universität freiburg

ROBOTICS SCIENCE AND SYSTEMS

JULY 15-19, 2024 | DELFT UNIVERSITY OF TECHNOLOGY, NETHERLANDS

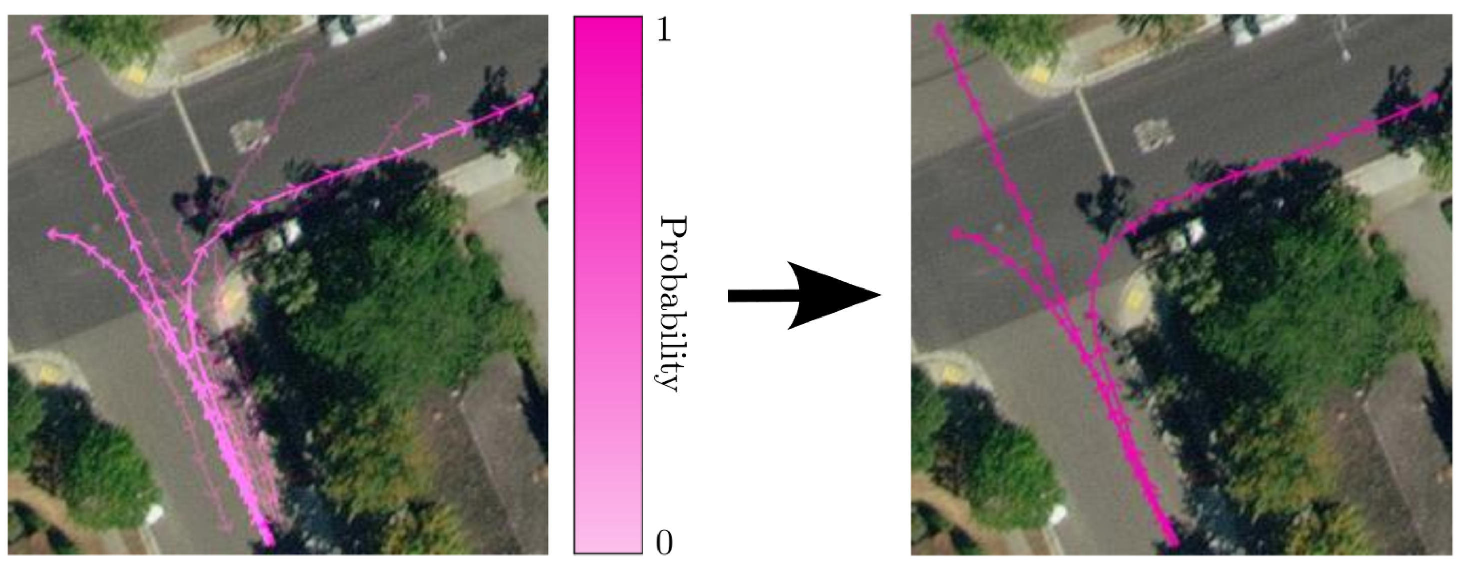
Learning Lane Graphs from Aerial Imagery Using Transformers

Martin Büchner*, Simon Dorer*, and Abhinav Valada

University of Freiburg   * Equal contribution

ALGT

Motivation



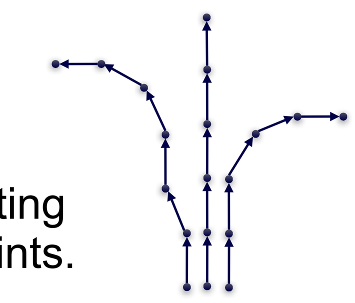
Path Predictions   **Successor Lane Graph**

