

Efficient DERT

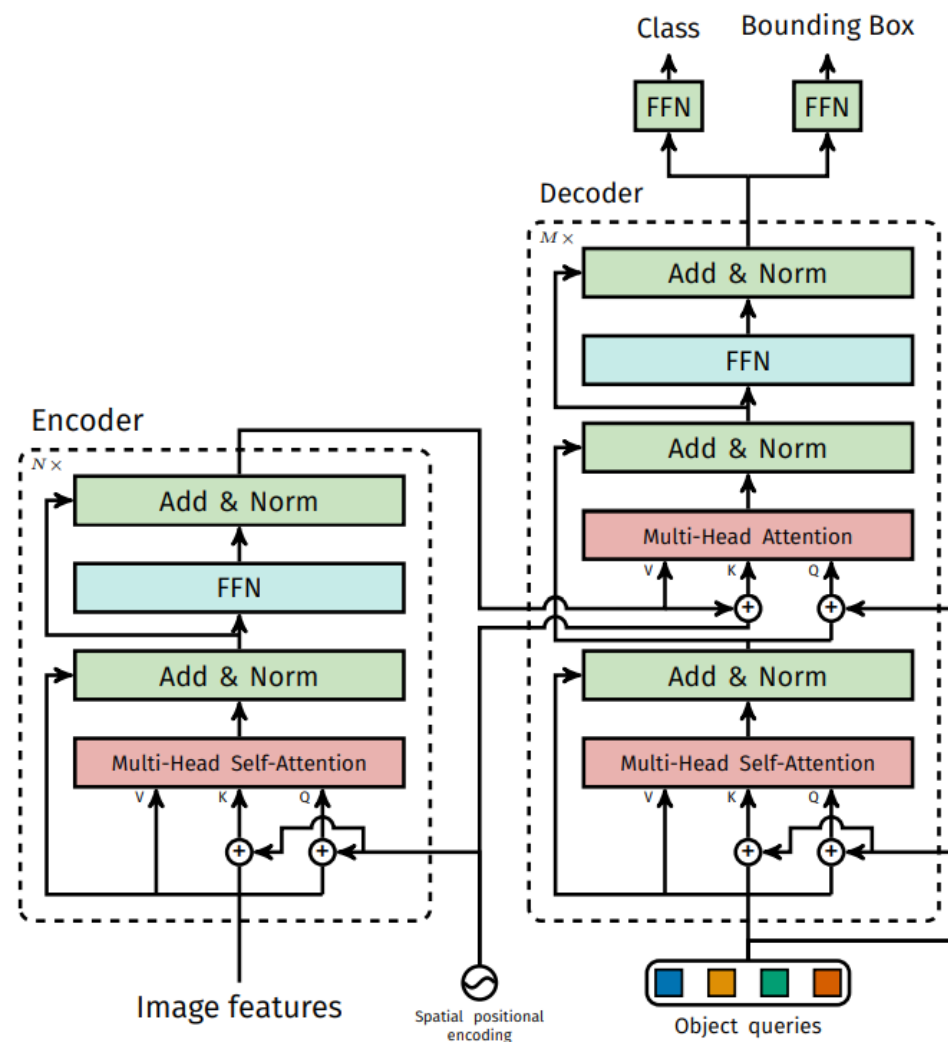
Paper Details

- Paper Title: Efficient DETR: Improving End-to-End Object Detector with Dense Prior
- Publication Date: 3 Apr 2021
- Publisher: Zhuyu Yao, Jiangbo Ai, Boxun Li, Chi Zhang
- Affiliation: Megvii Technology

Motivation

- both DETR and Deformable DETR have a 6-encoder and 6-decoder transformer architecture
- investigate the components of DETRs for understanding their mechanism
- find out random initialization can be improved
- Improve DERT with better initialization

Brief Retrospect of DETR



Exploring DETR

- Encoder and Decoder
- Initialization of object containers

Exploring DETR

- Encoder and Decoder

Encoder	Decoder	Decoding loss	AP	AP ₅₀	AP ₇₅
3	3	✓	41.5	60.4	44.3
3	1	✓	32.2	51.1	33.7
1	3	✓	39.8	58.2	42.5
1	3		28.3	47.4	29.3

Difference between Encoder and Decoder: Auxiliary loss

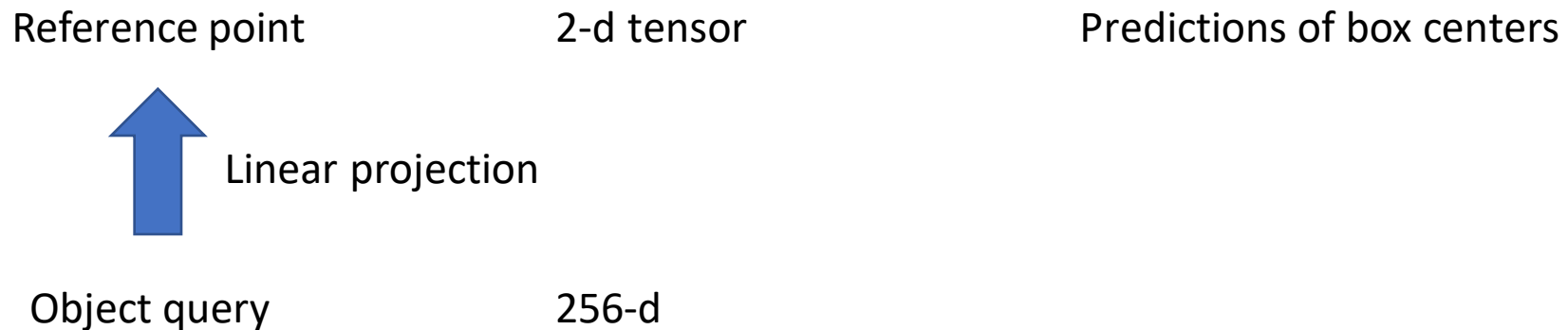
Exploring DETR

- Encoder and Decoder
- Decoder

Decoder layer	1	2	3	4	5	6
AP	32.1	39.4	41.7	42.4	42.7	42.4
AP ₅₀	50.8	58.4	60.6	61.5	61.8	61.8
AP ₇₅	34.0	41.9	44.4	45.5	45.9	45.6

