UrbanVGGT

Scalable Sidewalk Width Estimation from Street View Images

Kaizhen Tan[1] **Fan Zhang**[2]

[1]Heinz College of Information Systems and Public Policy, Carnegie Mellon University [2]Institute of Remote Sensing and Geographical Information System, Peking University

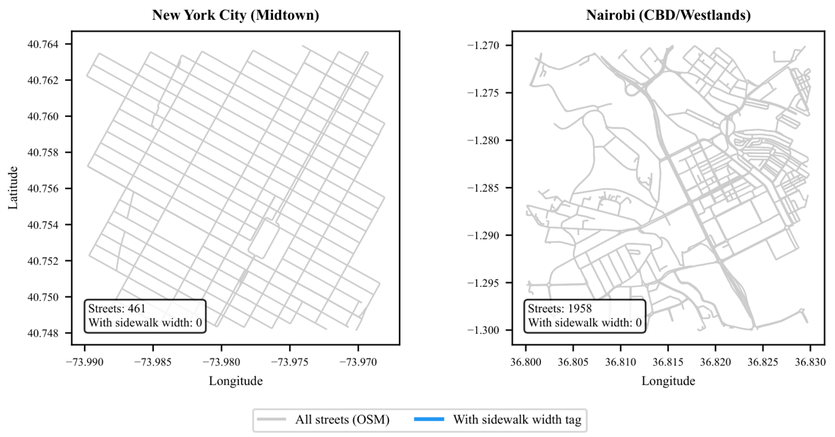
XXV ISPRS CONGRESS
TORONTO, CANADA | 4-11 JULY 2026
FROM IMAGERY TO UNDERSTANDING

01 · MOTIVATION — Why sidewalk width?

Sidewalk width is a micro-scale indicator of **pedestrian accessibility, comfort, and safety**, but metric width data remain scarce. Field surveys do not scale; overhead imagery is often occluded; prior street-view methods often require provider depth, image pairs, or flat-ground assumptions.

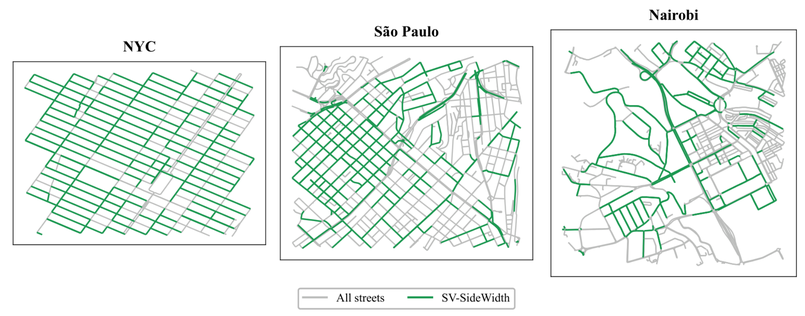
1. **Plane-constrained 3D formulation**
   Sidewalk-width measurement evaluated on a Washington, D.C. ground-truth benchmark.
2. **Controlled backbone comparison**
   Isolating the geometry representation across tested depth and reconstruction backbones.
3. **SV-SideWidth dataset deployment**
   Neighborhood-scale candidate OSM width attributes generated for three cities.

THE DATA GAP — Missing sidewalk width in OSM



Zero *sidewalk-width tags in extracted drivable segments:* **461** *in NYC (Midtown) and* **1,958** *in Nairobi (CBD).*

