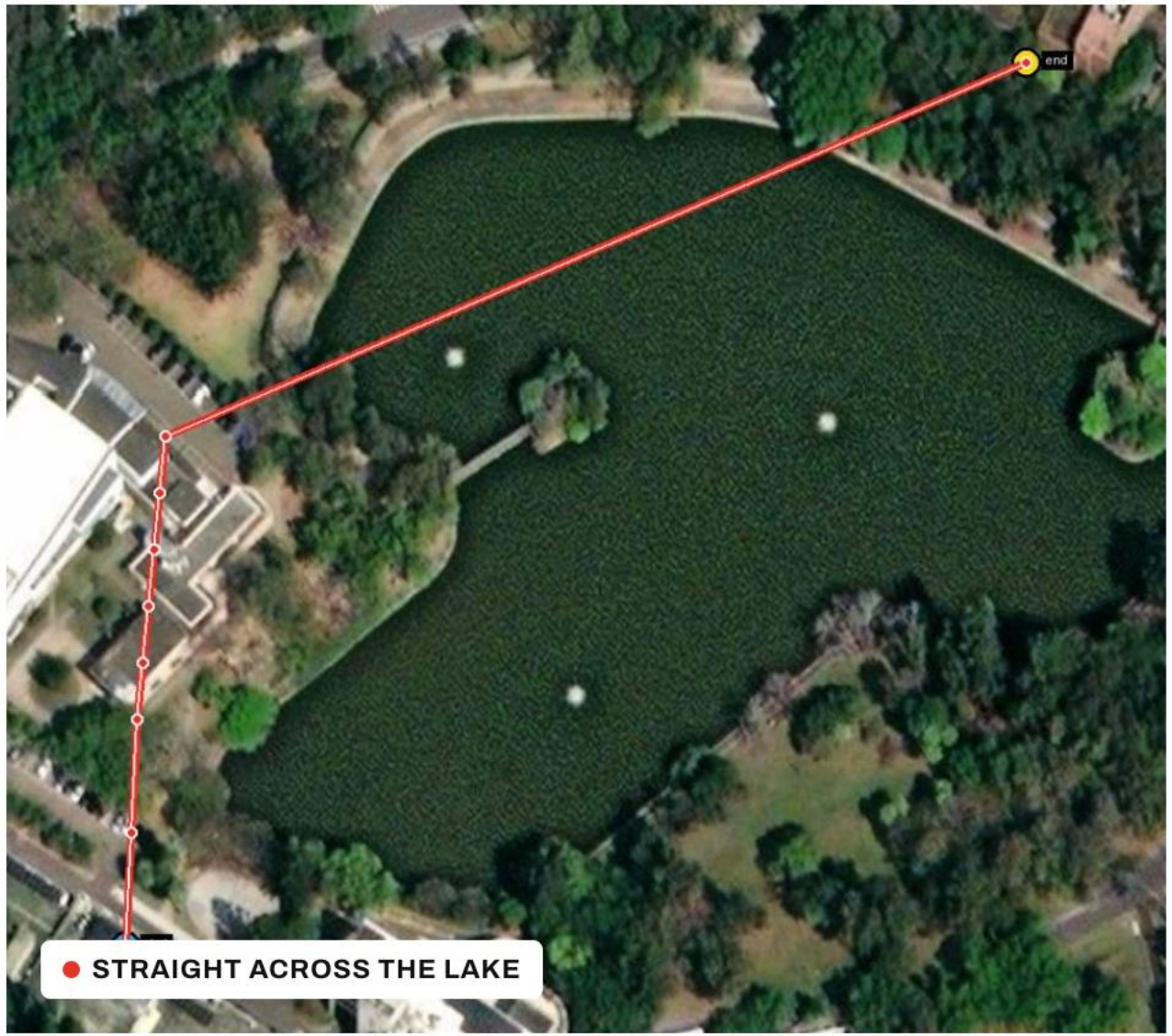
NTHU campus · Imagery: Esri, Maxar, Earthstar Geographics

Easy to say. Easy to get wrong.

> **“Go from the southwest path to the northeast path while surveying the lake shoreline. Do not fly over the water.”**

Great route for a duck. 🦆

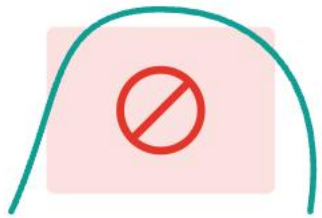
Not for a drone.



NTHU campus · route from a published planner (our re-run) · Imagery: Google

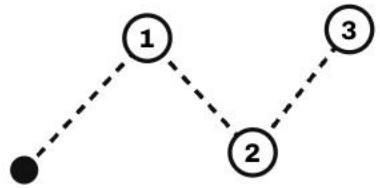
One sentence hides three kinds of rules.

“Survey the lake shoreline: south, east, then north, stay on paths where possible, and do not fly over the water.”



