GCPR 2024

Worldwide High-fidelity Road Extraction

from Aerial and Satellite Imagery enabled by Low-fidelity OpenStreetMap Labels

Corentin Henry [1]

[1] Remote Sensing Technology Institute
German Aerospace Center (DLR), Germany

Friedrich Fraundorfer [1,2]

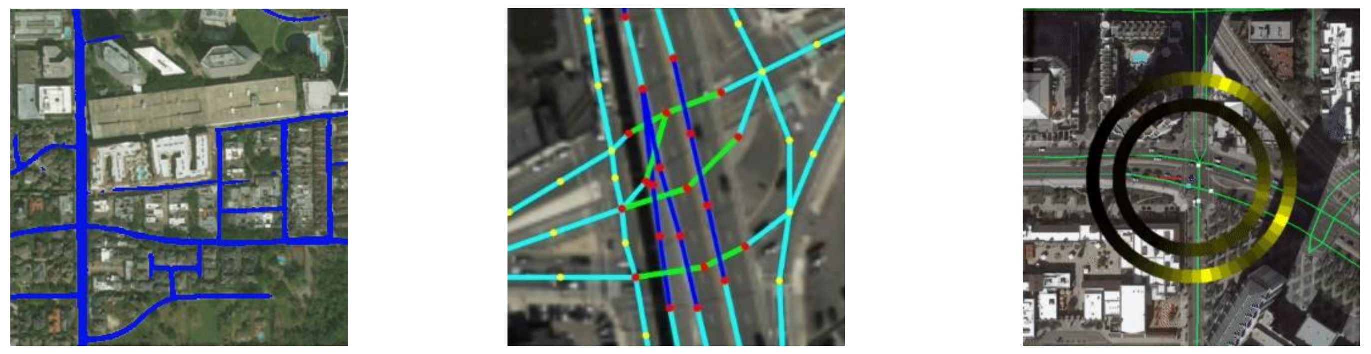
[1,2] Institute of Computer Graphics and Vision
Graz University of Technology (TUG), Austria

Road extraction – Why automate it?

Navigation systems   Urban monitoring   Emergency mapping

40-64M km of roads worldwide, ~17% of roads still unlabeled in the OSM database as of 2016 [1, 2]

Limitations of the current approaches



Segmentation	Graph-tensor Enc. [3]	Traversing agent [4]
+ Global prediction	+ Global prediction	– Local prediction
+ Fast inference	+ Fast inference	– Slow inference
+ Complete surface	– Centerline	– Centerline
– Connectivity	+ Connectivity	+ Connectivity
– Costly annotation	+ Cheap annotation	+ Cheap annotation
+ Easy to train	– Hard to train	+ Easy to train

Training a worldwide model with no manual labels

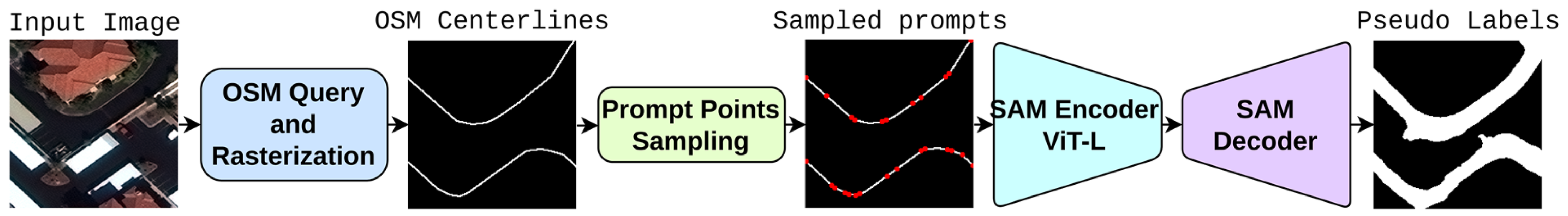
1. Generating high-fidelity pseudo-labels using the Segment Anything Model (SAM) [5]

- Training on 5 worldwide public road datasets
- Fast inference on new areas
- Requires only a low-fidelity inputs (e.g. OSM)

