CVPR
JUNE 17-21, 2024
SEATTLE, WA

Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection

Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Muhammad Zeshan Afzal

Poster: Arch 4A-E-99

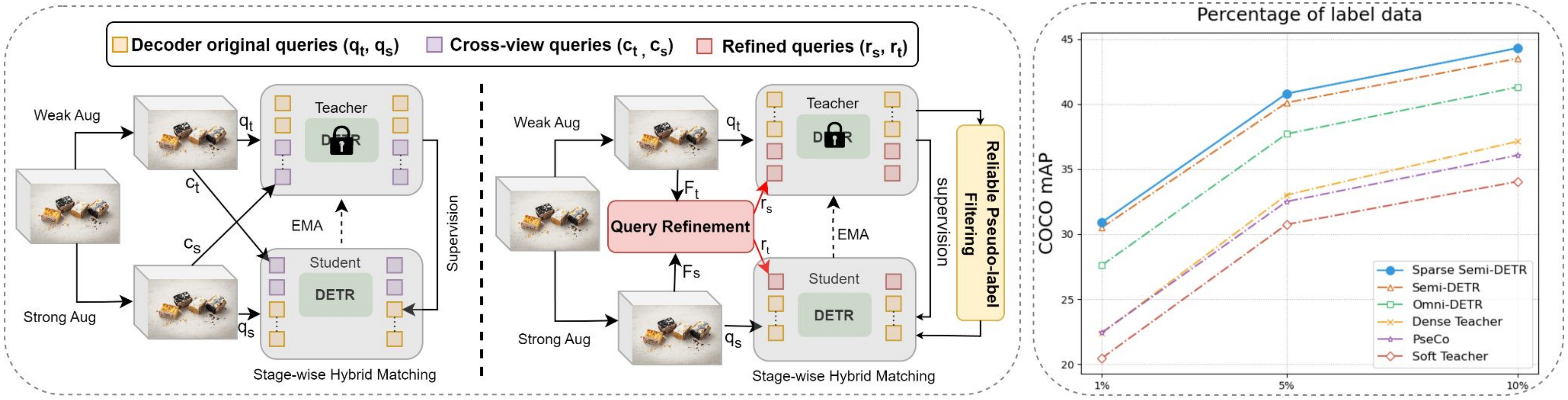
DFKI RPTU Kaiserslautern-Landau

dfki ai

RPTU Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

Sparse Semi-DETR Overview

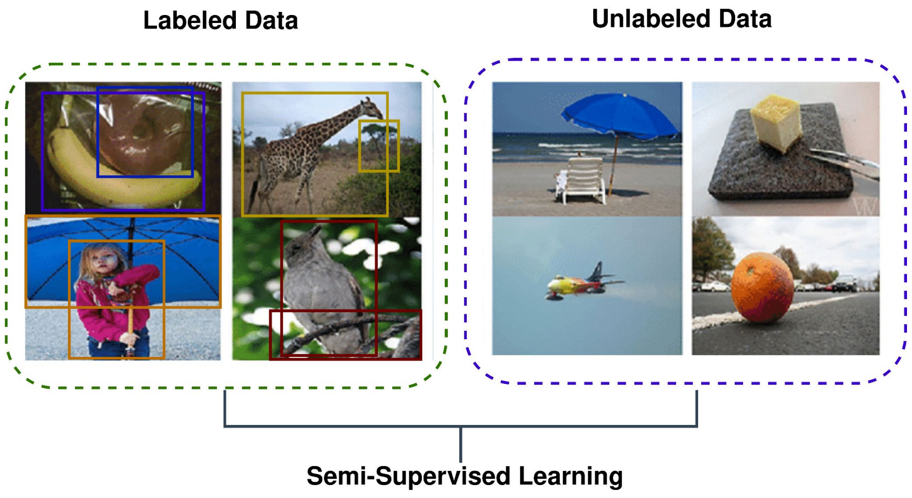
- Sparse Semi-DETR improves semi-supervised object detection for transformers, enhancing detection of small and obscured objects.
- Semi-DETR achieves SOTA on both COCO and PASCAL VOC datasets.



