dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

ICCV OCT 19-23, 2025 HONOLULU HAWAII

RPTU Rheinland-Pfälzische Technische Universität
Kaiserslautern
Landau

STEP-DETR: Advancing DETR-based Semi-Supervised Object Detection with Super Teacher and Pseudo-Label Guided Text Queries

Tahira Shehzadi

Khurram Azeem Hashmi

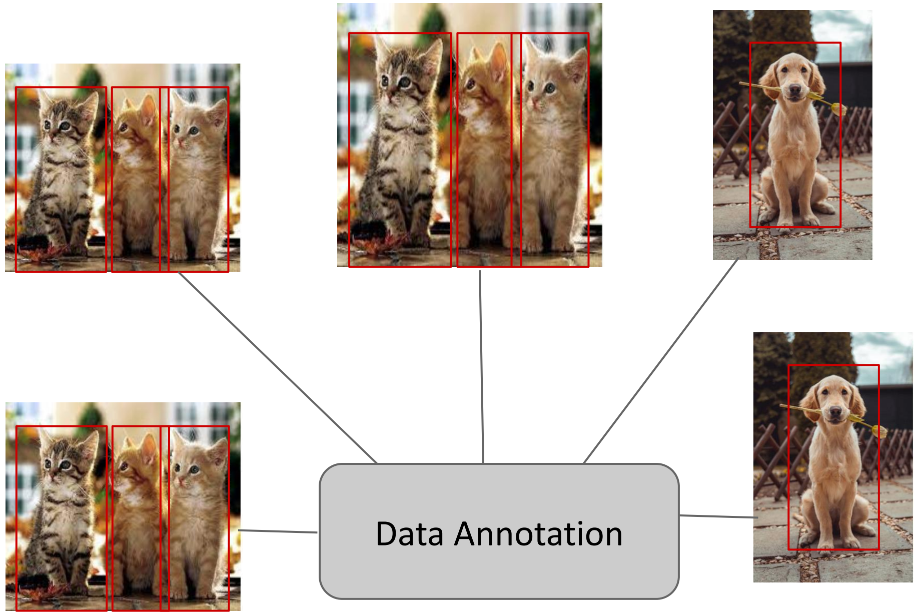
Shalini Sarode

Didier Stricker

Muhammad Zeshan Afzal

Problem Statement

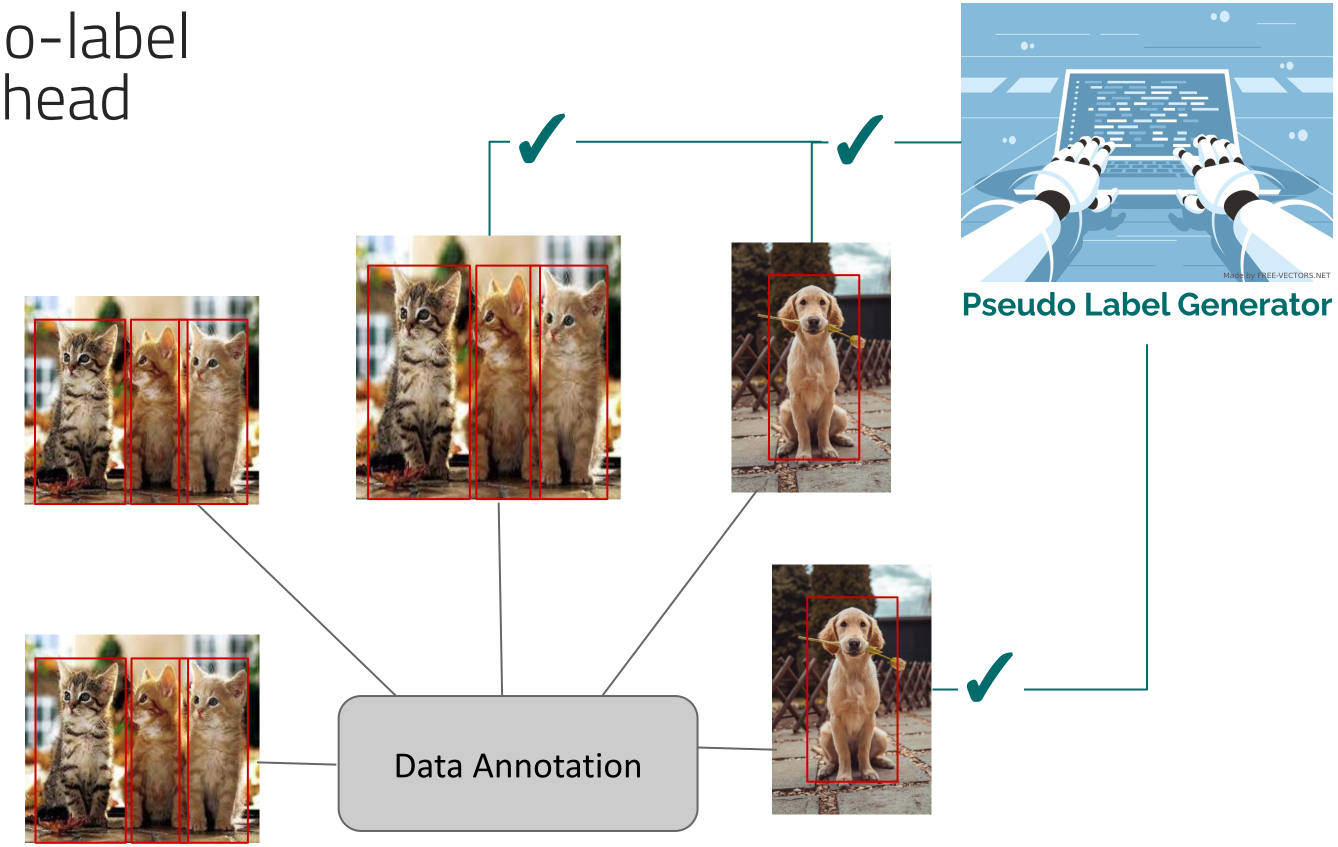
- Supervised table detection requires a lot of labeled data but annotation is expensive



Problem Statement

- Supervised table detection requires a lot of labeled data but annotation is expensive

- Semi-supervised methods use pseudo-label generation to reduce annotation overhead

