

Review of paper – Accelerating DETR Convergence via Semantic-Aligned Matching

Xuan Chen - Student #82286956

Paper link: <https://arxiv.org/abs/2203.06883>

Summary of the paper (3-4 sentences)

This paper presents SAM-DETR, a Semantic-Aligned-Matching DETR (DEtection TRansformer) that greatly accelerates DETR's convergence without sacrificing its accuracy, which is a good solution to slow convergence of DETR. Being like a plug and play, SAM-DETR complements existing convergence solutions well yet only introduces slight computational overhead. Extensive experiments show that the proposed SAM-DETR achieves superior convergence as well as competitive detection accuracy.

Main contributions (2-3 bullet points)

1. The paper proposes Semantic-Aligned-Matching DETR (SAM-DETR), which significantly accelerates DETR's convergence by innovatively interpreting its cross-attention as a 'matching and distillation' process and semantically aligning object queries with encoded image features to facilitate their matching.
2. It explicitly searches for objects' salient points with the most discriminative features and feed them to the cross-attention module for semantic aligned matching, which further boosts the detection accuracy and speeds up the convergence of our model.
3. The experiments validate that SAM-DETR achieves significantly faster convergence compared with the original DETR.
4. The approach only adds a plug-and-play module into the original DETR and leaves other operations mostly unchanged, the proposed SAM-DETR can be easily integrated with existing solutions that modify the attention mechanism to further improve DETR's convergence, leading to a comparable convergence speed with Faster R-CNN even within 12 training epochs.

Positive and negative points (2-3 points each)

1. Positive:
 - SAM-DETR projects object queries into the same embedding space as encoded image features, where the matching can be accomplished efficiently with aligned semantics.
 - SAM-DETR explicitly searches salient points with the most discriminative features for semantic-aligned matching, which further speeds up the convergence and boosts detection accuracy as well.
 - Without modifying the attention mechanism, SAM-DETR also eases the matching process between object queries and their corresponding target features.
2. Negative: Compared with Faster R-CNN, SAM-DETR inherits from DETR superior accuracy on large objects and degraded performance on small objects.

Unclear (2-3 points)

1. How can the model improve accuracy by leveraging multi-scale features?
2. Few discussion of predicted coordinates outside the reference boxes.