

موضع

X

RT-DETR (v3)

Real-Time Detection Transformer

**RT-DETRv3: Real-time End-to-End Object Detection with Hierarchical Dense
Positive Supervision**

Shuo Wang* Chunlong Xia* Feng Lv Yifeng Shi†

Baidu Inc, China

{wangshuo36, xiachunlong, lvfeng02, shiyifeng}@baidu.com

مزایای شبکه RT-DETR



علی‌رغم سرعت بالا، دقت آن نزدیک به مدل‌های سنگین‌مانند YOLOv8 و Faster R-CNN است.



دقت بالا

سرعت بلادرنگ همراه با دقت بالا



سرعت بلادرنگ

بدون نیاز به NMS، بدون anchor boxes



سادگی معماری

نسبت به DETR اصلی، سریع‌تر convergence می‌کند و نیاز به دیتای کمتر برای یادگیری دارد.



یادگیری پایدار

مناسب برای deployment روی دستگاه‌های edge با استفاده از TensorRT یا DeepStream.



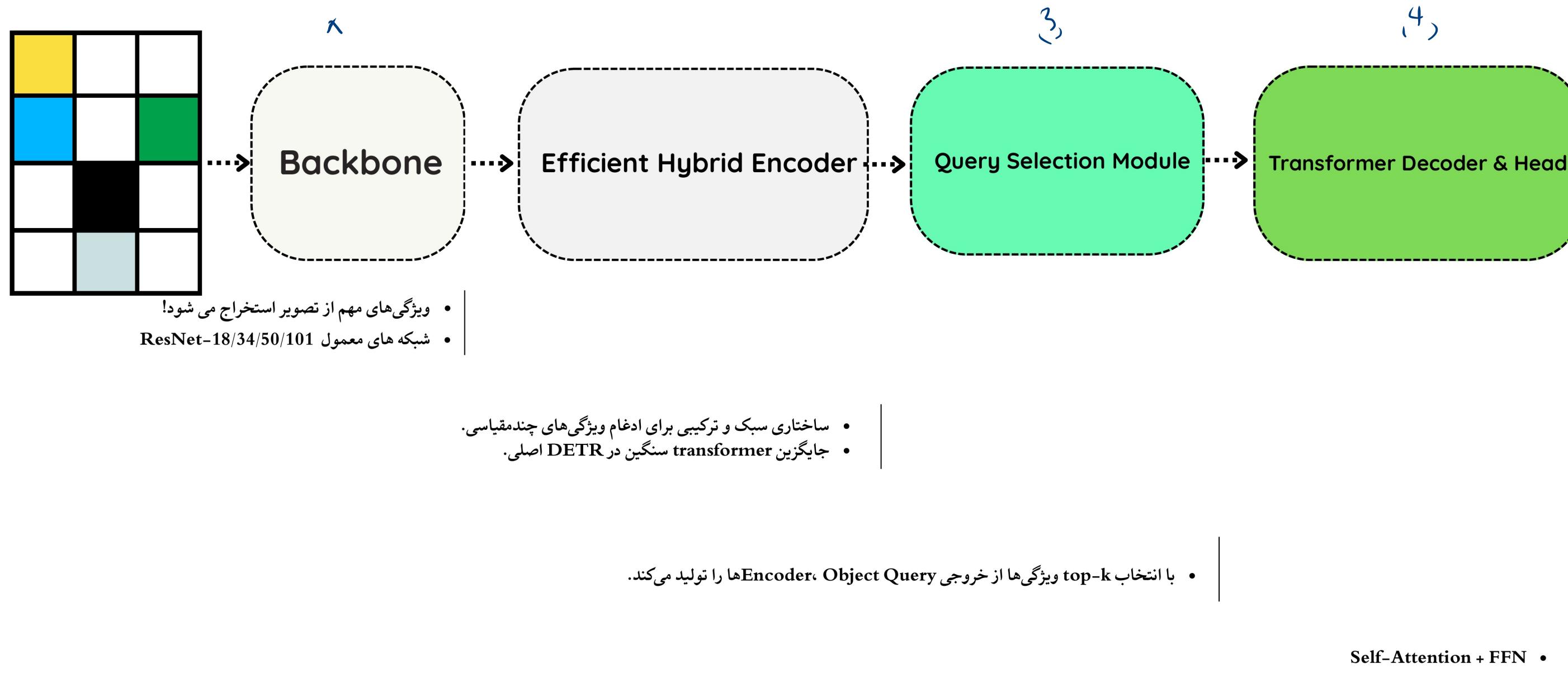
قابل انتقال به ONNX/TensorRT

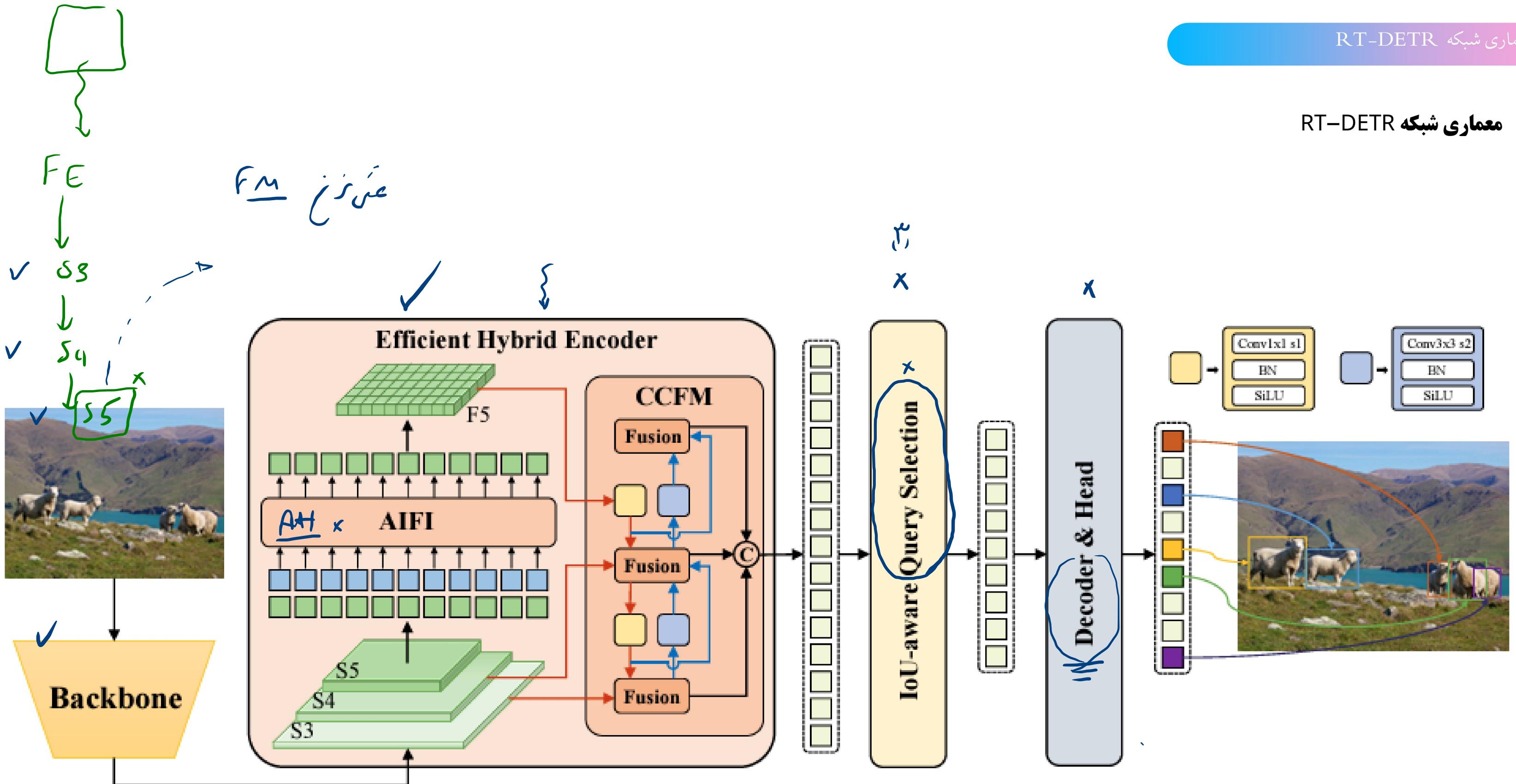
می‌توان آن را با backbone مختلف (مانند ResNet، Swin، ConvNeXt) ترکیب کرد.

