JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

SAP-DETR: Bridging the Gap between Salient Points and Queries-Based Transformer Detector for Fast Model Convergency

- Yang Liu
- Yang Liu[1,2], Yao Zhang[1,2], Yixin Wang[1,2], Jiang Tian[3], Zhongchao Shi[3], Jianping Fan[3], Zhiqiang He[1,2,4]
- [1]Institute of Computing Technology (ICT), Chinese Academy of Sciences
  [2]University of Chinese Academy of Sciences [3]AI Lab, Lenovo Research [4]Lenovo, Ltd.

- WED-PM-303

liuyang20c@mails.ucas.ac.cn

https://github.com/liuyang-ict/SAP-DETR

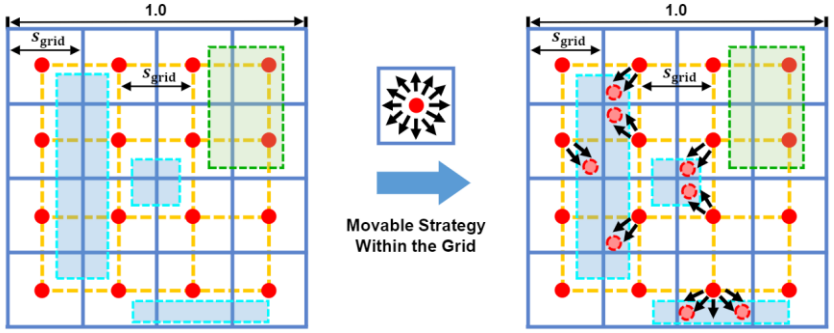
@CVPR23

Summary

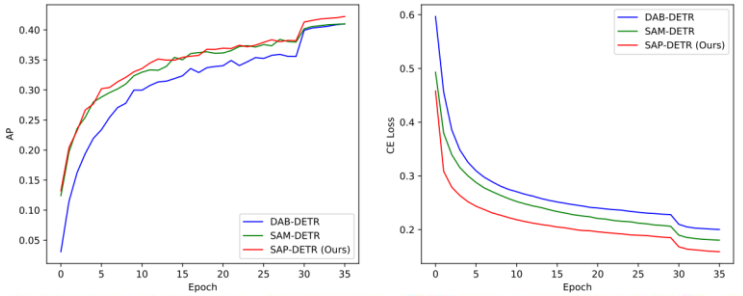
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

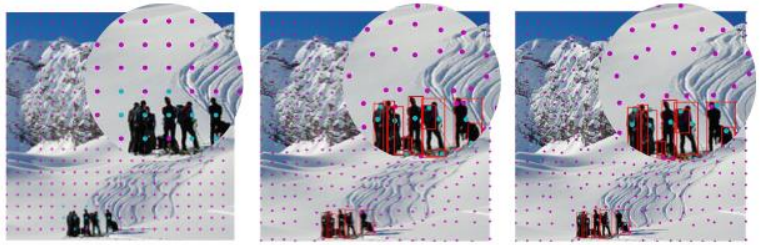
Introduction

We introduce the salient point concept into query-based Transformer detectors by assigning query-specific reference points to object queries. Unlike center-based methods with the confused central spatial prior, we restrict the reference location and define the point of the positive query as the salient one, hence enlarging the discrepancy of query as well as reducing the redundant predictions. SAP-DETR accelerates the convergency speed greatly, achieving competitive performance with ~30% fewer training epochs. The proposed movable strategy further boosts SAP-DETR to a new SoTA performance.



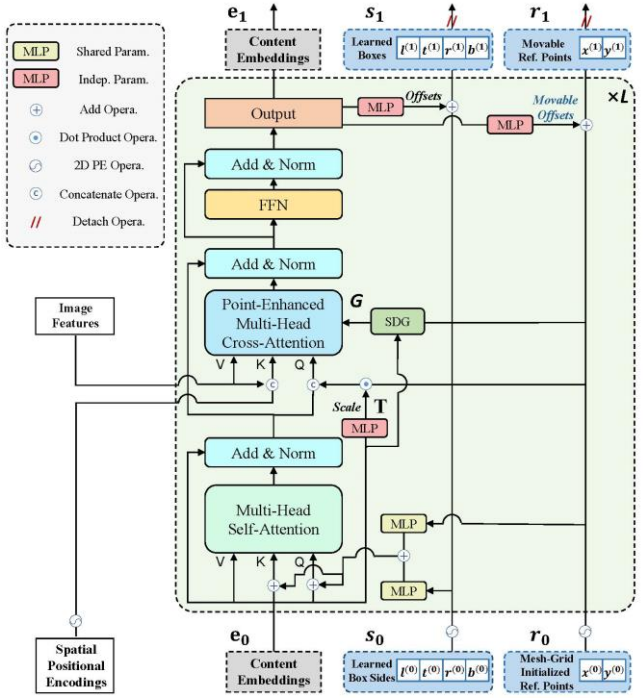
Comparison





Initialized Reference Point | Output Point from Layer #0 | Output Point from Layer #5

