Supplementary Document of RNGDet++

Zhenhua Xu, *Student Member, IEEE*, Yuxuan Liu, *Student Member, IEEE*,
Yuxiang Sun, *Member, IEEE*, Ming Liu, *Senior Member, IEEE*, and Lujia Wang, *Member, IEEE*

I. Instance Segmentation

The instance segmentation head predicts the heatmap of the ahead road of each vertex query. The instance segmentation head can help the network to better capture the information of road networks, which facilitates the training process and improves the reasoning ability of the network. The GT (Ground Truth) instance segmentation map marks the road ahead, which is a segment of the GT road. Suppose the agent current position is v_t, and the GT vertex in the next step is v^*_{t+1} (v^*_{t+1} is already on the GT road). Note that the GT instance segmentation is not the line connecting v_t and v^*_{t+1}. We first project v_t to the GT road as v^*_t, and then use the road ahead connecting v^*_t and v^*_{t+1} as the GT instance segmentation mask. In this way, the instance segmentation head can train the network to capture the correct road information when it deviates from the right track, which improves the final performance. An example is visualized in Fig. 1 to show how we define instance segmentation labels.

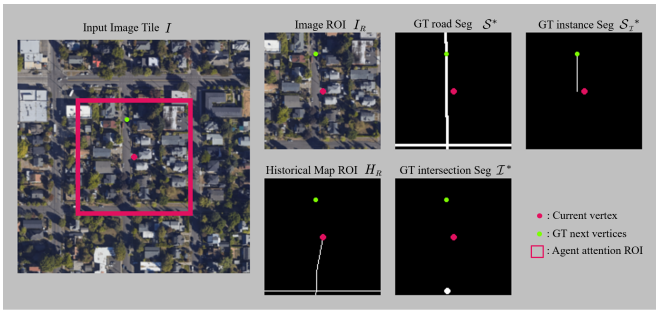


Fig. 1: Visualization of the GT instance segmentation mask (top right mask). When the agent is away from the road (pink point), the instance segmentation head still supervises the agent to capture the correct road information, which improves the final performance.

II. Training Data Sampling

We propose a BFS (Breath First Search) graph traversal algorithm as the imitation learning expert. The expert traversals the GT road network graph vertex by vertex. At each step, the expert generates a training sample $(I_R, H_R, \mathcal{S}^*, \mathcal{I}^*, \mathcal{S}_\mathcal{I}{}^*, \{(v^i_{t+1})*\}^N_{i=1})$, where I_R is cropped aerial image ROI, H_R is cropped historical image ROI, $\mathcal{S}^*$ is the GT road segment segmentation map, $\mathcal{I}^*$ is the GT road intersections segmentation map, $\mathcal{S}_\mathcal{I}{}^*$ is the GT instance segmentation map, and $\{(v^i_{t+1})*\}^N_{i=1})$ are the GT vertices in the next step. To enhance robustness, we add noise to the trajectory of the expert during the sampling phase. Some examples of sampled training data are visualized in Fig. 2.

III. RNGDet++ Structure

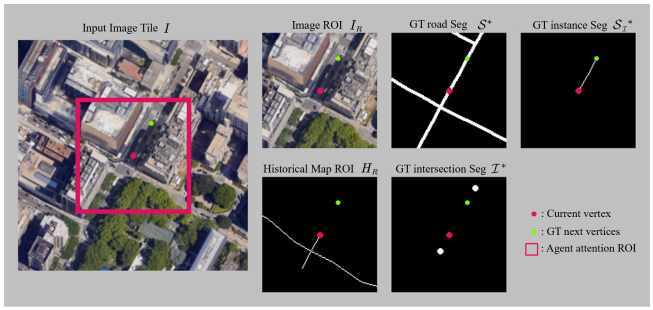
Compared with the previous SOTA approach RNGDet, RNGDet++ can utilize all feature layers of the backbone network. It also adds the instance segmentation head to facilitate the training of the network. The network structures of different models are visualized in Fig. 3.

