# Destr : Object Detection with Split Transformer

Liqiang He and Sinisa Todorovic

Oregon State University

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

**Motivation 1:**

The content of object queries is inferred for every image from scratch => slows training convergence.

How to improve initialization of the object representation in the decoder?

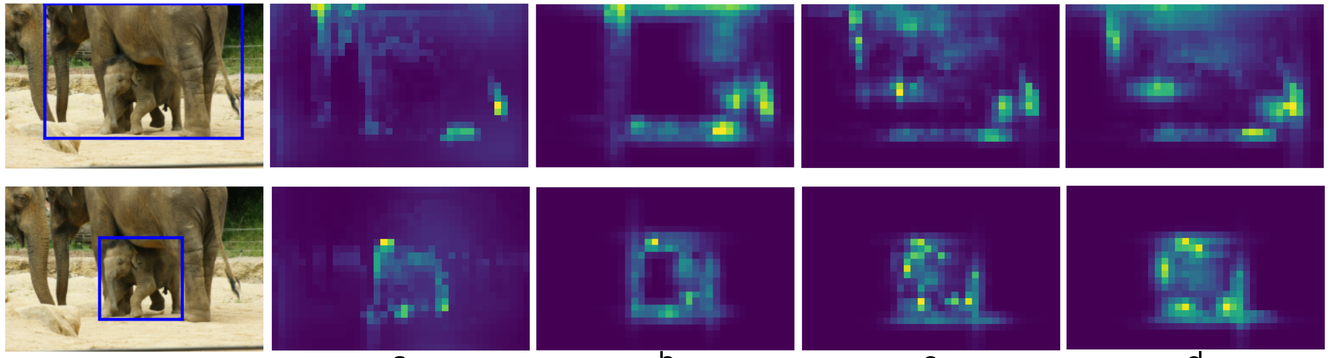
**Contribution 1:** Mini-detector is used to initialize the objects' classification, regression and positional embeddings.

**Motivation 2:**

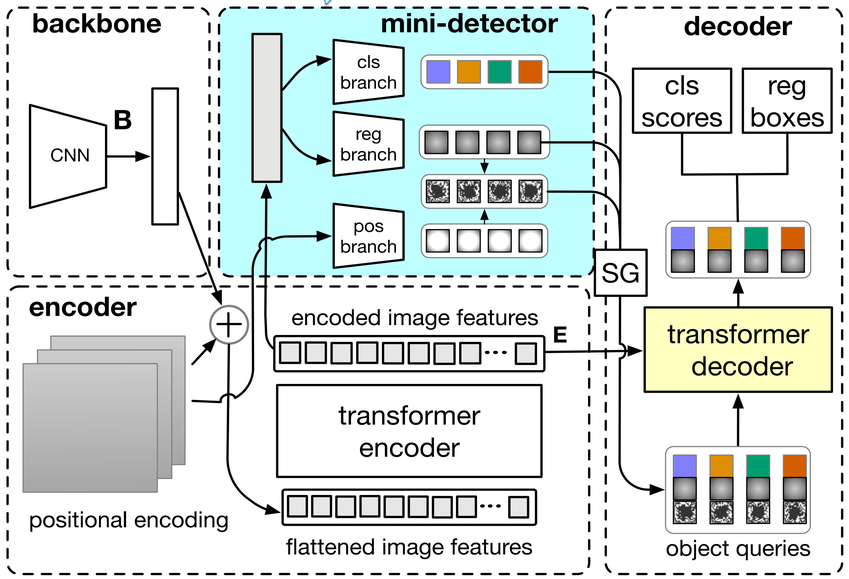
A single cross-attention is typically used for both classification and bounding-box regression.

These two tasks are different, and hence a single cross-attention may not be optimal for both.

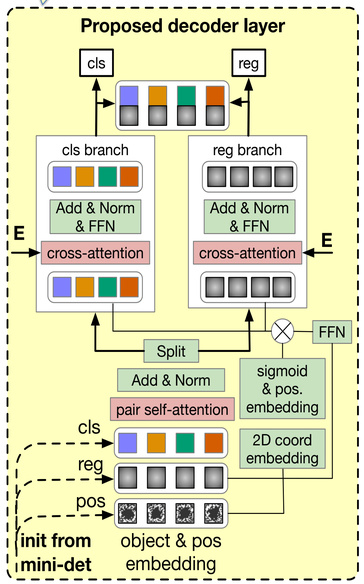
**Contribution 2:** Split the cross attention into two independent branches – one for classification and the other for box regression




Cross-attention maps estimated by: (a) DETER; (b) C-DETR; (c) DESTR's classification branch and (d) regression branch;




An overview of DESTR



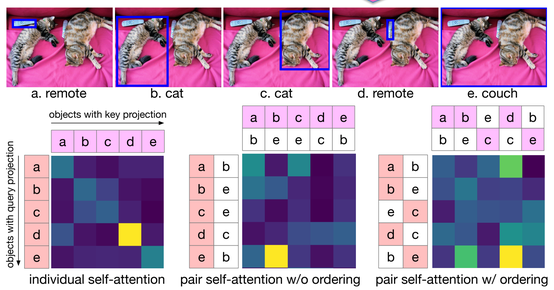

Decoder layer in DESTR

**Motivation 3:**

Partially occluded objects provide lower attention to the fully-visible objects.

Adjacent pairs of object queries may provide more important cues

**Contribution 3:** Combine two adjacent objects to form a "pair" to increase the attention.




Pair self-attention

| Model | #epochs | GFLOPs | #params (M) | AP |
|---|---|---|---|---|
| DETR-R50 [2] | 500 | 86 | 41 | 42.0 |
| Deform-DETR-R50-SS [30] | 50 | **78** | **34** | 39.4 |
| UP-DETR-R50 [4] | 150 | 86 | 41 | 40.5 |
| UP-DETR-R50 [4] | 300 | 86 | 41 | 42.8 |
| C-DETR-R50 [19] | 50 | 90 | 44 | 40.9 |
| Anchor DETR-R50 [25] | 50 | – | – | 42.1 |
| DESTR-R50 | 50 | 104 | 69 | **43.6** |
| DETR-DC5-R50 [2] | 500 | 187 | 41 | 43.3 |
| Deform-DETR-DC5-R50-SS [30] | 50 | **128** | **34** | 41.5 |
| C-DETR-DC5-R50 [19] | 50 | 195 | 44 | 43.8 |
| Anchor DETR-DC5-R50 [25] | 50 | 151 | – | 44.2 |
| DESTR-DC5-R50 | 50 | 232 | 69 | **45.3** |
| DETR-R101 [2] | 500 | **152** | **60** | 43.5 |
| C-DETR-R101 [19] | 50 | 156 | 63 | 42.8 |
| Anchor DETR-R101 [25] | 50 | – | – | 43.5 |
| DESTR-R101 | 50 | 171 | 88 | **44.6** |
| DETR-DC5-R101 [2] | 500 | **253** | **60** | 44.9 |
| C-DETR-DC5-R101 [19] | 50 | 262 | 63 | 45.0 |
| Anchor DETR-DC5-R101 [25] | 50 | – | – | 45.1 |
| DESTR-DC5-R101 | 50 | 299 | 88 | **46.4** |

Comparison with other DETR variants on COCO-val

Acknowledgement: DARPA MCS Award N66001-19-2-4035.