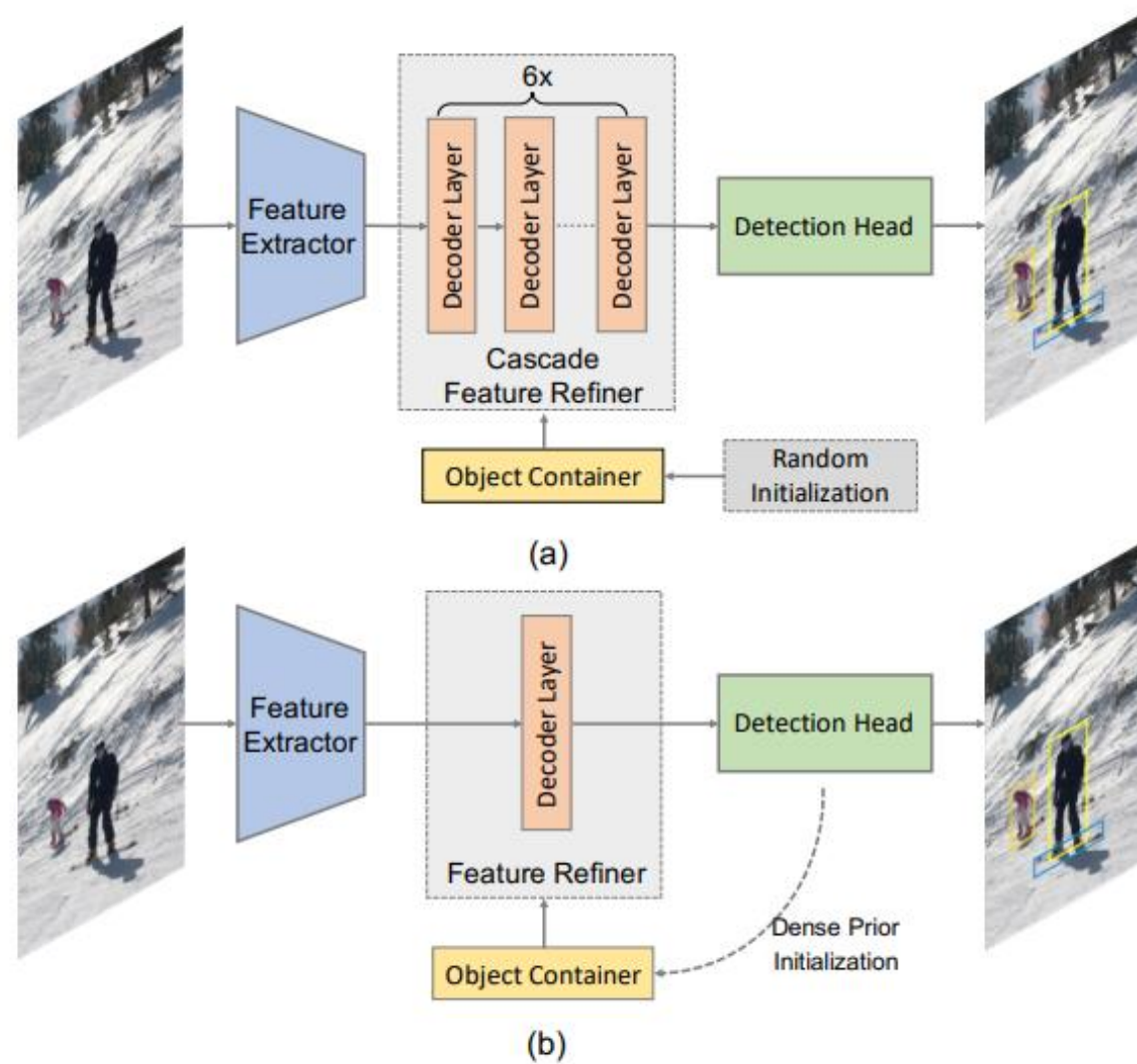
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



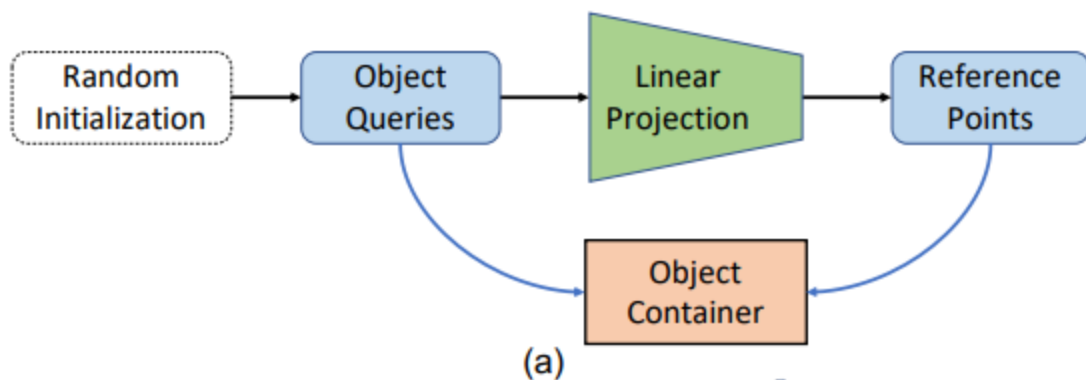
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