Avoid

water, roofs, no-fly areas



Visit, in order

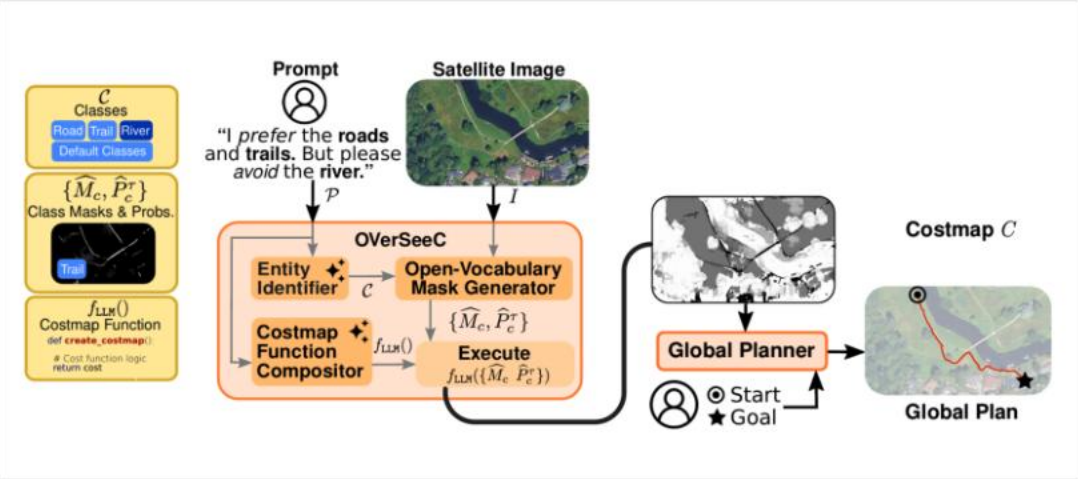
checkpoints, loops, surveys



Prefer

paths, roads, open ground

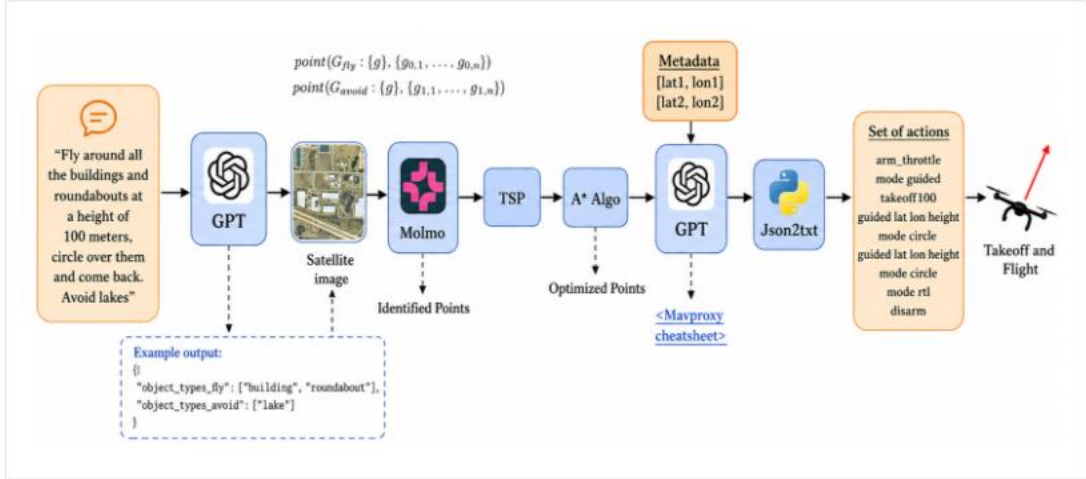
Two ways people do it today.



Cost map first

Language → terrain costs → shortest path.

OVerSeeC · Rana et al., 2026



Waypoints first

VLM picks points → order them → connect.

UAV-VLPA* · Sautenkov et al., 2025

They reach the goal.
Not the mission.

> “Circle the softball field once, clockwise, just inside the tree line.”



