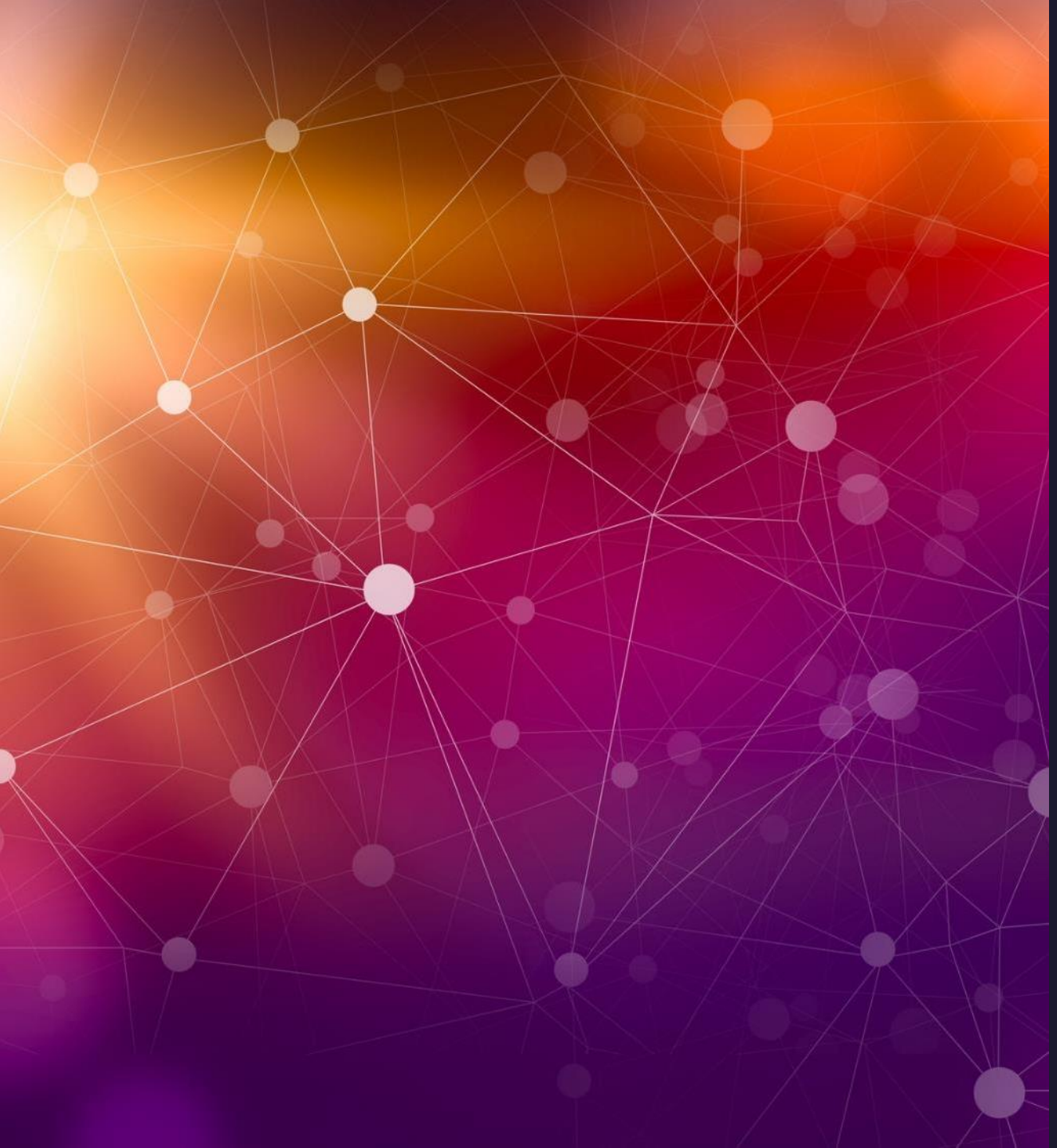


Hydra-SGG: Hybrid Relation Assignment for One-Stage Scene Graph Generation

Authors: Minghan Chen, Guikun Chen, Wenguan Wang, Yi Yang

Agenda

Motivation & Problem

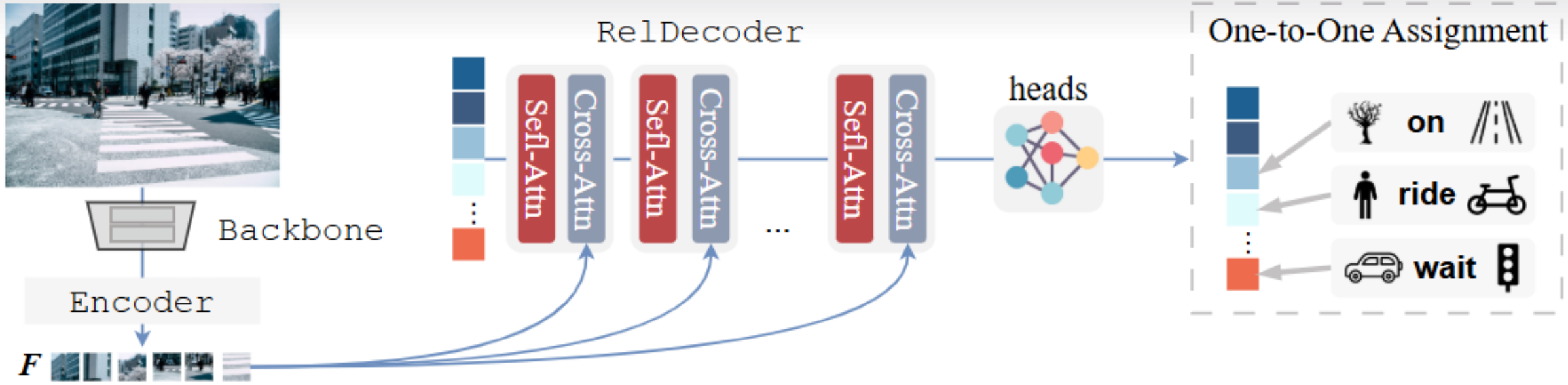
Method

Results

Motivation & Problem