Method



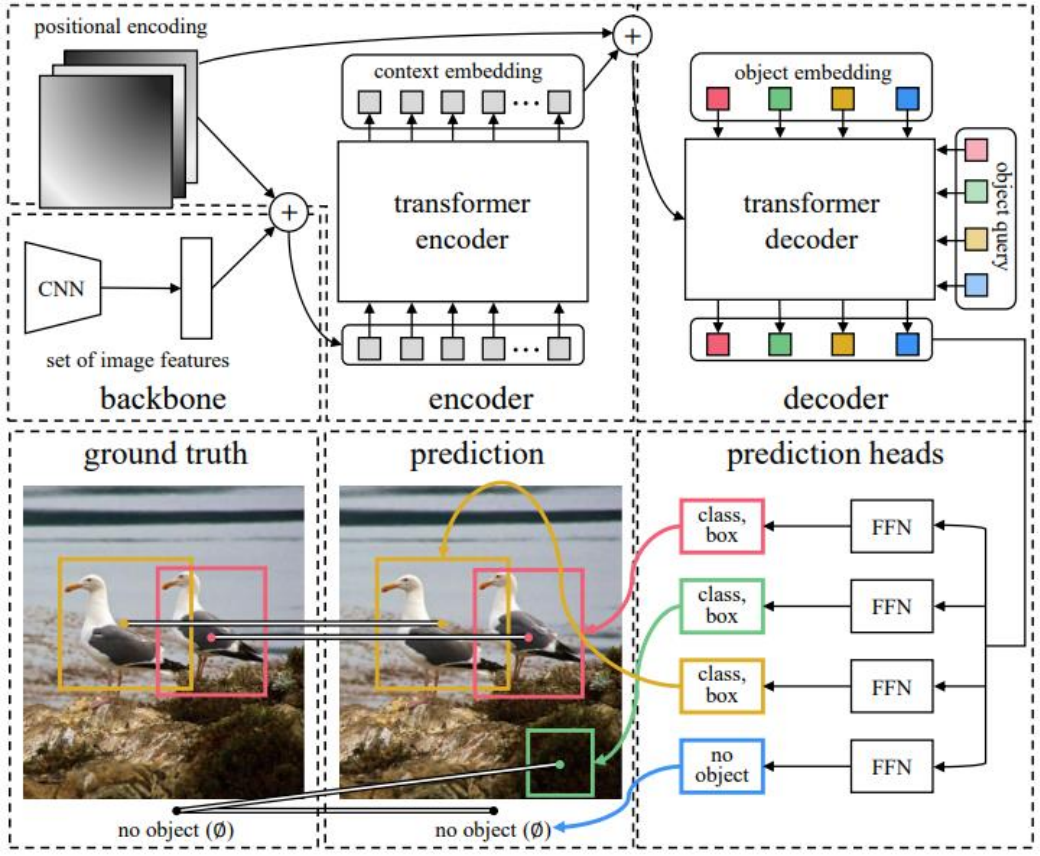
Experimental Results

Method	#Epochs	#Params(M)	GFLOPs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
ResNet-50 Backbone									
Faster RCNN-FPN-R50 [10, 18]	108	42	180	42.0	62.1	45.5	26.6	45.5	53.4
DETR-R50 [2]	500	41	86	42.0	62.4	44.2	20.5	45.8	61.1
Deformable DETR-R50 [31]	50	34	78	39.4	59.6	42.3	20.6	43.0	55.5
SMCA-DETR-R50 [5]	50	42	86	41.0	-	-	21.9	44.3	59.1
Conditional DETR-R50 [16]	50	44	90	40.9	61.8	43.3	20.8	44.6	59.2
Anchor DETR-R50 [25]	50	39	85	42.1	63.1	44.9	22.3	46.2	60.0
DAB-DETR-R50 [13]	50	44	90†	42.2	63.1	44.7	21.5	45.7	60.3
SAM-DETR-w/SMCA-R50 [26]	50	58	100	41.8	63.2	43.9	22.1	45.9	60.9
SAP-DETR-R50 (**Ours**)	50	47	92	**43.1**	**63.8**	**45.4**	**22.9**	**47.1**	**62.1**
ResNet-101 Backbone									
Faster RCNN-FPN-R101 [10, 18]	108	60	246	44.0	63.9	47.8	27.2	48.1	56.0
DETR-R101 [2]	500	60	152	43.5	63.8	46.4	21.9	48.0	61.8
Conditional DETR-R101 [16]	50	63	156	42.8	63.7	46.0	21.7	46.6	60.9
Anchor DETR-R101 [25]	50	58	150	43.5	64.3	46.6	23.2	47.7	61.4
DAB-DETR-R101 [13]	50	63	157†	43.5	63.9	46.6	23.6	47.3	61.5
SAP-DETR-R101 (**Ours**)	50	67	158	**44.4**	**64.9**	**47.1**	**24.1**	**48.7**	**63.1**

Backbone	Epoch	w/ SAP	DN-DETR [9]		DINO (Single-Scale) [28]	
			AP / AP_{50} / AP_{75}	AP_S / AP_M / AP_L	AP / AP_{50} / AP_{75}	AP_S / AP_M / AP_L
R50	12		38.3 / 58.6 / 40.5	18.4 / 41.6 / 57.1	39.7 / 58.3 / 42.4	19.1 / 43.7 / 57.1
		✓(**Ours**)	**39.5 / 59.7 / 41.5**	**18.7 / 42.8 / 59.0**	**40.0 / 60.1** / 42.1	**20.2** / 43.4 / **58.5**
R101	12		40.5 / 60.8 / 43.0	19.3 / 44.3 / 59.6	41.9 / 60.8 / 44.4	22.5 / 46.3 / 59.5
		✓(**Ours**)	**41.0 / 61.2 / 43.4**	**19.8 / 45.3 / 60.0**	41.5 / **61.4** / 43.6	20.3 / 45.2 / **60.0**
R50-DC	12		41.7 / 61.4 / 44.1	21.2 / 45.0 / 60.2	43.6 / 61.4 / 47.0	24.8 / 47.3 / 59.5
		✓(**Ours**)	**43.6 / 62.5 / 46.2**	**23.3 / 47.3 / 61.0**	**44.0 / 63.1** / 46.5	**24.8 / 47.3 / 61.1**
R101-DC	12		43.4 / 61.9 / 47.2	24.8 / 46.8 / 59.4	45.4 / 63.5 / 49.2	26.4 / 49.5 / 61.1
		✓(**Ours**)	**44.6 / 63.9 / 48.0**	**25.5 / 48.9 / 62.5**	**45.6 / 64.5** / 48.7	25.0 / **49.7 / 62.5**

Background

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



One object
- Semantic class [*cls*] 91
- Bounding box coordinates [*cx,cy,w,h*] or [*x1,y1,x2,y2*] 4