Cost map first ✕ no loop

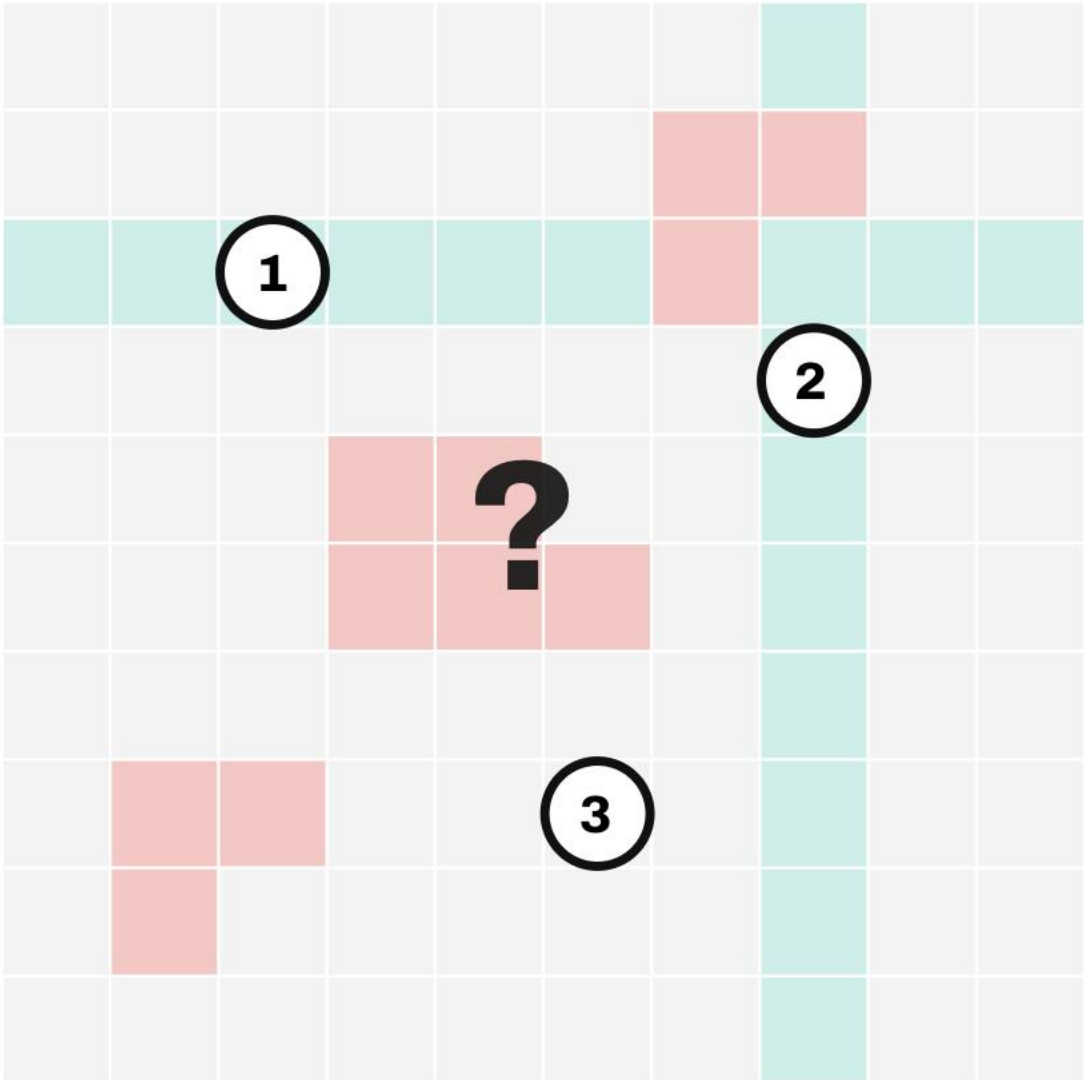


Waypoints first ✕ no loop

Our re-runs of OVerSeeC and UAV-VLPA* · Imagery: Google

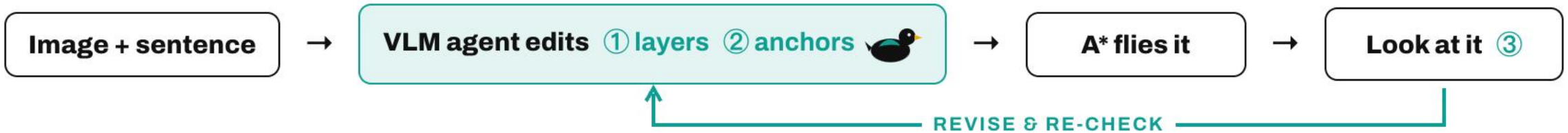
A cost map knows *where*. Not *in what order*.

- ✓ prefer paths
- ✓ avoid roofs
- ✗ visit 1, then 2, then 3
- ✗ close the loop
- ✗ …and nobody checks the result



Let the VLM draw the map — with an order, and a second look.

WHAT WAS MISSING		WHAT WE ADD
✓ **where to fly** — prefer paths, avoid roofs	→	**① Semantic cost layers** — keep what cost maps do well
✗ **in what order** — visit 1, 2, 3 · close the loop	→	**② Ordered anchors** — S → 1 → 2 → G, A* between each pair
✗ **nobody checks the result**	→	**③ Look, then revise** — inspect the route, fix, re-run (≤ 6 tries)