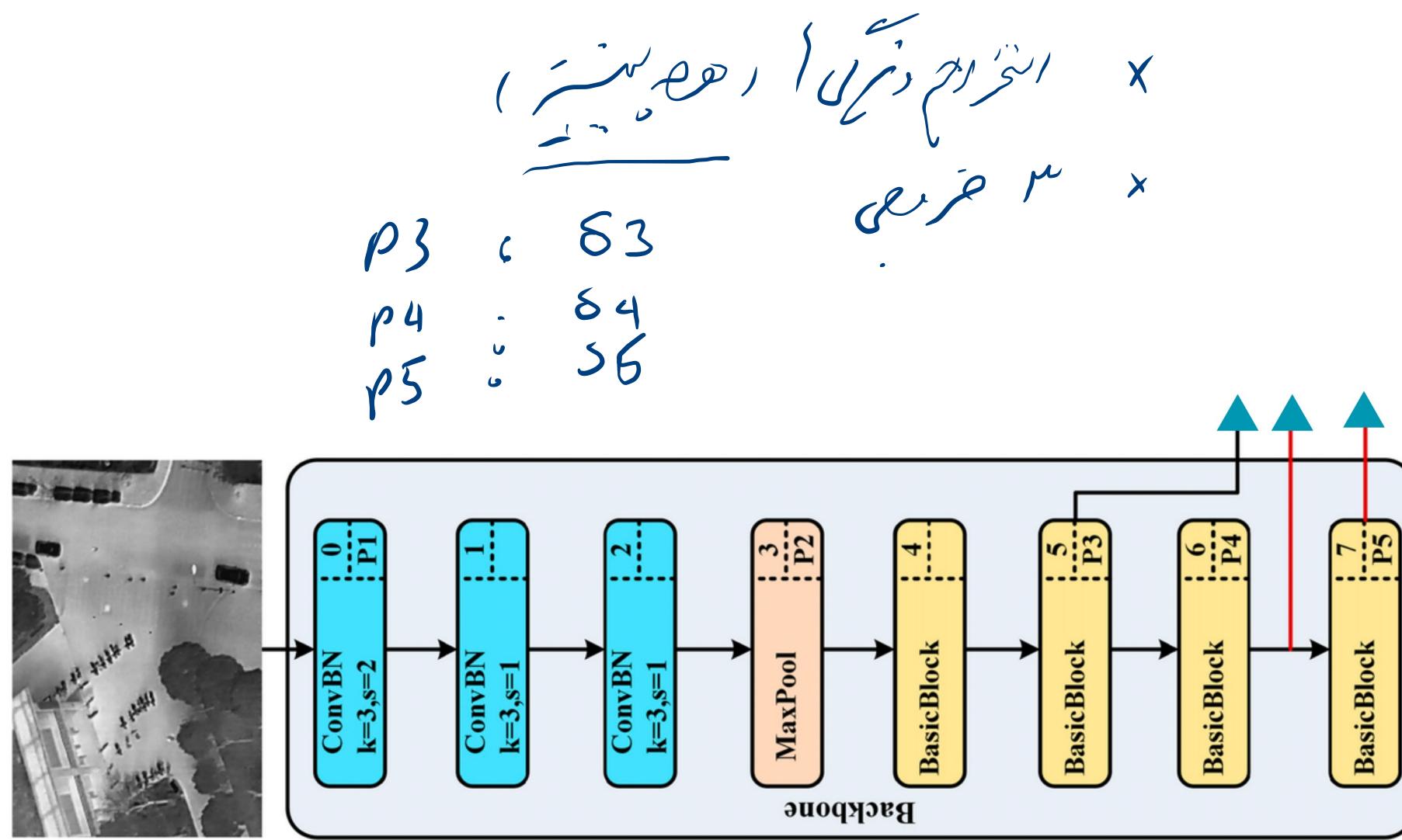


ماژولار و منعطف

معماری سطح بالا RT-DETR

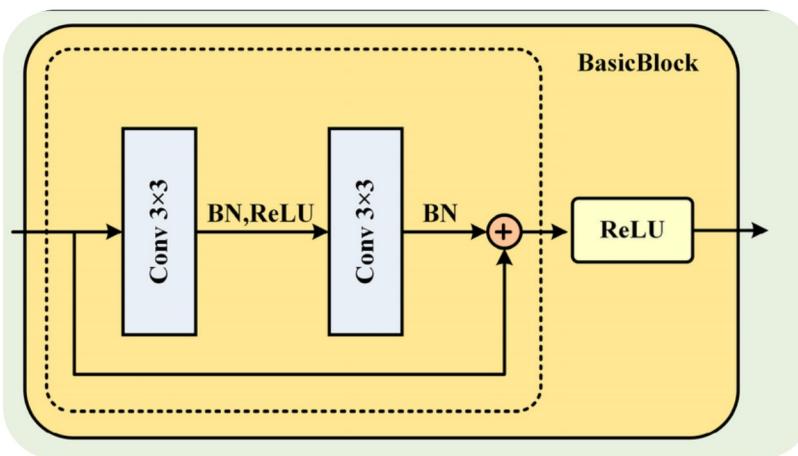




Backbone

Head	Feature Map Size	Channels
C2f (x2)	p_3 <u>80×80</u>	128
C2f (x3)	p_4 <u>40×40</u>	256
C2f (x4)	p_5 <u>20×20</u>	512

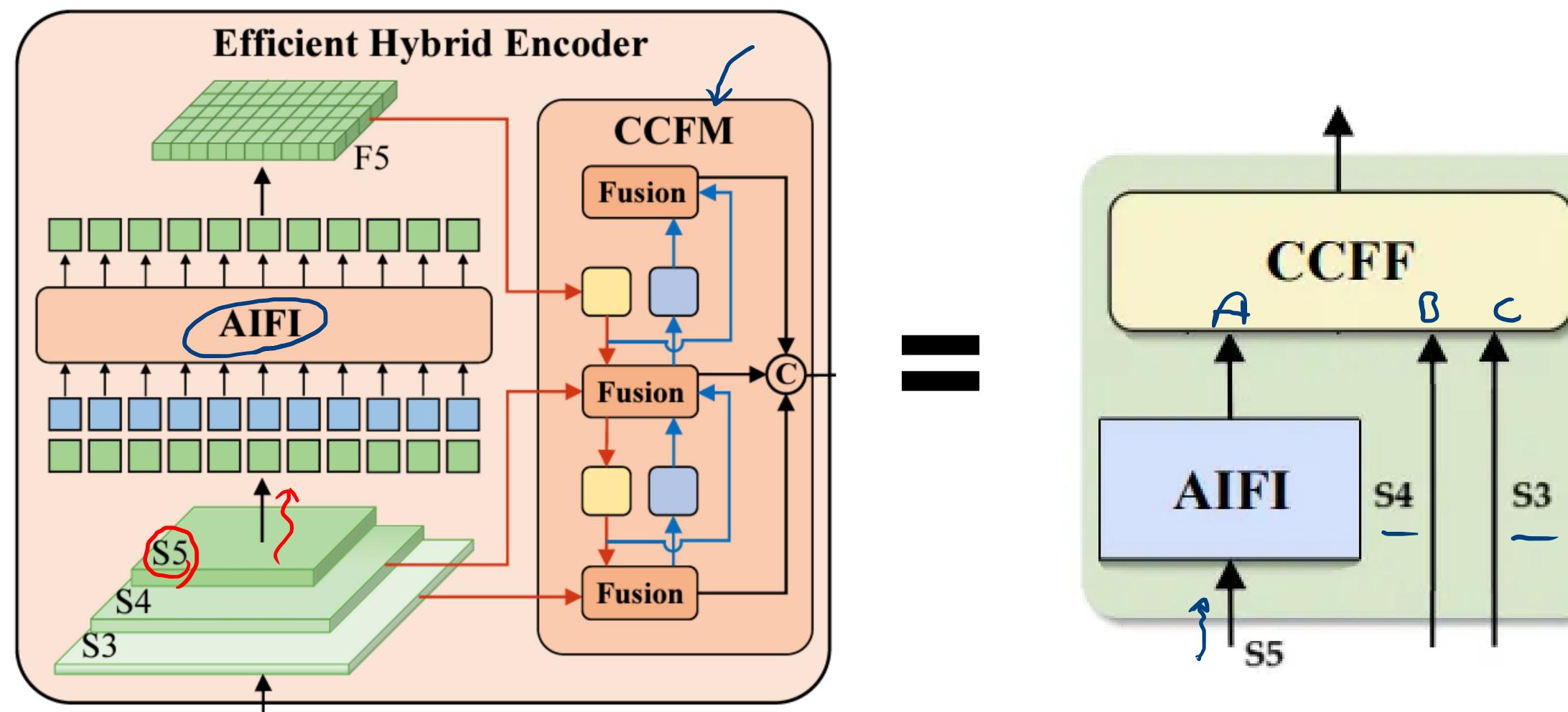
Brackets on the left group the outputs into three categories: C2f (x2), C2f (x3), and C2f (x4).