Object vector [*cls,cx,cy,w,h*] 95

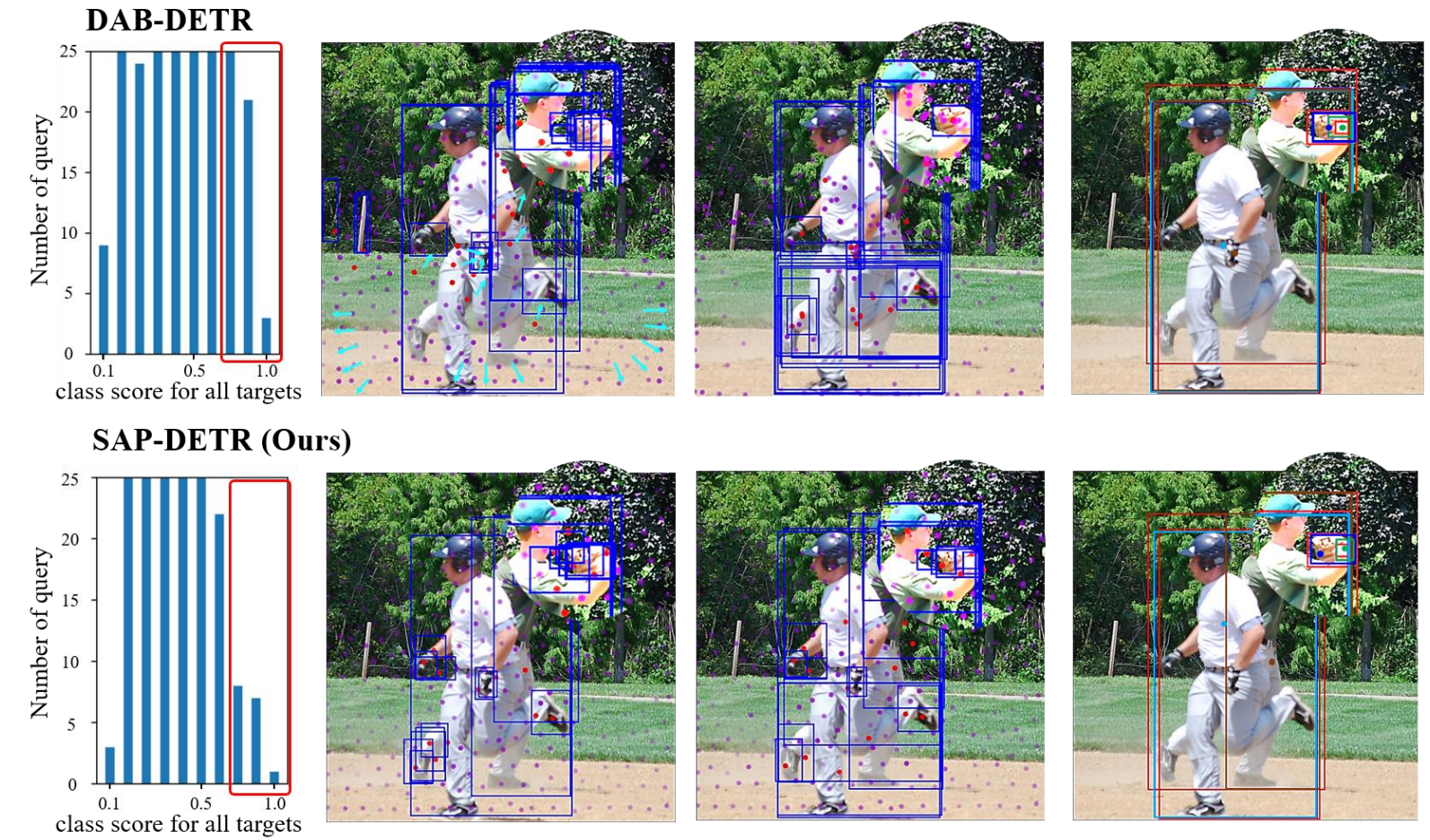
Object query [*q1,q2,q3,…,qC*] C

Add

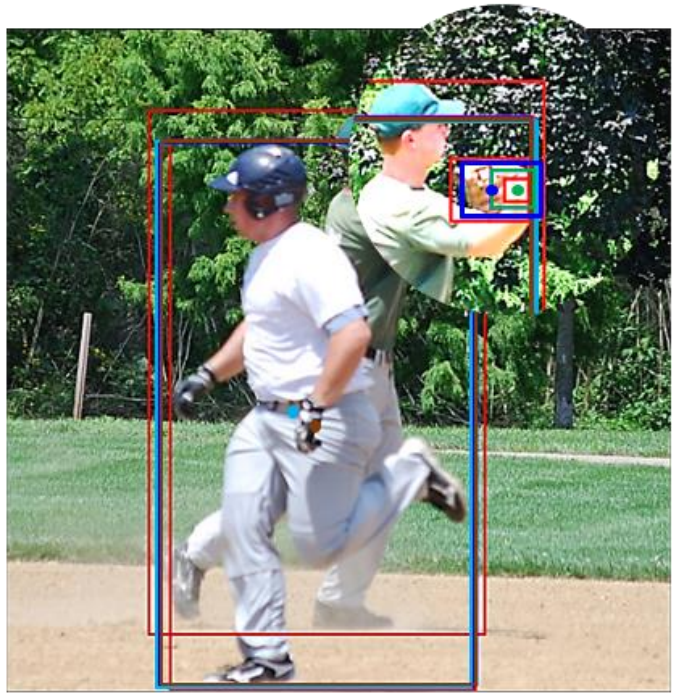
Learnable Positional Encoding [*pe1,pe2,pe3,…,peC*] C

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Problem Description

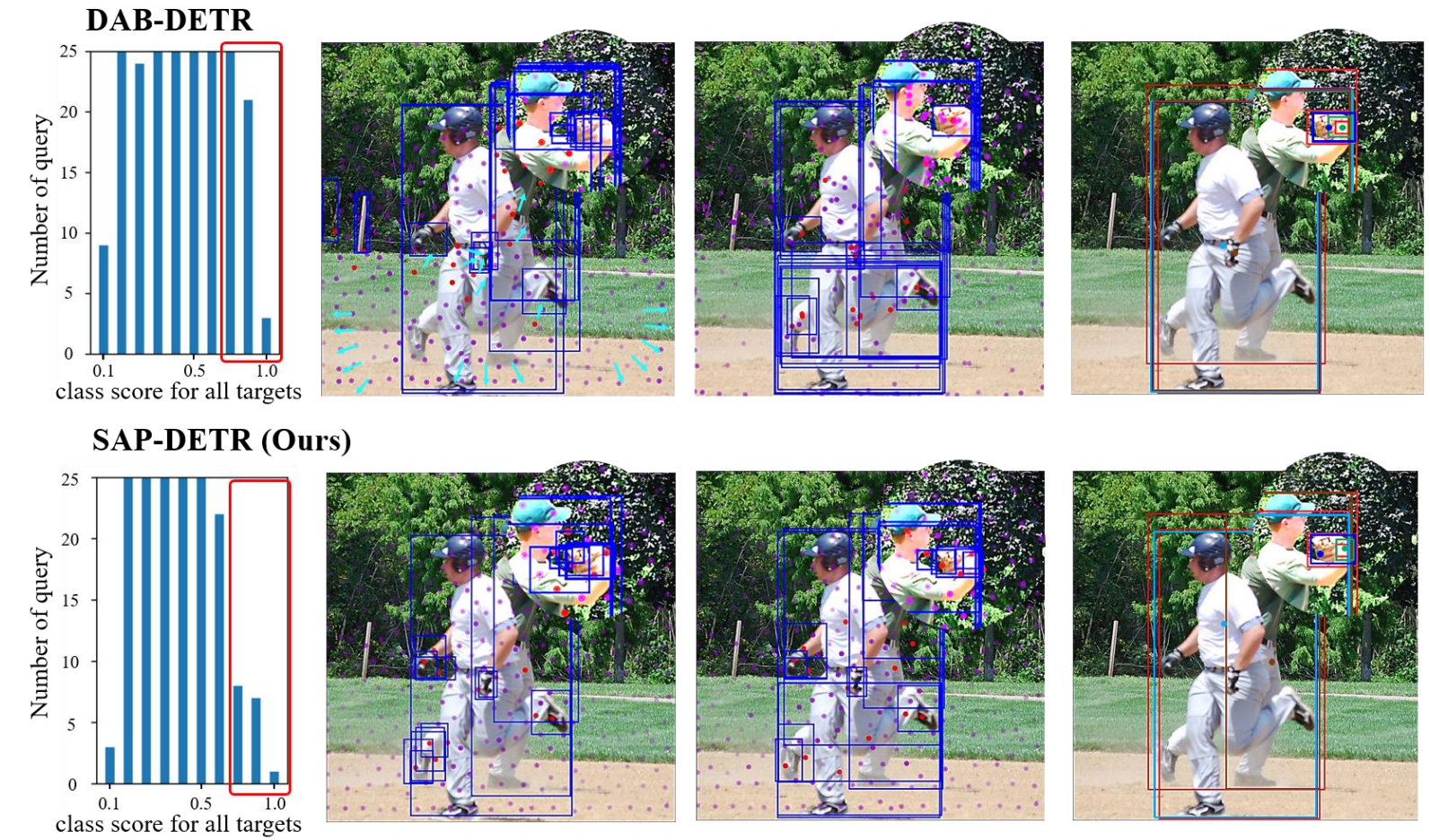


(a) Score Distribution (b) Layer 0 Top-20 (c) Layer 2 Top-20 (d) Layer 2 Output&GT