(a) Semi-DETR (b) Sparse Semi-DETR (c) Comparison of SSOD Methods

Semi-Supervised Object Detection (SSOD)

- Background
  - Problem Statement

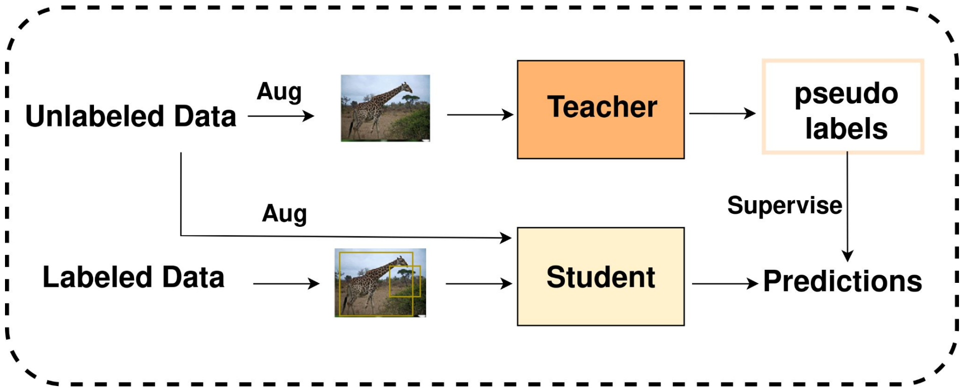


Settings:

- 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)

- Background
  - Current Research



- Soft-Teacher (Two-Stage Detector),
- Dense-Teacher (One-Stage Detector)

Problem: Anchor generation, Label assignment by various rules, NMS …

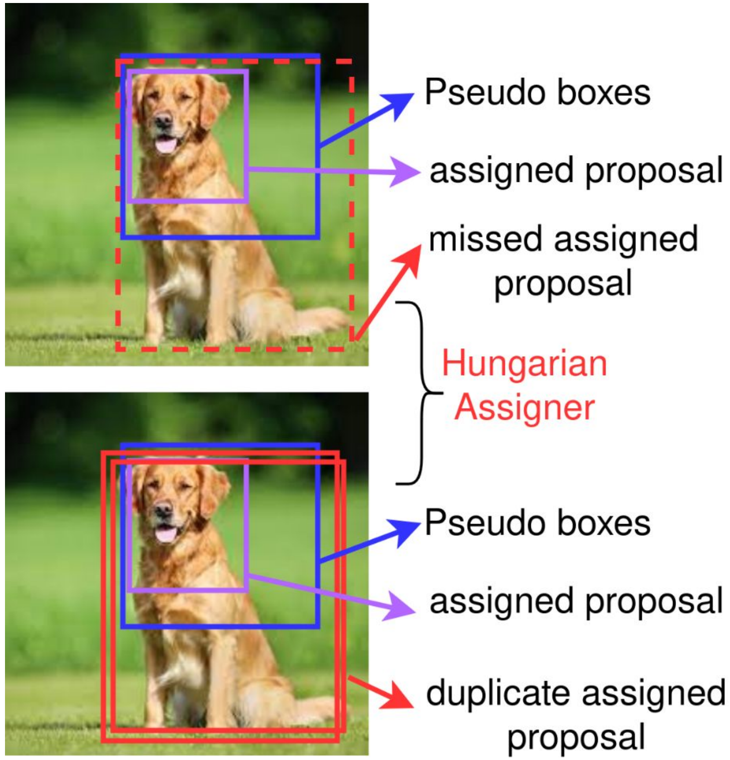
- Semi-DETR (End-to-End)

Problem: low performance on small objects, object queries quality.

[Xu 2021] End-to-End Semi-Supervised Object Detection with Soft Teacher ICCV2021
[Zhou 2022] Dense Teacher: Dense Pseudo-Labels for Semi-supervised Object Detection ECCV2022
[zhang 2023] Semi-DETR: Semi-Supervised Object Detection with Detection Transformers CVPR2023

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Motivation
  - Bipartite matching makes NMS-free but causes learning inefficiency.
  - Input query features in one-to-many assignment strategy needs to be refined to improve performance for small and partially obscured objects.