Efficient Hybrid Encoder

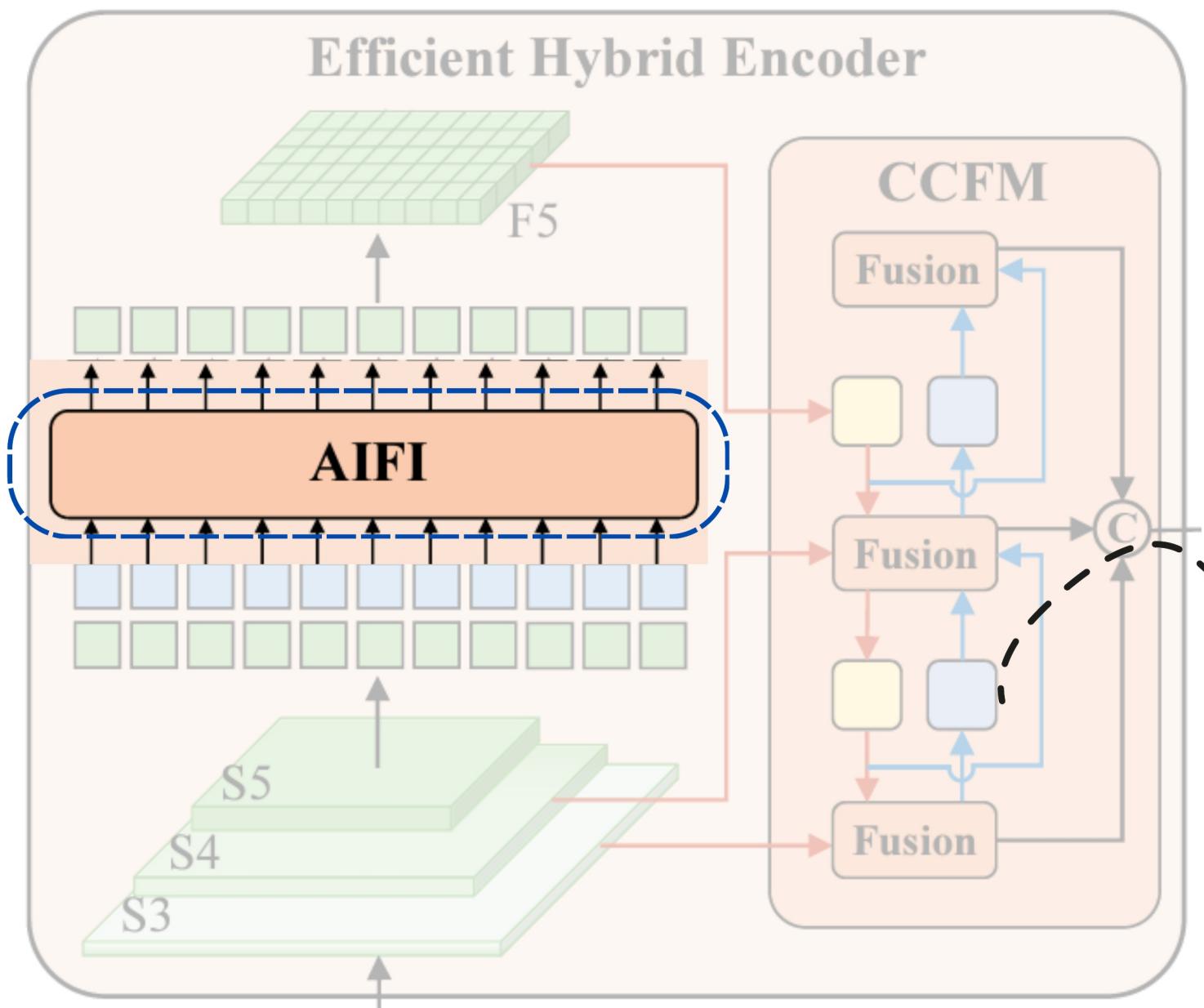
۰ پردازش عمیق تر خروجی Backbone : ویژگی های چندمقیاسی (S3,S4, S5)

- ۰ پردازش ویژگی های چندمقیاسی به صورت کارآمد و سریع، بدون استفاده مستقیم از ساختارهای سنگین تر ViT
- ۰ ارتباط بین پیکسل هایش را بهبود می دهد
- ۰ ویژگی های مهم را تقویت می کند!

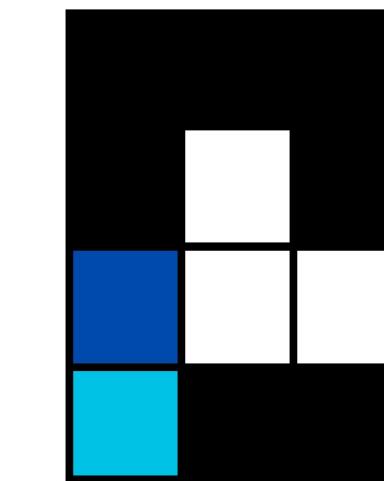
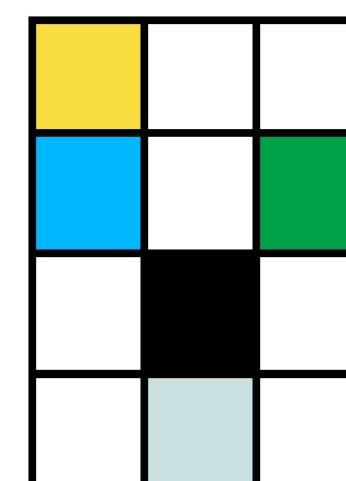