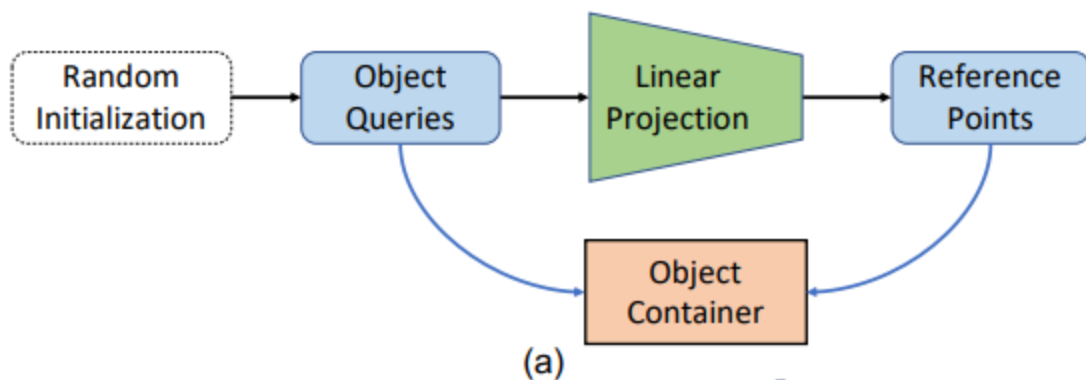
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



Exploring DETR

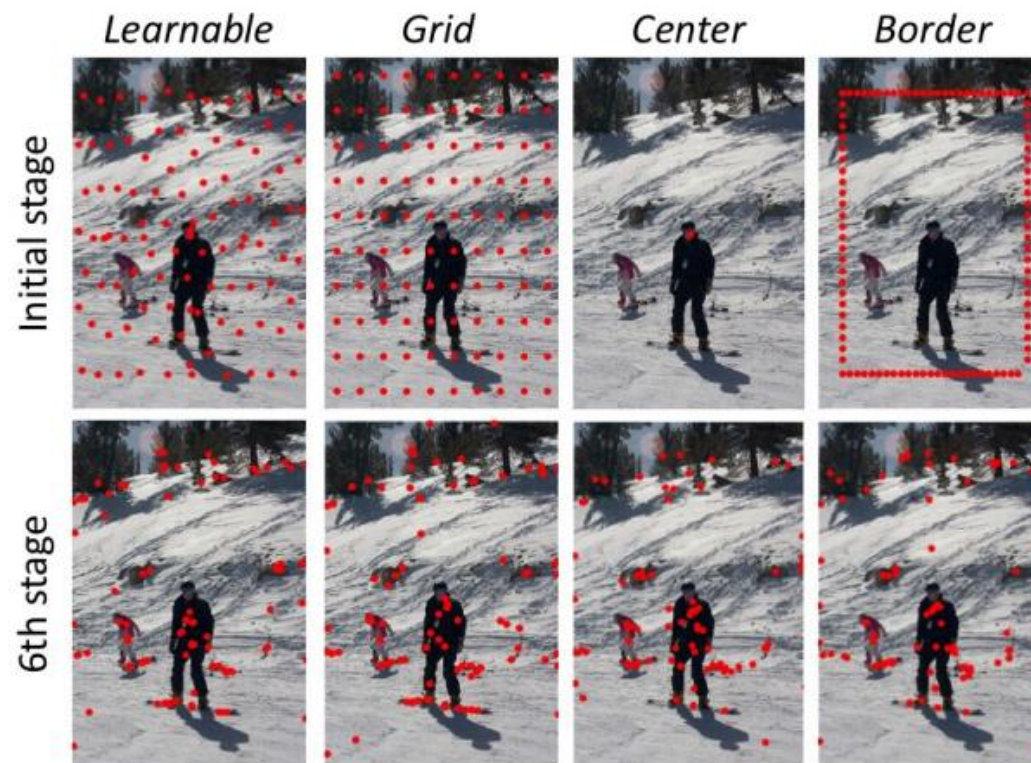
- Encoder and Decoder
- Decoder
- Initialization of object queries



Exploring DETR

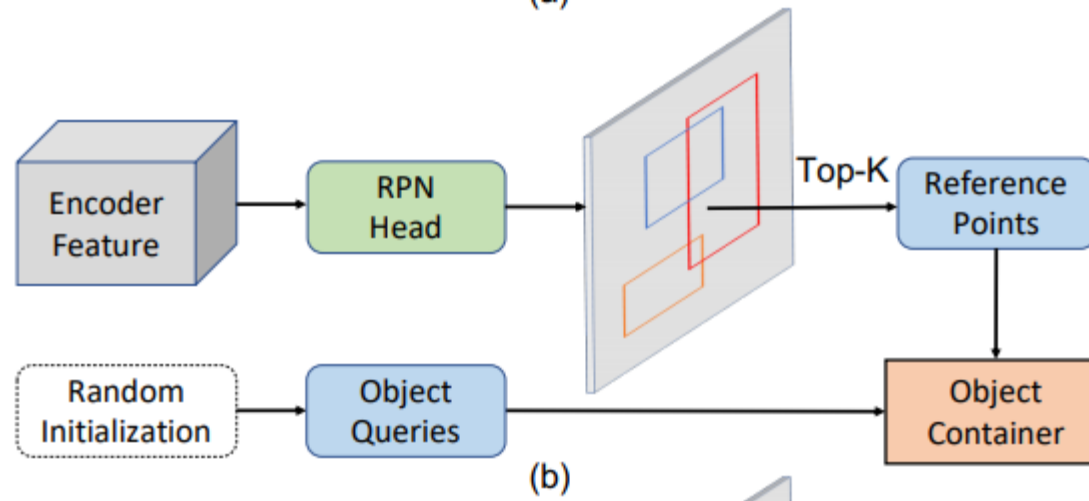
- Encoder and Decoder
- Decoder
- Initialization of object queries

