

Fig.R 1: Cross-Attention map of global relation query, subject query, and object query in Decoder.

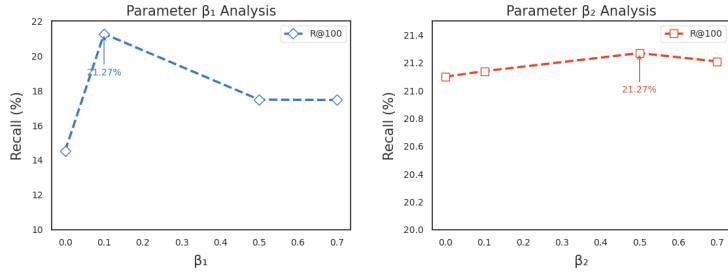


Fig.R 2: Ablation study of  $\beta$  in loss function.

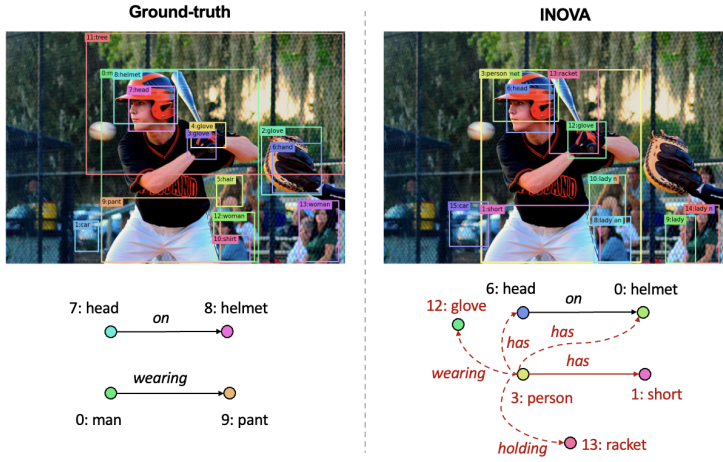


Fig.R 3: Analysis of bad cases.



Fig.R 4: Mismatched relation triplets examples.

Table.R 1: Experimental results of OvR-SGG setting on PSG test set.

Method	Joint Base+Novel			Novel (Rel)		
	R@20	R@50	R@100	R@20	R@50	R@100
SGTR CVPR'22	-	14.2	18.2	-	-	-
PGSG CVPR'24	-	18.0	20.2	-	-	-
OvSGTR ECCV'24	15.14	17.76	19.50	5.32	6.93	8.08
<b>INOVA (Ours)</b>	<b>16.69</b>	<b>20.01</b>	<b>21.71</b>	<b>6.78</b>	<b>8.78</b>	<b>9.70</b>

Table.R 2: Experimental results of OvR-SGG setting on the VG test set. \* and \* denote pretrained with MegaSG data and VG caption data.

Method	Joint Base+Novel			Novel (Rel)		
	R@20	R@50	R@100	R@20	R@50	R@100
OvMotifs MMM'25	-	25.77	30.57	-	8.74	22.89
OvSGTR* ECCV'24	21.09	27.92	32.74	16.59	22.86	27.73
OvSGTR* ECCV'24	20.96	28.19	32.98	15.30	23.39	28.97
<b>INOVA* (Ours)</b>	<b>22.00</b>	<b>29.22</b>	<b>33.77</b>	<b>26.90</b>	<b>34.64</b>	<b>39.68</b>

Table.R 3: Experimental results of Fully-supervised Closed-World setting on VG test set.

Method	R@20	R@50	R@100	R@20	mR@50	mR@100
SGTR CVPR'22	-	24.6	28.4	-	-	-
VS CVPR'23	27.3	36.0	40.9	4.4	6.5	7.8
OvSGTR ECCV'24	27.0	35.8	41.3	5.0	7.2	8.8
RAHP AAAI'25	-	34.25	40.40	-	7.21	10.45
OvMotifs MMM'25	-	30.9	36.9	-	7.0	9.0
<b>INOVA (Ours)</b>	<b>27.63</b>	<b>36.40</b>	<b>42.01</b>	<b>5.31</b>	<b>7.51</b>	<b>9.12</b>

Table.R 4: Experimental results of Weakly-supervised setting on VG test set.