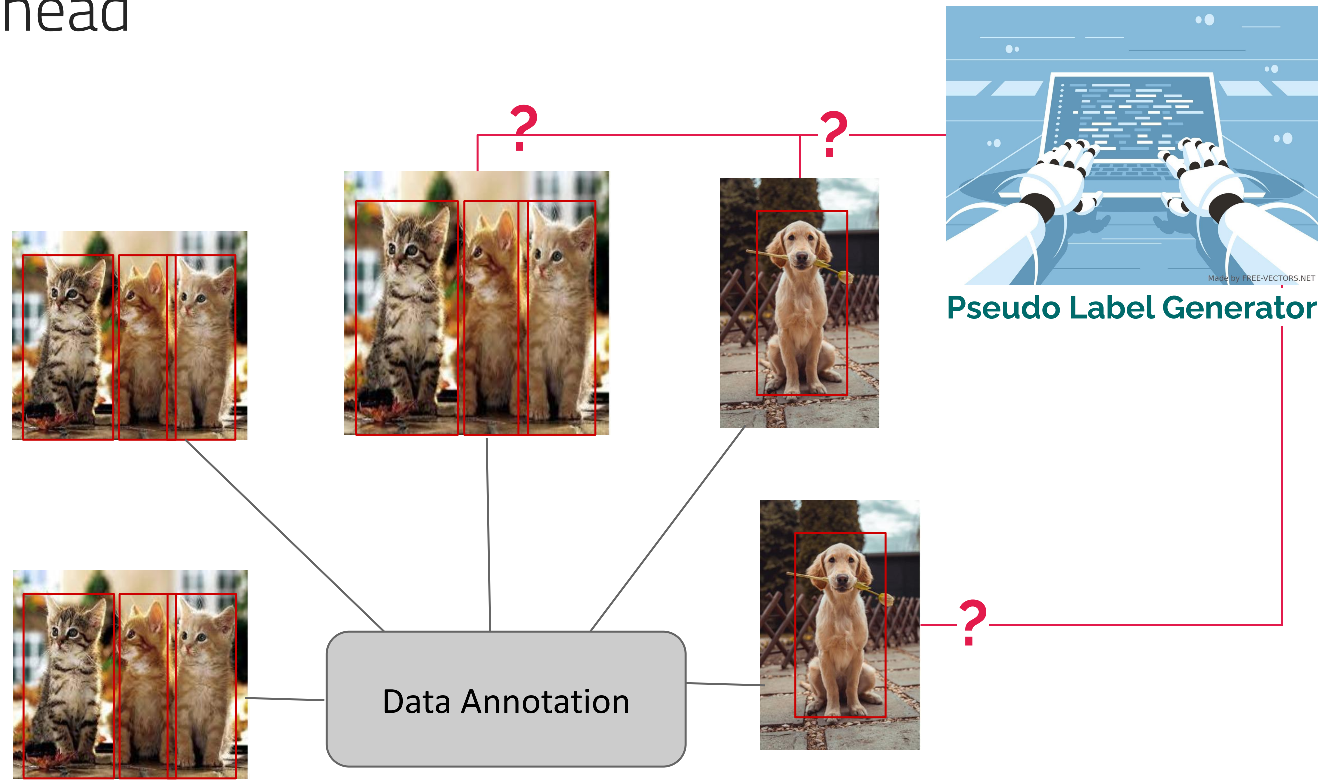


Problem Statement

- Supervised table detection requires a lot of labeled data but annotation is expensive

- Semi-supervised methods use pseudo-label generation to reduce annotation overhead

- But the quality of pseudo-labels is often suboptimal!

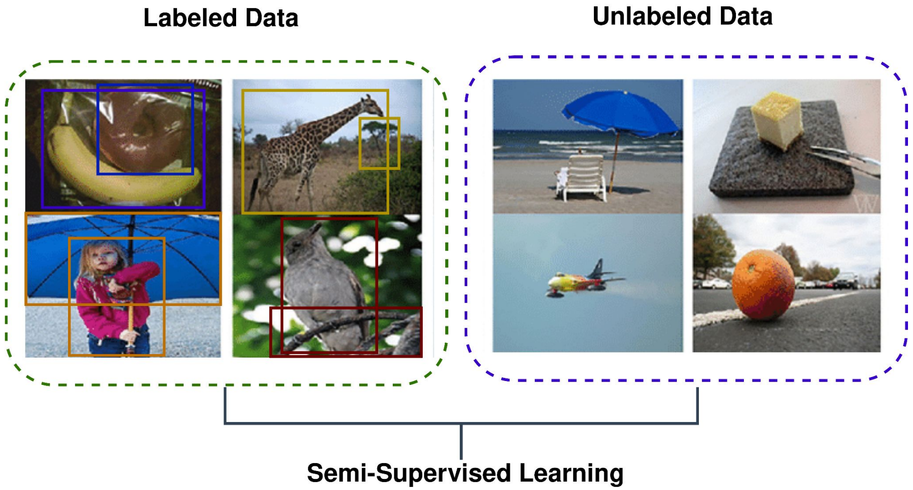


Data Annotation

Pseudo Label Generator

Semi-Supervised Object Detection (SSOD)

- Background
  - Problem Statement

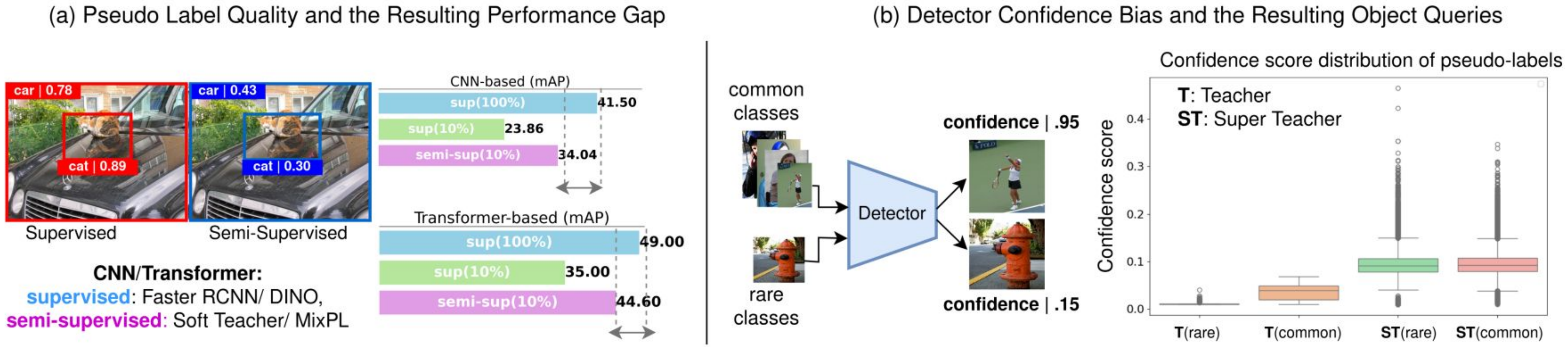


Settings:

- labeled data is limited: Taking 10% coco as labeled data, and the rest as unlabeled data.
- labeled data is abundant: Taking full coco (118k images) as labeled data, and unlabeled (123k images) as unlabeled data.