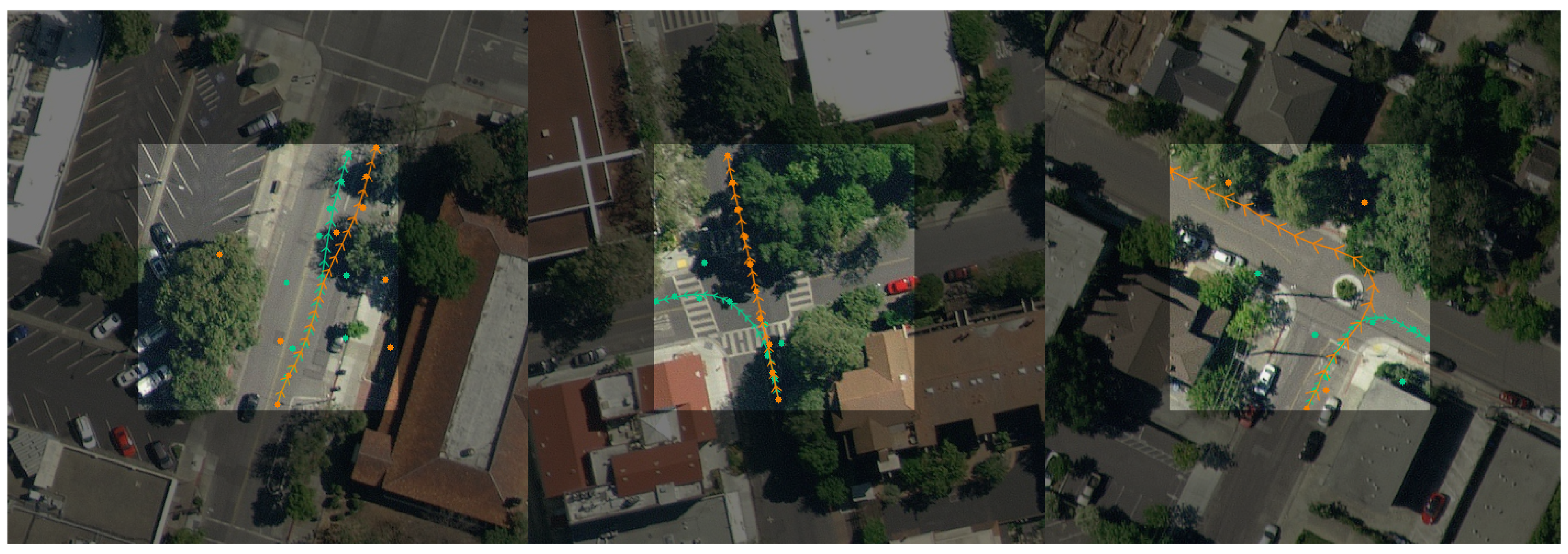
- How can we leverage transformers for aerial lane graph prediction?
- What are effective image encoding backbones?
- What is the most suitable path parametrization for set-level lane graph prediction?