Initialization	Learnable	Grid	Center	Border	Dense
$AP_{1-dec.}$	32.1	32.0	21.0	26.0	39.0
$AP_{6-dec.}$	42.4	42.7	42.8	42.8	-



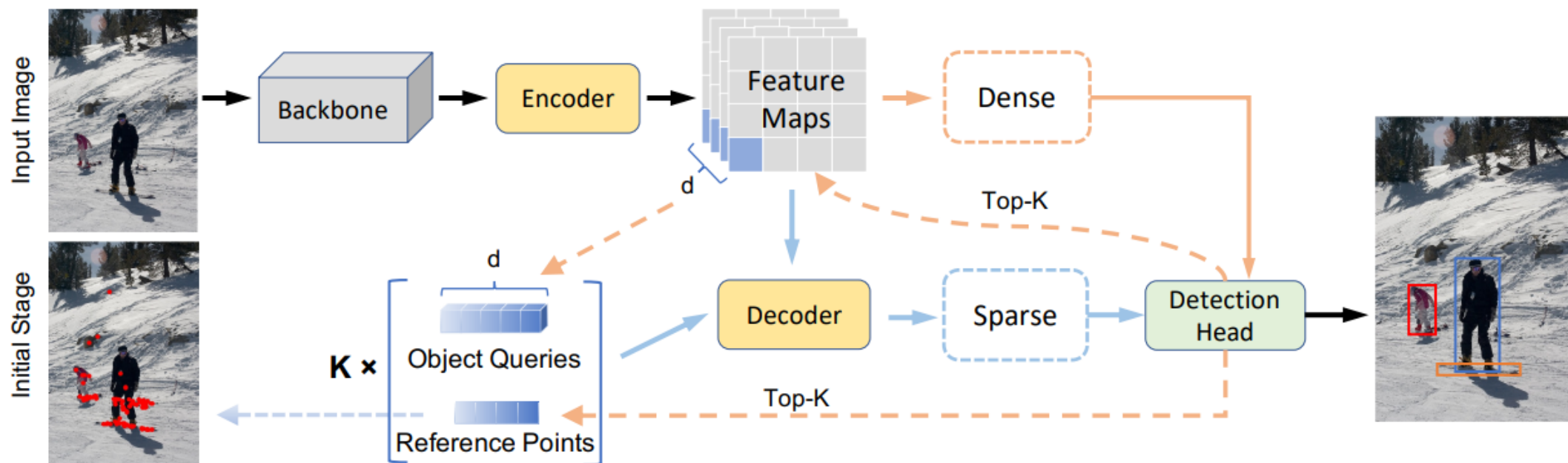
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



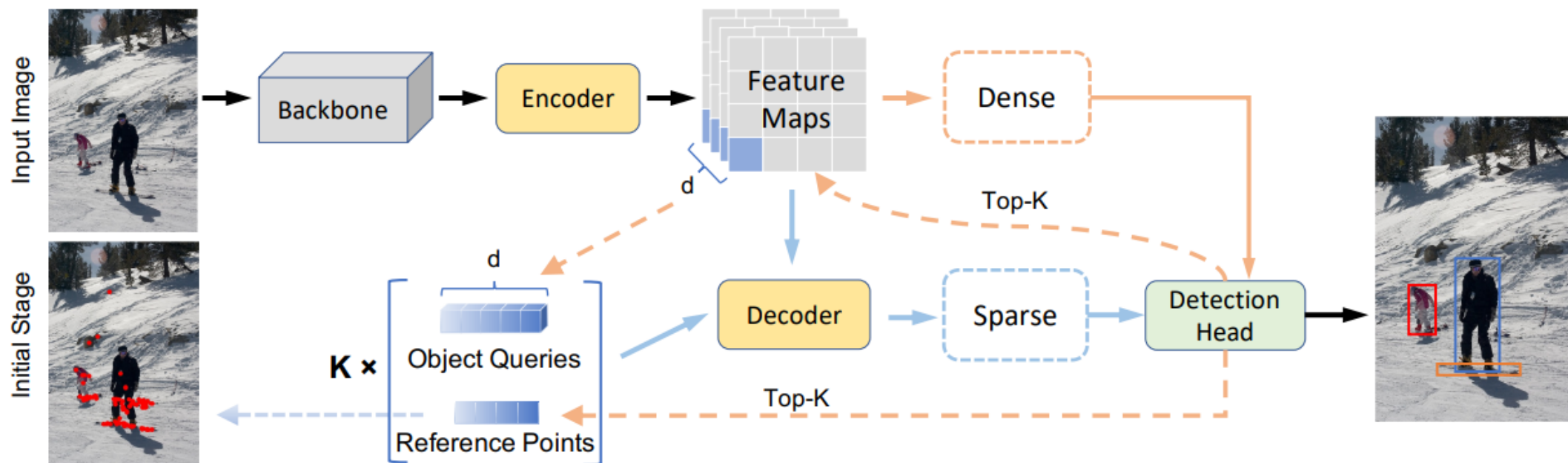
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



