

# Driving Scene Understanding with Traffic Scene-Assisted Topology Graph Transformer

## Appendix

Paper ID: 4230

### ACM Reference Format:

Paper ID: 4230. 2024. Driving Scene Understanding with Traffic Scene-Assisted Topology Graph Transformer Appendix. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## A ADDITIONAL MODULE DETAILS

**Traffic Scene-Assisted Reasoning Module.** In this section, we primarily elaborate on the traffic rule matrix  $M_{tr}$  within the traffic scene-assisted reasoning module. The traffic rule matrix  $M_{tr}$  is constructed based on the categories of traffic elements and the directions of lanes. In the Openlane-V2 dataset, excluding the semantically unobservable *unknown* category, valid elements are classified into two groups: the traffic light class and the road sign class. The categories of the traffic light class include three types: *red*, *green* and *yellow*, which serve as indicators for all directions of lanes within the control range. The road sign class comprises nine categories: *go\_straight*, *turn\_left*, *turn\_right*, *no\_left\_turn*, *no\_right\_turn*, *u\_turn*, *no\_u\_turn*, *slight\_left*, *slight\_right*, which only indicate the lanes with the corresponding directions within the control range. For instance, the *turn\_left* and *no\_left\_turn* traffic elements only associate with the lanes of turning left. The complete traffic rule matrix  $M_{tr}$  is illustrated in Table 1. By feeding the detected categories of traffic elements and inferred lane directions into  $M_{tr}$ , the correct traffic rule constraint topology result can be obtained.

## B ADDITIONAL EXPERIMENTAL STUDIES

**The other Image Backbones.** In this part, we present the experimental results of training and inference by employing VOV and Swin-B as the image backbone of TSTGT, and compare them fairly with TopoMLP. The result of the comparison on Openlane-V2 *subset\_A* are reported in Table 2. It can be observed that when utilizing the VOV backbone, our approach achieves an OLS score of 43.0, compared to 40.1 of TopoMLP. Notably, employing Swin-B as backbone yields an OLS score of 44.6, representing a 2.4 OLS score improvement over TopoMLP. Furthermore, when the training epoch is set to 48, our model demonstrates superior performance with an OLS score of 46.9, outpacing TopoMLP by 3.2 OLS score under identical conditions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
Conference'17, July 2017, Washington, DC, USA

© 2024 Association for Computing Machinery.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00  
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Table 1: The detailed information of the traffic rule matrix.

$\begin{matrix} C_{tr} \\ D_{lane} \end{matrix}$	left	straight	right
<i>unknown</i>	0	0	0
<i>red</i>	1	1	1
<i>green</i>	1	1	1
<i>yellow</i>	1	1	1
<i>go_straight</i>	0	1	0
<i>turn_left</i>	1	0	0
<i>turn_right</i>	0	0	1
<i>no_left_turn</i>	1	0	0
<i>no_right_turn</i>	0	0	1
<i>u_turn</i>	1	0	0
<i>no_u_turn</i>	1	0	0
<i>slight_left</i>	1	0	0
<i>slight_right</i>	0	0	1

Table 2: Comparison with state-of-the-art methods on the Openlane-V2 *subset\_A* dataset using other backbones.

Method	Backbone	Epoch	DET <sub>l</sub>	DET <sub>t</sub>	TOP <sub>ll</sub>	TOP <sub>lt</sub>	OLS
TopoMLP	VOV	24	29.7	52.1	7.9	25.6	40.1
TSTGT	VOV	24	<b>32.0</b>	<b>50.6</b>	<b>14.3</b>	<b>26.5</b>	<b>43.0</b>
TopoMLP	Swin-B	24	30.7	<b>54.3</b>	9.5	28.3	42.2
TSTGT	Swin-B	24	<b>32.6</b>	53.7	<b>15.1</b>	<b>28.4</b>	<b>44.6</b>
TopoMLP	Swin-B	48	32.5	53.5	11.9	29.4	43.7
TSTGT	Swin-B	48	<b>35.4</b>	<b>55.1</b>	<b>18.2</b>	<b>29.6</b>	<b>46.9</b>

Table 3: Comparison with state-of-the-art methods on the Openlane-V2 *subset\_B* dataset using other backbones.