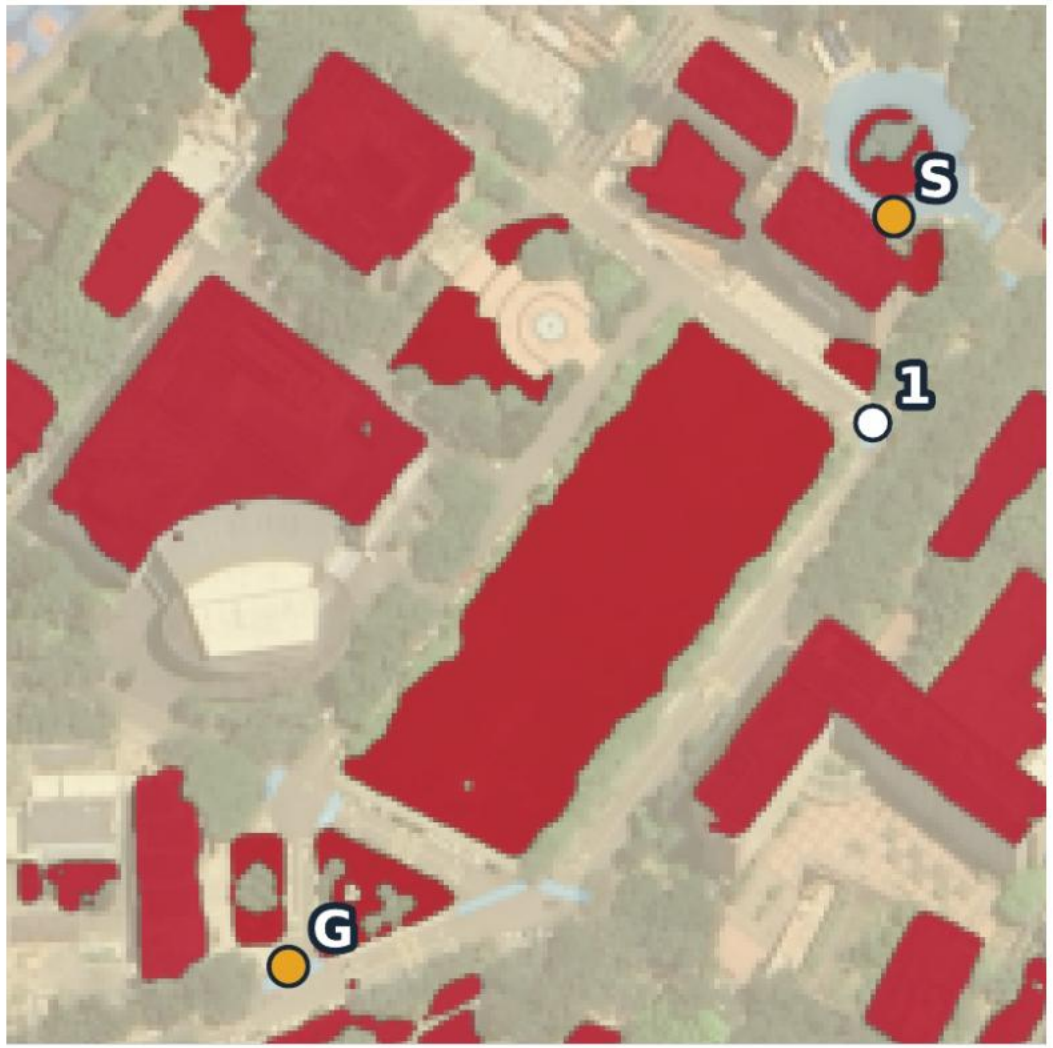


① Read. ② See.

“Avoid roofs and maintained grass, prefer asphalt, and visit crossing 1.”

Every mission element is linked back to the exact words it came from.
SAM 3 then segments the concepts the agent asks for.





③ Paint the costs.

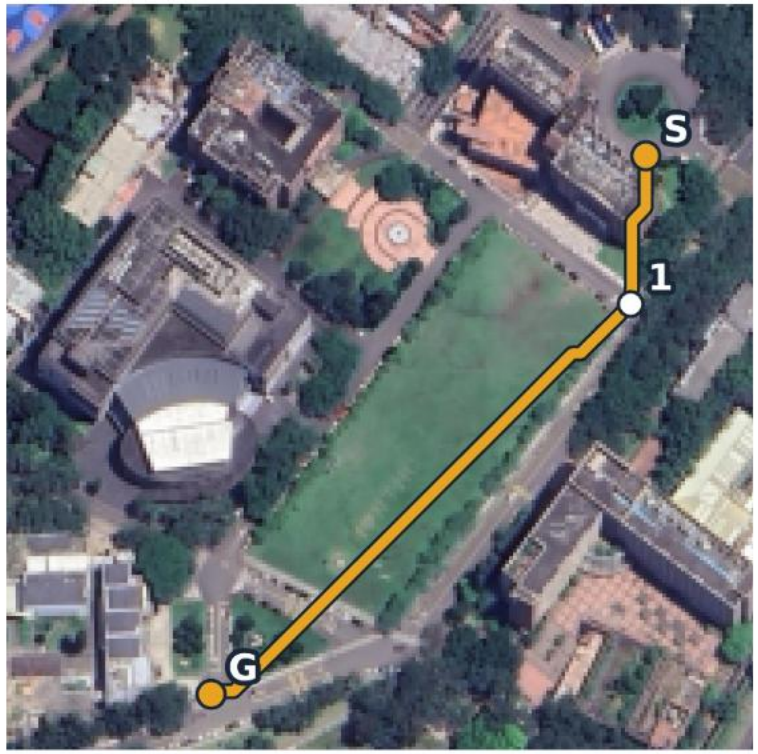
"Avoid roofs and maintained grass, prefer asphalt, and visit crossing 1."

Roofs	hard block
Grass	hard block
Asphalt	negative weight → cheaper

$$C(x) = clip(0.5 + \Sigma_k\, w_k\, M_k(x),\ 0.01,\ 1)$$

Imagery: Google

④ Pin the order. ⑤ Let A* fly.