Attention based Intra Scale Feature Interaction

حفظ سرعت بالا در عین افزایش دقیقت
استخراج نواحی مهم

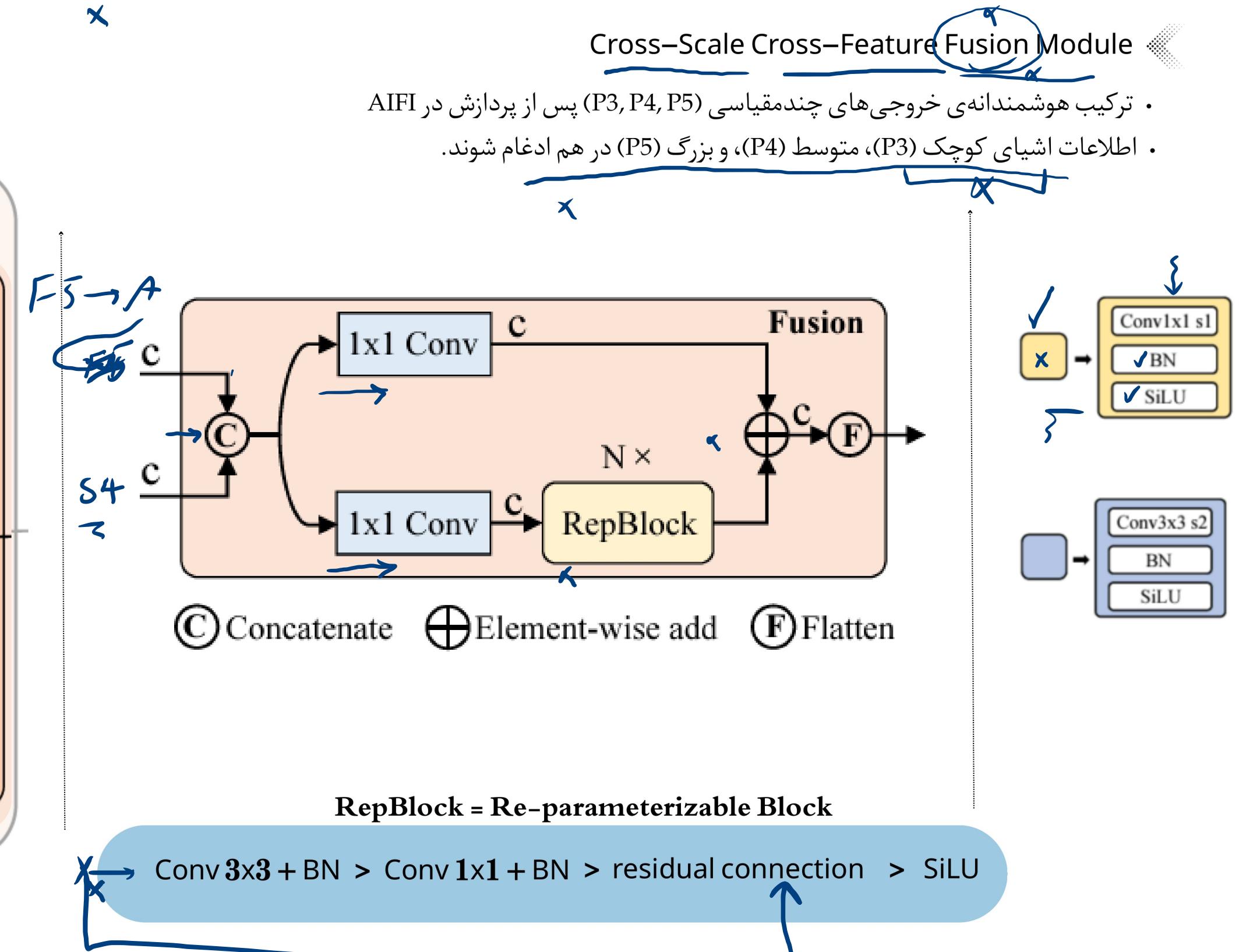
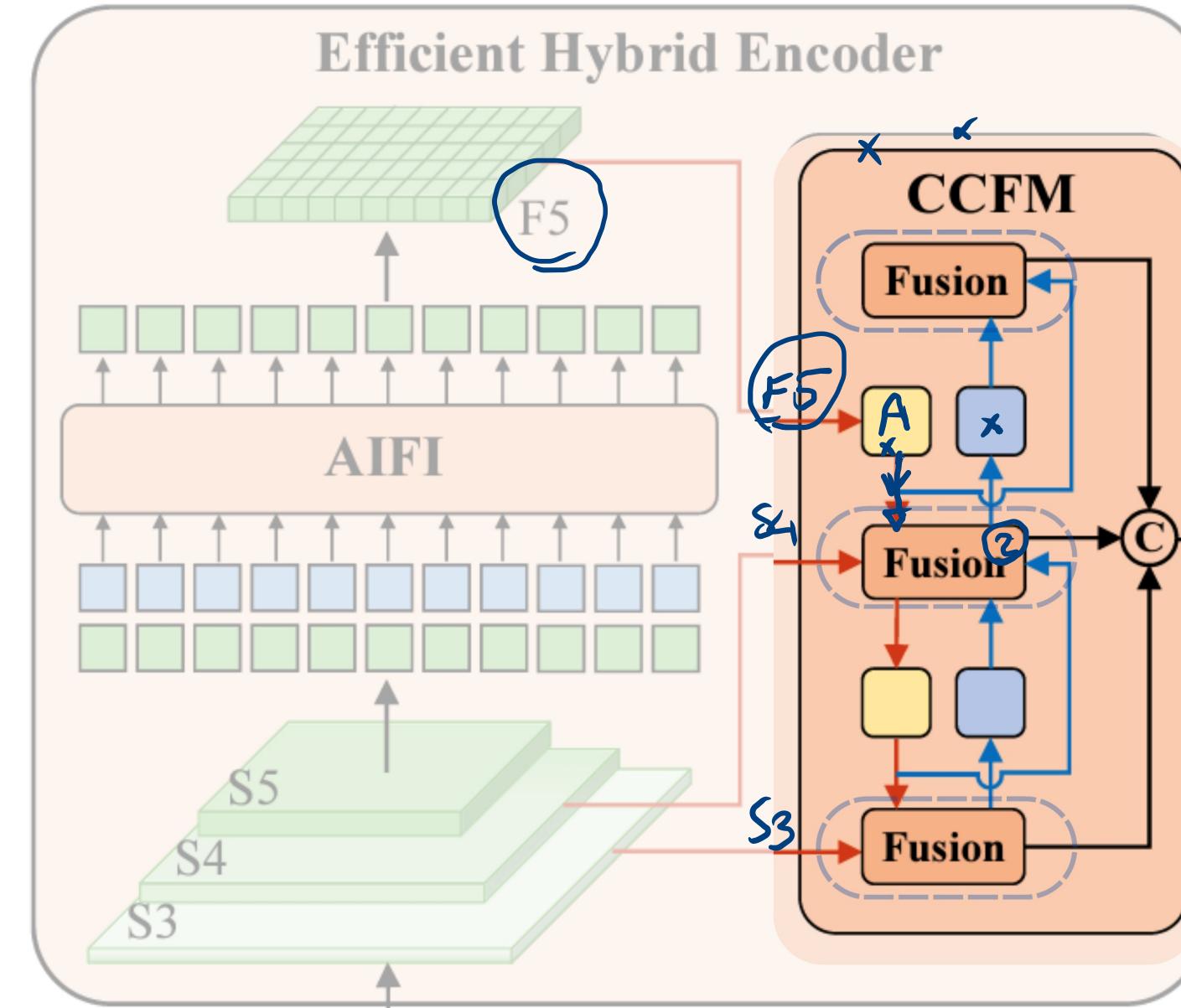


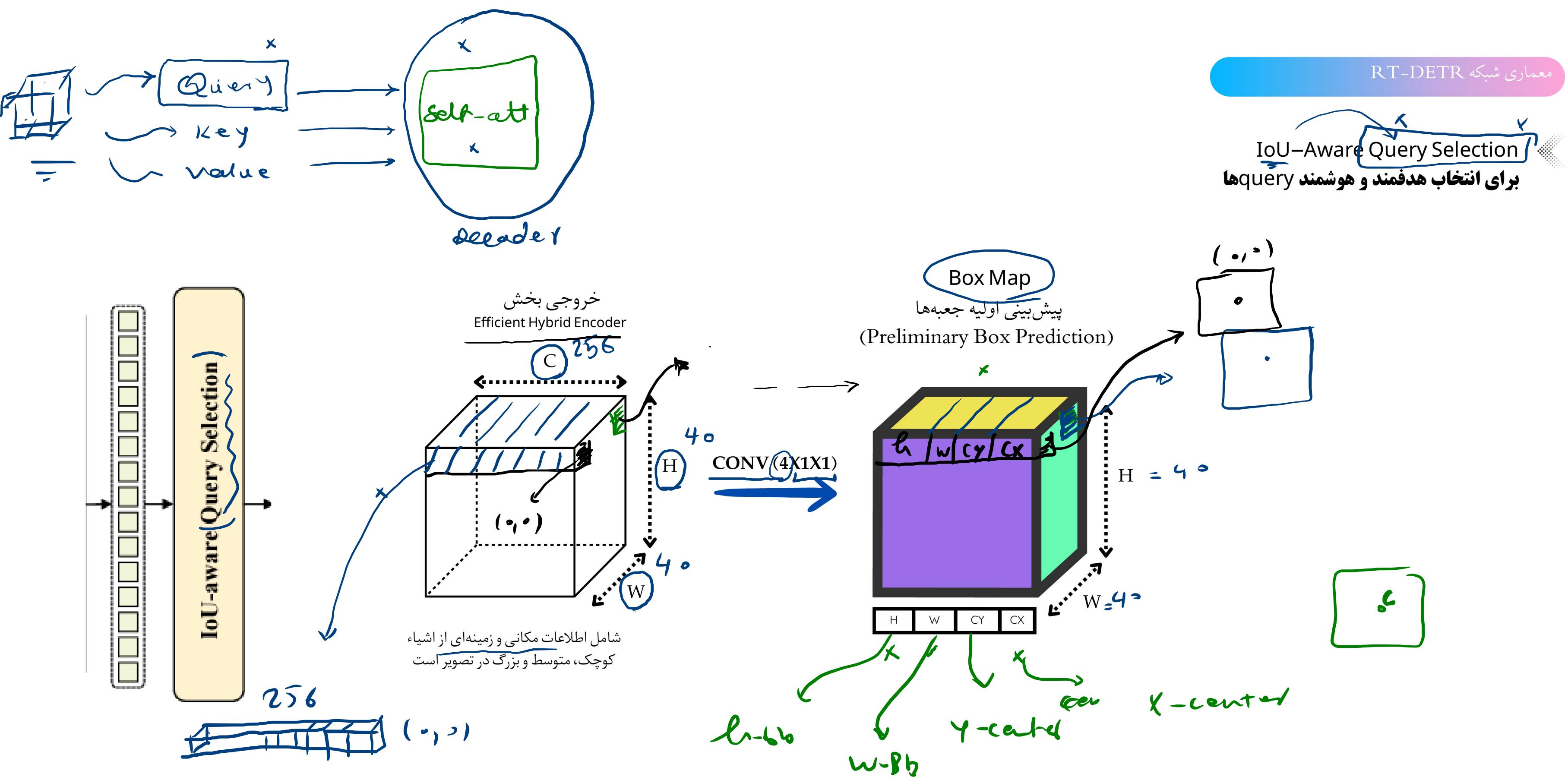
$$F_{\text{attn}} = \text{FeatureMap} \times \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(\text{FeatureMap})))$$