Method	Backbone	Epoch	DET <sub>l</sub>	DET <sub>t</sub>	TOP <sub>ll</sub>	TOP <sub>lt</sub>	OLS
TopoMLP	VOV	24	29.6	<b>62.2</b>	8.9	20.5	41.7
TSTGT	VOV	24	<b>32.4</b>	61.4	<b>18.2</b>	<b>21.1</b>	<b>45.6</b>
TopoMLP	Swin-B	24	32.3	65.5	10.5	23.2	44.6
TSTGT	Swin-B	24	<b>34.4</b>	<b>68.0</b>	<b>21.0</b>	<b>24.2</b>	<b>49.3</b>

In the experimental results of Openlane-V2 *subset\_B* recorded in Table 3, it is evident that when using VOV as the backbone, TSTGT surpasses TopoMLP by a 3.9 OLS score, reaching an OLS score of 45.6. Furthermore, employing the more powerful Swin-B as the backbone, our model achieves an OLS score of 49.3. The utilization of diverse backbones across distinct data subsets consistently yields exceptional outcomes, thus affirming the efficacy of our approach.

**Table 4: The ablation experiments of DCTGT components, where LL part represents lane-lane part of DCTGT, LT part represents lane-traffic part of DCTGT and TSR represents traffic scene-assisted reasoning module.**

LL part	LT part	TSR	DET <sub>l</sub>	DET <sub>t</sub>	TOP <sub>ll</sub>	TOP <sub>lt</sub>	OLS
			28.9	50.4	8.2	23.2	39.0
✓			28.4	50.2	11.3	23.1	40.1
	✓		28.9	<b>51.6</b>	8.2	23.2	39.3
✓	✓		28.5	50.9	11.6	23.3	40.4
✓	✓	✓	<b>29.0</b>	50.5	<b>12.1</b>	<b>23.5</b>	<b>40.7</b>

**Divide-and-Conquer Topology Graph Transformer.** Here, we conduct a more detailed component analysis to demonstrate the effectiveness of each part within the divide-and-conquer topology graph Transformer (DCTGT). The experiments utilize ResNet-50 as the backbone and are conducted on the Openlane-V2 *subset\_A* dataset. The experimental results are presented in Table 4. The initial setup involves the TSTGT only using the point-wise matching strategy, with the original topology reasoning part replaced by the MLP structure similar to that in TopoMLP. This initial setup achieves an OLS score of 39.0. When incorporating the lane-lane part and lane-traffic part of DCTGT separately, the model reaches OLS scores of 40.1 and 39.3, respectively, resulting in gains of 1.1 and 0.3 OLS scores. When applying both components to the initial model simultaneously, the model attains an OLS score of 40.4, indicating a 1.4 OLS score improvement compared to the initial setup. Finally, the inclusion of the traffic scene-assisted reasoning module results in the highest performance, obtaining an OLS score of 40.7.

**Traffic Scene-Assisted Reasoning Module.** In this division, we explore the relevant parameter settings of the traffic scene-assisted reasoning module. Experiments are employed using ResNet-50 as the backbone on the Openlane-V2 *subset\_A* dataset. We first investigate the control range threshold  $\tau$ , which is set to 0.1, 0.2, and 0.5, respectively. The experimental results are shown in Table 5. As we can observe that the TSTGT model achieves the best accuracy by setting  $\tau$  as 0.2, while a broader constraint, such as 0.5, can even degrade the performance. Therefore, we choose  $\tau=0.2$  as the control range threshold of the traffic scene-assisted reasoning module.

We also investigate the impact of different settings for the lane direction threshold  $\sigma$  on the model's performance. We set the lane direction threshold to 0.05, 0.1, and 0.2, respectively, and the outcomes are shown in Table 5. It can be seen that when the threshold is set to 0.1, the model can more accurately distinguish lanes of different directions. As this threshold increases, some turning lanes may be incorrectly identified as straight lanes, thereby affecting the accuracy of prediction results. Therefore, we set this threshold to 0.1 to ensure optimal model performance.

**MPNN Layer.** In this part, we discuss the influence of different configurations of the Message Passing Neural Network (MPNN) layer in the divide-and-conquer topology graph Transformer (DCTGT) on the model performance. We primarily employ two prevalent MPNN architectures: GINE and GatedGCN. Node and edge features are utilized in the MPNN for local information interaction and propagation. GatedGCN updates both node and edge features,