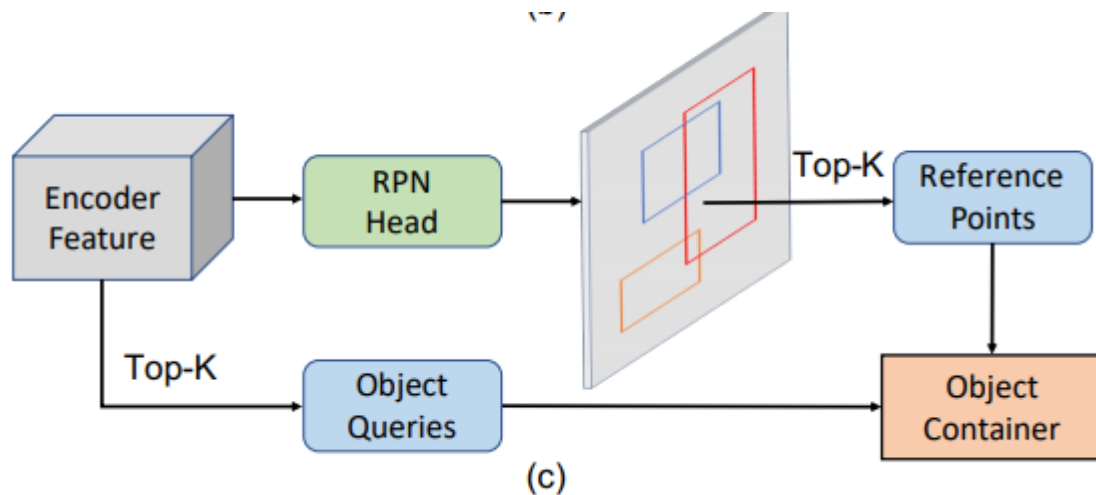
Exploring DETR

- Encoder and Decoder
- Decoder
- Initialization of object queries



Exploring DETR

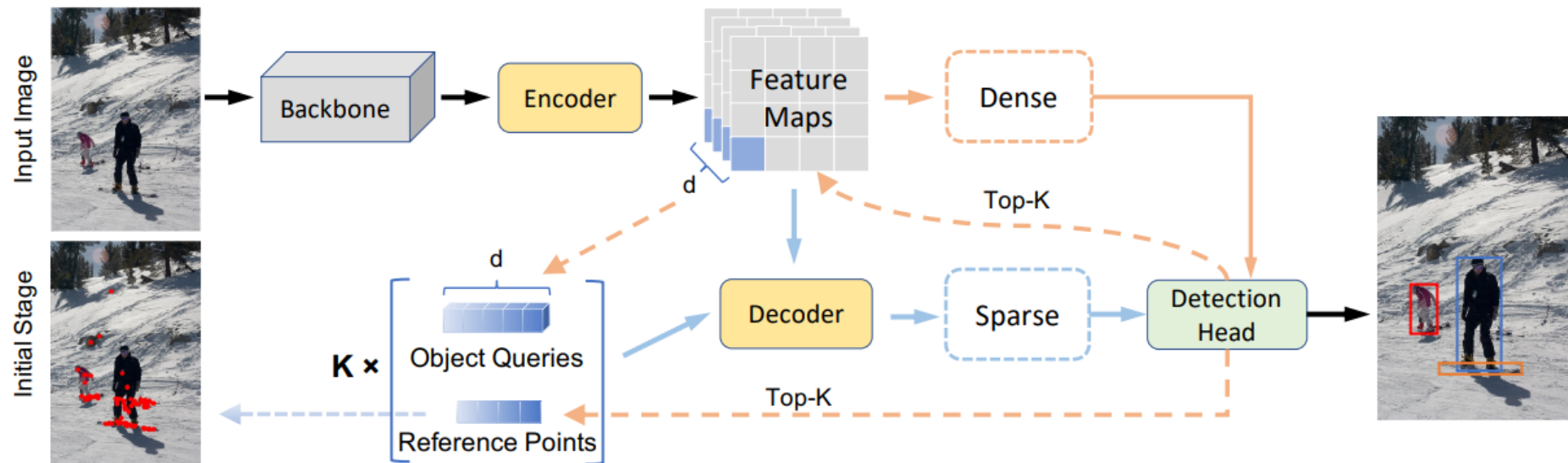
- Encoder and Decoder
- Decoder
- Initialization of object queries



2-d Ref.	4-d Ref.	Object query	AP	AP ₅₀	AP ₇₅
			32.1	50.1	34.0
✓			39.0	57.0	42.7
		✓	41.1	60.8	45.0
✓		✓	42.0	60.5	45.6
	✓	✓	43.0	60.9	46.5

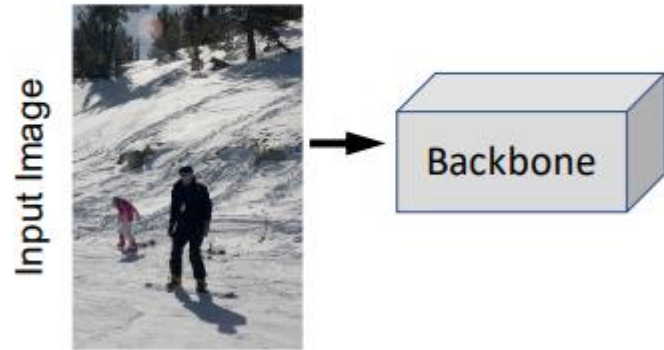
Table 4. Initializing reference point and object query with dense prior. 2-d Ref. denotes reference points which are 2-d coordinates and initialized with the center points of region proposals. 4-d Ref. denotes reference points which are 4-d bounding boxes and directly initialized with region proposals. Object queries are initialized with 256-d proposal feature vectors from encoder feature maps.