DEPLOYMENT — SV-SideWidth Prototype

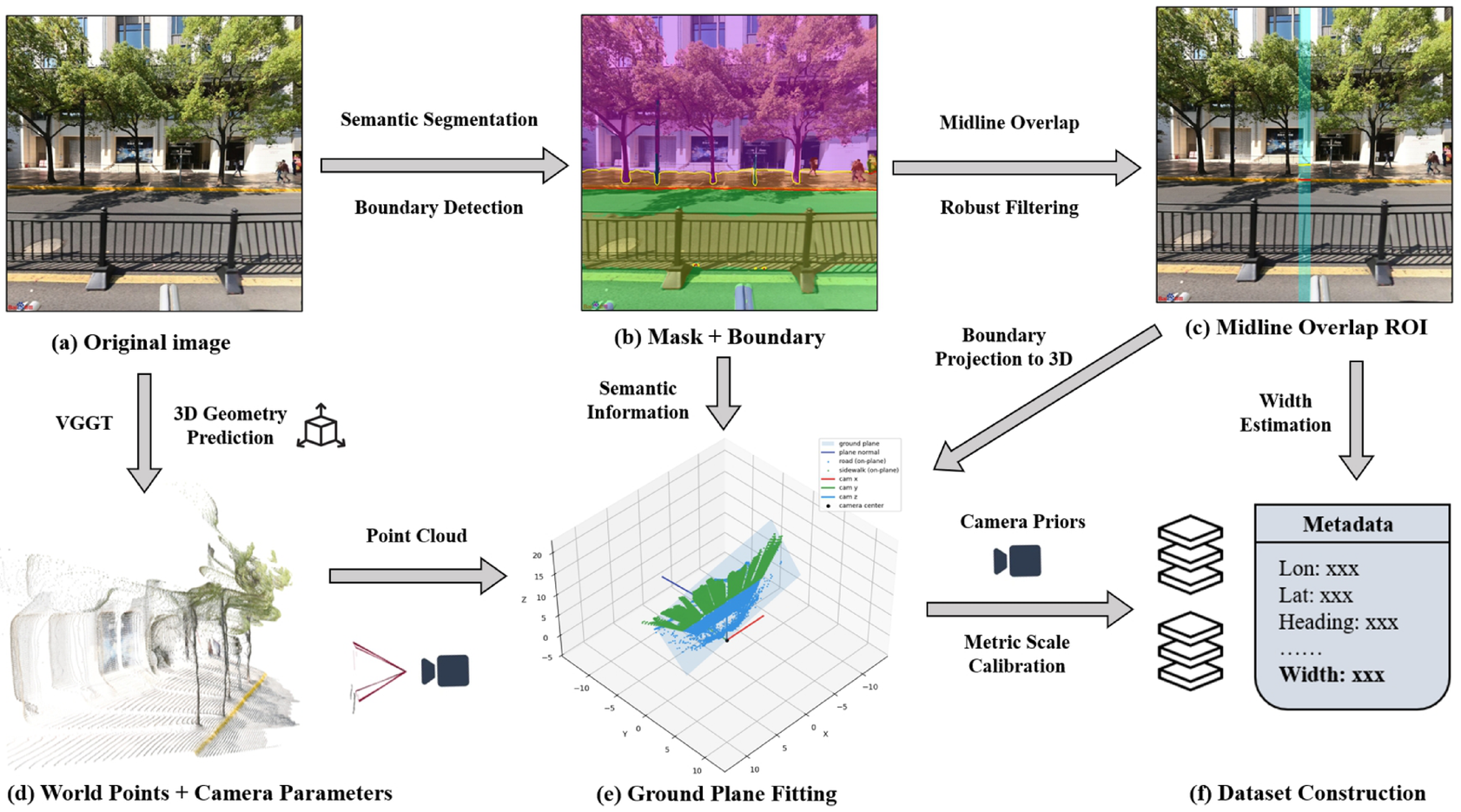


Candidate sidewalk-width estimates linked to OSM segments; local validation pending.

City	Covered	Total	Coverage	Median width
New York	176	461	38.2%	2.58 m
São Paulo	203	1,539	13.2%	2.64 m
Nairobi	148	1,958	7.6%	2.28 m

02 · METHOD — The UrbanVGGT pipeline

`Single image → dense 3D → fitted ground plane → metric sidewalk width`



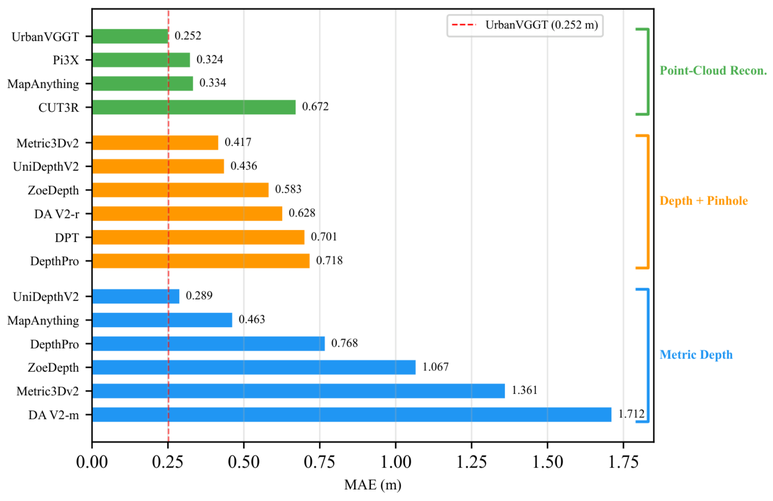
Why VGGT?

VGGT (CVPR 2025) can predict a dense 3D point map and camera pose from a single image in one forward pass, enabling fast, scalable measurement.

