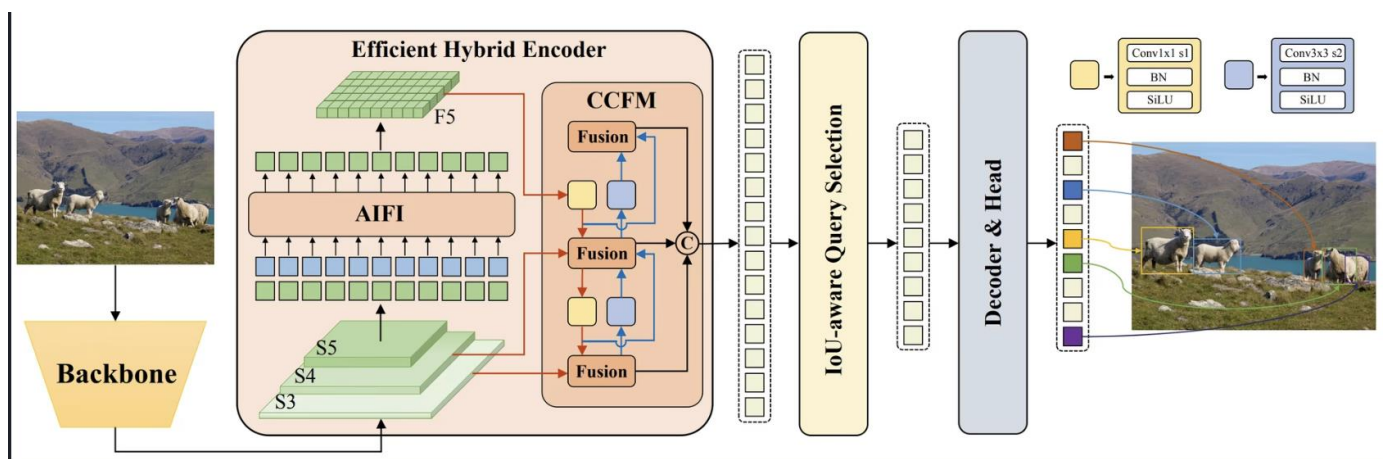


Model Architecture:

RTDETR (Real-Time DETection Transformer) is an evolution of DETR (DEtection Transformer), designed to address its predecessor's limitations. The key differences lie in RTDETR's focus on speed and efficiency. RTDETR incorporates multi-scale feature processing and an improved attention mechanism, enabling faster image processing and better handling of varied object sizes. It also uses a decoupled head design and introduces auxiliary losses, leading to quicker training convergence.

Unlike DETR, RTDETR simplifies post-processing by eliminating the need for complex non-maximum suppression. These improvements allow RTDETR to achieve real-time performance while maintaining or even surpassing DETR's accuracy, making it more suitable for practical, time-sensitive applications.



Implementation Details:

Backbone : RTDETRv2_r101vd_6x_coco

Feature Extractor : HybridEncoder

Detection Head : RTDETRTransformerv2

Loss Function :

Varifocal Focal Loss (VFL): Used for classification, weight = 1

L1 Loss : Used for bounding box regression, weight = 5

GloU Loss : Used for bounding box optimization, weight = 2

Focal Loss parameters : alpha = 0.75, gamma = 2.0

Hungarian matcher is used for target assignment, considering class cost, bounding box cost, and GloU cost.

Parameter Settings :

Hidden dimension: 256

Number of decoder layers: 6

Number of queries: 300

Number of denoising queries: 100

Optimizer Settings :

Optimizer: AdamW

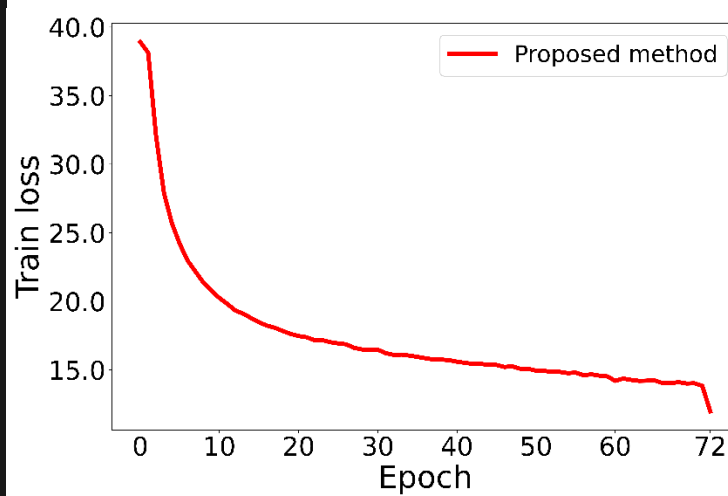
Base learning rate: 0.0001

Adam parameters: beta = [0.9, 0.999]

Weight decay: 0.0001

Performance :

Class	AP50	AP75
0	0.9255	0.8632
1	0.8221	0.6554
2	0.4704	0.3538
3	0.9016	0.8181
4	0.4706	0.3093
5	0.7227	0.5407
6	0.2103	0.1637
7	0.3893	0.2836
8	0.6925	0.5312
9	0.6465	0.5212
10	0.7577	0.7099
11	0.8556	0.7415
12	0.9169	0.8514
13	0.3525	0.2696
14	0.6531	0.4798
15	0.3732	0.3199
16	0.5625	0.4788
mAP(50-95): 0.7183		



Result :



training results show a steady decrease in loss over 72 epochs, indicating good learning progress. While AP50 scores are generally above 0.5, there's a notable performance disparity across classes, suggesting potential long-tail effects. This is evident in the stark contrast between high-performing classes (like 0 and 12) and underperforming ones (such as 14 and 15). The inconsistency between AP50 and AP75 scores for some classes further hints at challenges in precise detection of possibly rare objects.