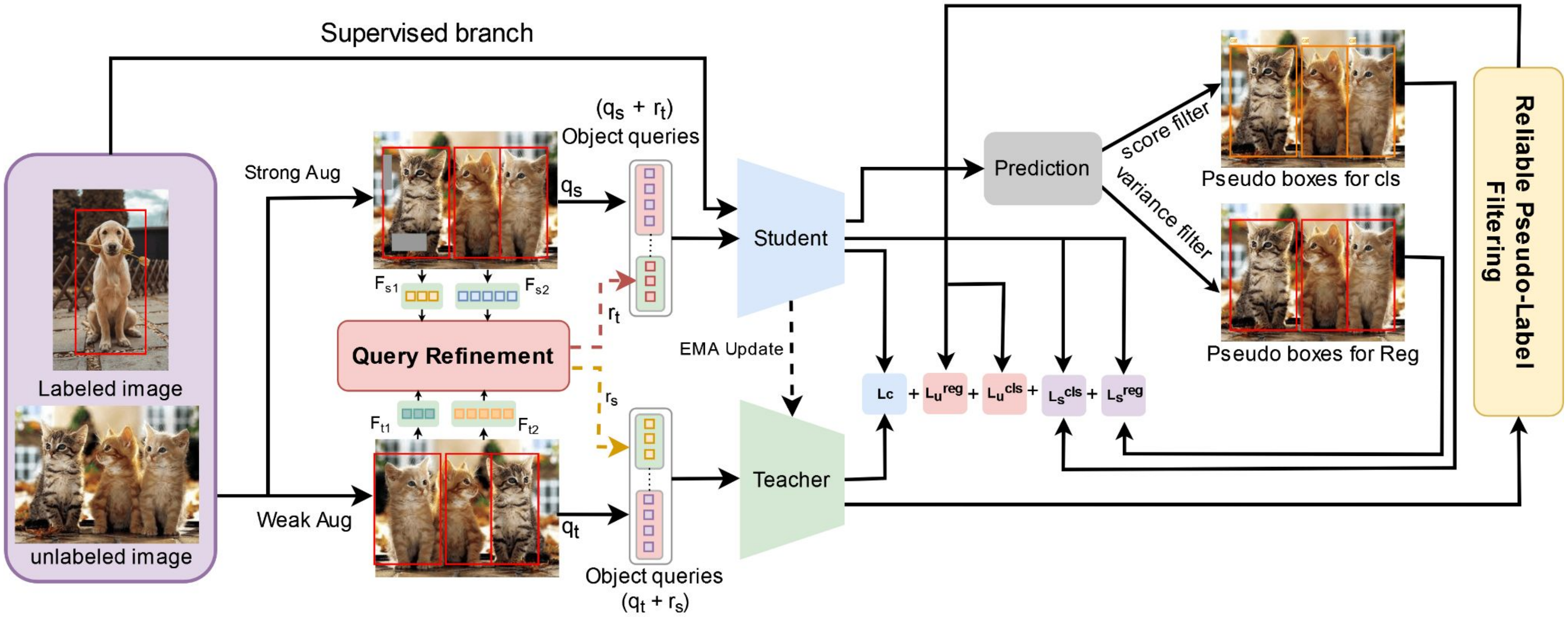


one-to-one assignment assigns a single positive proposal for each inaccurate pseudo box and missing potential better proposals.

one-to-many assignment assigns many positive proposals for each inaccurate pseudo box and results in duplicate proposals.

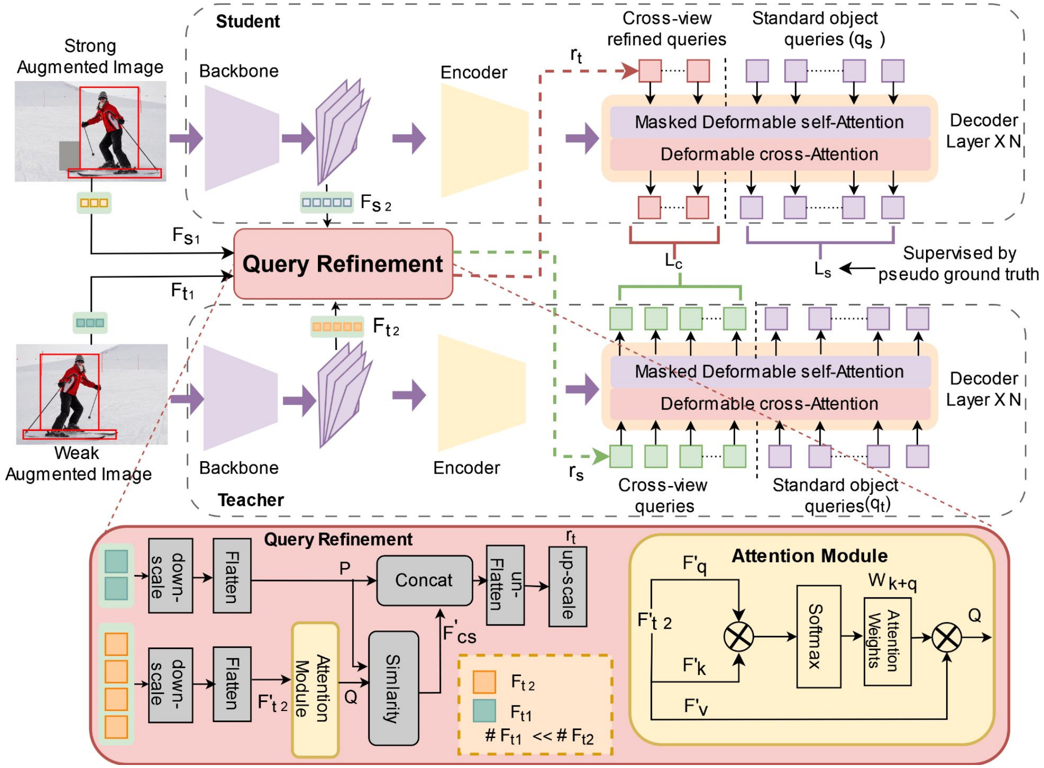
Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Approach



Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Approach
  - **Query Refinement:** Refine one-to-many assignment and combine it with one-to-one assignment to enable more efficient training for small and partially occluded objects.



Stage 1:
one-to-many query assignment with refines queries.

refined queries encoded features

$$\hat{o}_t, o_t = \mathrm{Dec}_t(\boxed{r_s}, \boxed{q_t}, \boxed{E_t}\boxed{A})$$
$$\hat{o}_s, o_s = \mathrm{Dec}_s(\boxed{r_t}, \boxed{q_s}, \boxed{E_s}\boxed{A})$$

attention mask

object queries

Stage 2:
one-to-one query assignment

encoded features

$$\hat{o}_t, o_t = \mathrm{Dec}_t(\boxed{q_t}, \boxed{E_t}\boxed{A})$$
$$\hat{o}_s, o_s = \mathrm{Dec}_s(\boxed{q_s}, \boxed{E_s}\boxed{A})$$

attention mask

object queries

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Approach
  - **Reliable Pseudo-Label Filtering:** The m groups of ground truths for one-to-many assignment strategy and select the top-k predictions. m is set to 6.

$$\hat{g} = \{\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_m\}$$

$$\hat{\sigma}_{one2many} = \left\{ \underset{\sigma_j \in C_M^N}{\arg\min} \sum_{k}^{M} \mathcal{L}_{\text{match}} \left(\hat{y}_j^t, \hat{g}_{\sigma_j(k)}^s \right) \right\}_j^{|\hat{y}^t|}$$

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Results

Methods	Reference	COCO-Partial		
		1%	5%	10%
FCOS [43] (Supervised)	-	8.43 ± 0.03	17.01 ± 0.01	20.98 ± 0.01
DSL [3]	CVPR22	22.03 ± 0.28 (+13.98)	30.87 ± 0.24 (+13.86)	36.22 ± 0.18 (+15.24)
Unbiased Teacher v2 [29]	CVPR22	22.71 ± 0.42 (+14.28)	30.08 ± 0.04 (+13.07)	32.61 ± 0.03 (+11.63)
Dense Teacher [58]	ECCV22	22.38 ± 0.31 (+13.95)	33.01 ± 0.14 (+16.00)	37.13 ± 0.12 (+16.15)
Faster RCNN [36] (Supervised)	-	9.05 ± 0.16	18.47 ± 0.22	23.86 ± 0.81
Humble Teacher [42]	CVPR22	16.96 ± 0.38 (+7.91)	27.70 ± 0.15 (+9.23)	31.61 ± 0.28 (+7.75)
Instant-Teaching [59]	CVPR21	18.05 ± 0.15 (+9.00)	26.75 ± 0.05 (+8.28)	30.40 ± 0.05 (+6.54)
Soft Teacher [52]	ICCV21	20.46 ± 0.39 (+11.41)	30.74 ± 0.08 (+12.27)	34.04 ± 0.14 (+10.18)
PseCo [20]	ECCV22	22.43 ± 0.36 (+13.38)	32.50 ± 0.08 (+14.03)	36.06 ± 0.24 (+12.2)
DINO [56] (Supervised)	-	18.00 ± 0.21	29.50 ± 0.16	35.00 ± 0.12
Omni-DETR [47] (DINO)	CVPR22	27.60 (+9.60)	37.70(+8.20)	41.30 (+6.30)
Semi-DETR [57] (DINO)	CVPR23	30.5 ± 0.30 (+12.50)	40.10 ± 0.15 (+10.6)	43.5 ± 0.10 (+8.5)
Sparse Semi-DETR	-	**30.9 ± 0.23** (+12.90)	**40.8 ± 0.12** (+11.30)	**44.3 ± 0.01** (+9.30)

COCO Partial

Methods	Labels	COCO-Partial		
		AP_S	AP_M	AP_L
Semi-DETR [57]	1%	13.6	31.2	40.8
	5%	23.0	43.1	53.7
	10%	25.2	46.8	58.0
Sparse Semi-DETR	1%	14.8	32.5	41.4
	5%	23.9	44.2	54.2
	10%	26.9	48.0	59.6