**Table 5: Model analysis of different settings in traffic scene-assisted reasoning module.**

Method	Settings	DET <sub>l</sub>	DET <sub>t</sub>	TOP <sub>ll</sub>	TOP <sub>lt</sub>	OLS
<b>Control Range Threshold</b>						
TSTGT	0.1	27.6	49.4	11.6	23.6	39.9
TSTGT	0.2	29.0	50.5	12.1	23.5	<b>40.7</b>
TSTGT	0.5	28.4	49.4	11.6	23.9	40.2
<b>Lane Direction Threshold</b>						
TSTGT	0.05	28.5	50.2	11.7	22.9	40.2
TSTGT	0.1	29.0	50.5	12.1	23.5	<b>40.7</b>
TSTGT	0.2	28.3	50.0	11.6	22.5	39.9

**Table 6: Model analysis of different settings in MPNN layers of divide-and-conquer topology graph Transformer.**

Method	LL part	LT part	DET <sub>l</sub>	DET <sub>t</sub>	TOP <sub>ll</sub>	TOP <sub>lt</sub>	OLS
TSTGT	GINE	GINE	28.9	49.6	11.9	23.5	40.4
TSTGT	GatedGCN	GatedGCN	28.7	49.6	12.0	23.1	40.2
TSTGT	GatedGCN	GINE	28.6	50.3	11.7	24.0	40.6
TSTGT	GINE	GatedGCN	29.0	50.5	12.1	23.5	<b>40.7</b>

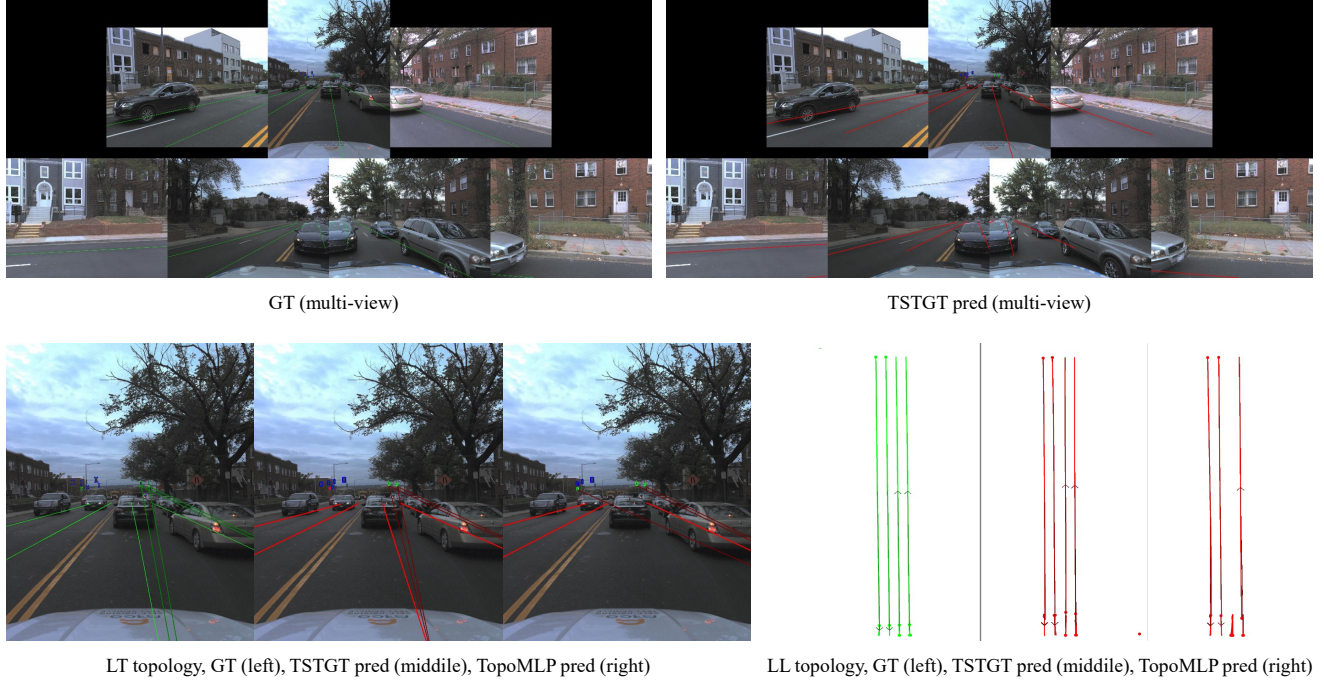
while GINE only updates node features. We explore various MPNN configurations for the lane-lane part and lane-traffic part of DCTGT, as elaborated in Table 6. From the experimental results, it is evident that employing GINE for the MPNN layer of lane-lane part of DCTGT and GatedGCN for the lane-traffic part yields the most optimal experimental outcomes. Our model adopts this configuration.