2. Fine-tuning a OneFormer model [6] on unseen regions

- Training on 5 worldwide public road datasets
- Fast fine-tuning with high-fidelity pseudo-labels only
- High performance and fast inference speed

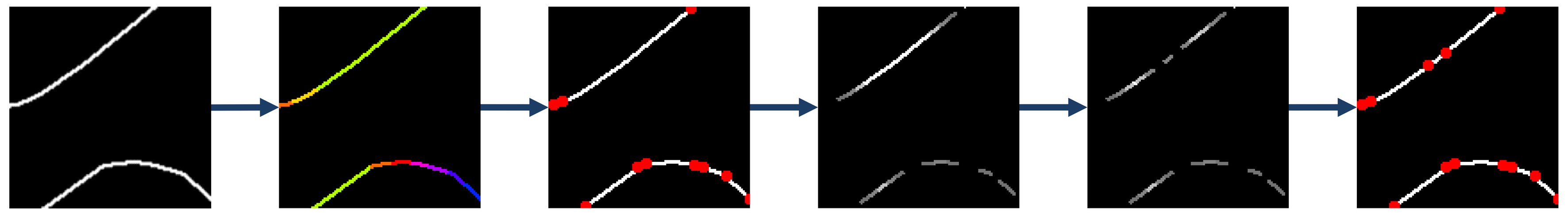
Generating high-fidelity labels



Step 1 – Getting road centerline labels

1. Query OSM database for "highway" objects, filter keys and tags
2. Rasterize vector shapefiles into 1-px-wide road rasters

Step 2 – Sampling evenly spaced prompt points



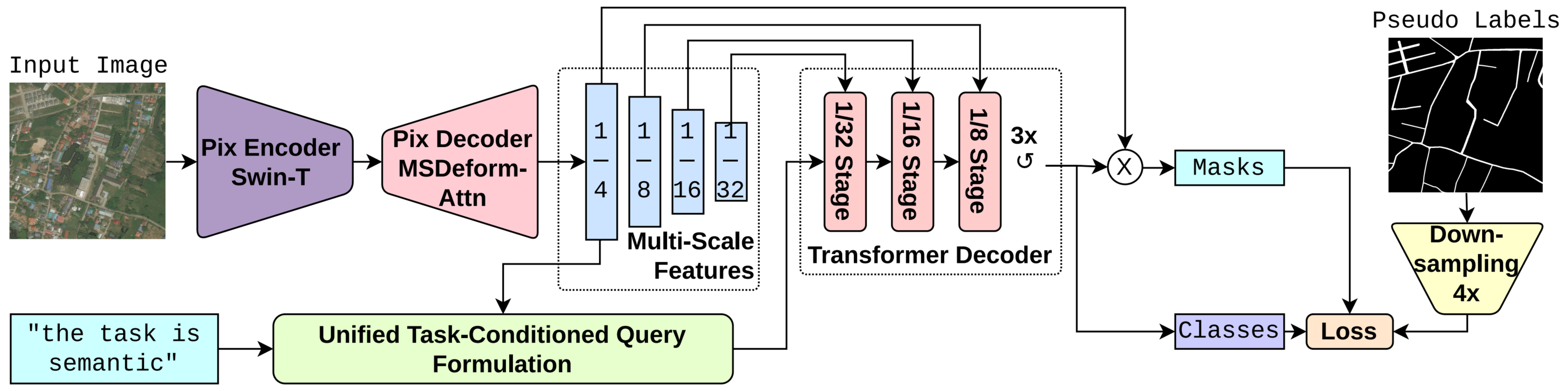
Centerline   Angle [0°, 179°]   Endpoints   Geodetic Dist.   Distance % 40   Midpoints

Step 3 – Generate pseudo-labels with SAM

1. Give endpoints and midpoints as positive prompt points
2. Predict 3 masks with 3 IoU scores per prompt point, keep the best
3. Merge all masks as one binary road mask