Method	Supervision	R@20	R@50	R@100
LSWS CVPR'21	COCO Caption	-	3.85	4.04
SGNLS ICCV'21		-	3.80	4.46
Li et al MM'22		-	6.40	7.33
VS CVPR'23		6.04	8.15	9.90
OvSGTR ECCV'24		6.88	9.30	11.48
LLM4SGG CVPR'24		-	8.91	10.43
<b>INOVA (Ours)</b>		-	11.61	14.33
VS CVPR'23	VG Caption	10.98	15.51	19.75
OvSGTR ECCV'24		16.36	22.14	26.20
LLM4SGG CVPR'24		-	18.40	22.28
<b>INOVA (Ours)</b>		<b>18.93</b>	<b>24.70</b>	<b>28.49</b>

Table.R 5: Experimental results of OvR-SGG setting on VG test set trained with VG caption. † denotes based on the VS framework

Method	Joint Base+Novel			Novel (Rel)		
	R@20	R@50	R@100	R@20	R@50	R@100
VS CVPR'23	-	7.61	9.60	-	4.06	5.58
<b>INOVA† (Ours)</b>	<b>5.53</b>	<b>8.95</b>	<b>12.28</b>	<b>3.23</b>	<b>6.15</b>	<b>9.03</b>

Table.R 6: Comparison of Large Model utilization under OvR-SGG setting on VG test set. ‡ denotes counter-action generation with **Pattern**.

Method	Large Model	Joint Base+Novel		
		R@20	R@50	R@100
VS CVPR'23	GLIP	-	15.50	17.37
OvSGTR ECCV'24	Grounding DINO	-	20.46	23.86
RAHP AAAI'25	GPT-3.5-turbo, Grounding DINO	-	20.50	25.74
<b>INOVA (Ours)</b>	Llama2, Grounding DINO	<b>17.49</b>	<b>23.22</b>	<b>27.40</b>
<b>INOVA‡ (Ours)</b>	Grounding DINO	<b>17.36</b>	<b>22.98</b>	<b>27.14</b>

Table.R 7: Ablation study on the large model size under OvD+R-SGG setting on VG test set.

Method	Large Model	Size	Joint Base+Novel		
			R@20	R@50	R@100
<b>INOVA (Ours)</b>	Llama2	7B	<b>13.50</b>	<b>18.88</b>	<b>23.19</b>
<b>INOVA (Ours)</b>	Qwen2.5	0.5B	<b>13.64</b>	<b>18.99</b>	<b>23.43</b>
<b>INOVA‡ (Ours)</b>	Pattern (Python Lib)	-	<b>13.36</b>	<b>18.56</b>	<b>22.64</b>
OvSGTR ECCV'24	Grounding DINO-T	174M	10.02	13.50	16.37
<b>INOVA (Ours)</b>	Grounding DINO-T	174M	<b>12.61</b>	<b>17.43</b>	<b>21.27</b>
OvSGTR ECCV'24	Grounding DINO-B	224M	12.37	17.14	21.03
<b>INOVA (Ours)</b>	Grounding DINO-B	224M	<b>13.50</b>	<b>18.88</b>	<b>23.19</b>

Table.R 8: Inference costs under OvD+R-SGG setting on VG test set.

Method	Inference Costs ( s / I )	Joint Base+Novel		
		R@20	R@50	R@100
OvSGTR ECCV'24	2.2231161964684725	10.02	13.50	16.37
<b>INOVA (Ours)</b>	<b>2.2574067325145006</b>	<b>13.34</b>	<b>18.76</b>	<b>23.01</b>