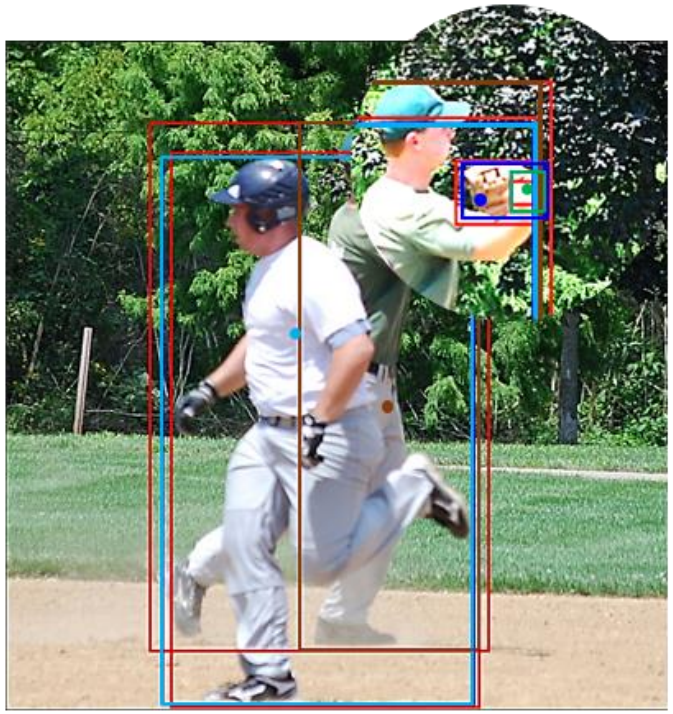


liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Problem Description

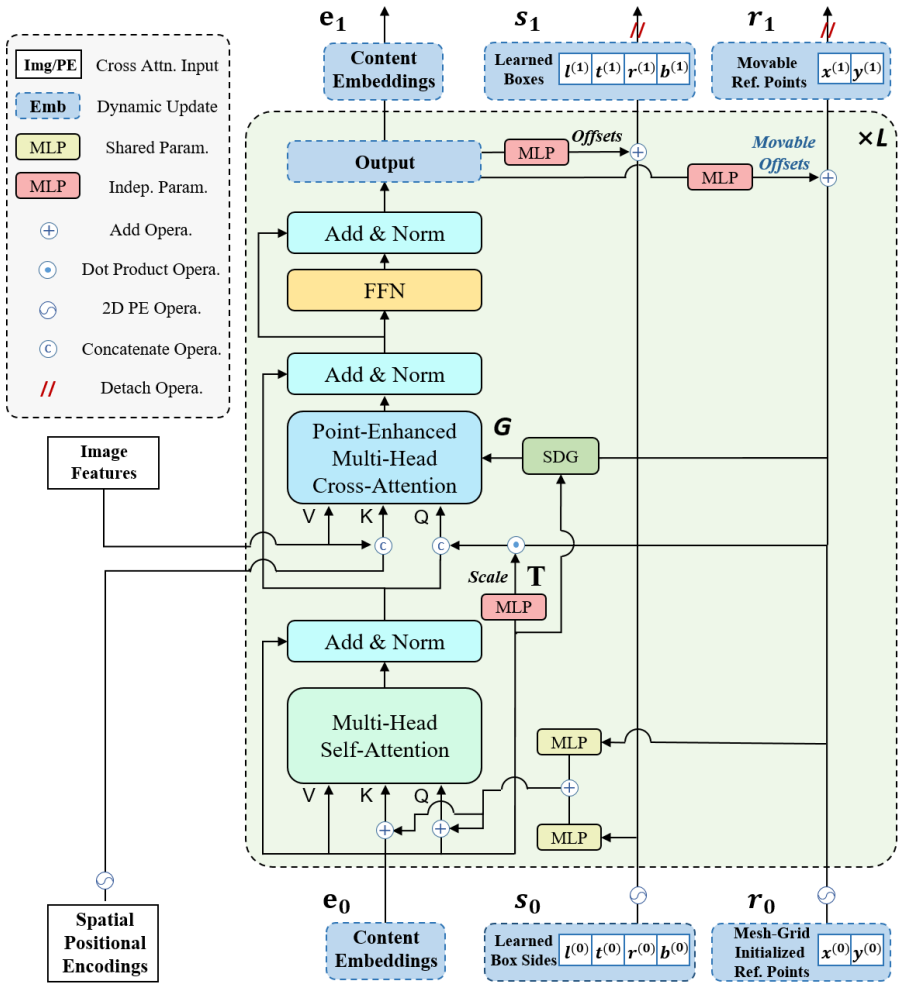


(a) Score Distribution (b) Layer 0 Top-20 (c) Layer 2 Top-20 (d) Layer 2 Output&GT

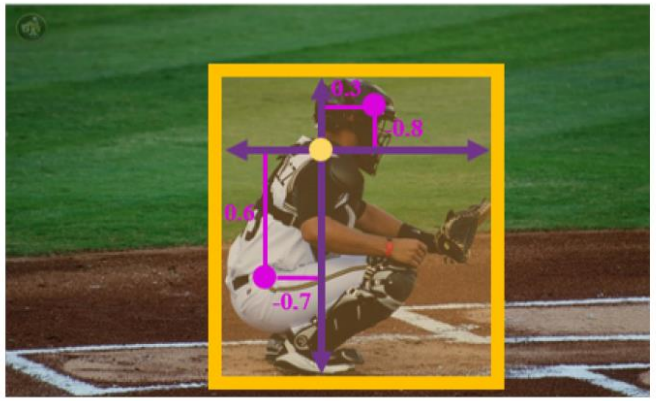


liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Method



Overview



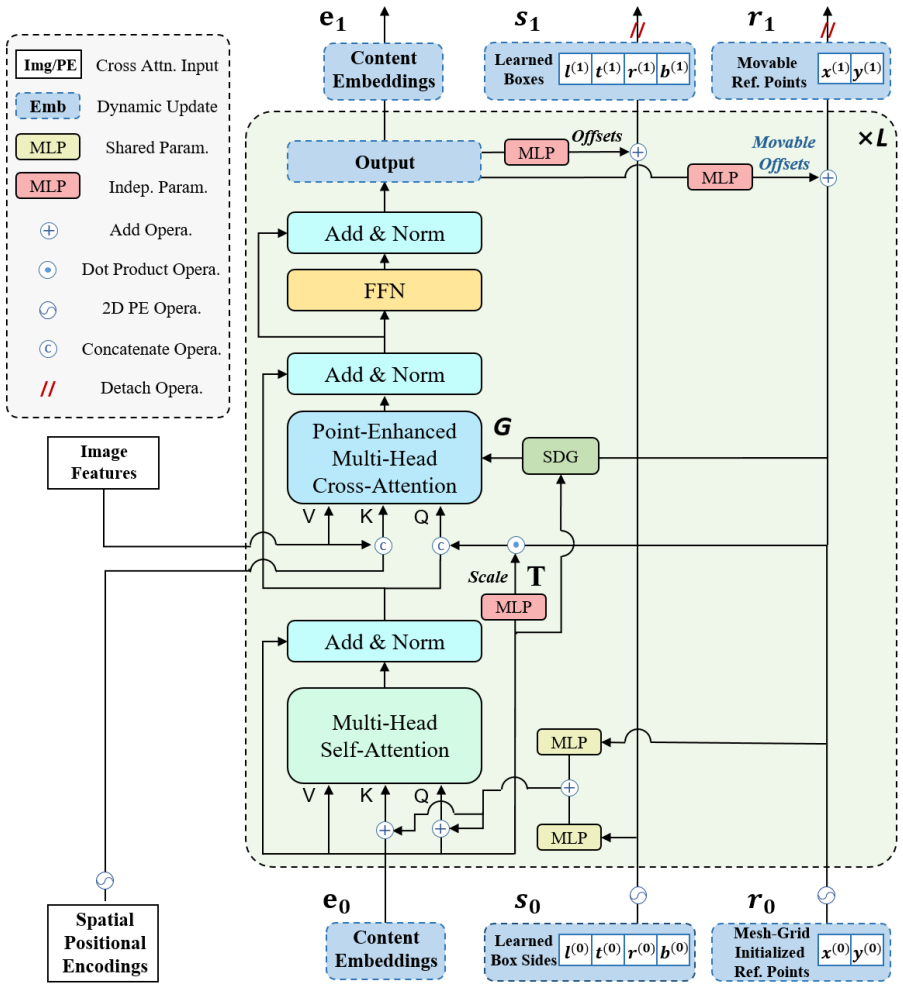
$$\boldsymbol{r}=\{x,y\}\in[0,1]^2 \qquad \boldsymbol{s}=\{\ell,t,r,b\}\in[0,1]^4$$

$$\boldsymbol{q}=\{\mathbf{e};\boldsymbol{r},\boldsymbol{s}\},\ \mathbf{e}\in\mathbb{R}^d \qquad \hat{\boldsymbol{b}}=\{\hat{x}-\hat{\ell},\hat{y}-\hat{t},\hat{x}+\hat{r},\hat{y}+\hat{b}\}$$

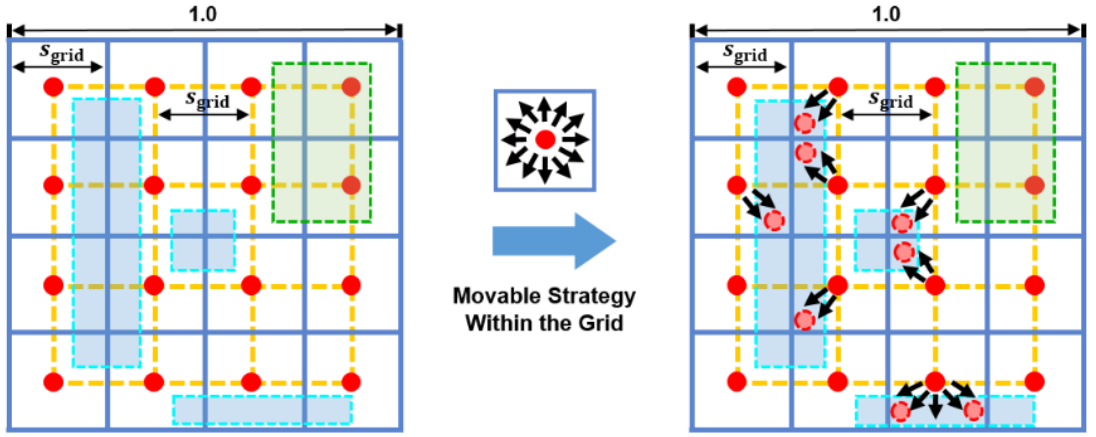
$$\Delta \boldsymbol{s}_l=\mathrm{BoxHead}_l(\boldsymbol{s}_{l-1},\mathbf{e}_{l-1},\boldsymbol{r}_{l-1}),$$
$$\hat{\boldsymbol{s}}_l=\sigma(\sigma^{-1}(\boldsymbol{s}_{l-1})+\Delta \boldsymbol{s}_l), \quad \boldsymbol{s}_l=\mathrm{Detach}(\hat{\boldsymbol{s}}_l),$$
$$\boldsymbol{r}_l=\hat{\boldsymbol{r}}_l=\boldsymbol{r}_{l-1}, \quad \hat{\boldsymbol{b}}_l=\{\hat{\boldsymbol{r}}_l-\hat{\boldsymbol{s}}_l[:2],\hat{\boldsymbol{r}}_l+\hat{\boldsymbol{s}}_l[2:]\},$$