Sparse supervision: Each image only has around 5 relations, but the model uses 300 relation queries

False negatives: Some predictions are close to the ground truth but are wrongly treated as negatives, hurting training.



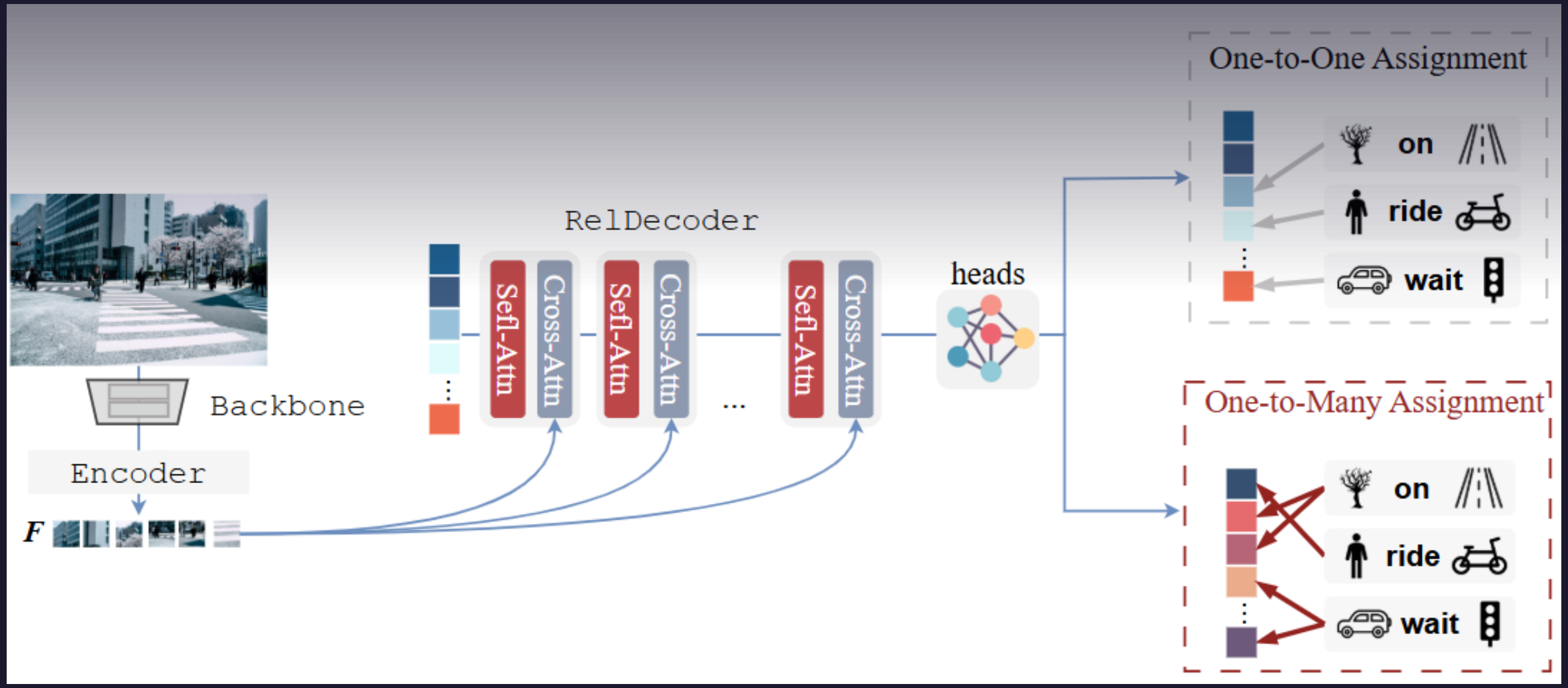
Method

To solve this, we propose **Hydra-SGG**, a one-stage SGG model with two main ideas:

