Visual Traffic Knowledge Graph Generation from Scene Images
Supplementary Material

Abstract

This supplementary material is organized as follows: 1) Section 1: detailed statistics, annotation rules, and the structure of the visual traffic knowledge graphs of RS10K; 2) Section 2: more visualization results of our proposed method.

1. Dataset

1.1. Statistics

The categories of each element, the number of each category, and the total number of elements are shown in Table 1. For relations, we report the number of the three types of relations, i.e., the S-S, C-C, and A-T relations, among which the C-C relation is composed of the association relation and the pointing relation defined by [1]. We list the categories of all components in Table 2. Notably, the arrow symbols (the first 21 symbols) and arrowheads (only 8 categories) are different. The arrow symbols include 8 simple arrows (the first 8 symbols) and compound arrows (9th to 21st symbols) that may own one or more arrowheads.

Annotations	Type	Category	Number	Total Number
Element	Road	Current	10066	21058
		Left	3827	
		Right	3842	
		Front	589	
		Crosswalk	2734	
	Lane	Current	9798	41891
		Non-Current	32093	
	Sign	Rectangle	18787	42923
		Circle	5592	
		Triangule	565	
		Blur	17979	
	Component	Text	46505	86027
		Symbol	24087	
		Arrowhead	15435	
Relation	S-S	-	3116	104428
	C-C	Association	46085	
		Pointing	15327	
	A-T	-	39900	

Table 1. Statistics of different annotations in RS10K. Categories and numbers of different elements and relations are displayed in the table. Be aware that the elements with †are not involved in VTKGG.

Group	Components
Text	Text
Symbol	Forward, Front right, Front left, Left rear, Right rear, To left, To right, Back, Turn left, Turn right, Turn left or go forward, Turn right or go forward, Turn around, Turn left or turn around, Go forward or turn around, Turn left or right, Other arrows, Crossroads, Interchange, Confluence, Circular curve, Roundabout, White dotted line, White solid line, Yellow dotted line, Yellow solid line, Crosswalk, Parking, Overpass, Underground passage, Camera, Information, Passing-bay, Van, Bus, Bicycle, Motorcycle, Vehicle, Car, Airport, Train station, Pedestrian, Disabled, Weighing station, Other vehicles, Dangerous chemicals vehicle, Gas station, Restaurant, Rest area, Serving station, Freeway starting point, Freeway terminal, Toilet, Shopping, Boiled water, Charge, Accommodation, Circular sign, Triangular sign
Arrowhead	Forward, Front right, Front left, Left rear, Right rear, To left, To right, Back

Table 2. Component categories defined in RS10K.

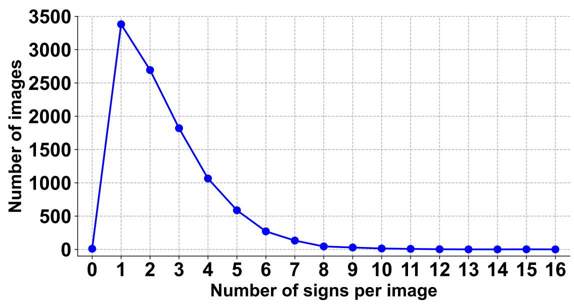


Figure 1. Quantity distribution of signs within an image.

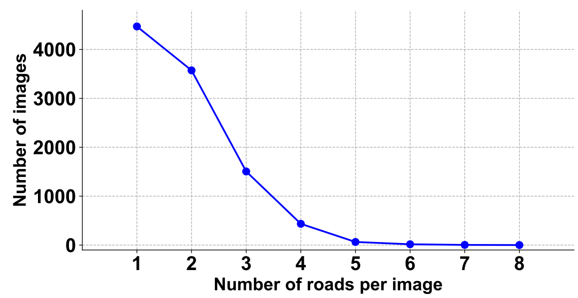


Figure 2. Quantity distribution of roads within an image.

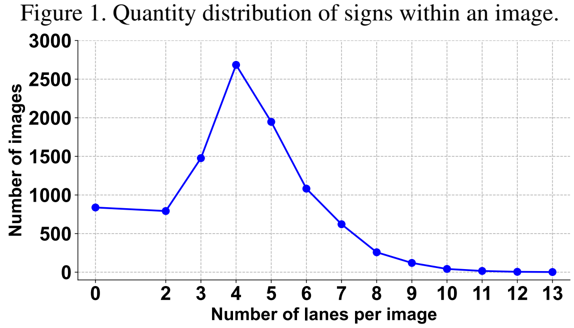


Figure 3. Quantity distribution of lanes within an image.

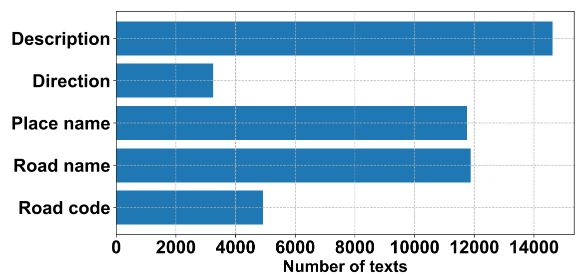


Figure 4. Quantity distribution of text attributes.