Semi-Supervised Object Detection (SSOD)

- Challenges in existing methods:
  - Noisy pseudo-labels
  - Confidence bias
  - Inefficient query generation for rare categories



(a) Pseudo Label Quality and the Resulting Performance Gap

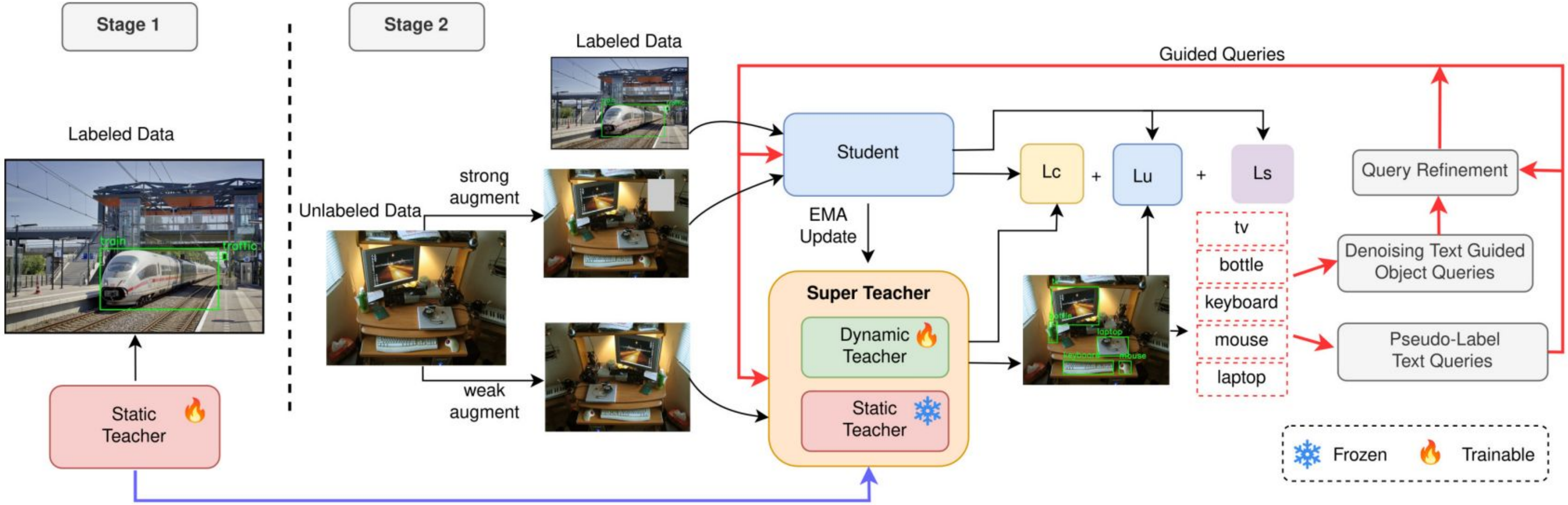
(b) Detector Confidence Bias and the Resulting Object Queries

STEP-DETR - Motivation

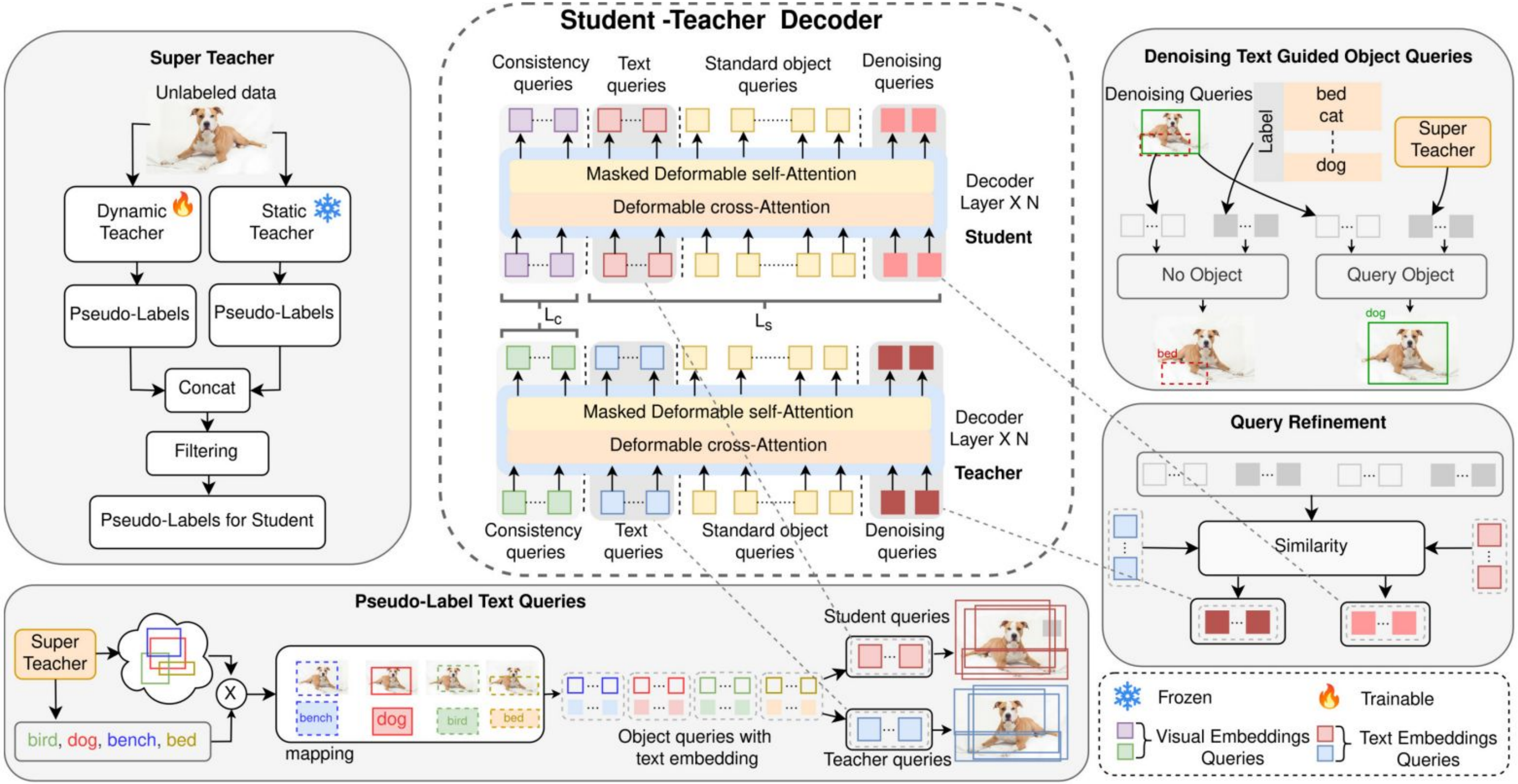
- Bipartite matching makes NMS-free but causes learning inefficiency.
- Need a framework that:
    - Generates high-quality pseudo-labels.
    - Balances confidence across common and rare categories.
    - Efficiently differentiates objects from background.