Same anchors, flat map
cuts straight across the grass

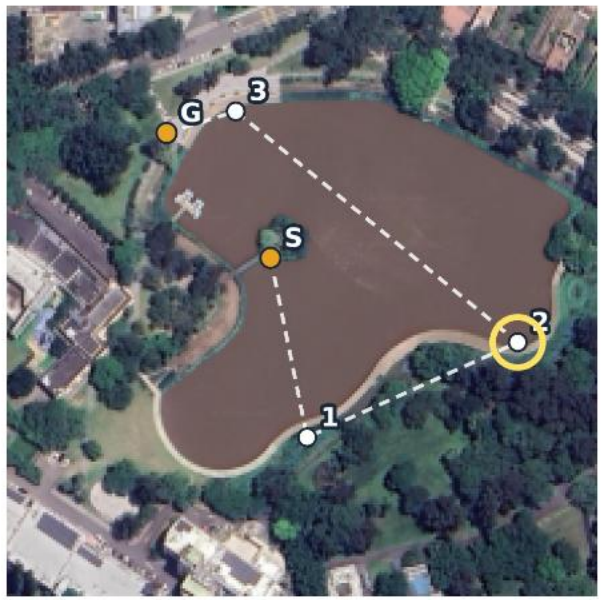


Same anchors, our map
around the blocked grass and roofs, on the road

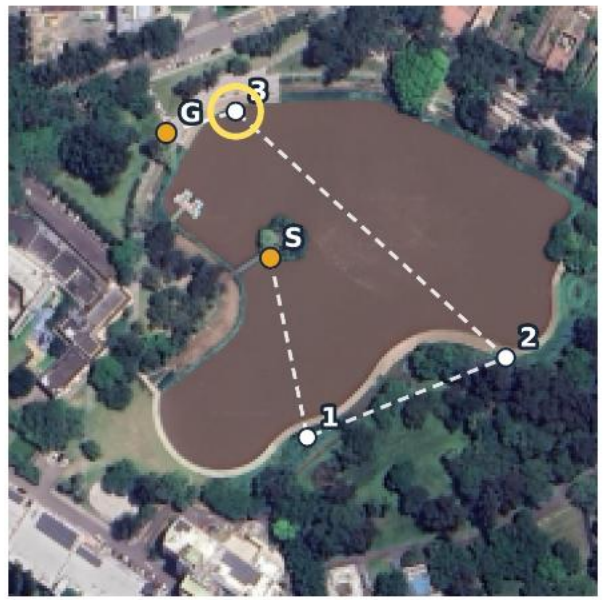
Left panel: illustrative uniform-cost control, not a benchmark run · Imagery: Google

⑥ Look. Fix. Fly again.

"Survey the shoreline through the south, east, then north."



Eval 1 ✕ anchor 2 is in the water



Eval 2 ✕ moved it · anchor 3 still blocked



Eval 3 ✓ connected · no layers changed

Dashed = requested anchor order · Imagery: Google

Four ways to judge a route.

✓ visit crossing 1
✓ avoid roofs
✗ avoid grass
✓ end at G

1 / 4 = 25%

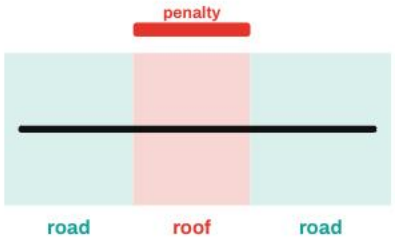
Checklist violation rate ↓

Share of the mission's requirements the route breaks. Main metric.



Distance to reference ↓

Mean gap to a reviewed reference route, ÷ image diagonal.



Terrain regret (RRPI) ↓

Average penalty for flying over terrain the mission ranked lower.

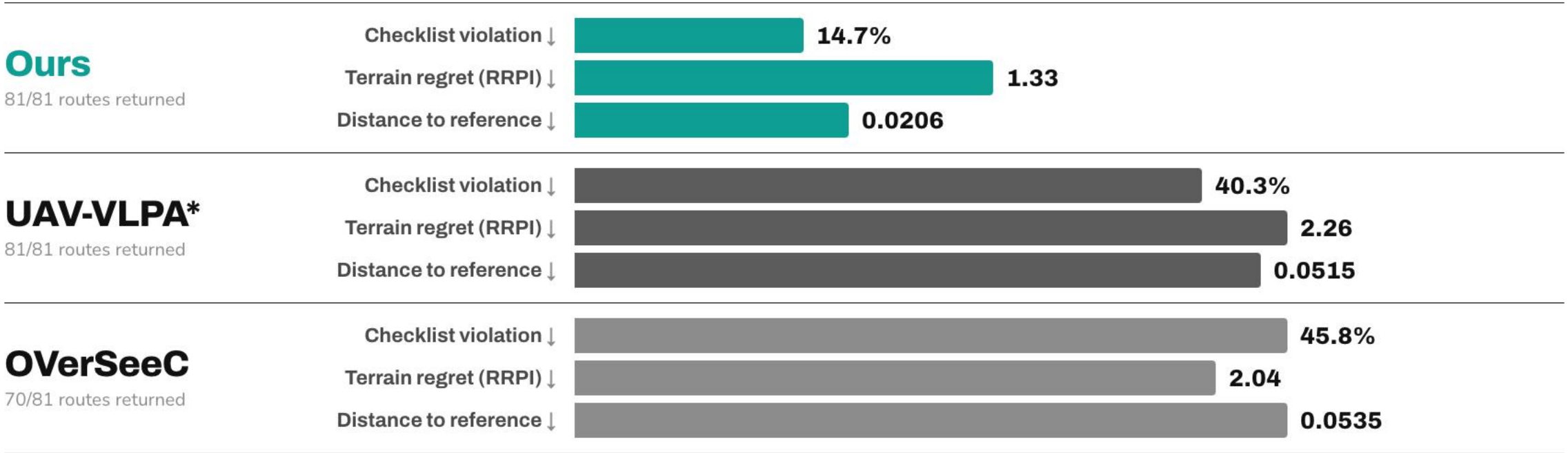
222 m

Path length

Reported as a trade-off, not a score: visiting every checkpoint takes distance.

References and checklist judging are model-assisted, then reviewed · RRPI follows OVerSeeC

Ours follows the mission.
About ⅓ the violations.