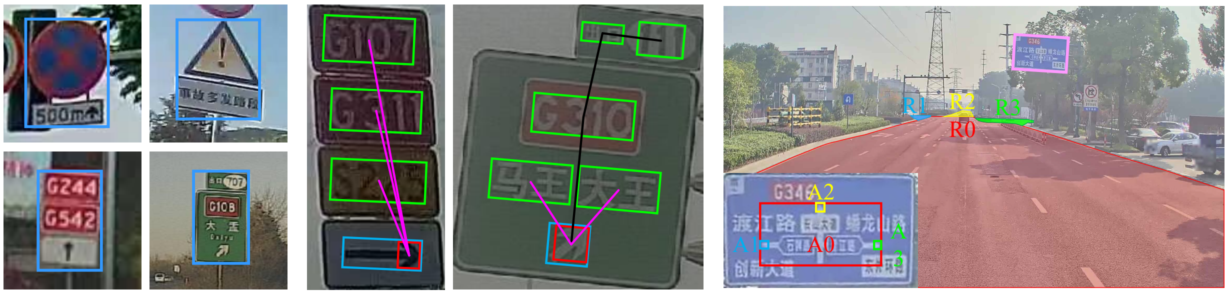
The quantity distribution of signs, roads, lanes, and text attributes are shown in Figure 1-4. Because the images without traffic signs are of little significance to us, most of the scene images we collect contain traffic signs. The distribution of roads and lanes conforms to the real situation. All texts are annotated with an attribute, and their distribution is shown in Figure 4.

1.2. Relation Annotations

Here we explain the details of S-S, C-C, and A-T relations with some visual examples.

The S-S relation indicates that one traffic sign is a supplement to another. In some cases, if there are multiple signs associated with each other, they together deliver a piece of traffic information. As shown in Figure 5(a), we show four common situations of S-S relations.

The definition of C-C relations originates from [1], which defines two types of relations between components in traffic signs, namely, association relation and pointing relation. The association relation (black lines in Figure 5(b)) refers to the relation between a specific place name or road name and its attached texts or symbols. The pointing relation (pink lines in Figure 5(b)) means the relation between the arrowhead and the place name or road name in the arrowhead's direction,



(a) (b) (c)

Figure 5. Visualization of some relation examples. (a) the S-S relations, which are demonstrated by all signs inside the blue boxes of an image. (b) C-C relations that span multiple traffic signs, where the texts are in green boxes, symbols are in blue boxes, arrowheads are in red boxes, the black lines highlight the association relation, and the pink lines denote pointing relation. (c) a typical example to illustrate the A-T relation. The panel inside the pink box is adopted for representation, where the arrow elements and their traffic roads are in the same color. The correspondence can be represented by (A0, R0), (A1, R1), (A2, R2), and (A3, R3).

including their accessory components. Figure 5(b) shows the C-C relation across two or more signs, which connects all components on multiple signs and makes it possible to parse these signs completely.

For the A-T relation, a typical case is shown in Figure 5(c), the traffic panel marked with the pink box is taken as an example. It contains a crossroad symbol A0, which includes three arrowheads in different directions: A1, A2, and A3. Therefore, this symbol corresponds to the current road $R0$. Its three arrowheads correspond to the three branches of the crossroad. In this way, we connect the contents of traffic signs and traffic elements to build visual traffic knowledge graphs. Furthermore, due to such a strong correlation, we can replace these arrow symbols or arrowheads with corresponding roads or lanes. In this case, the relations between other components and the traffic elements (C-T) are equivalent to the relations between other components and the arrow elements, that is why **C-T relations can be inferred from C-C and A-T relations**.

1.3. Visual Traffic Knowledge Graph

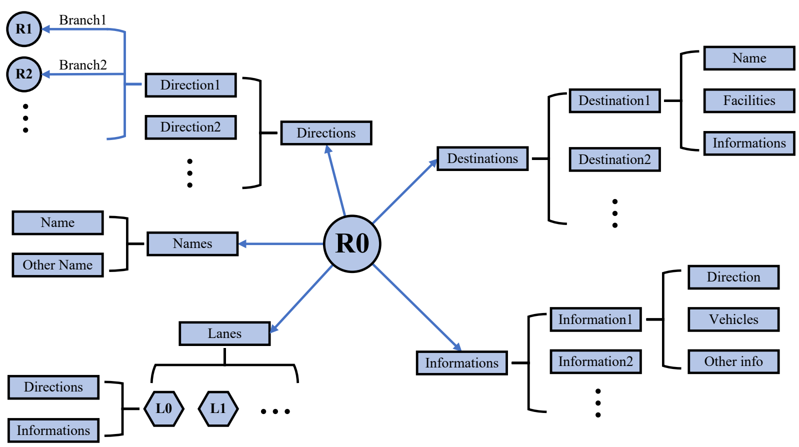


Figure 6. The structure of a complete knowledge tree in the visual traffic knowledge graph.