IV. Additional Visualizations

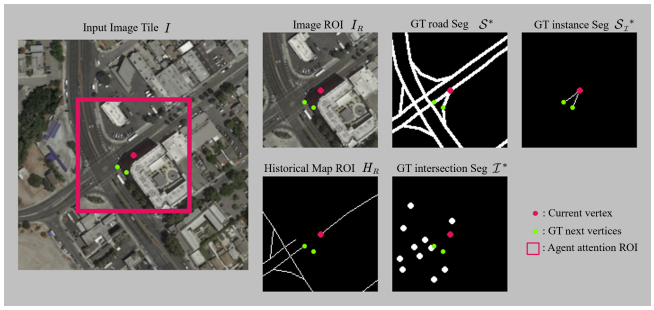
References

[1] Z. Xu, Y. Liu, L. Gan, Y. Sun, X. Wu, M. Liu, and L. Wang, "Rngdet: Road network graph detection by transformer in aerial images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

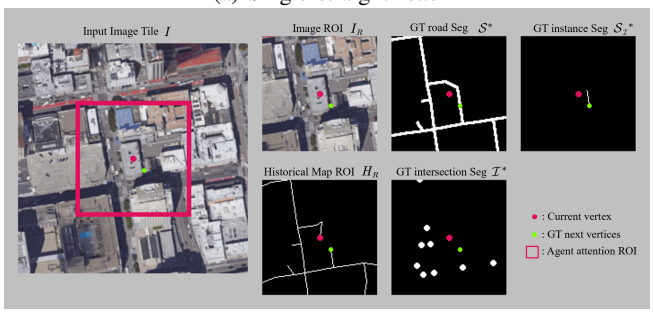
[2] S. He, F. Bastani, S. Jagwani, M. Alizadeh, H. Balakrishnan, S. Chawla, M. M. Elshrif, S. Madden, and M. A. Sadeghi, "Sat2graph: road graph extraction through graph-tensor encoding," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*. Springer, 2020, pp. 51–67.