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Method



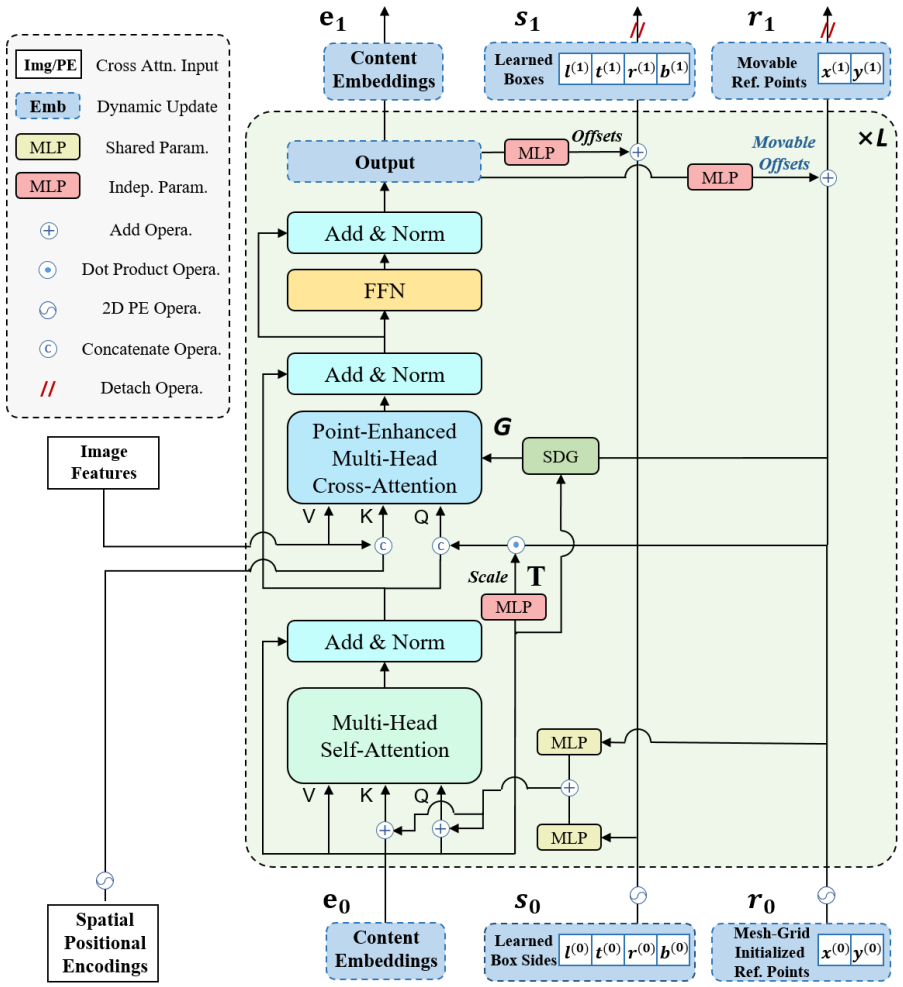
Movable Reference Point



$$\Delta \boldsymbol{r}_l' = \text{PointHead}_l(\boldsymbol{s}_{l-1}, \mathbf{e}_{l-1}, \boldsymbol{r}_{l-1}),$$
$$\Delta \boldsymbol{r}_l = \sigma(\sigma^{-1}(\boldsymbol{r}_{l-1} - \boldsymbol{r}_0) + \Delta \boldsymbol{r}_l'),$$
$$\hat{\boldsymbol{r}}_l = \boldsymbol{r}_0 + \Delta \boldsymbol{r}_l \cdot \boldsymbol{s}_{\text{grid}}, \quad \boldsymbol{r}_l = \text{Detach}(\hat{\boldsymbol{r}}_l),$$

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Method



Point-Enhanced Cross-Attention (PECA)

$$\mathbf{A}_{\text{peca}} = \mathbf{e}_q \mathbf{e}_k^\top + \mathbf{T}\text{PE}(\boldsymbol{r}_q)\text{PE}(\boldsymbol{r}_k)^\top + \mathbf{T}g(\text{PE}(\boldsymbol{r}_q - \{\ell, t\}, \boldsymbol{r}_q + \{r, b\}))\text{PE}(\boldsymbol{r}_k)^\top$$

Side-Directed Gaussian (SDG)

Input: Content embedding $\mathbf{e}$, reference point $\boldsymbol{r}$ and box $\boldsymbol{s}$.

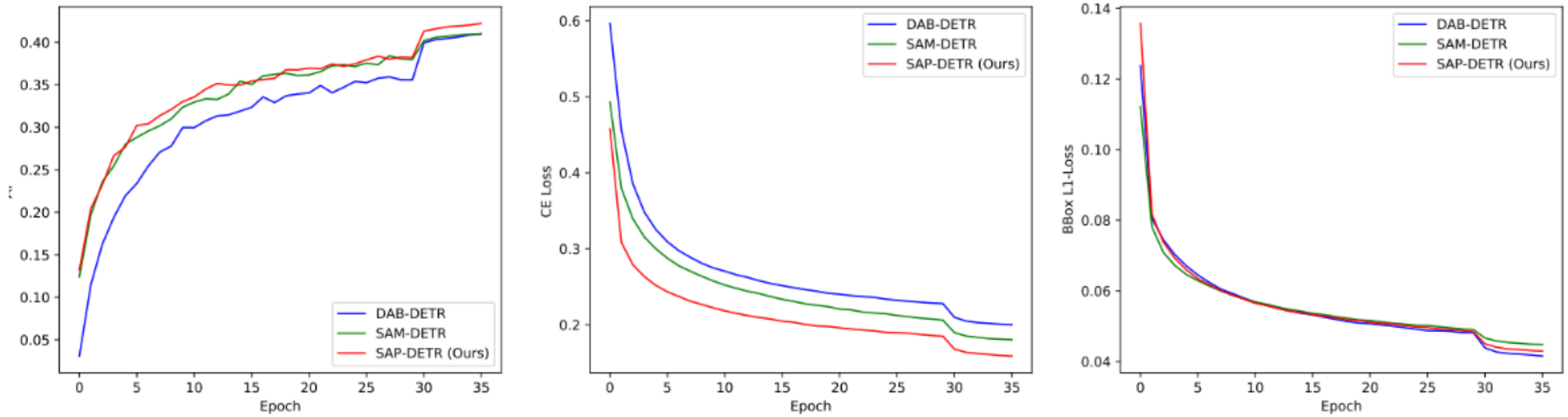
Output: Head-specific points $\mathbf{c} = \{(c_{w,i}, c_{h,i}) | i \in H\}$ and head-specific attention $\mathbf{v} = \{(v_{w,i}, v_{h,i}) | i \in H\}$.