STEP-DETR Overview

- Super Teacher
- Pseudo-Label Text Queries
- Denoising Text Guided Object Queries
- Query Refinement Module

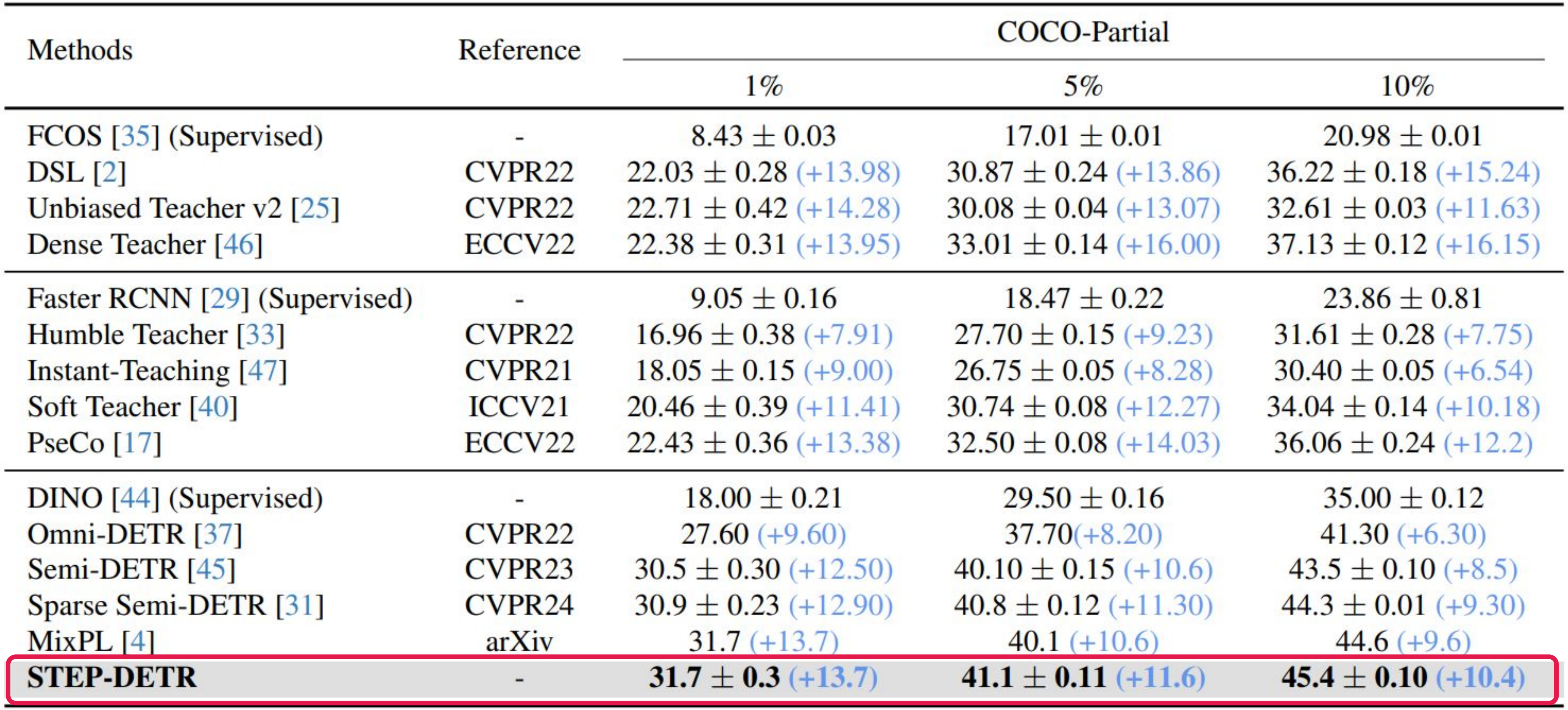


STEP-DETR Overview



Results

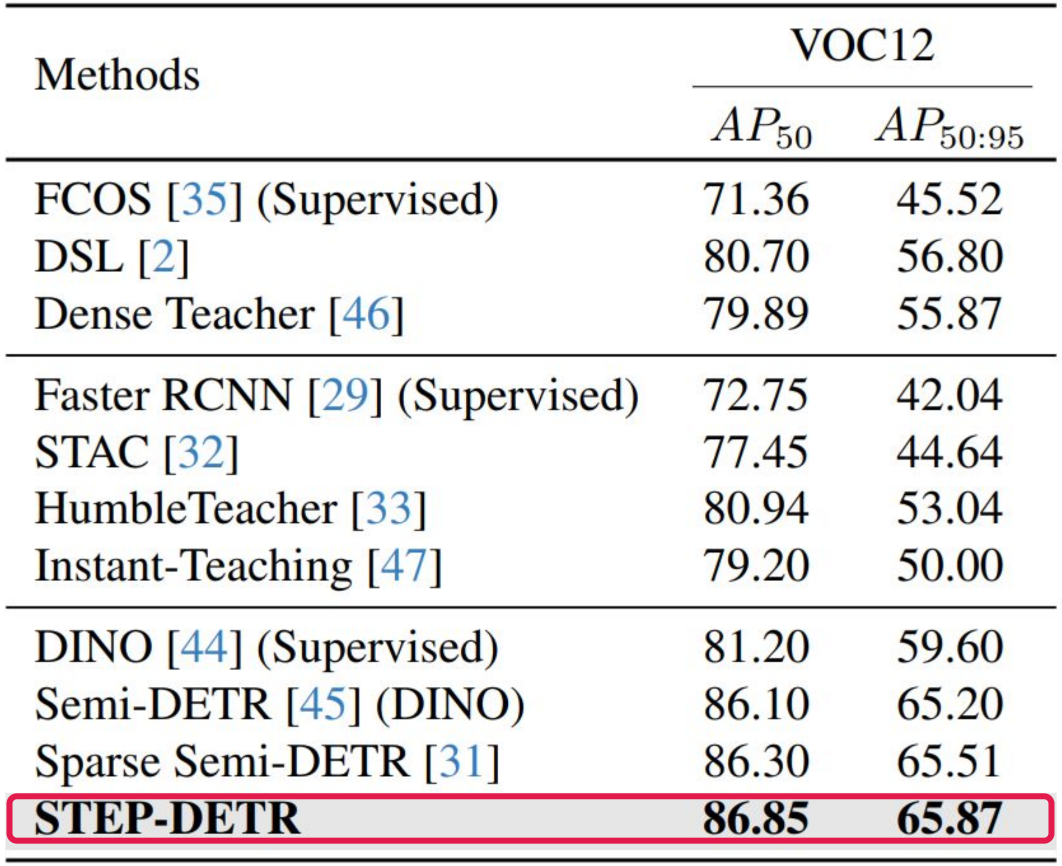
- We evaluate our approach on MS-COCO & Pascal VOC.
- Evaluation of STEP-DETR against existing approaches on the COCO-Partial setting.



Methods	Reference	COCO-Partial		
		1%	5%	10%
FCOS [35] (Supervised)	-	8.43 ± 0.03	17.01 ± 0.01	20.98 ± 0.01
DSL [2]	CVPR22	22.03 ± 0.28 (+13.98)	30.87 ± 0.24 (+13.86)	36.22 ± 0.18 (+15.24)
Unbiased Teacher v2 [25]	CVPR22	22.71 ± 0.42 (+14.28)	30.08 ± 0.04 (+13.07)	32.61 ± 0.03 (+11.63)
Dense Teacher [46]	ECCV22	22.38 ± 0.31 (+13.95)	33.01 ± 0.14 (+16.00)	37.13 ± 0.12 (+16.15)
Faster RCNN [29] (Supervised)	-	9.05 ± 0.16	18.47 ± 0.22	23.86 ± 0.81
Humble Teacher [33]	CVPR22	16.96 ± 0.38 (+7.91)	27.70 ± 0.15 (+9.23)	31.61 ± 0.28 (+7.75)
Instant-Teaching [47]	CVPR21	18.05 ± 0.15 (+9.00)	26.75 ± 0.05 (+8.28)	30.40 ± 0.05 (+6.54)
Soft Teacher [40]	ICCV21	20.46 ± 0.39 (+11.41)	30.74 ± 0.08 (+12.27)	34.04 ± 0.14 (+10.18)
PseCo [17]	ECCV22	22.43 ± 0.36 (+13.38)	32.50 ± 0.08 (+14.03)	36.06 ± 0.24 (+12.2)
DINO [44] (Supervised)	-	18.00 ± 0.21	29.50 ± 0.16	35.00 ± 0.12
Omni-DETR [37]	CVPR22	27.60 (+9.60)	37.70(+8.20)	41.30 (+6.30)
Semi-DETR [45]	CVPR23	30.5 ± 0.30 (+12.50)	40.10 ± 0.15 (+10.6)	43.5 ± 0.10 (+8.5)
Sparse Semi-DETR [31]	CVPR24	30.9 ± 0.23 (+12.90)	40.8 ± 0.12 (+11.30)	44.3 ± 0.01 (+9.30)
MixPL [4]	arXiv	31.7 (+13.7)	40.1 (+10.6)	44.6 (+9.6)
STEP-DETR	-	**31.7 ± 0.3** (+13.7)	**41.1 ± 0.11** (+11.6)	**45.4 ± 0.10** (+10.4)