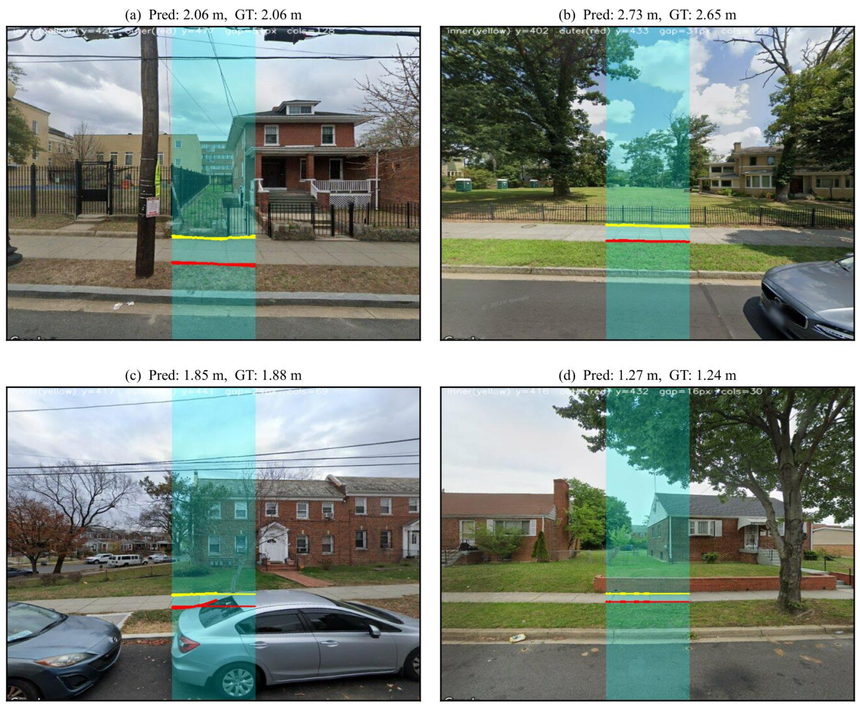
1. **Semantic segmentation**
   SegFormer-B5 (Cityscapes) keeps the sidewalk + road masks and ground support.
2. **3D reconstruction · VGGT**
   Feed-forward dense point map + camera pose from one image, in a single forward pass.
3. **Ground-plane fitting**
   SVD seed → MAD-adaptive RANSAC threshold → SVD refit for a robust support plane.
4. **Metric scale calibration**
   Recover absolute scale from the camera-height prior:
   `s = h_cam / h_pred`
5. **Column-wise width**
   Robust per-column boundary distance along the across-sidewalk direction on the plane.

03 · RESULTS — Best among compared geometry backbones

Same downstream pipeline; only the geometry backbone changes. UrbanVGGT achieves the lowest MAE across 16 evaluation configurations: 0.252 m, with 95.5% of estimates within 0.50 m.

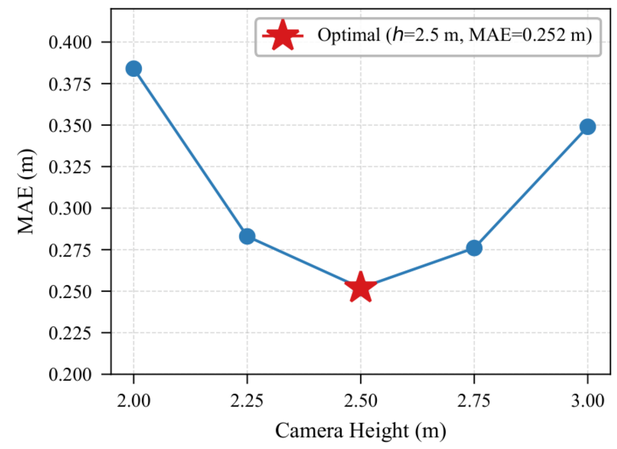


QUALITATIVE — Examples in Washington, D.C.



ABLATION & SENSITIVITY

Camera-height prior



Component ablation

Variant	MAE (m)	RMSE (m)	<0.25 m (%)	<0.50 m (%)
Full pipeline	**0.252**	**0.293**	49.9	**95.5**
− Scale calibration	1.571	1.599	0.0	0.0
Pinhole only	1.096	1.203	4.5	11.0
Full image width	0.265	0.351	**54.2**	88.5

Takeaways

UrbanVGGT estimates **metric sidewalk width from a single street-view image** without stereo pairs, LiDAR, or provider-supplied depth maps. It is best viewed as a scalable candidate-width generation workflow, not an authoritative sidewalk inventory.

Limitations & next steps

Validation is currently limited to one D.C. benchmark. Accuracy may degrade under non-planar ground, occlusion, narrow sidewalks, complex curb geometry, or provider-specific camera-height variation. Next: cross-city ground truth, multi-view fusion, and uncertainty calibration.