معماری تقریبی

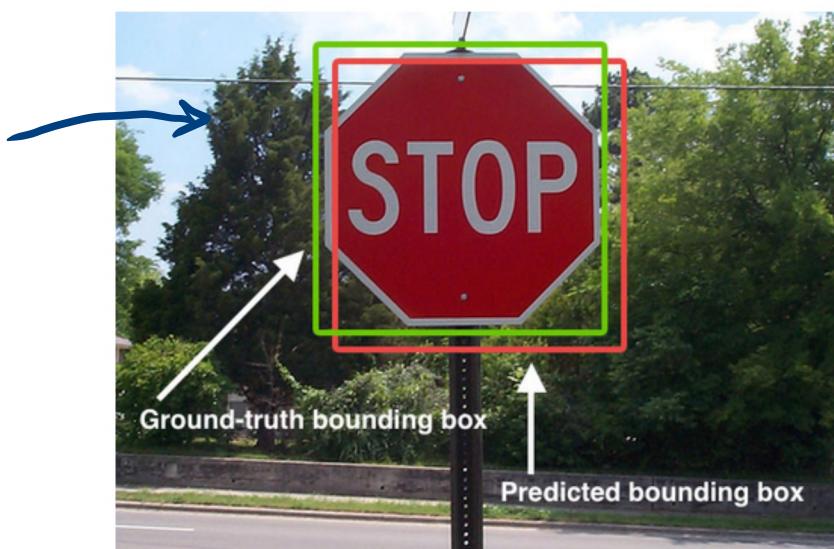
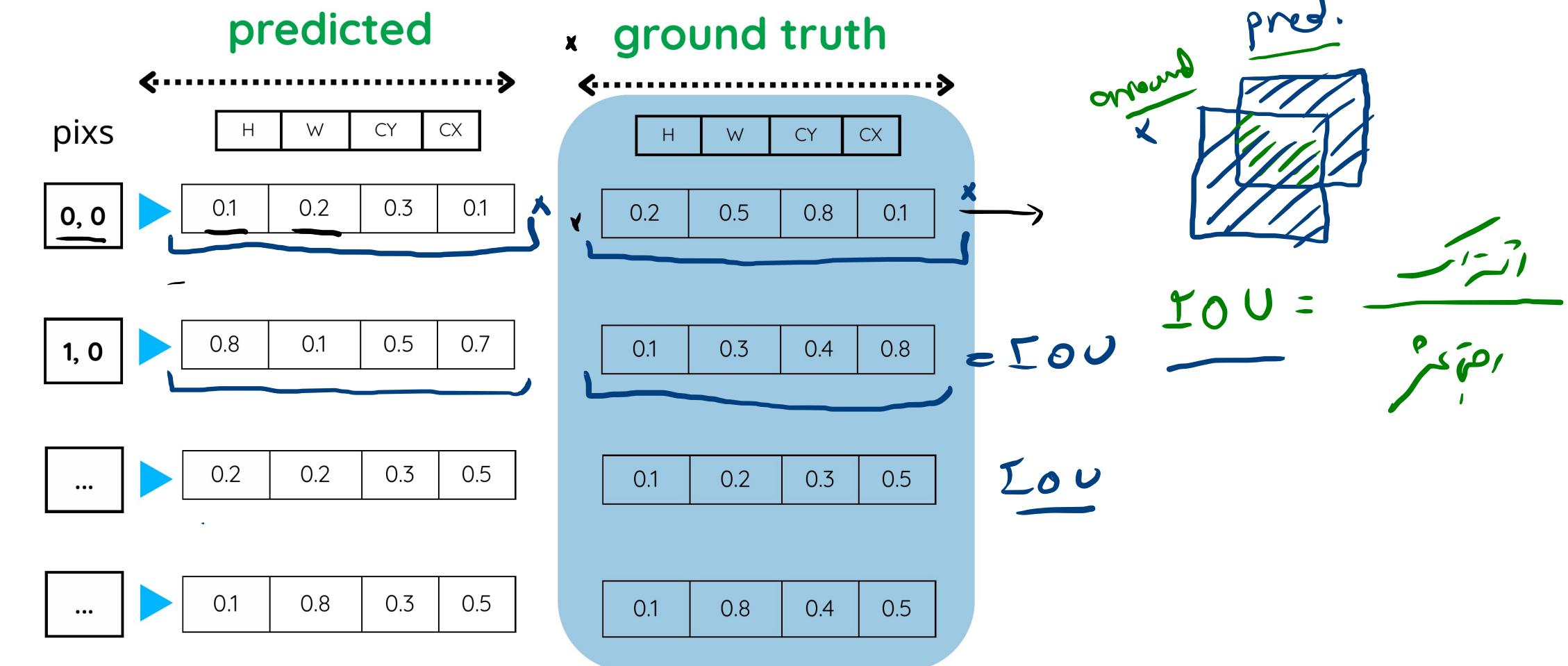
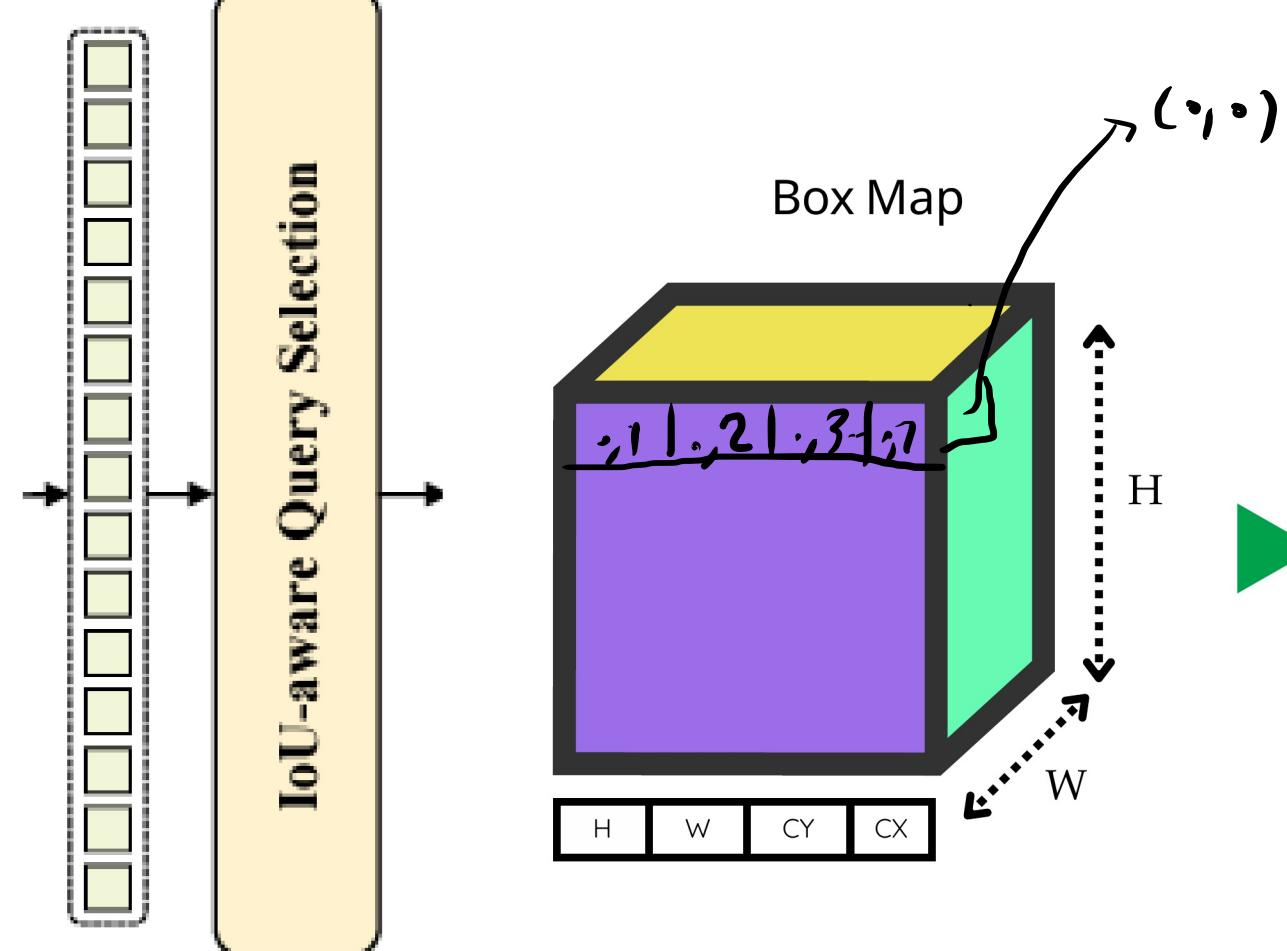
ملاحت \rightarrow سایر نتایج را ازین مرتبه! سایر دامنه‌های تغییر نشوند!