Training a worldwide road extraction model

OneFormer model for semantic segmentation



Training on 5 worldwide public benchmark datasets combined

RoadNet 8km² @ 21cm/px   DeepGlobe18 1600km² @ 50cm/px   Mass. Roads 2600km² @ 1m/px   SpaceNet 3 & 5 630km² @ 30cm/px   City Scale 750km² @ 1m/px

Making losses aware of label noise with a ground truth trust coefficient β

Bootstrapped (binary) cross-entropy loss

$$-\sum_i^N [\beta y_{ik} + (1-\beta) p_{ik}] \log(p_{ik})$$

Bootstrapped dice coefficient loss

$$1 - \frac{1 + \sum_i^N 2[\beta y_i + (1-\beta) p_i] p_i}{1 + \sum_i^N [\beta y_i + (1-\beta) p_i]^2 + p_i^2}$$

Results on benchmark datasets

Testing on custom test set from DeepGlobe18

Experiment \ Training	City Scale		DeepGlobe18		Mass. Roads		RoadNet		SpaceNet 3-5		All datasets	
	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.
Trained w/ hi-fi labels	30.09	32.87	**66.05**	**76.05**	29.30	35.24	31.53	34.42	43.78	48.04	63.46	71.05
Fine-t. w/ hi-fi labels	52.98	63.40	–	–	52.40	60.42	57.23	64.67	60.71	69.03	–	–
Fine-t. w/ dOSM labels	48.35	64.91	–	–	47.40	63.55	48.63	62.65	52.65	70.30	–	–
Fine-t. w/ SAM labels	53.50	63.61	–	–	53.59	63.02	58.62	66.23	60.44	68.62	59.57	68.79

Testing on a custom test set from SpaceNet 3 & 5

Experiment \ Training	City Scale		DeepGlobe18		Mass. Roads		RoadNet		SpaceNet 3-5		All datasets	
	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.
Trained w/ hi-fi labels	29.98	30.82	36.77	34.34	15.65	19.91	17.56	13.61	**63.17**	**66.81**	58.71	63.22
Fine-t. w/ hi-fi labels	51.66	54.26	58.14	62.51	49.62	51.97	55.06	63.81	–	–	–	–
Fine-t. w/ dOSM labels	32.16	44.47	41.04	48.83	23.64	33.76	43.70	51.22	–	–	–	–
Fine-t. w/ SAM labels	52.18	55.11	55.42	55.99	50.73	53.93	53.08	62.41	–	–	53.03	59.16

Testing on a custom test set from City Scale

Experiment \ Training	City Scale		DeepGlobe18		Mass. Roads		RoadNet		SpaceNet 3-5		All datasets	
	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.	IoU	Qual.
Trained w/ hi-fi labels	**54.47**	**67.90**	44.98	55.23	42.30	52.92	36.90	41.09	21.75	26.13	44.62	57.73
Fine-t. w/ hi-fi labels	–	–	55.98	67.77	54.78	67.09	51.92	61.71	54.75	66.35	–	–
Fine-t. w/ dOSM labels	–	–	56.23	66.58	54.38	66.54	51.52	60.13	55.63	66.23	–	–
Fine-t. w/ SAM labels	–	–	50.28	60.17	48.69	59.80	48.45	57.13	50.48	60.69	48.88	60.97

Fine-tuning on unseen areas using SAM pseudo-labels

Palu, Indonesia.
Satellite imagery
© 2019 Maxar Technologies

Niigata, Japan.
Satellite imagery
© 2023 Maxar Technologies

Detected before and after fine-tuning
Removed after fine-tuning
Added after fine-tuning

References

[1] Wikipedia, "List of countries by road network size", 2024-09-07
[2] Barrington-Leigh et al., "The world's user-generated road map is more than 80% complete", PLOS ONE 12(8), 1–20 (08 2017)
[3] He et al., "Sat2graph: Road graph extraction through graph-tensor encoding", ECCV, 2020
[4] Bastani et al., "RoadTracer: Automatic Extraction of Road Networks from Aerial Images", CVPR, 2018
[5] Kirillov et al., "Segment anything", arXiv:2304.02643, 2023
[6] Jain et al., "Oneformer: One transformer to rule universal image segmentation", CVPR, 2023