Efficient DETR Architecture



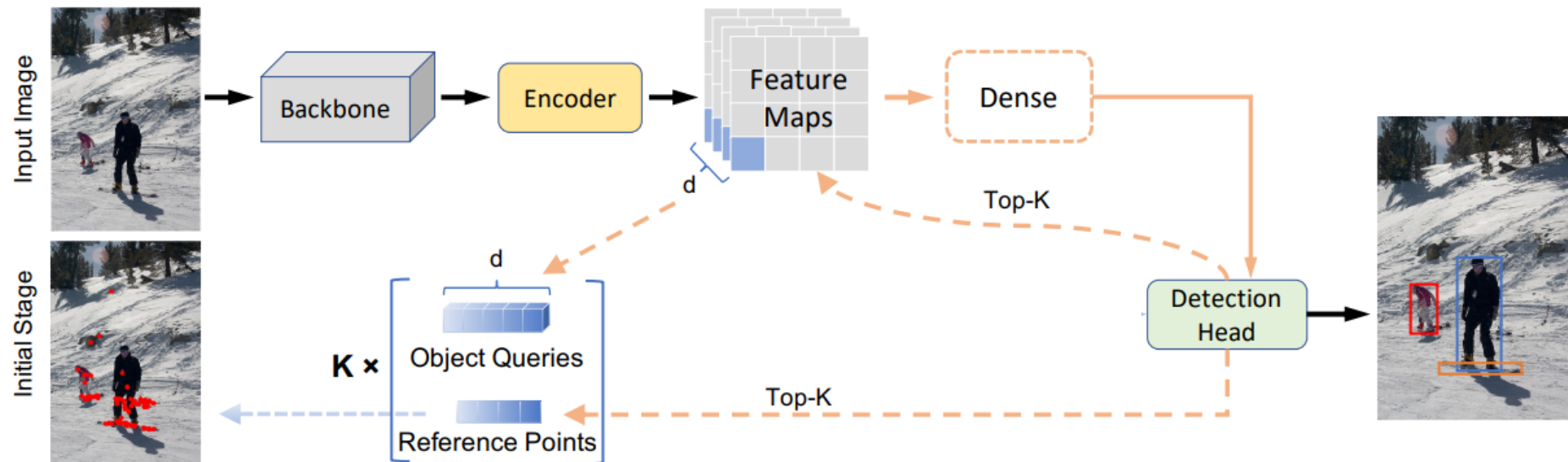
Efficient DETR Architecture

- Backbone



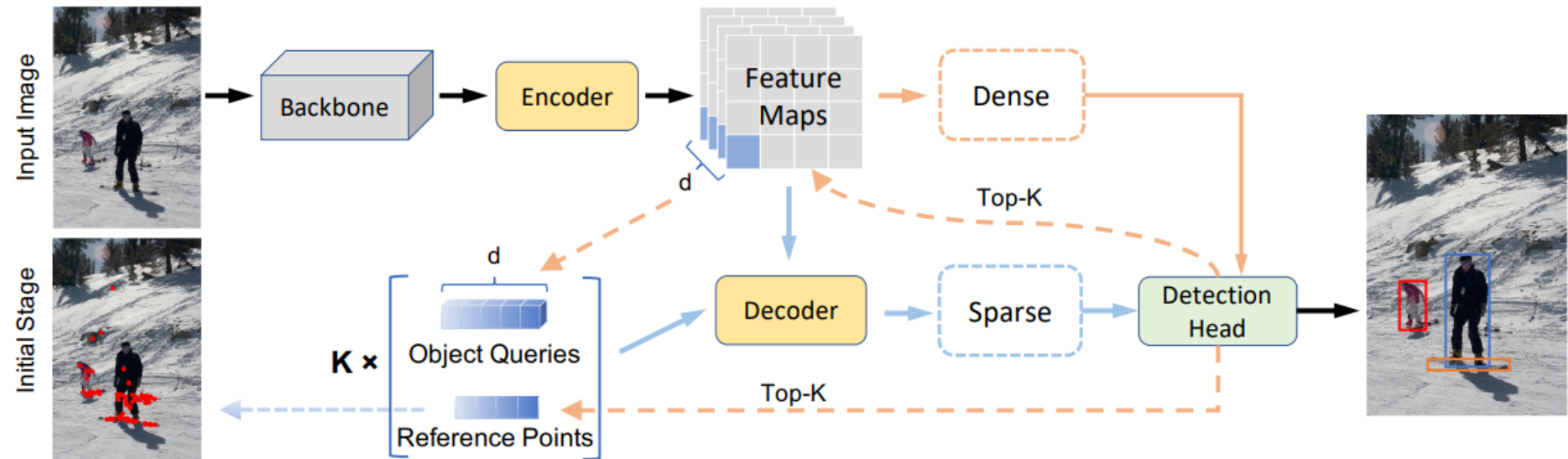
Efficient DETR Architecture

- Backbone
- Dense part



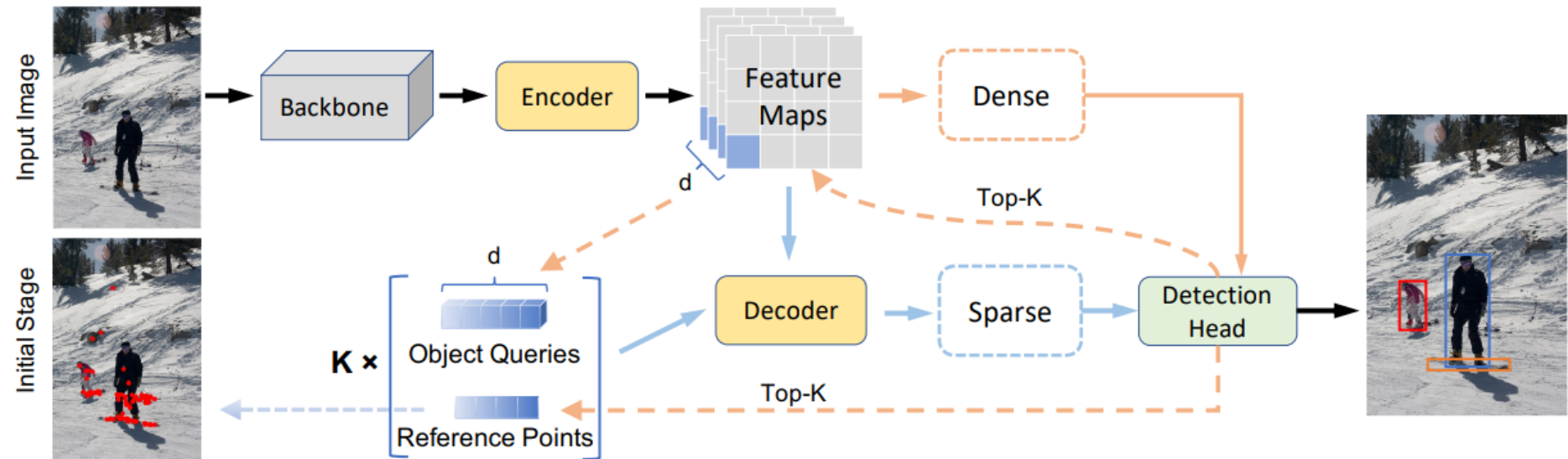
Efficient DETR Architecture

- Backbone
- Dense part
- Sparse part



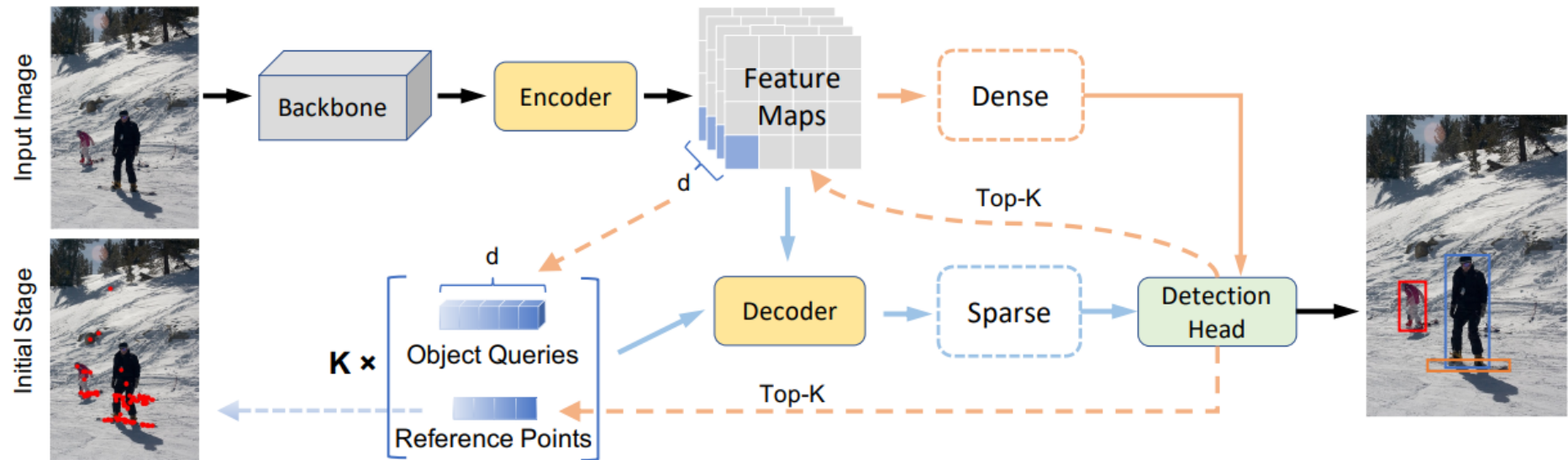
Efficient DETR Architecture

- Backbone
- Dense part
- Sparse part



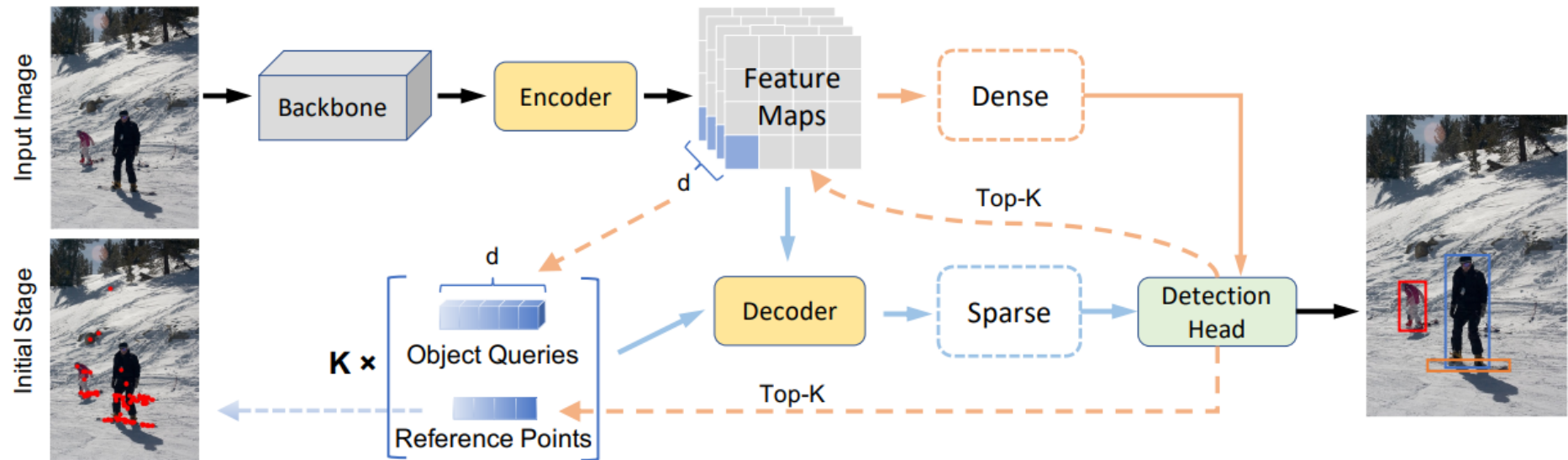
Efficient DETR Architecture

- Backbone
- Dense part
- Sparse part
- Proposals



Efficient DETR Architecture

- Backbone
- Dense part
- Sparse part
- Proposals
- Loss



$$\mathcal{L} = \lambda_{cls} \cdot \mathcal{L}_{cls} + \lambda_{L1} \cdot \mathcal{L}_{L1} + \lambda_{giou} \cdot \mathcal{L}_{giou}$$

Efficient DETR Results

Model	Epochs	GFLOPs	Params (M)	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