Results

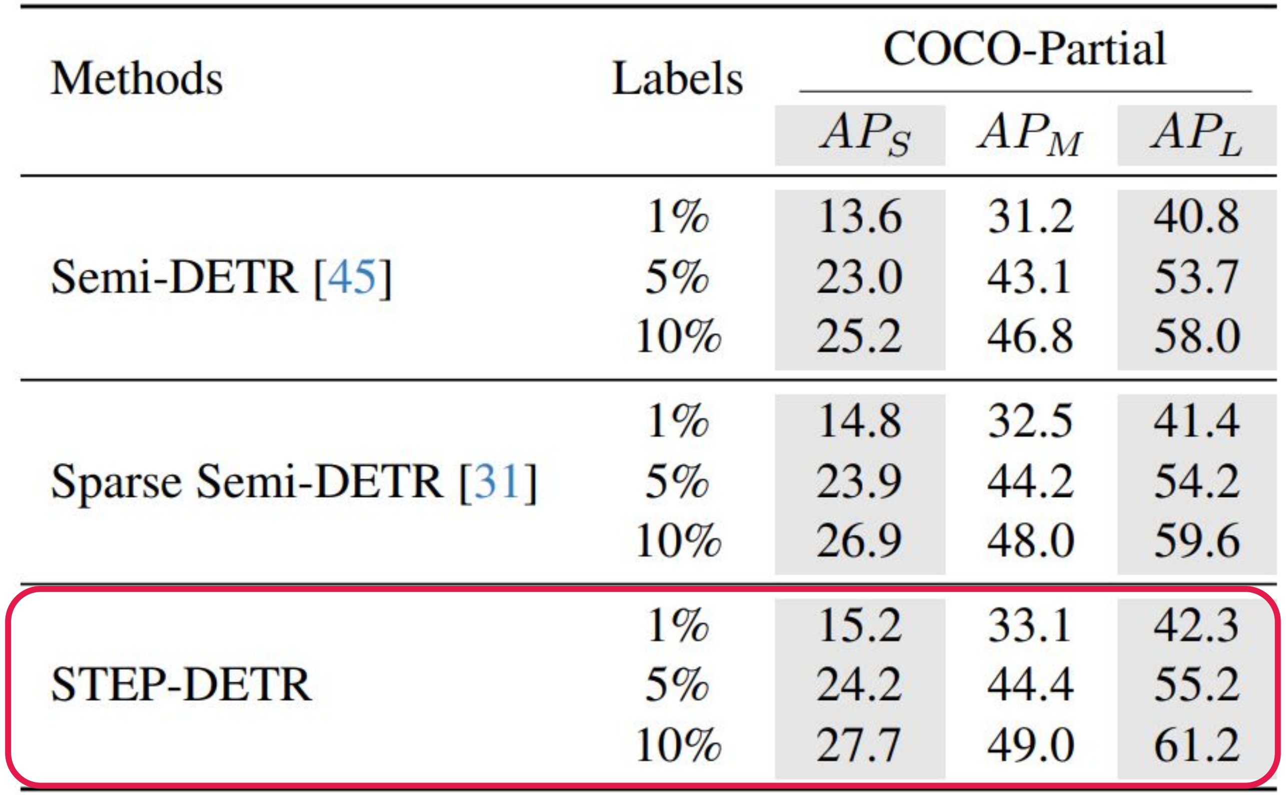
- Results on Pascal VOC



Methods	VOC12	
	AP_{50}	$AP_{50:95}$
FCOS [35] (Supervised)	71.36	45.52
DSL [2]	80.70	56.80
Dense Teacher [46]	79.89	55.87
Faster RCNN [29] (Supervised)	72.75	42.04
STAC [32]	77.45	44.64
HumbleTeacher [33]	80.94	53.04
Instant-Teaching [47]	79.20	50.00
DINO [44] (Supervised)	81.20	59.60
Semi-DETR [45] (DINO)	86.10	65.20
Sparse Semi-DETR [31]	86.30	65.51
STEP-DETR	**86.85**	**65.87**

Results