1: Predict offset scale and attention scale based on content embedding, $\mathbf{o} = \tanh(\text{MLP}(\mathbf{e}))$, $\mathbf{v} = \text{MLP}(\mathbf{e})$;
2: **for** $h \leftarrow 1 \in H$ **do**
3: Select the index of direction guided by the sign of offset scale, $\{a, b\} = \text{sgn}(\mathbf{o}_i) + \{1, 2\}, \quad a, b \in \{0, 1, 2, 3\}$;
4: According to the index of direction, predict head-specific point, $\mathbf{c}_i = \mathbf{o}_i \cdot \boldsymbol{s}[a, b] + \boldsymbol{r}$;
5: **end for**
6: **return** $\mathbf{c}_i, \mathbf{v}_i, \ \forall i = 1, ..., H$

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Comparison

Method	Spatial Prior	Reference Coordinate	Target Prediction	Cross-Attn.	Reference Prior Update	Discriminative PE
DETR	No	No	$[cx, cy, w, h]$	Standard		✓
Deformable DETR	Implicit	4D	$[dcx, dcy, w, h]$	Deformable Points		✓
SMCA-DETR	Implicit	4D	$[\Delta cx, \Delta cy, w, h]$	Gaussian Points		
Conditional DETR	Implicit	2D	$[\Delta cx, \Delta cy, w, h]$	Conditional		
Anchor DETR	Explicit	2D	$[\Delta cx, \Delta cy, w, h]$	Standard		✓
DAB-DETR	Explicit	4D	$[\Delta cx, \Delta cy, \Delta w, \Delta h]$	Conditional	✓	
SAM-DETR-w/SMCA	Explicit	4D	$[\Delta cx, \Delta cy, \Delta w, \Delta h]$	Gaussian Points	✓	
SAP-DETR (Ours)	Explicit	2D+4D	$[\Delta x, \Delta y, \Delta \ell, \Delta t, \Delta r, \Delta b]$	Conditional Side	✓	✓



liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Experimental Results

Method	#Epochs	#Params(M)	GFLOPs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
3-Layer Encoder-Decoder Transformer Neck with ResNet-50 Backbone									
DETR-R50 [2]	36	33	82	15.8	28.0	15.4	5.3	16.7	24.6
Deformable DETR-R50 [31]	36	30	77	37.1	57.6	39.4	18.3	40.8	51.6
SMCA-DETR-R50 [5]	12 / 36	-	-	28.8 / 37.7	48.1 / 58.7	29.9 / 40.1	13.8 / 19.4	31.3 / 40.5	41.3 / 54.8
Conditional DETR-R50 [16]	12 / 36	40	82	29.6 / 37.1	48.7 / 57.9	30.7 / 39.0	13.0 / 17.6	32.3 / 40.3	43.1 / 55.0
Anchor DETR-R50 [25]	12 / 36	31	79	30.8 / 37.6	51.1 / 58.7	31.8 / 39.7	14.3 / 18.8	34.1 / 41.5	44.3 / 53.5
DAB-DETR-R50 [13]	12 / 36	34	83	32.3 / 39.0	51.3 / 58.6	34.0 / 41.8	15.7 / 20.0	35.2 / 42.5	45.7 / 56.0
SAM-DETR-w/SMCA-R50 [26]	12 / 36	41	89	35.1 / 40.4	54.7 / 60.7	36.7 / 42.7	16.0 / 20.2	38.4 / 44.4	52.1 / 58.3
SAP-DETR-R50 (Ours)	12 / 36	36	84	**36.2 / 41.2**	**56.2 / 61.6**	**37.9 / 43.4**	**16.4 / 21.0**	**39.5 / 44.5**	**53.8 / 60.1**
6-Layer Encoder-Decoder Transformer Neck with ResNet-50 Backbone									
DETR-R50 [2]	36	42	89	14.0	24.4	14.0	4.2	13.7	22.5
Deformable DETR-R50 [31]	36	34	81	38.0	58.2	40.4	18.5	41.7	54.2
SMCA-DETR-R50 [5]	12 / 36	-	-	32.4 / 40.1	52.3 / 61.4	34.0 / 42.8	15.5 / 20.3	34.9 / 43.3	47.7 / 57.1
Conditional DETR-R50 [16]	12 / 36	44	90	33.1 / 40.2	53.0 / 61.0	34.8 / 42.4	14.5 / 19.9	35.9 / 43.5	49.2 / 58.8
Anchor DETR-R50 [25]	12 / 36	37	85	33.7 / 39.7	54.5 / 60.5	35.1 / 41.9	15.6 / 19.9	37.3 / 43.5	49.8 / 57.3
DAB-DETR-R50 [13]	12 / 36	44	92	34.9 / 41.0	55.5 / 61.7	36.4 / 43.4	16.2 / 21.3	38.4 / 44.7	51.5 / 58.9
SAM-DETR-w/SMCA-R50 [26]	12 / 36	59	105	36.2 / 40.9	57.2 / 62.2	37.4 / 43.1	16.1 / 20.1	39.8 / 44.7	55.3 / 60.7
SAP-DETR-R50 (Ours)	12 / 36	47	94	**37.5 / 42.2**	**58.5 / 62.7**	**39.2 / 44.6**	**17.3 / 22.6**	**40.6 / 45.7**	**55.4 / 60.8**

Table 1. Comparison between Transformer necks. Based on ResNet-50 backbone, all models are trained by the official source codes with their original settings and evaluated on COCO `val2017`. All models uses 400 queries except Anchor DETR, while Anchor DETR uses 200 queries with 2 pattern embeddings. GFLOPs and Params are measured by Detectron2[1].

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Experimental Results

Method	#Epochs	#Params(M)	GFLOPs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	Infer. Time(s/img)†
ResNet-50 Backbone										
Faster RCNN-FPN-R50 [10, 18]	108	42	180	42.0	62.1	45.5	26.6	45.5	53.4	0.039
DETR-R50 [2]	500	41	86	42.0	62.4	44.2	20.5	45.8	61.1	0.040
Deformable DETR-R50 [31]	50	34	78	39.4	59.6	42.3	20.6	43.0	55.5	0.043
SMCA-DETR-R50 [5]	50	42	86	41.0	-	-	21.9	44.3	59.1	0.045
Conditional DETR-R50 [16]	50	44	90	40.9	61.8	43.3	20.8	44.6	59.2	0.057
Anchor DETR-R50 [25]	50	39	85	42.1	63.1	44.9	22.3	46.2	60.0	0.050
DAB-DETR-R50 [13]	50	44	90†	42.2	63.1	44.7	21.5	45.7	60.3	0.059
SAM-DETR-w/SMCA-R50 [26]	50	58	100	41.8	63.2	43.9	22.1	45.9	60.9	0.065
SAP-DETR-R50 (**Ours**)	50	47	92	**43.1**	**63.8**	**45.4**	**22.9**	**47.1**	**62.1**	0.063
ResNet-101 Backbone										
Faster RCNN-FPN-R101 [10, 18]	108	60	246	44.0	63.9	47.8	27.2	48.1	56.0	0.050
DETR-R101 [2]	500	60	152	43.5	63.8	46.4	21.9	48.0	61.8	0.066
Conditional DETR-R101 [16]	50	63	156	42.8	63.7	46.0	21.7	46.6	60.9	0.070
Anchor DETR-R101 [25]	50	58	150	43.5	64.3	46.6	23.2	47.7	61.4	0.068
DAB-DETR-R101 [13]	50	63	157†	43.5	63.9	46.6	23.6	47.3	61.5	0.072
SAP-DETR-R101 (**Ours**)	50	67	158	**44.4**	**64.9**	**47.1**	**24.1**	**48.7**	**63.1**	0.078
DC5-ResNet-50 Backbone										
DETR-DC5-R50 [2]	500	41	187	43.3	63.1	45.9	22.5	47.3	61.1	0.087
Conditional DETR-DC5-R50 [16]	50	44	195	43.8	64.4	46.7	24.0	47.6	60.7	0.093
Anchor DETR-DC5-R50 [25]	50	39	151	44.2	64.7	47.5	24.7	48.2	60.6	0.069
DAB-DETR-DC5-R50 [13]	50	44	194†	44.5	65.1	47.7	25.3	48.2	62.3	0.094
SAM-DETR-w/SMCA-DC5-R50 [26]	50	58	210	45.0	65.4	47.9	26.2	49.0	**63.3**	0.126
SAP-DETR-DC5-R50 (**Ours**)	50	47	197	**46.0**	**65.5**	**48.9**	**26.4**	**50.2**	62.6	0.116
DC5-ResNet-101 Backbone										
DETR-DC5-R101 [2]	500	60	253	44.9	64.7	47.7	23.7	49.5	62.3	0.101
Conditional DETR-DC5-R101 [16]	50	63	262	45.0	65.6	48.4	26.1	48.9	62.8	0.105
Anchor DETR-DC5-R101 [25]	50	58	227	45.1	65.7	48.8	25.8	49.4	61.6	0.083
DAB-DETR-DC5-R101 [13]	50	63	263†	45.8	65.9	49.3	27.0	49.8	63.8	0.110
SAP-DETR-DC5-R101 (**Ours**)	50	67	266	**46.9**	**66.7**	**50.5**	**27.9**	**51.3**	**64.3**	0.130

