Destr: Object Detection with Split Transformer

Liqiang He and Sinisa Todorovic

Motivation 1:
The content of object queries is inferred for every image from scratch => slows training convergence.

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.

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

Motivation 3:
These two tasks are different, and hence a single cross-attention may not be optimal for both.

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

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