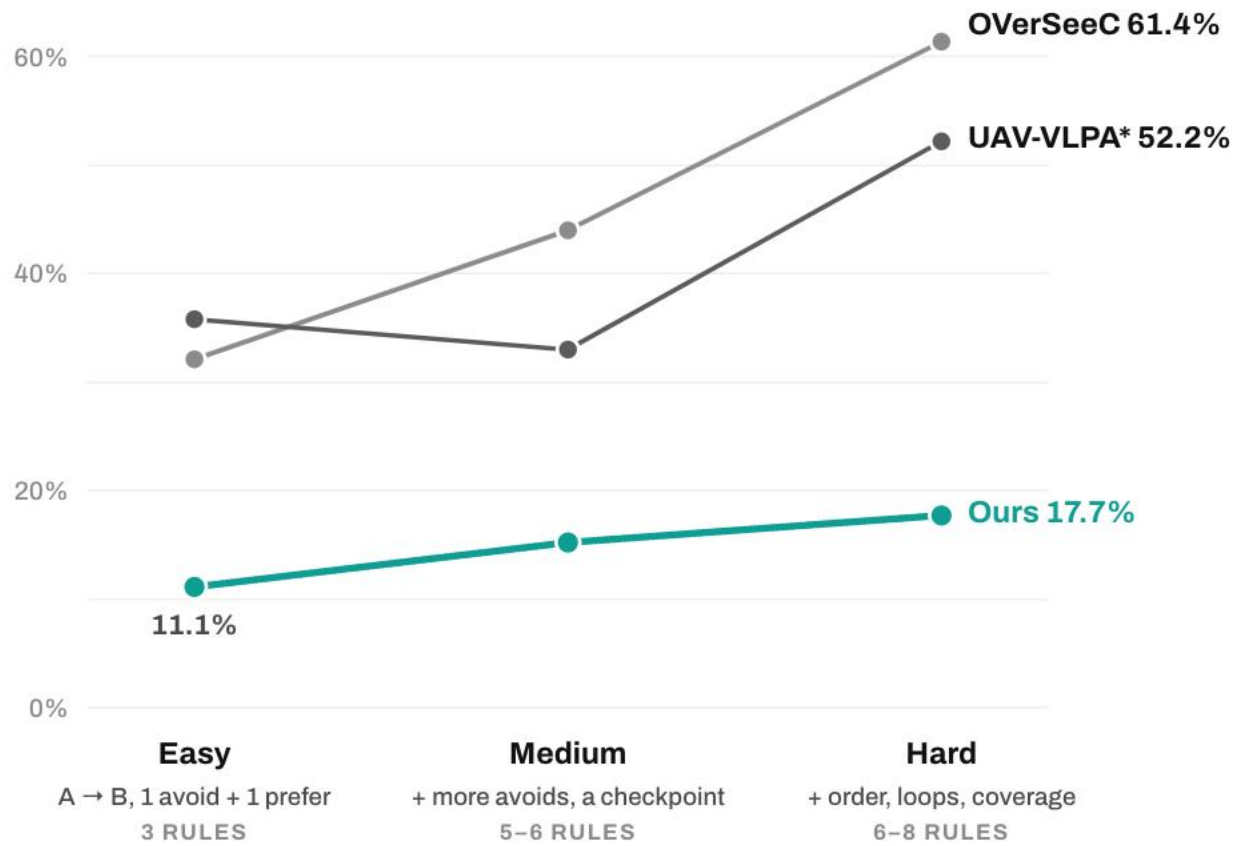


Lower is better · each metric is drawn on its own scale (longest bar = worst method) · means over 81 runs per method (OVerSeeC geometry over its 70 returned routes)

243 runs · 3 locations × 3 scales × 3 difficulties = 27 tasks, × 3 methods × 3 repeats · same image, sentence and endpoints for every method

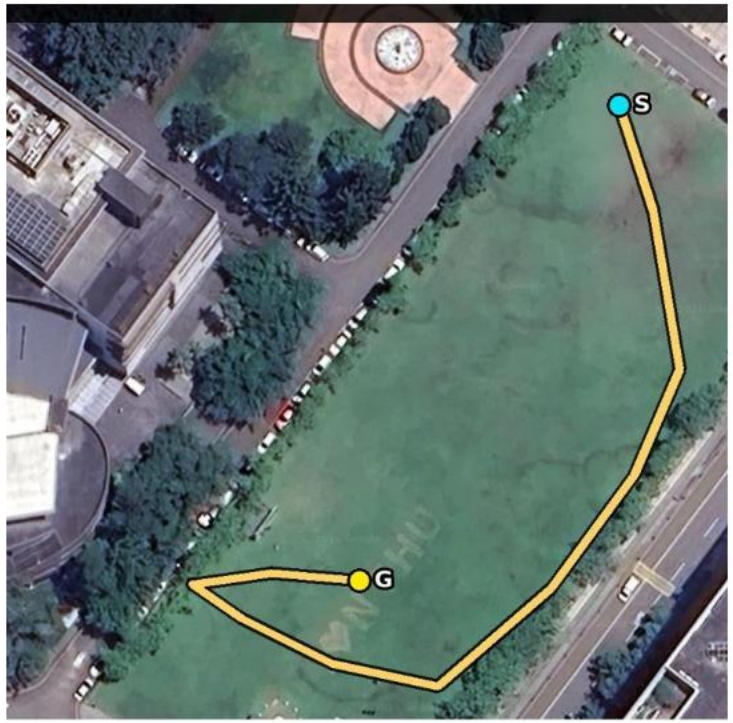
Harder missions, smaller slip.

Easy → Hard, violations rise
+6.6 pts for ours, **+29.3 pts** for OVerSeeC.