DETR-R50 [2]	500	86	41	42.0	62.4	44.2	20.5	45.8	61.1
DETR-DC5-R50 [2]	500	187	41	43.3	63.1	45.9	22.5	47.3	61.1
Faster RCNN-FPN-R50 [2]	36	180	42	40.2	61.0	43.8	24.2	43.5	52.0
Deformable DETR-R50 [43]	50	173	40	43.8	62.6	47.7	26.4	47.1	58.0
TSP-FCOS-R50 [32]	36	189	-	43.1	62.3	47.0	26.6	46.8	55.9
TSP-RCNN-R50 [32]	36	188	-	43.8	63.3	48.3	28.6	46.9	55.7
SMCA-R50 [8]	50	152	40	43.7	63.6	47.2	24.2	47.0	60.4
Sparse R-CNN-R50 [31]	36	-	-	44.5	63.4	48.2	26.9	47.2	59.5
Efficient DETR-R50	36	159	32	44.2	62.2	48.0	28.4	47.5	56.6
Efficient DETR*-R50	36	210	35	45.1	63.1	49.1	28.3	48.4	59.0
DETR-R101 [2]	500	152	60	43.5	63.8	46.4	21.9	48.0	61.8
DETR-DC5-R101 [2]	500	253	60	44.9	64.7	47.7	23.7	49.5	62.3
Faster RCNN-FPN-R101 [2]	36	256	60	42.0	62.1	45.5	26.6	45.4	53.4
TSP-FCOS-R101 [32]	36	255	-	44.4	63.8	48.2	27.7	48.6	57.3
TSP-RCNN-R101 [32]	36	254	-	44.8	63.8	49.2	29.0	47.9	57.1