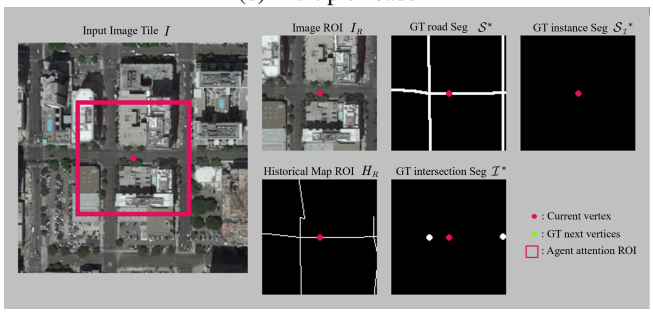
(a) Single straight road



(b) Multiple roads

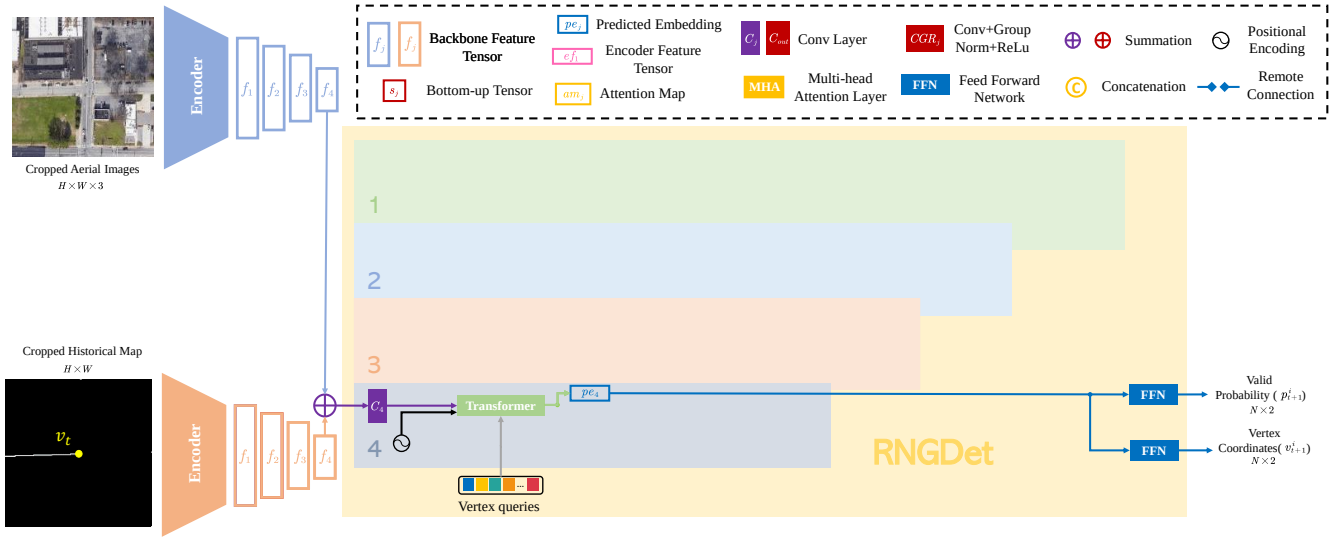


(c) Road loop

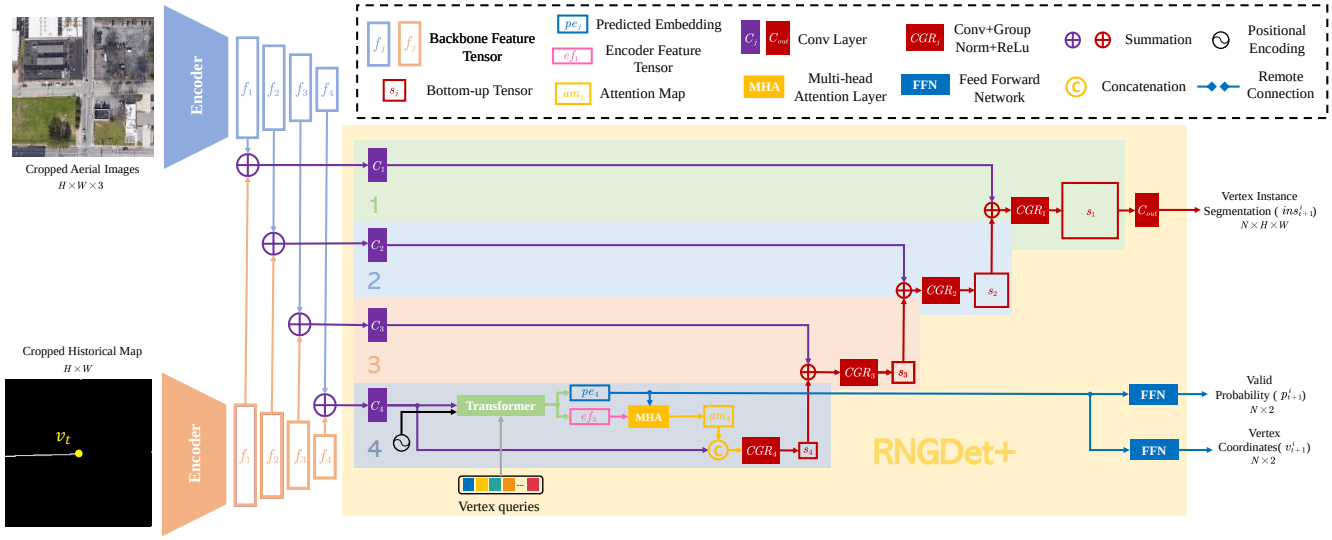


(d) Stop

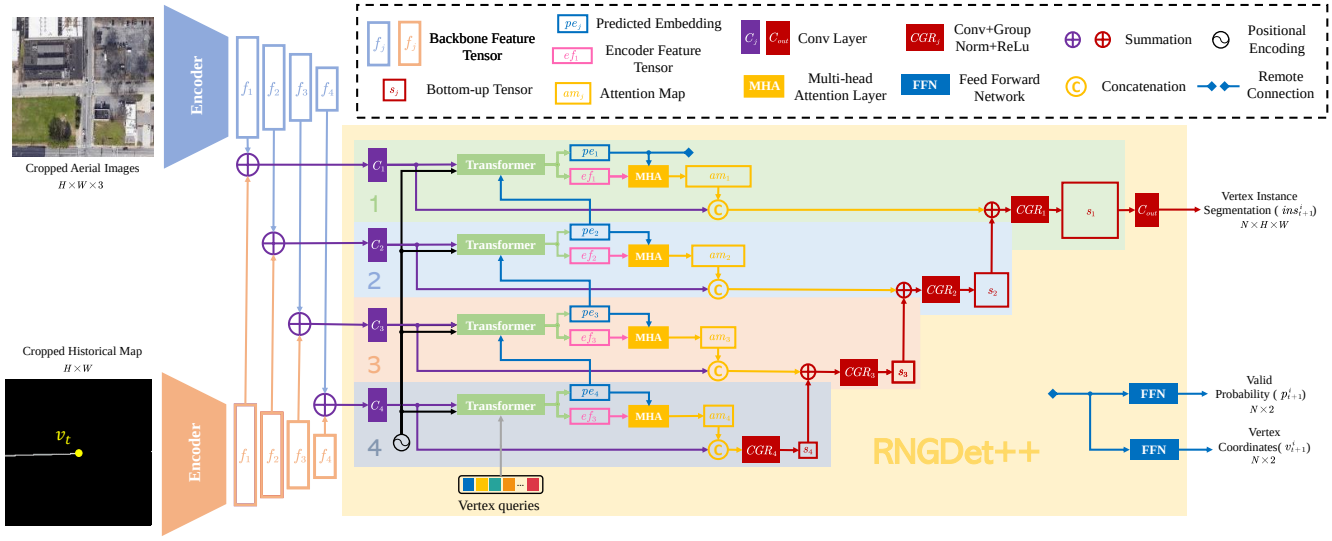
Fig. 2: Visualizations of sampled training data. For better visualization, we add the instance segmentation labels of all queries into a single mask (top right). Note that this is the visualization of the expert sampling process, not the inference results. (a) A sample on the straight road. (b) A sample on intersections. There are two vertices in the next step. (c) A sample on the loop road. The agent should predict and align with previously predicted vertices to close the road loop. (d) A sample on stop. There is no unexplored road ahead, thus the agent should stop. The proposed expert sampling algorithm can generate correct expert trajectories for imitation learning with high efficiency.