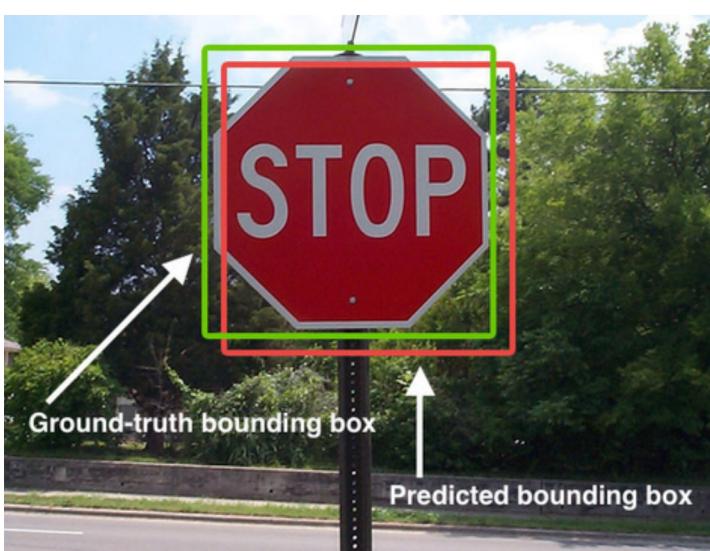
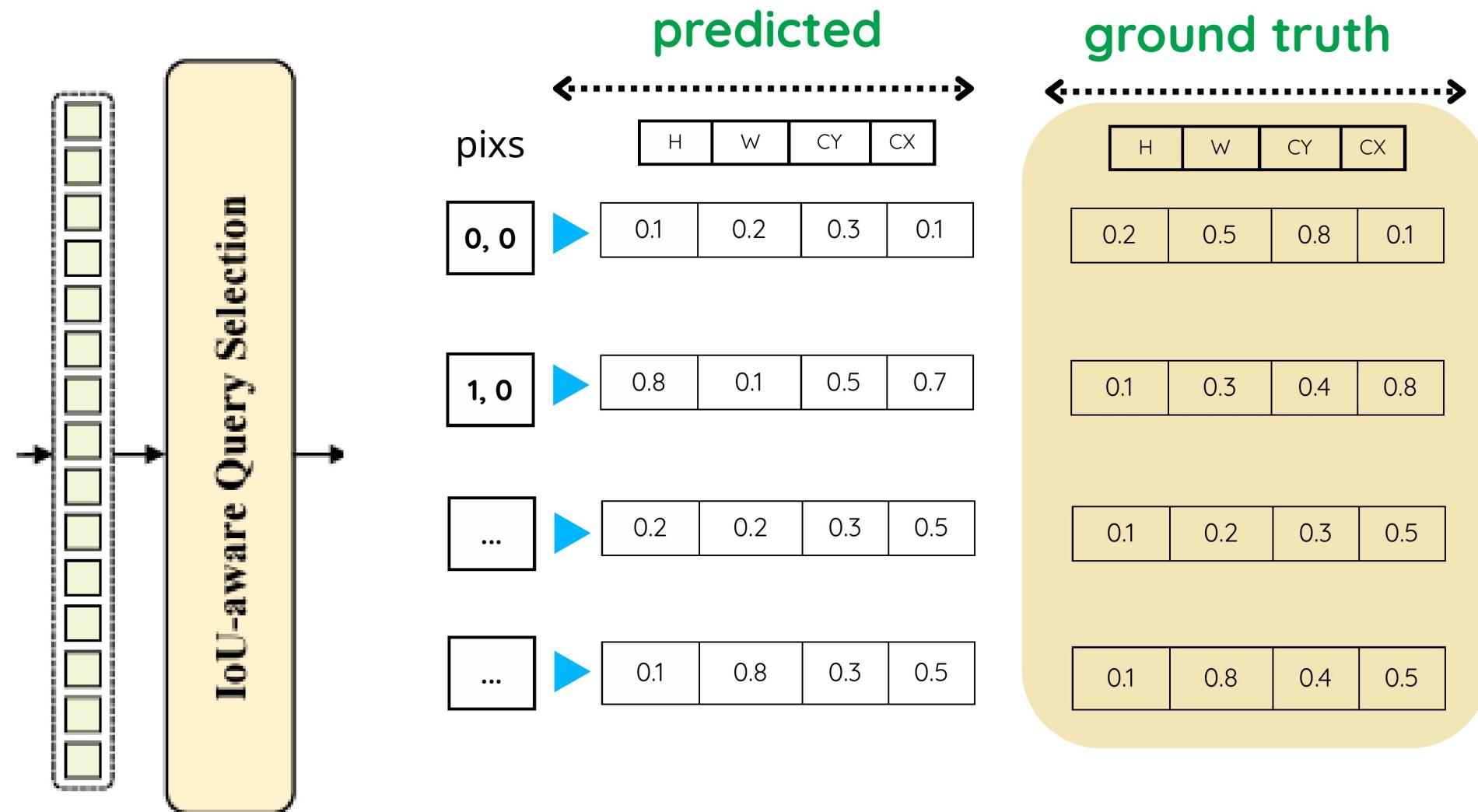
برای انتخاب هدفمند و هوشمند query ها