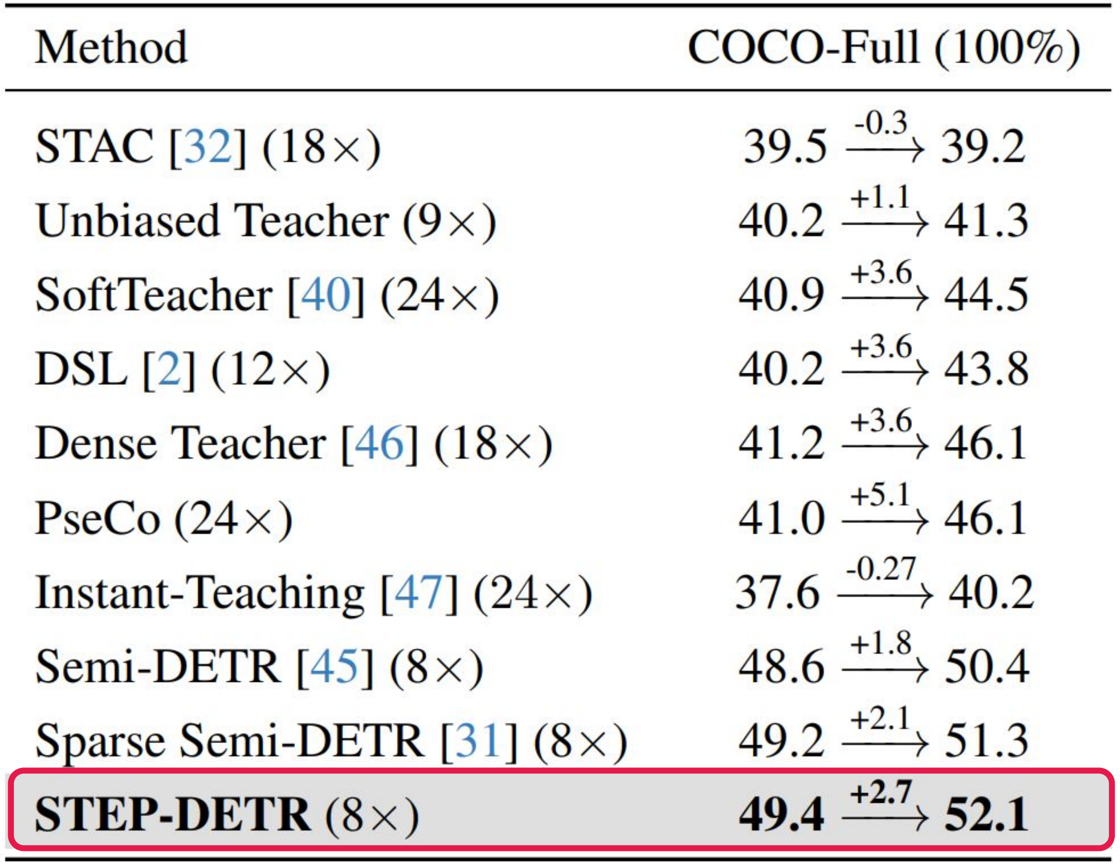
- Results on the COCO-partial setting for objects of different sizes.



Methods	Labels	COCO-Partial		
		AP_S	AP_M	AP_L
Semi-DETR [45]	1%	13.6	31.2	40.8
	5%	23.0	43.1	53.7
	10%	25.2	46.8	58.0
Sparse Semi-DETR [31]	1%	14.8	32.5	41.4
	5%	23.9	44.2	54.2
	10%	26.9	48.0	59.6
STEP-DETR	1%	15.2	33.1	42.3
	5%	24.2	44.4	55.2
	10%	27.7	49.0	61.2

Results

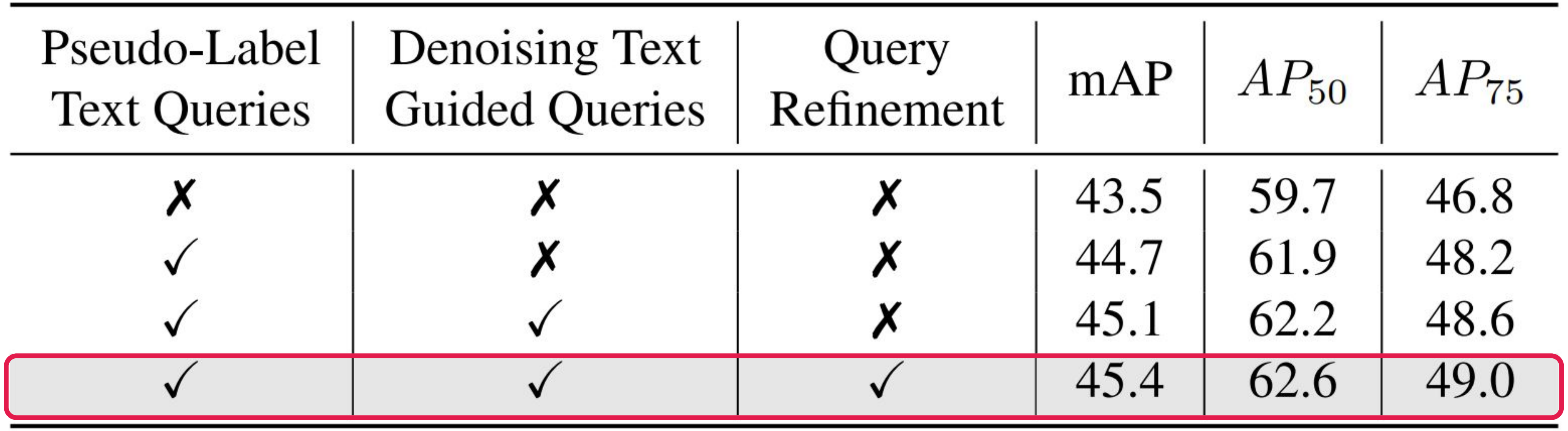
- Performance comparison on COCO-Full.