SMCA-R101 [8]	50	218	58	44.4	65.2	48.0	24.3	48.5	61.0
Sparse R-CNN-R101 [31]	36	-	-	45.6	64.6	49.5	28.3	48.3	61.6
Efficient DETR-R101	36	239	51	45.2	63.7	48.8	28.8	49.1	59.0
Efficient DETR*-R101	36	289	54	45.7	64.1	49.5	28.2	49.1	60.2

Efficient DETR Results

Proposals		100	300	500	1000
Fixed		43.8	44.2	44.1	44.0
Linear Decrease		-	44.2	44.0	44.0
Agnostic	Specific	Share Head	AP	AP ₅₀	AP ₇₅
✓			43.0	60.9	46.5
✓	✓		43.8	62.0	47.5
	✓		43.8	61.9	47.4
	✓	✓	43.8	61.9	47.4

More Experiments

- CrowdHuman Data (Crowded Scenes)

Method	Proposals	AP ₅₀	mMR	Recall
Faster-RCNN [27]	-	85.0	50.4	90.24
RetinaNet [20]	-	81.7	57.6	88.6
FCOS [33]	-	86.1	55.2	94.3
ATSS [37]	-	87.1	50.1	94.0
POTO+3DMF+Aux [36]	-	89.2	49.6	96.6
DETR [2]	400	66.12	80.62	-
Deformable DETR [43]	400	86.74	53.98	92.51
Efficient DETR	400	90.68	49.80	97.99
Efficient DETR	100	90.75	48.98	97.94