Methods	VOC12	
	AP_{50}	$AP_{50:95}$
FCOS [43] (Supervised)	71.36	45.52
DSL [3]	80.70	56.80
Dense Teacher [58]	79.89	55.87
Faster RCNN [36] (Supervised)	72.75	42.04
STAC [40]	77.45	44.64
HumbleTeacher [42]	80.94	53.04
Instant-Teaching [59]	79.20	50.00
DINO [56] (Supervised)	81.20	59.60
Semi-DETR [57] (DINO)	86.10	65.20
Sparse Semi-DETR	**86.30**	**65.51**

COCO Full

Method	COCO-Full (100%)
STAC [40] (18×)	$39.5 \xrightarrow{-0.3} 39.2$
Unbiased Teacher (9×)	$40.2 \xrightarrow{+1.1} 41.3$
SoftTeacher [52] (24×)	$40.9 \xrightarrow{+3.6} 44.5$
DSL [3] (12×)	$40.2 \xrightarrow{+3.6} 43.8$
Dense Teacher [58] (18×)	$41.2 \xrightarrow{+3.6} 46.1$
PseCo (24×)	$41.0 \xrightarrow{+5.1} 46.1$
Instant-Teaching [59] (24×)	$37.6 \xrightarrow{-0.27} 40.2$
Semi-DETR [57] (8×)	$48.6 \xrightarrow{+1.8} 50.4$
Sparse Semi-DETR (8×)	$\mathbf{49.2} \xrightarrow{+2.1} \mathbf{51.3}$

PASCAL VOC

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Ablation Study

Query Refinement	Pseudo-Label Filtering	mAP	AP_{50}	AP_{75}
✗	✗	43.5	58.9	46.0
✓	✗	43.8	61.1	47.3
✗	✓	43.9	60.5	46.3
✓	✓	44.3	61.7	47.6

Effect of Individual Component.

Teacher	Student	mAP	AP_{50}	AP_{75}
✗	✗	43.5	58.9	46.0
✓	✗	43.8	61.2	47.2
✗	✓	44.3	61.7	47.6
✓	✓	42.8	59.7	45.8

Effect of QR on Student and Teacher module.

Method	mAP	AP_{50}	AP_{75}
Single-view Queries	43.0	59.3	46.3
Cross-view Queries	43.5	58.9	46.0
Query Refinement	44.3	61.7	47.6

Effectiveness of Attentional module in QR.

Method	mAP	AP_{50}	AP_{75}
Simple Concat	43.4	58.8	46.1
Cosine Similarity	43.7	60.3	46.1
Attention Module	44.3	61.7	47.6

Effect of different variants of queries.

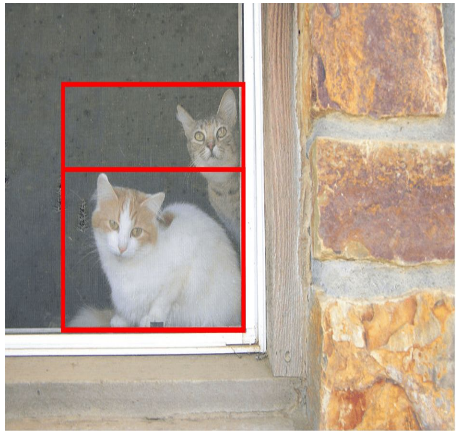
Approach	Training time (min)
Semi-DETR	38.56
Sparse Semi-DETR	**34.38** +4.18

Training time for 1k iterations in one-to-many assignment strategy.

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

- Visualization
  - Sparse Semi-DETR improves Semi-DETR on the detection of small and partially obscured objects.

Semi-DETR



Sparse Semi-DETR

Conclusion

- We present Sparse Semi-DETR, a novel approach in semi-supervised object detection with Sparse Learnable Queries for detection transformers.

- We introduce a novel query refinement module designed to improve object query features, particularly in complex detection scenarios such as identifying small or partially obscured objects.

- We introduce a Reliable Pseudo-Label Filtering Module specifically designed to reduce the effect of noisy pseudo-labels. This module is designed to efficiently identify and extract reliable pseudo boxes from unlabeled data using augmented ground truths, enhancing the consistency of the learning process.

- On the MS-COCO and Pascal VOC object detection benchmarks, Sparse Semi-DETR achieves improvements in performance over current state-of-the-art methods.

Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

CVPR JUNE 17-21, 2024 SEATTLE, WA

Sparse Semi-DETR: Sparse Learnable Queries for SSOD with Detection Transformers

Thanks a lot for Your Attention!