Method	COCO-Full (100%)
STAC [32] (18×)	$39.5 \xrightarrow{-0.3} 39.2$
Unbiased Teacher (9×)	$40.2 \xrightarrow{+1.1} 41.3$
SoftTeacher [40] (24×)	$40.9 \xrightarrow{+3.6} 44.5$
DSL [2] (12×)	$40.2 \xrightarrow{+3.6} 43.8$
Dense Teacher [46] (18×)	$41.2 \xrightarrow{+3.6} 46.1$
PseCo (24×)	$41.0 \xrightarrow{+5.1} 46.1$
Instant-Teaching [47] (24×)	$37.6 \xrightarrow{-0.27} 40.2$
Semi-DETR [45] (8×)	$48.6 \xrightarrow{+1.8} 50.4$
Sparse Semi-DETR [31] (8×)	$49.2 \xrightarrow{+2.1} 51.3$
STEP-DETR (8×)	$\mathbf{49.4} \xrightarrow{\mathbf{+2.7}} \mathbf{52.1}$

Results

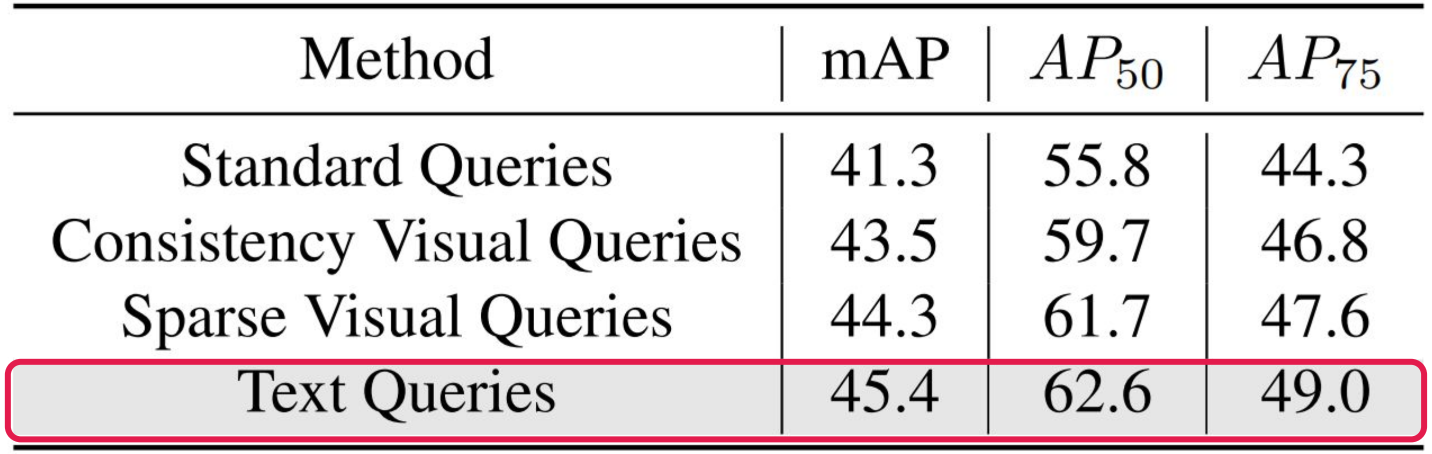
- Effect of Individual Module



Pseudo-Label Text Queries	Denoising Text Guided Queries	Query Refinement	mAP	AP_{50}	AP_{75}
✗	✗	✗	43.5	59.7	46.8
✓	✗	✗	44.7	61.9	48.2
✓	✓	✗	45.1	62.2	48.6
✓	✓	✓	45.4	62.6	49.0

Results

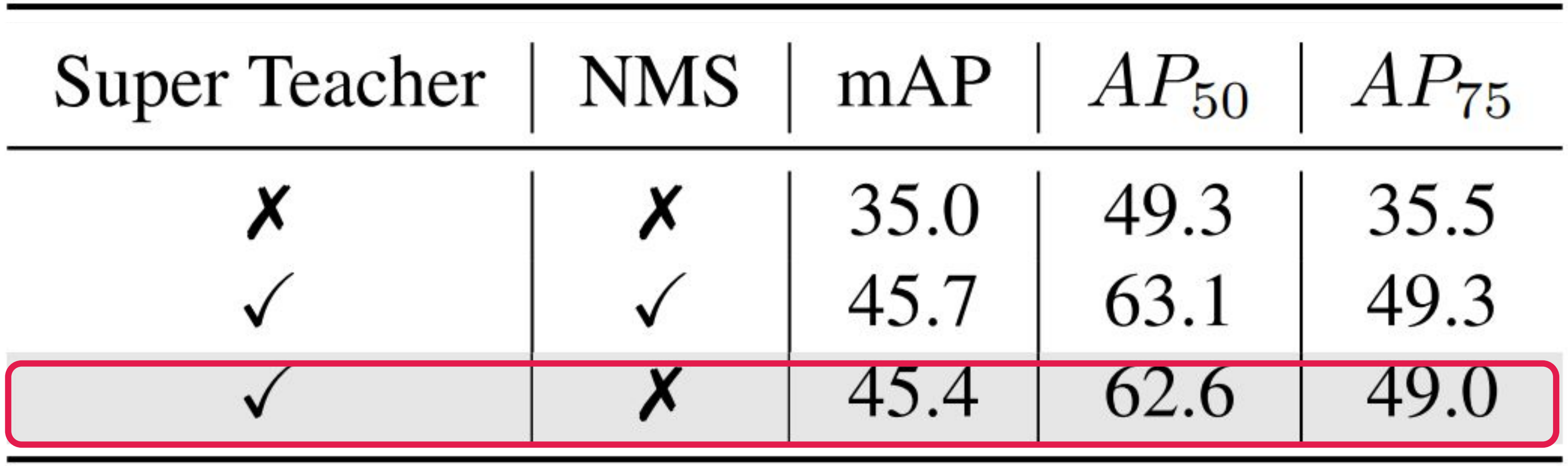
- Effect of different variants of queries.