- **Hybrid Relation Assignment**: Combines traditional One-to-One matching with a new **IoU-based One-to-Many assignment** to exploit more positive queries.
- **Hydra Branch**: An auxiliary decoder without self-attention that promotes duplicate predictions—which actually helps the One-to-Many training.

Vanilla Hydra-SGG

Our hybrid strategy gives 60% more positive samples per step.



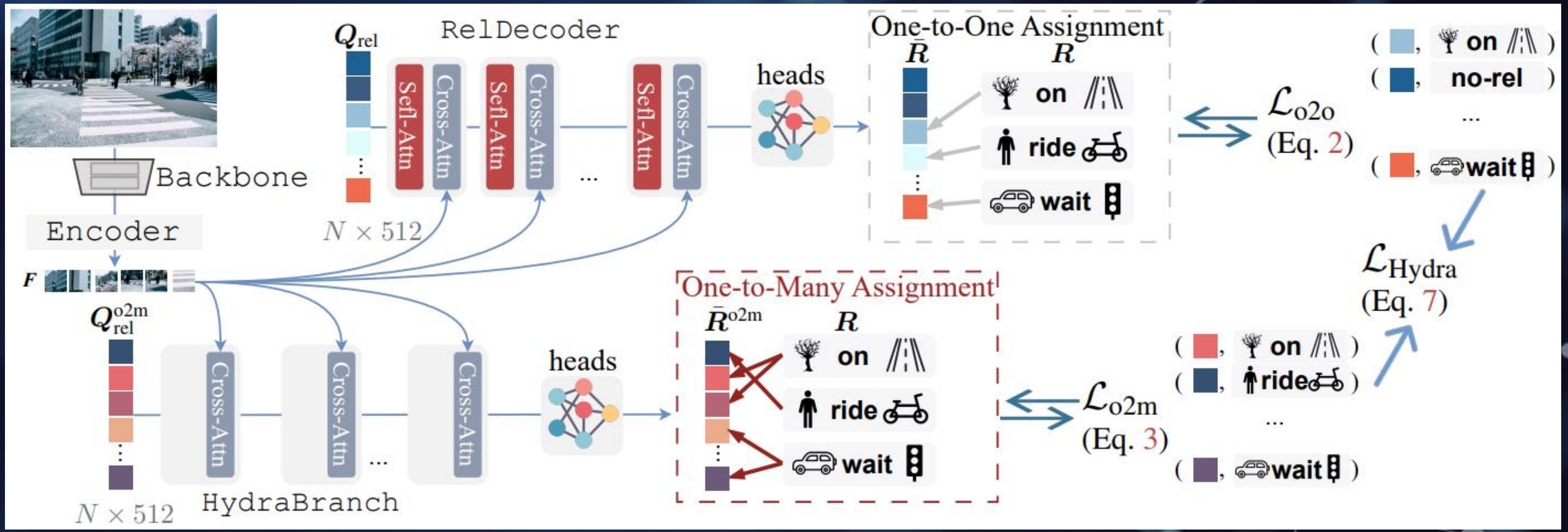
Insights

Previous studies indicate that self-attention helps inhibit duplicate predictions

Query ID	*w* Self-attn		*w/o* Self-attn
41	on		above
61	with	Converge ⟼	above
191	near		above
209	on		above

Query ID	*w* Self-attn		*w/o* Self-attn
42	of		has
52	wear	Converge ⟼	has
67	has		has
126	on		has

Hydra-SGG


Backbone
Encoder
F
Q_{rel}
$N \times 512$
RelDecoder
Sefl-Attn
Cross-Attn
heads
One-to-One Assignment
$\bar{R}$
R
on
ride
wait
$\mathcal{L}_{o2o}$
(Eq. 2)
no-rel
$\mathcal{L}_{Hydra}$
(Eq. 7)
Q^{o2m}_{rel}
HydraBranch
One-to-Many Assignment
$\bar{R}^{o2m}$
$\mathcal{L}_{o2m}$
(Eq. 3)

Main Results

Table 1: SGDet evaluation on VG150 [27] `test` (§4.2). $^{+}$: detector pre-trained on VG150. FPS (Frames Per Second) indicates inference speed. F-Recall of Hydra-SGG is calculated based on the best results.