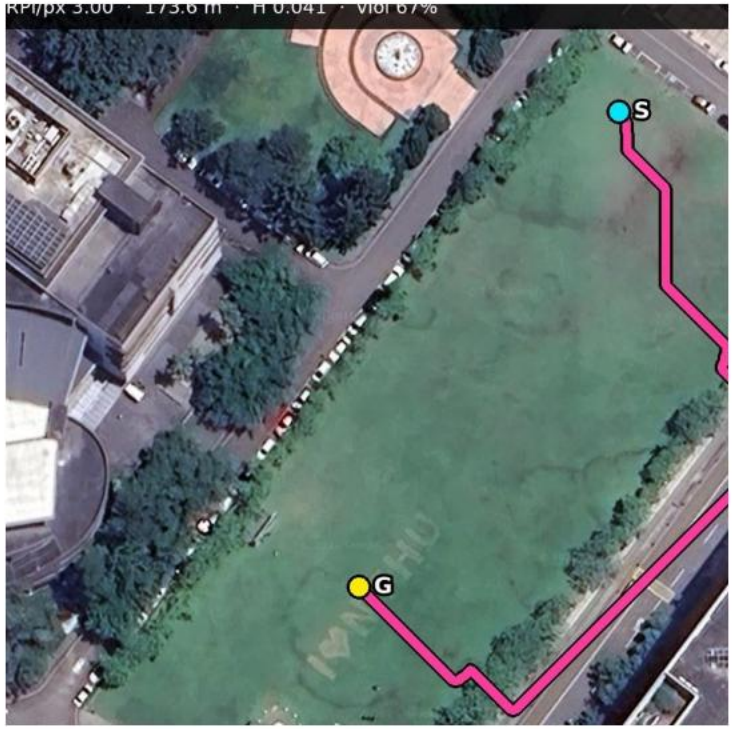


Longer routes — on purpose.

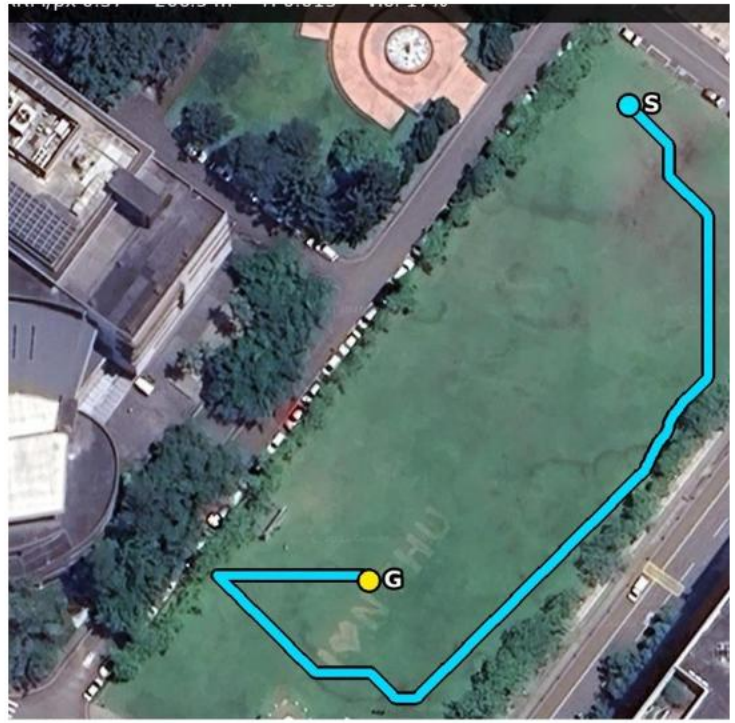
"...visit the east, south-edge and west checkpoints in order, then stop at the large lettering."