IoU-Aware Query Selection

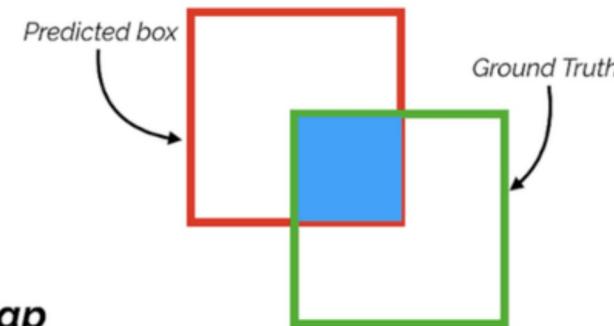


برای انتخاب هدفمند و هوشمند query‌ها

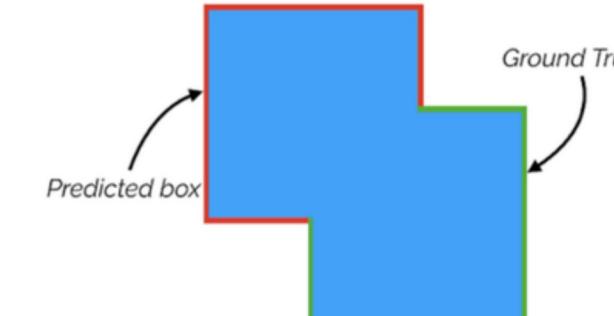
IoU–Aware Query Selection



محاسبه IOU



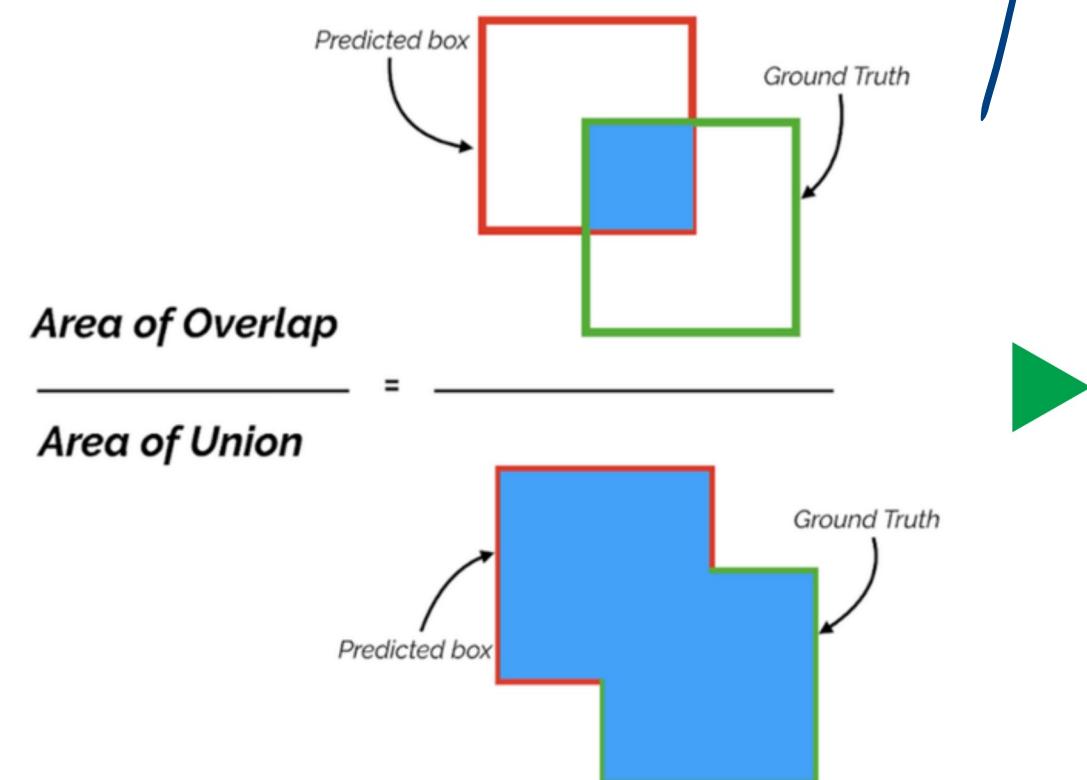
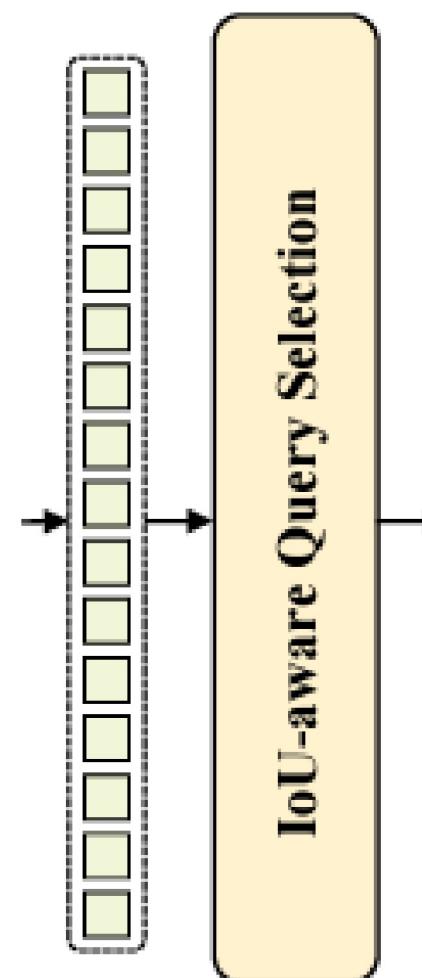
$$\frac{\text{Area of Overlap}}{\text{Area of Union}} =$$