Method	Backbone	# Epoch	FPS	# Param	R@20/50/100	mR@20/50/100	F@20/50/100
Two-stage methods							
MOTIFS [85] [CVPR2018]	ResNeXt101-FPN	-	-	369.9M	25.1 / 32.1 / 36.9	4.1 / 5.5 / 6.8	7.1 / 9.2 / 11.7
VCTree-TDE [65] [CVPR2020]	ResNeXt101-FPN	-	-	361.3M	14.3 / 19.6 / -	6.3 / 9.3 / 11.1	8.8 / 12.4 / -
BGNN [39] [CVPR2021]	ResNeXt101-FPN	-	-	341.9M	23.3 / 31.0 / 35.8	7.5 / 10.7 / 12.7	11.3 / 15.5 / 19.0
PE-Net [91] [CVPR2023]	ResNeXt101-FPN	32^{+}	-	-	- / 30.7 / 35.2	- / 12.4 / 14.5	- / 17.7 / 21.2
IS-GGT [29] [CVPR2023]	ResNet101	70	-	-	- / - / -	- / 9.1 / 11.3	- / - / -
VETO [61] [ICCV2023]	ResNeXt101-FPN	33^{+}	-	-	- / 27.5 / 31.5	- / 8.1 / 9.5	- / 12.5 / 14.6
UniVRD [90] [ICCV2023]	CLIP ViT-B	-	-	-	- / - / -	- / 9.6 / 12.1	- / - / -
DRM [32] [CVPR2024]	ResNeXt101-FPN	-	-	-	- / 34.0 / 38.9	- / 9.0 / 11.2	- / 14.2 / 17.4
One-stage methods							
SGTR [40] [CVPR2022]	ResNet101	123^{+}	-	117.1M	- / 25.1 / 26.6	- / 12.0 / 14.6	- / 16.2 / 18.9
SSR-CNN [67] [CVPR2022]	ResNet101	-	-	274.3M	25.8 / 32.7 / 36.9	6.1 / 8.4 / 10.0	9.9 / 13.4 / 15.7
ISG [24] [NeurIPS2022]	ResNet101	52	-	93.5M	- / 29.5 / 32.1	- / 7.4 / 8.4	- / 11.8 / 13.3
RelTR [7] [TPAMI2023]	ResNet50	150	6.5	63.7M	21.2 / 27.5 / 30.7	6.8 / 10.8 / 12.3	10.3 / 15.5 / 17.6
DSGG [13] [CVPR2024]	-	60	-	-	- / 32.9 / 38.5	- / 13.0 / 17.3	- / 18.6 / 23.9
SpeaQ [26] [CVPR2024]	ResNet101	52	-	-	- / 32.9 / 36.0	- / 11.8 / 14.1	- / 17.4 / 20.3
EGTR [18] [CVPR2024]	ResNet50	275^{+}	7.7	42.5M	23.5 / 30.2 / 34.3	5.5 / 7.9 / 10.1	8.9 / 12.5 / 15.6
Ours							
Hydra-SGG [ICLR2025]	ResNet50	12	5.3	67.6M	21.9 / 28.6 / 33.4 ±0.1 / ±0.2 / ±0.3	**10.3 / 15.9 / 19.4** ±0.2 / ±0.2 / ±0.2	**14.0 / 20.5 / 24.7**

Table 2: Evaluation on Open Images V6 [30] `test` (§4.2). $^{+}$: detector pre-trained on Open Images V6.

Method	Backbone	# Epoch	# Param	R@50	wmAP_{rel}	wmAP_{phr}	score_{wtd}
Two-stage methods							
Motifts [85] [CVPR2018]	ResNeXt101-FPN	-	369.9M	71.6	29.9	31.6	38.9
BGNN [39] [CVPR2021]	ResNeXt101-FPN	-	341.9M	75.0	35.5	34.2	42.1
PE-Net [91] [CVPR2023]	ResNeXt101-FPN	-	-	76.5	35.4	34.9	44.9
One-stage methods							
SGTR [40] [CVPR2022]	ResNet101	123^{+}	117.1M	59.9	37.0	38.7	42.3
RelTR [7] [TPAMI2023]	ResNet50	150	63.7M	71.7	34.2	37.5	43.0
EGTR [18] [CVPR2024]	ResNet50	275^{+}	42.5M	75.0	42.0	41.9	48.6
Ours							
Hydra-SGG [ICLR2025]	ResNet50	7	67.6M	$76.0_{\pm 0.2}$	$\mathbf{42.8}_{\pm 0.2}$	$\mathbf{44.1}_{\pm 0.2}$	$\mathbf{50.0}_{\pm 0.2}$

Thank you

Minghan Chen

Minghan.Chen@student.uts.edu.au