## C ADDITIONAL VISUALIZATION RESULTS

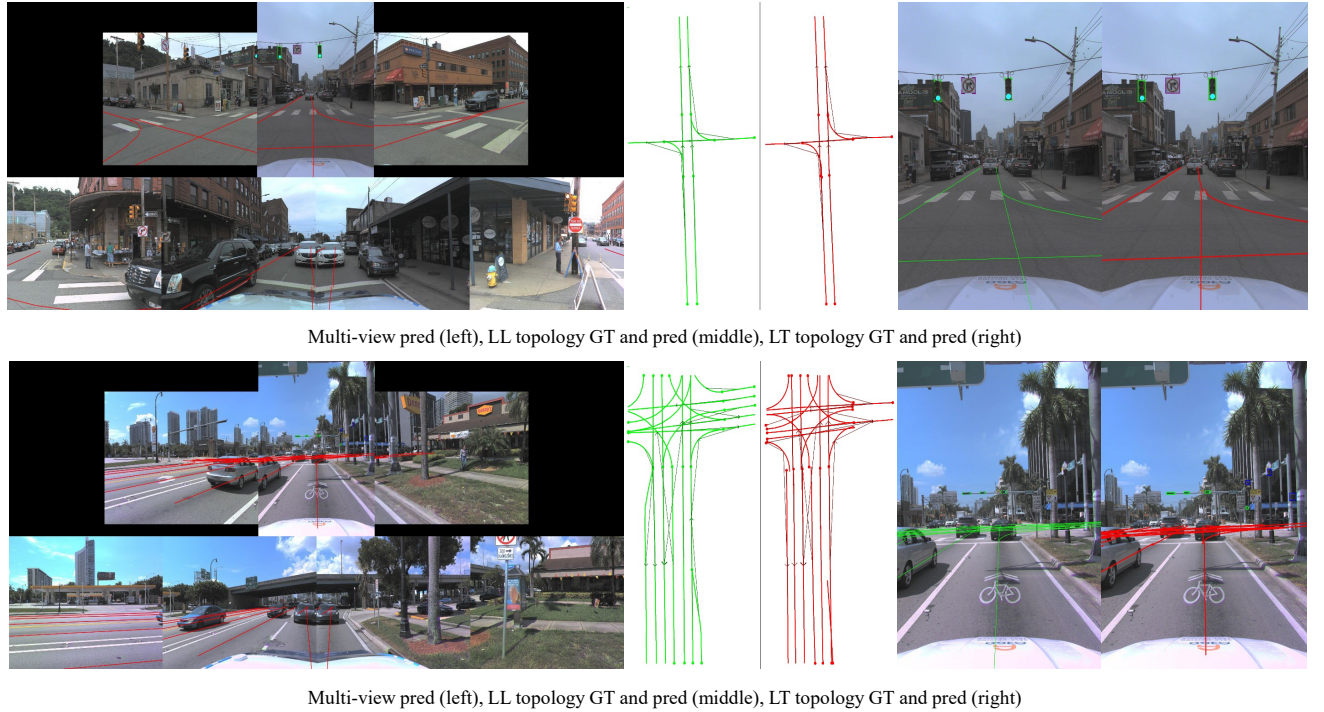
In this section, we present more additional visualization results on Openlane-V2 *subset\_A*, which include the visualization comparisons between our proposed TSTGT and the baseline TopoMLP, and the visualization of the detection and reasoning results of our TSTGT.

The visualization comparisons between TopoMLP and TSTGT are shown in Fig.1. From the detection and reasoning results in the figures, it can be observed that for lane detection and lane-lane topology reasoning, TopoMLP misses some lane and predicts incorrect lane-lane topology relationships. In contrast, our TSTGT yields highly accurate predictions for both lane and their topology relationships. Regarding the lane-traffic topology reasoning, TopoMLP erroneously associates the left traffic light with the lane on the right side of the image, contrary to our understanding that traffic elements only indicate the lanes within their control range. Alternatively, our TSTGT predicts the right lane-traffic topology. These results intuitively demonstrate the effectiveness of our traffic scene-assisted topology graph Transformer and point-wise matching strategy.

The visualization of the detection and reasoning results of our TSTGT is illustrated in Fig.2. It can be observed that our model can accurately detect lanes and traffic elements and infer their topology relationships, which intuitively present the high performance of our proposed TSTGT.



**Figure 1: Visualization comparisons between TopoMLP and TSTGT on Openlane-V2 subset\_A.**



**Figure 2: Visualization results of our proposed TSTGT on Openlane-V2 subset\_A.**