$$4^0 \times 4^0 = \underline{1600} \rightarrow \underline{50}$$

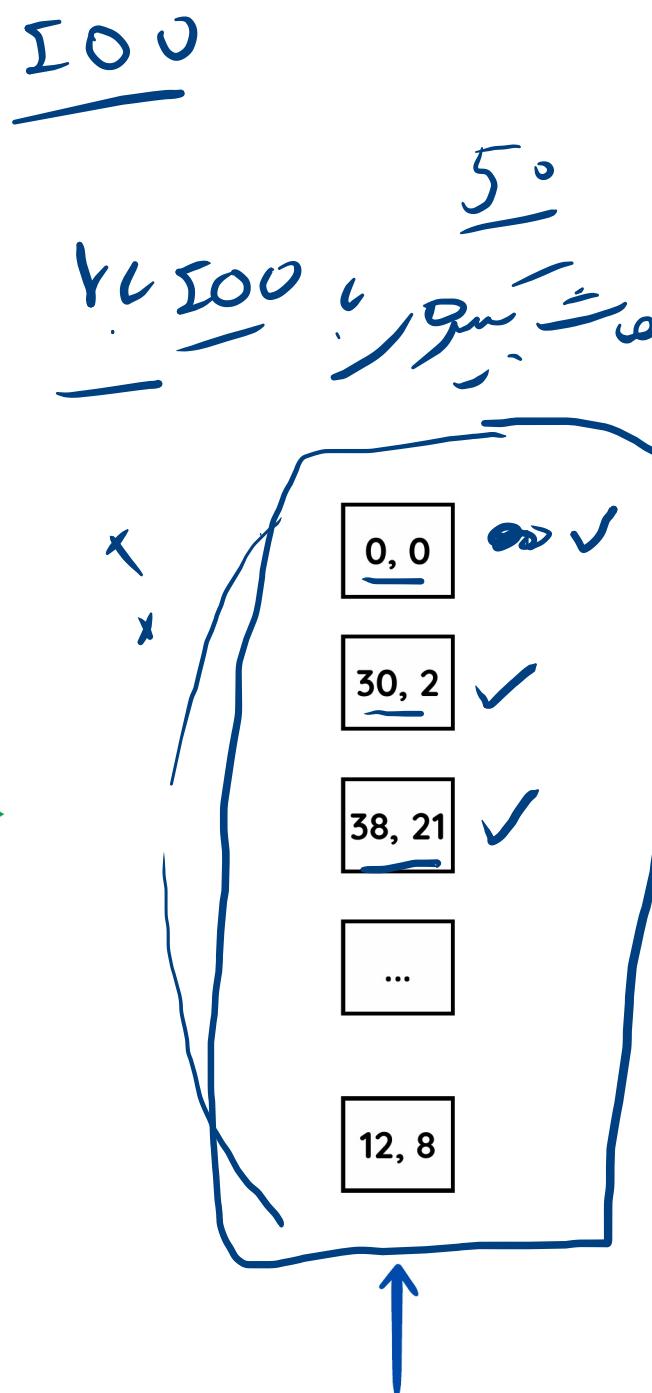
100501600

فناز سر بر

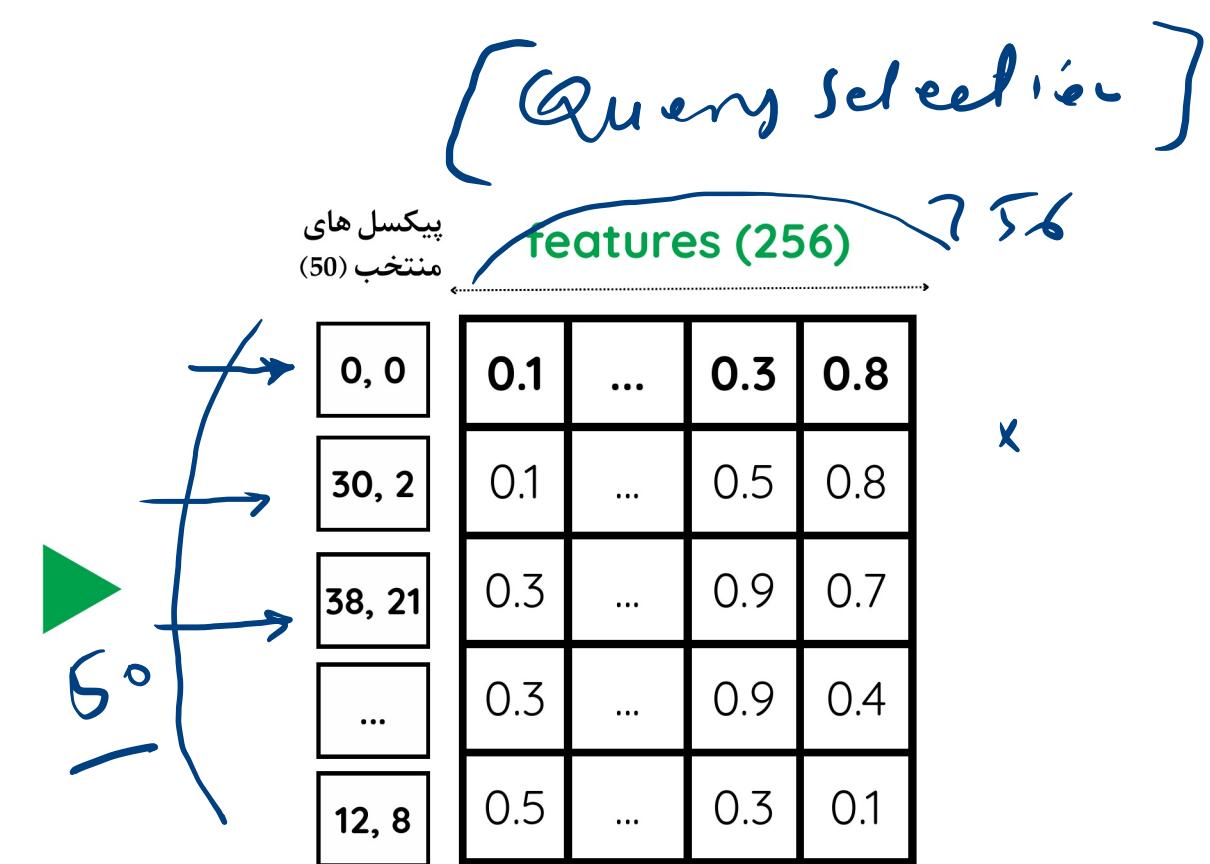


محاسبه IOU

انتخاب K تا از anchorها با بیشترین IoU
متلا ۵۰ تا

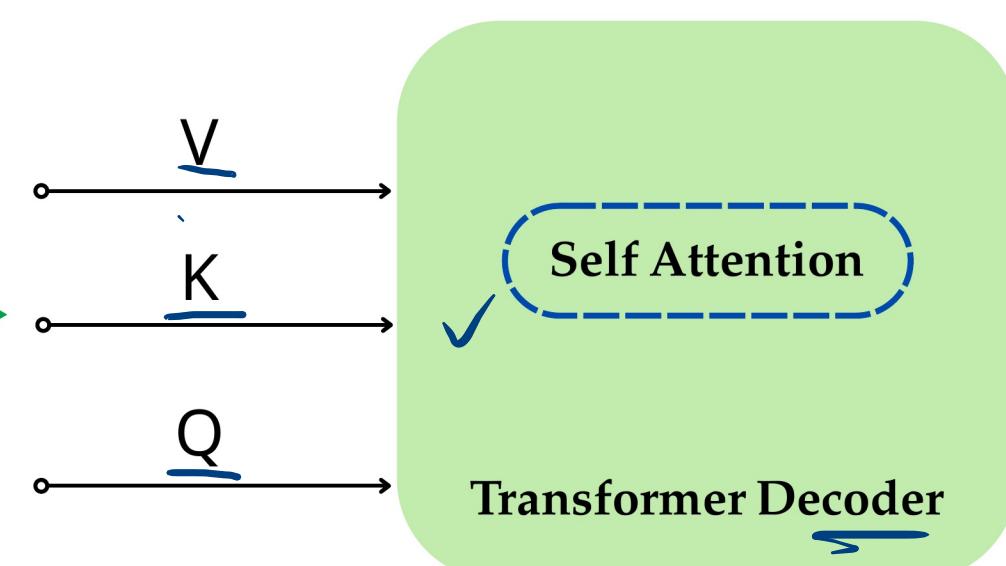
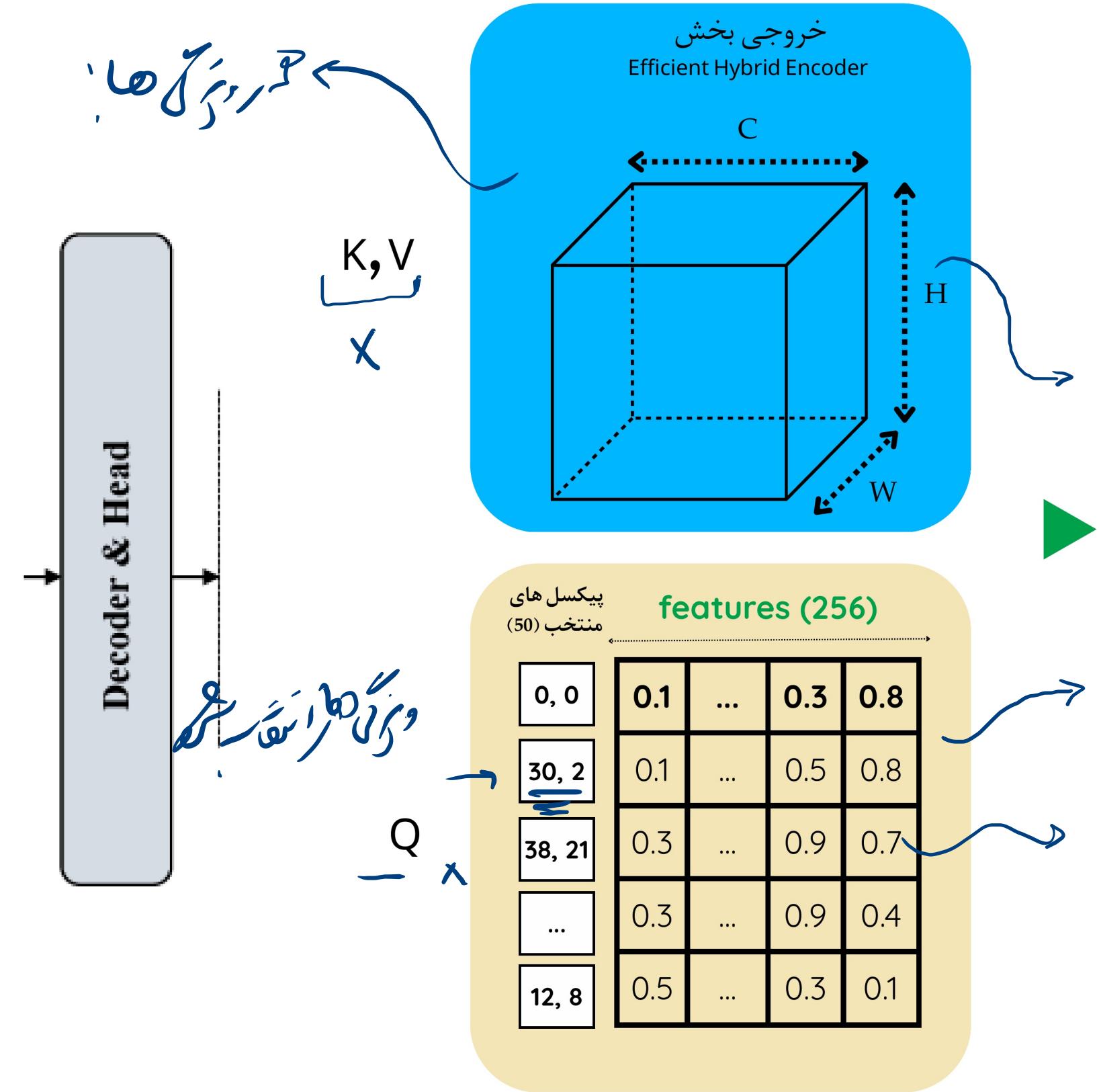


برای انتخاب هدفمند و هوشمند
IoU-Aware Query Selection



تشکیل ماتریس
برای ورود به دیکودر ترنسفورمر

object queries



the end

RT - Detr