Method

Lane Graph Representation

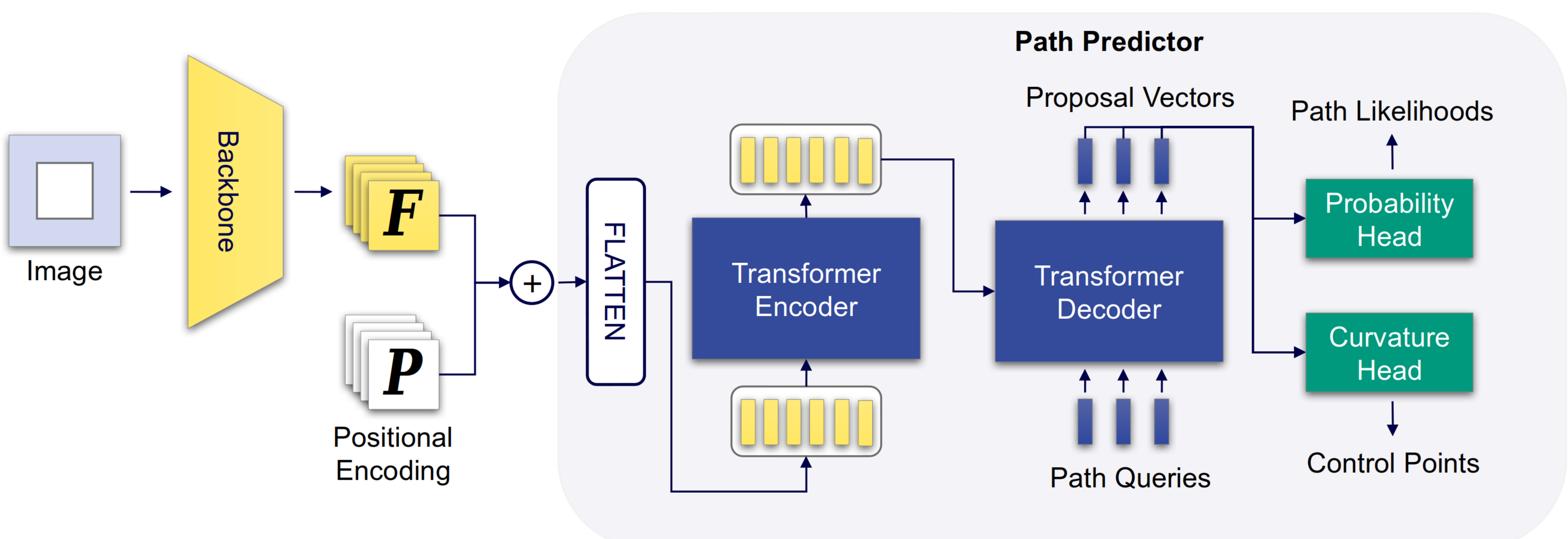
- Following MapTR [1], we represent successor lane graphs as maximal-length paths to enable set-level predictions.
- We parametrize traversal paths as standard polylines consisting of straight segments or Bézier curves using 10-20 control points.





Aerial Lane Graph Transformer (ALGT)

- As proposed in the UrbanLaneGraph [2] dataset we take aerial crops including context and embed them using various image feature encoders.
- We add fixed two-dimensional sinusoidal positional encodings, flatten the feature maps and feed them to a transformer encoder.
- Based on set of fixed-size vector path queries the transformer decoder produces proposal vectors representing lane graph paths. Thus we do not utilize any positional encoding within the decoder.
- We regress path probabilities and associated control points using MLPs.



- Training: We minimize a composite loss consisting of the MSE of the Hungarian matching objective between predicted and GT paths using the Manhattan distance and the predicted path probability:

$$\mathcal{L} = \alpha \cdot \sum_{Y_i \in Y} \mathcal{L}_{mse}(Y_i, \hat{Y}_{\sigma^*(i)}) + \beta \cdot \mathcal{L}_{bce}$$

- Aggregation: We disregard paths not meeting a certain minimum likelihood threshold. The remaining paths are fused based on Euclidean node-to-node merging to obtain cohesive successor lane graphs.

Results

Ablation Study on UrbanLaneGraph – Palo Alto Split

Variant	TOPO P/R	GEO P/R	APLS	SDA_{20}	SDA_{50}	Graph IoU
Path Representation						
Bézier	0.395/0.339	0.567/0.527	0.619	0.191	0.405	0.290
Polyline	**0.479/0.420**	**0.639/0.594**	**0.664**	**0.251**	**0.479**	**0.338**
Backbone						
ResNet-50	0.268/0.223	0.433/0.396	0.509	0.158	0.367	0.200
ViT-B-16	0.253/0.212	0.418/0.383	0.465	0.114	0.348	0.195
PSPNet	0.479/**0.420**	0.639/**0.594**	0.664	**0.251**	**0.479**	0.338
PSPNet + ResNet-50	**0.485**/0.414	**0.644**/0.587	**0.665**	0.237	0.447	**0.345**
Architecture						
$(1, 1, 64, 10)$	0.432/0.353	0.598/0.536	0.643	0.193	0.344	0.304
$(2, 2, 128, 20)$	0.474/0.402	0.631/0.575	0.651	0.223	0.420	0.331
$(4, 4, 128, 10)$	**0.479/0.420**	**0.639/0.594**	**0.664**	**0.251**	**0.479**	**0.338**