(a) RNGDet

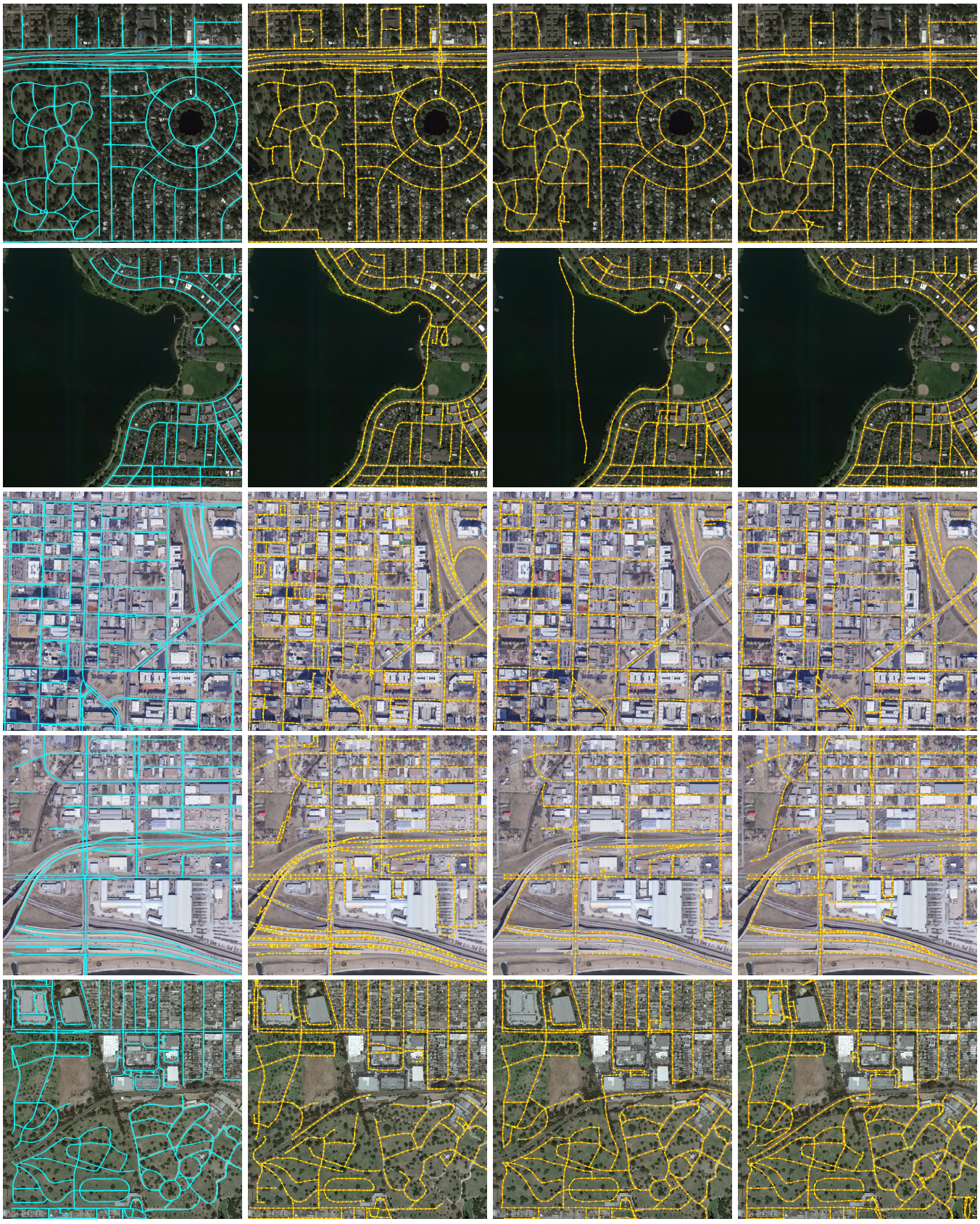


(b) RNGDet++ with a single feature layer



(c) RNGDet++

Fig. 3: Network structures of different approaches. (a) RNGDet [1]. RNGDet only leverages the deepest feature layer and does not have the instance segmentation head. (b) RNGDet++ with a single feature layer (the deepest layer). Compared with RNGDet, this model adds the instance segmentation module. (c) RNGDet++. It can utilize multi-scale features for road network detection, and can facilitate the training process by the instance segmentation head.



(a) Ground truth (b) Sat2Graph [2] (c) RNGDet [1] (d) RNGDet++

Fig. 4: Qualitative visualization. (a) Ground truth road networks (Cyan lines). (b) Road network graph detected by Sat2Graph. (c) Road network graph detected by RNGDet. (d) Road network graph detected by RNGDet++. For (b)-(d), yellow points represent graph vertices and orange lines represent graph edges. For the visualization, it is found that RNGDet++ can output road network graph with more accurate structure and correctness compared with previous works. This figure is best viewed in color. Please zoom in for details.