All visual traffic knowledge graphs are composed of one or more knowledge trees, whose root nodes are real roads or virtual roads. The virtual road is represented by an arrow element whose traffic element is invisible. Figure 6 shows the structure of a complete knowledge tree. Figure 7 display several examples. In the first image in Figure 7, since the roads that the two traffic elements S0-A1 and S0-A2 refer to are invisible in the image, they are both represented by the two arrowheads. Such a visual traffic knowledge graph can integrate almost all traffic information, and has good scalability.

2. Experiments

More visualization results, involving visualization of S-S, C-C, and A-T relations, are shown in Figure 7, which demonstrate the effectiveness of our method to generate visual traffic knowledge graphs and its applicability to various traffic scenes.

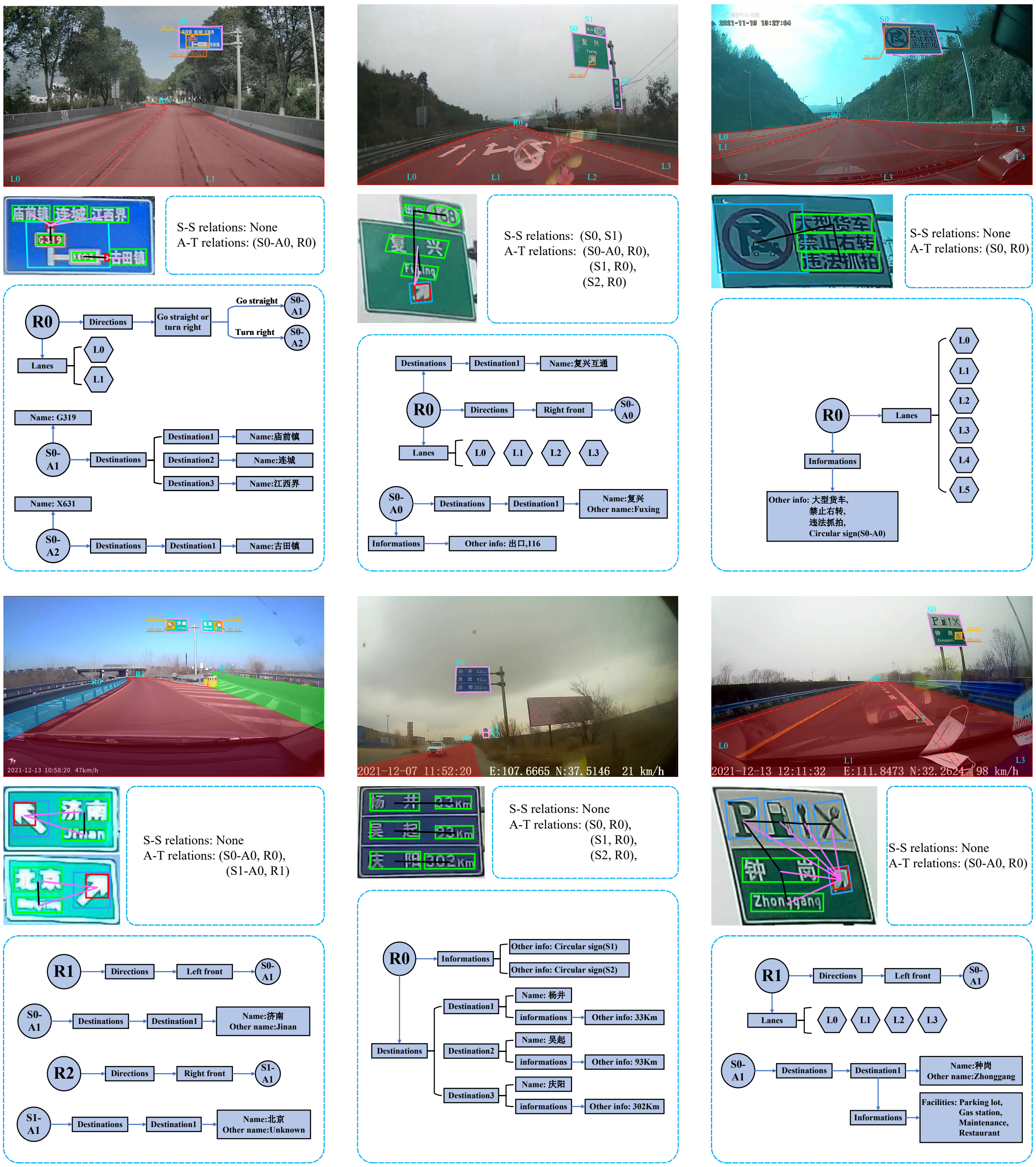


Figure 7. The scene images, visual relation results, and the generated visual traffic knowledge graphs. In the scene images, the current road is in red, the left road is in blue, and the right road is in green. The lane dividing lines are highlighted by red dashed lines inside the road areas. Under the images are the C-C relations displayed on the panels and the S-S and A-T relations on the right. Next are the visual traffic knowledge graphs, in which the roads or virtual roads are denoted by circles, and lanes are hexagons. "Unknown" in the graph represents a fuzzy text that cannot be recognized.

References

[1] Yunfei Guo, Wei Feng, Fei Yin, Tao Xue, Shuqi Mei, and Cheng-Lin Liu. Learning to understand traffic signs. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pages 2076–2084, 2021.