Table 2. Comparison of Transformer necks with 300 queries on COCO `val2017`. All results are reported from their original paper. All models uses 300 queries except Anchor DETR, while Anchor DETR uses 100 queries with 3 pattern embeddings. All inference speeds are measured by a single Nvidia A100 GPU. † denotes the results are measured by ourselves.

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Experimental Results

Backbone	Epoch	w/ SAP	DN-DETR [9]		DINO (Single-Scale) [28]		Group DETR [4]	
			AP / AP_{50} / AP_{75}	AP_S / AP_M / AP_L	AP / AP_{50} / AP_{75}	AP_S / AP_M / AP_L	AP / AP_{50} / AP_{75}	AP_S / AP_M / AP_L
R50	12		38.3 / 58.6 / 40.5	18.4 / 41.6 / 57.1	39.7 / 58.3 / 42.4	19.1 / 43.7 / 57.1	39.1 / - / -	19.7 / 42.5 / 56.8
		✓(Ours)	**39.5 / 59.7 / 41.5**	**18.7 / 42.8 / 59.0**	**40.0 / 60.1** / 42.1	**20.2** / 43.4 / **58.5**	**39.8 / 60.2 / 42.0**	**20.2 / 43.5 / 58.6**
R101	12		40.5 / 60.8 / 43.0	19.3 / 44.3 / 59.6	41.9 / 60.8 / 44.4	22.5 / 46.3 / 59.5	- / - / -	- / - / -
		✓(Ours)	**41.0 / 61.2 / 43.4**	**19.8 / 45.3 / 60.0**	41.5 / **61.4** / 43.6	20.3 / 45.2 / **60.0**	**41.1 / 61.5 / 43.4**	**20.5 / 45.5 / 59.4**
R50-DC	12		41.7 / 61.4 / 44.1	21.2 / 45.0 / 60.2	43.6 / 61.4 / 47.0	24.8 / 47.3 / 59.5	41.9 / - / -	23.3 / 45.6 / 58.4
		✓(Ours)	**43.6 / 62.5 / 46.2**	**23.3 / 47.3 / 61.0**	**44.0 / 63.1** / 46.5	**24.8 / 47.3 / 61.1**	**43.9 / 63.2 / 46.8**	**24.5 / 47.6 / 61.3**
R101-DC	12		43.4 / 61.9 / 47.2	24.8 / 46.8 / 59.4	45.4 / 63.5 / 49.2	26.4 / 49.5 / 61.1	- / - / -	- / - / -
		✓(Ours)	**44.6 / 63.9 / 48.0**	**25.5 / 48.9 / 62.5**	**45.6 / 64.5** / 48.7	25.0 / **49.7 / 62.5**	**44.4 / 63.9 /47.4**	**25.9 / 48.5 / 61.4**

Table 3. Comparison with denoised methods on COCO dataset based on the 12-epoch training schedule and 300 object queries.

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Ablation Study

Comment	Movable	Inner Loss	PECA	SDG	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
SAP-DETR (**Ours**)	✓	✓	✓	✓	**36.2**	56.2	**37.9**	**16.4**	**39.5**	**53.8**
−SDG	✓	✓	✓		35.6	56.2	36.9	16.3	38.9	52.7
−PECA	✓	✓		✓	34.8	55.5	36.0	15.7	37.3	52.0
−PECA & SDG	✓	✓			34.0	54.9	35.3	15.0	36.7	51.5
−Movable		✓	✓	✓	35.2	55.4	36.8	15.8	38.5	**53.8**
−Inner Loss	✓		✓	✓	35.9	**56.3**	37.4	16.2	39.3	52.5
DAB-DETR (Baseline)	-	-	-	-	32.3	51.3	34.0	15.7	35.2	45.7
+Salient Point Concept	-	-	-	-	33.5	54.3	35.1	14.3	36.5	51.0

Table 4. Ablation on each components

Inner Cost ($\mathcal{L}_{inner}$)	Movable within Grid (s_{grid})	AP	AP_S	AP_M	AP_L
		35.9	**17.0**	38.8	52.7
✓		26.3	11.3	28.0	39.5
✓	✓	**36.2**	16.4	**39.5**	**53.8**

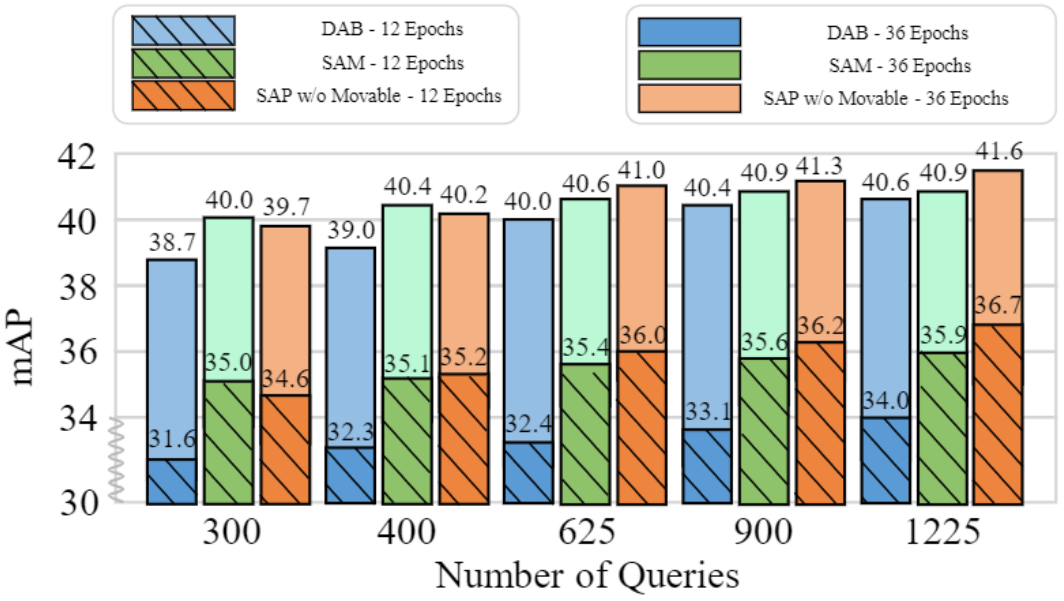
Table 5. Ablation on scaling factor of grid

PECA	Scaling Factor of SDG	AP	AP_S	AP_M	AP_L
		33.6	14.7	36.0	50.7
✓		35.7	**17.5**	38.8	52.6
✓	✓	**36.2**	16.4	**39.5**	**53.8**