Reference



OVerSeeC · 174 m · 67% violated

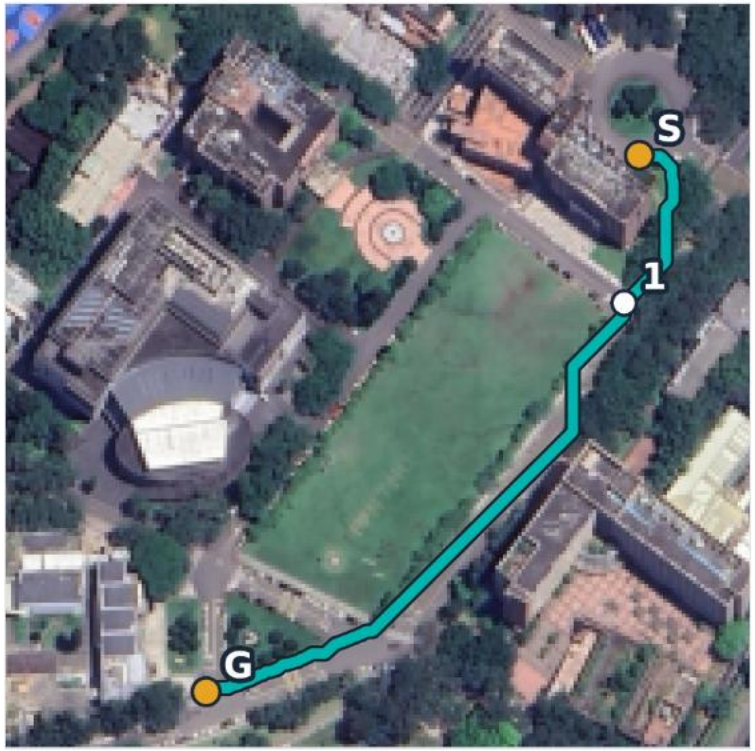


Ours · 207 m · 17% violated

Mean length: ours 222 m vs 143 m / 164 m · Imagery: Google

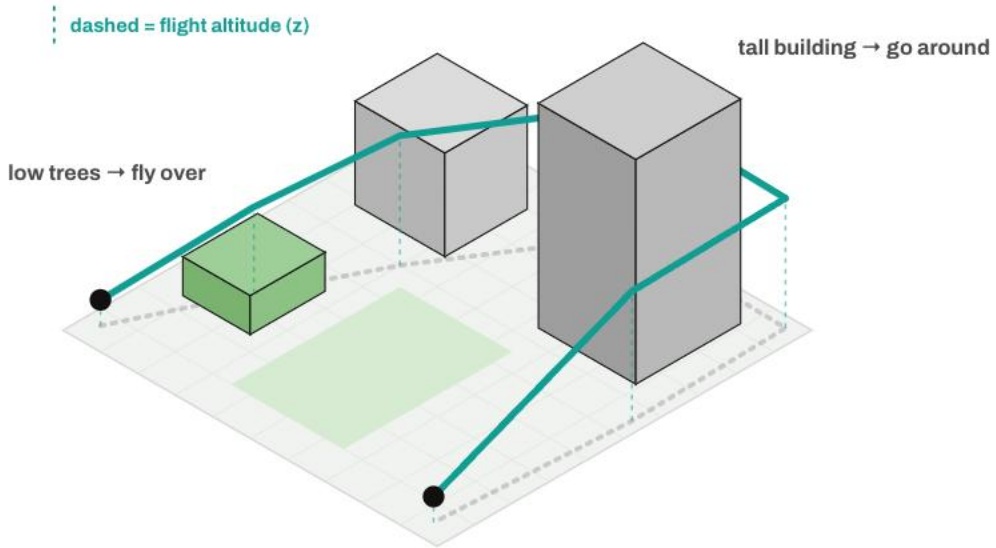
Today: 2D. Next: 3D.

TODAY · X, Y



Planned on a top-down image.
Heights unknown: altitude is left to the operator.

NEXT · X, Y, Z



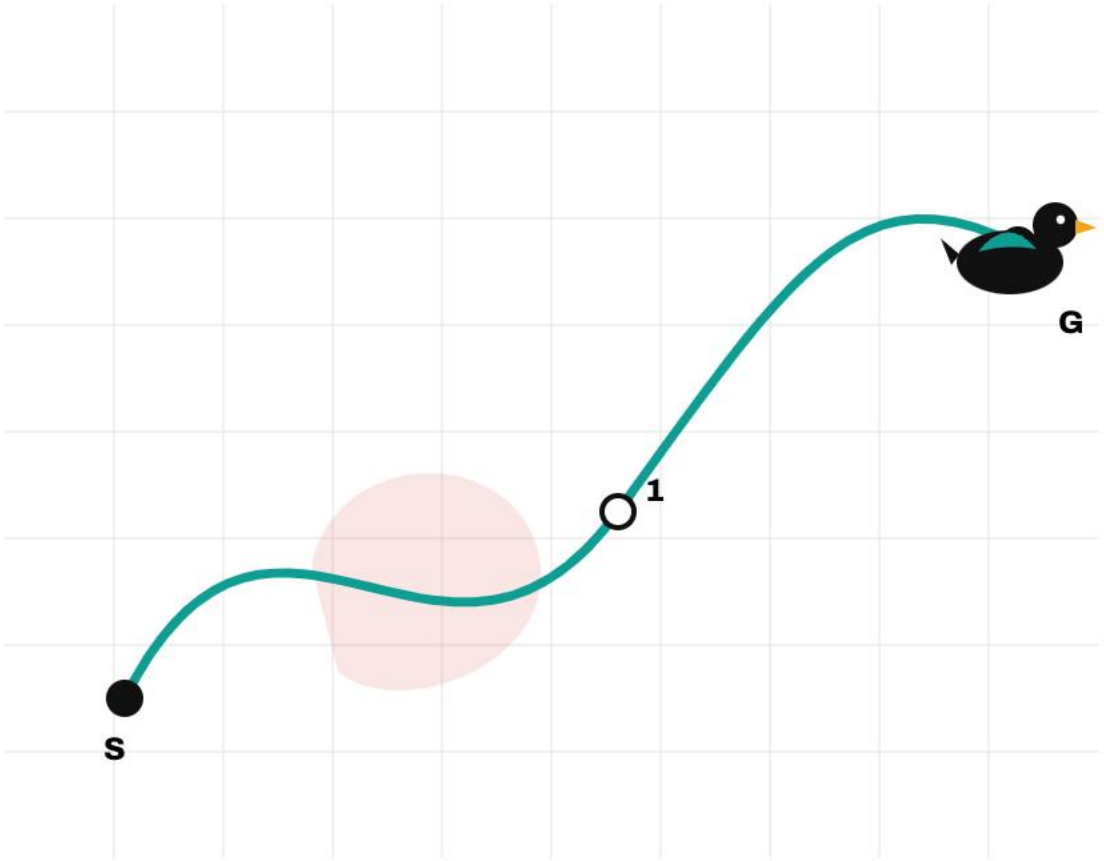
Bring in 3D geometry (height maps, point clouds)
so the agent plans in 3D: over low trees, around tall buildings.

Left: Imagery: Google · Right: illustration

TAAI 2026 · DOMESTIC ORAL SESSION 4

Thank you!

Mapping Intent · agentic UAV route planning with semantic cost maps



Jeng Wen Joshua Lean[1], Ting-Yu Yen[1], Kang-Chiao Lin[2], Xu-Peng He[2], Shih-Hsuan Hung[1], Hung-Kuo Chu[1]

[1]National Tsing Hua University · [2]Industrial Technology Research Institute

Paper #208

21 Nov 2026