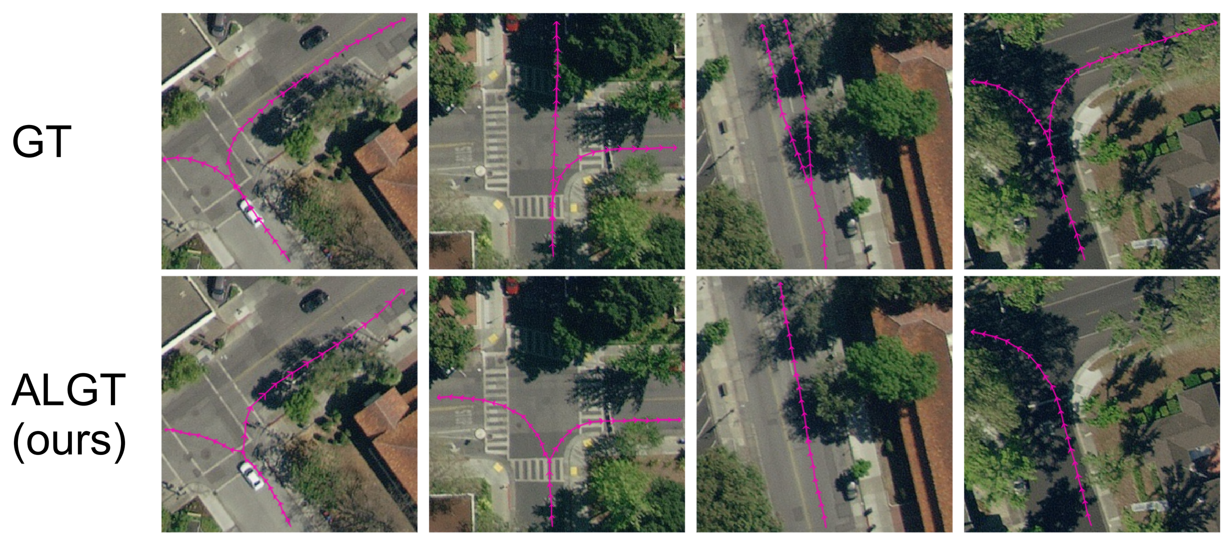
- Polyline representations outperform Bézier curves by a large margin.
- Segmentation backbones such as PSPNet are most suitable for lane graph prediction from aerial imagery.
- Detection backbones such as ViT or ResNet do not yield significant performance gains.

Comparison against LaneGNN on UrbanLaneGraph (Palo Alto)

Method	TOPO P/R	GEO P/R	APLS	SDA_{20}	SDA_{50}	Graph IoU
LaneGNN	**0.584/0.744**	0.582/**0.739**	0.177	0.220	0.367	**0.378**
ALGT	0.481/0.437	**0.645**/0.606	**0.714**	**0.224**	**0.497**	0.343

- ALGT shows competitive performance on the ULG benchmark when comparing against LaneGNN.
- The ALGT model vastly outperforms LaneGNN on the APLS and SDA metrics. Nonetheless, the topological accuracy of LaneGNN [2] is higher.
- Our approach does not suffer from inaccurate node positions and resolves the limitation of sampled node manifolds.

Failure Cases



- Topological errors mostly stem from missing entire paths.
- Similar to previous findings [2], inferring intersection rules from aerial views remains hard.

Conclusion

- We presented a novel successor lane graph prediction approach that generates highly accurate paths while not suffering from node initialization errors.
- Polyline path representations seem to outperform Bézier parametrizations.
- Future work could address the learned temporal aggregation of transformer-based predictions as well as the out-of-distribution problem inherent to large-scale lane graph prediction.

References

[1] Liao *et al.*, "MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction," *ICLR*, 2023.

[2] Buechner et al., "Learning and Aggregating Lane Graphs for Urban Automated Driving," CVPR, 2023.