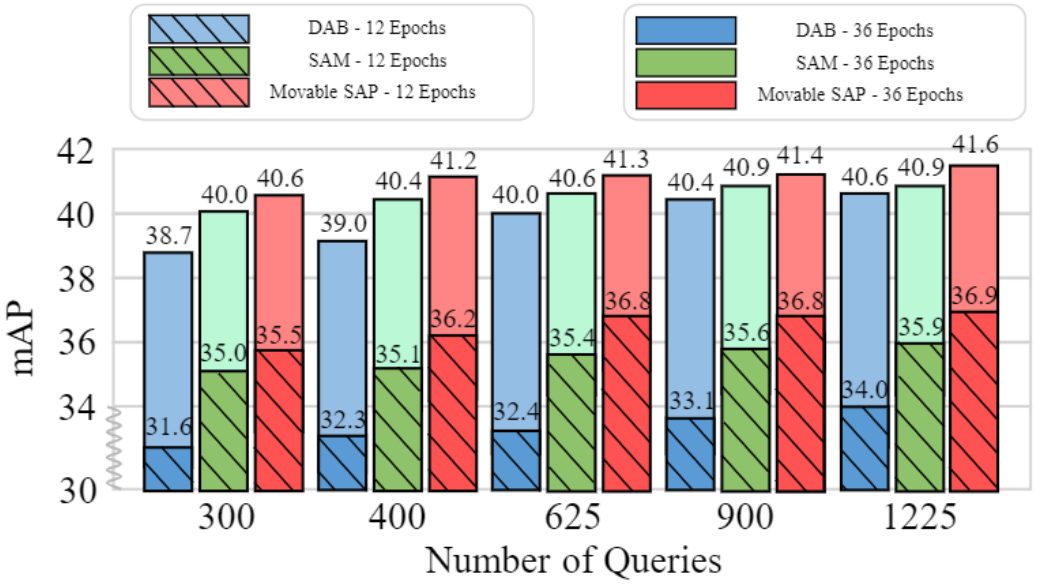
Table 6. Ablation on scaling factor of SDG

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Ablation Study



(a) w/o Movable Reference Point



(b) Movable Reference Point

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Visualization

Figure 10. Movable point update for COCO validation image #3255.

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Visualization

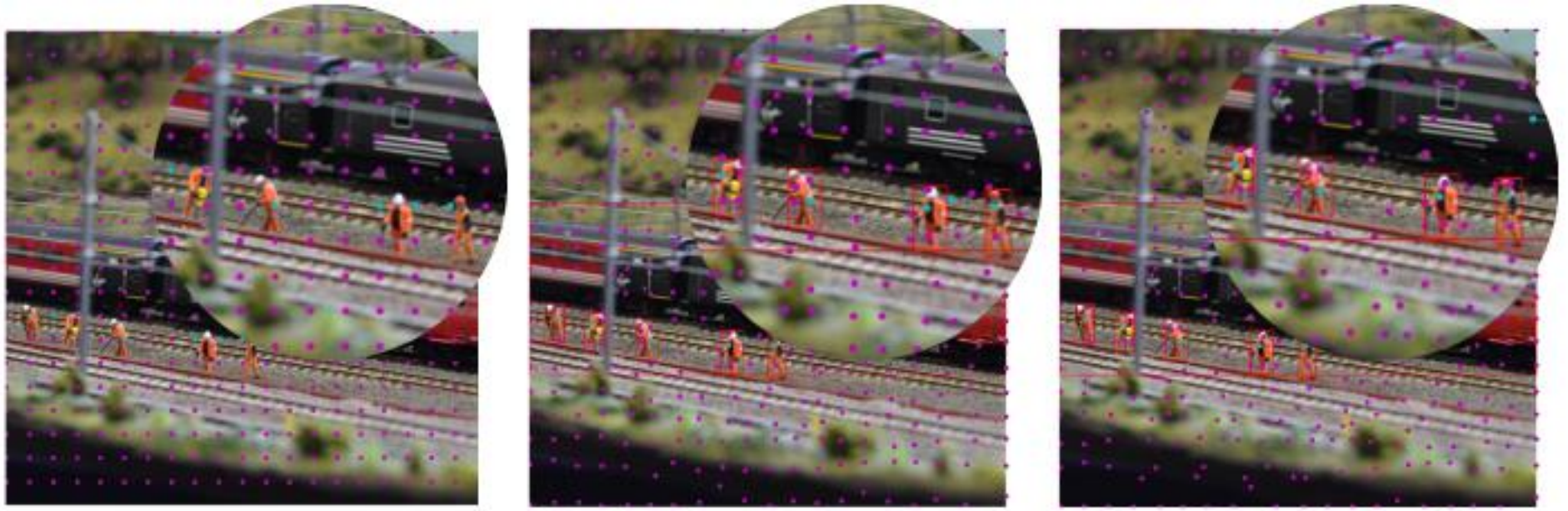


Figure 11. Movable point update for COCO validation image #14473.

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Visualization

(a) Ground Truth of COCO validation image #785

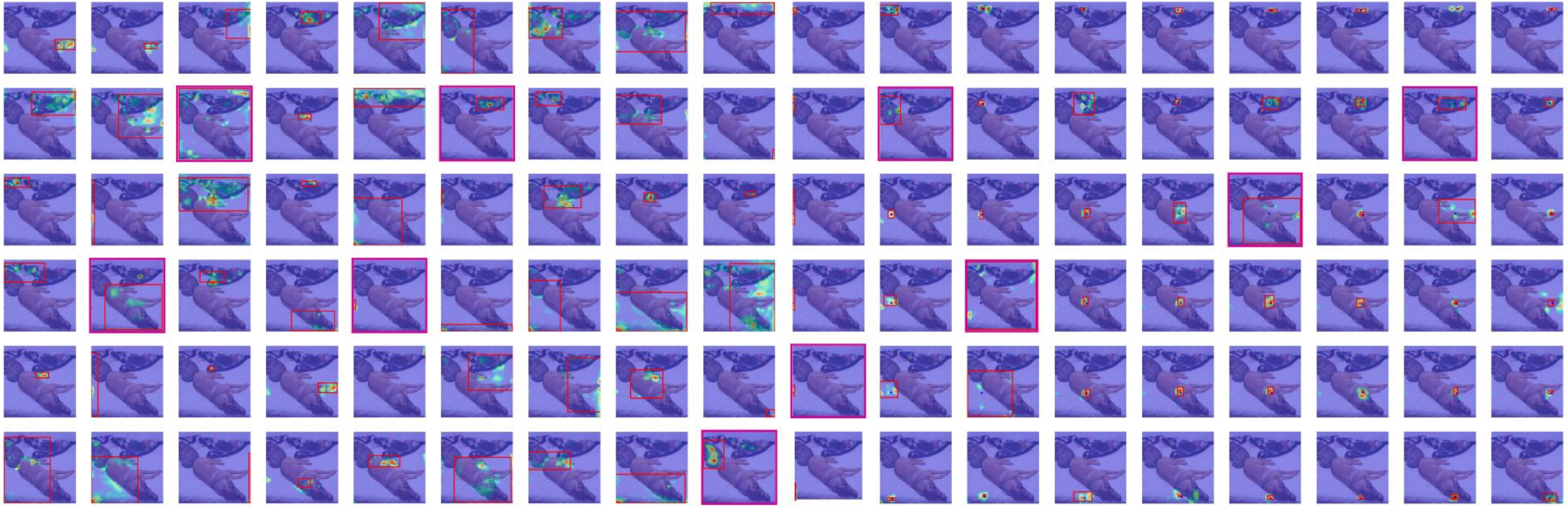
(b) DAB-DETR for COCO validation image #785

(c) SAP-DETR for COCO validation image #785

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

Visualization

(a) Ground Truth of COCO validation image #71226



(b) DAB-DETR for COCO validation image #71226

(c) SAP-DETR for COCO validation image #71226

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR

THANK YOU

DAKUJEM DANK BEDANKT MERCI תודה TAKK 谢谢
ありがとう СПАСИБО GRACIAS DZIĘKUJĘ DANKE
OBRIGADO БЛАГОДАРЯ GRAZIE धन्यवाद GRACIAS

中国科学院大学
University of Chinese Academy of Sciences

liuyang20c@mails.ucas.ac.cn
https://github.com/liuyang-ict/SAP-DETR