Method	mAP	AP_{50}	AP_{75}
Standard Queries	41.3	55.8	44.3
Consistency Visual Queries	43.5	59.7	46.8
Sparse Visual Queries	44.3	61.7	47.6
Text Queries	45.4	62.6	49.0

Results

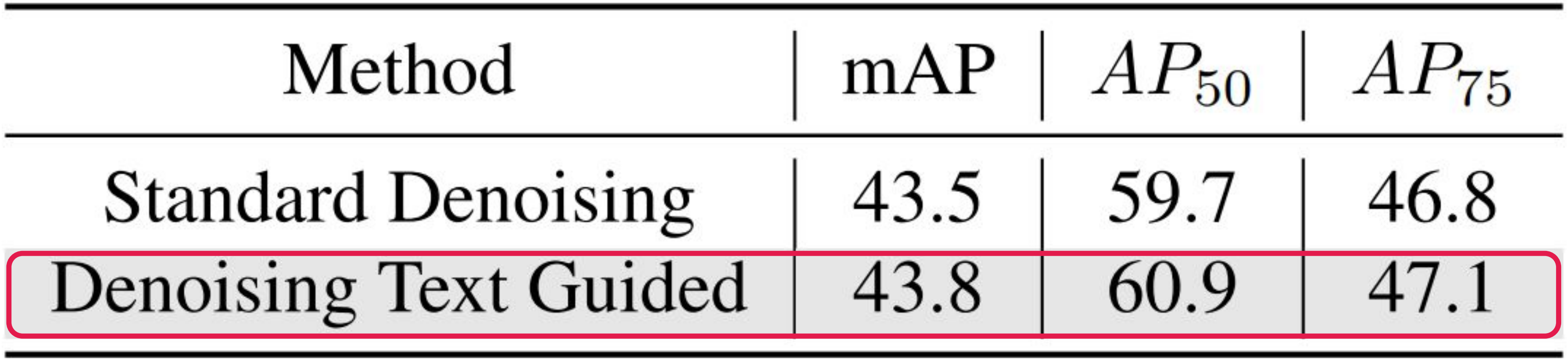
- Effect of Super Teacher



Super Teacher	NMS	mAP	AP_{50}	AP_{75}
✗	✗	35.0	49.3	35.5
✓	✓	45.7	63.1	49.3
✓	✗	45.4	62.6	49.0

Results

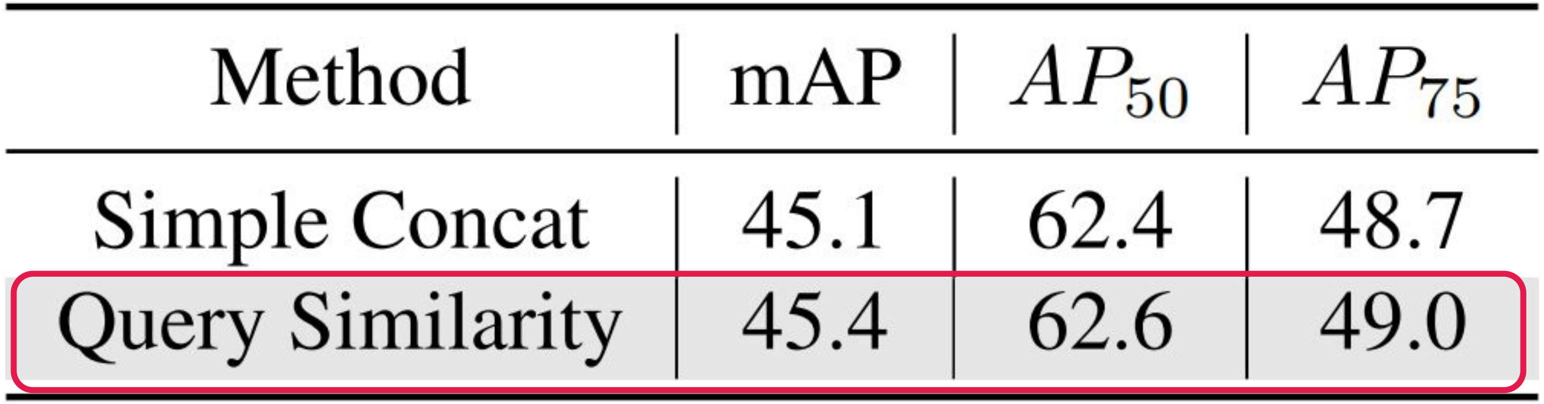
- Effect of Denoising Text Guided Queries.



Method	mAP	AP_{50}	AP_{75}
Standard Denoising	43.5	59.7	46.8
Denoising Text Guided	43.8	60.9	47.1

Results

- Effectiveness of Query Refinement



Method	mAP	AP_{50}	AP_{75}
Simple Concat	45.1	62.4	48.7
Query Similarity	45.4	62.6	49.0

Conclusion

- Semi-supervised object detection still struggles with noisy pseudo-labels, confidence bias, and inefficient queries.

- STEP-DETR addresses these issues by:
  - Generating reliable pseudo-labels with Super Teacher.
  - Incorporating text-guided queries for rare and common categories.
  - Refining queries to reduce noise and redundancy.

- Experiments on MS-COCO and Pascal VOC demonstrate that STEP-DETR outperforms existing methods, delivering state-of-the-art performance even with limited labeled data.

dfki ai
Deutsches Forschungszentrum
für Künstliche Intelligenz
German Research Center for
Artificial Intelligence

ICCV
OCT 19-23, 2025
HONOLULU
HAWAII

RPTU
Rheinland-Pfälzische
Technische Universität
Kaiserslautern
Landau

STEP-DETR: Advancing DETR-based Semi-Supervised Object Detection with Super Teacher and Pseudo-Label Guided Text Queries

Thanks a lot for your attention!