

DETRs With Collaborative Hybrid Assignments Training

What is the problem?

- Sparse Supervision: - One-to-one set matching assigns only one positive query per ground-truth box. - Resulting in limited positive samples for training.
- Impact on Learning: - Hinders encoder's ability to learn discriminative features. - Affects decoder's attention learning due to insufficient object information.

What has been done earlier?

- Traditional Object Detection Methods:
 - Rely on dense supervision techniques (e.g., Faster R-CNN).
- Attempts to Improve DETR:
 - Modifications to the matching process.
 - Integrating additional supervision signals.
- Limitations: - Many methods still struggle with the inherent sparsity of positive samples.

- What are the remaining challenges? What novel solution proposed by the authors to solve the problem?
- Remaining Challenges-
 1. Insufficient Positive Samples: - Limited queries lead to poor feature learning.
 2. Instability in Training: - Hungarian matching process can be unstable with few positive queries.
 3. Scalability: - Difficulty in scaling methods to larger datasets and complex scenarios.
- Co-DETR: Proposed Solution-

Collaborative Hybrid Assignments Training: - Introduces auxiliary heads with one-to-many label assignments.

Benefits: - Enriches supervision on the encoder. - Generates a feature pyramid for multi-scale operations. - Provides denser spatial supervision to enhance feature learning.