

موضع

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

RT-DETR مزایای شبکه



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

✓ دقت بالا ✗

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

✓ سرعت بلادرنگ ✗

yolo

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 ✗

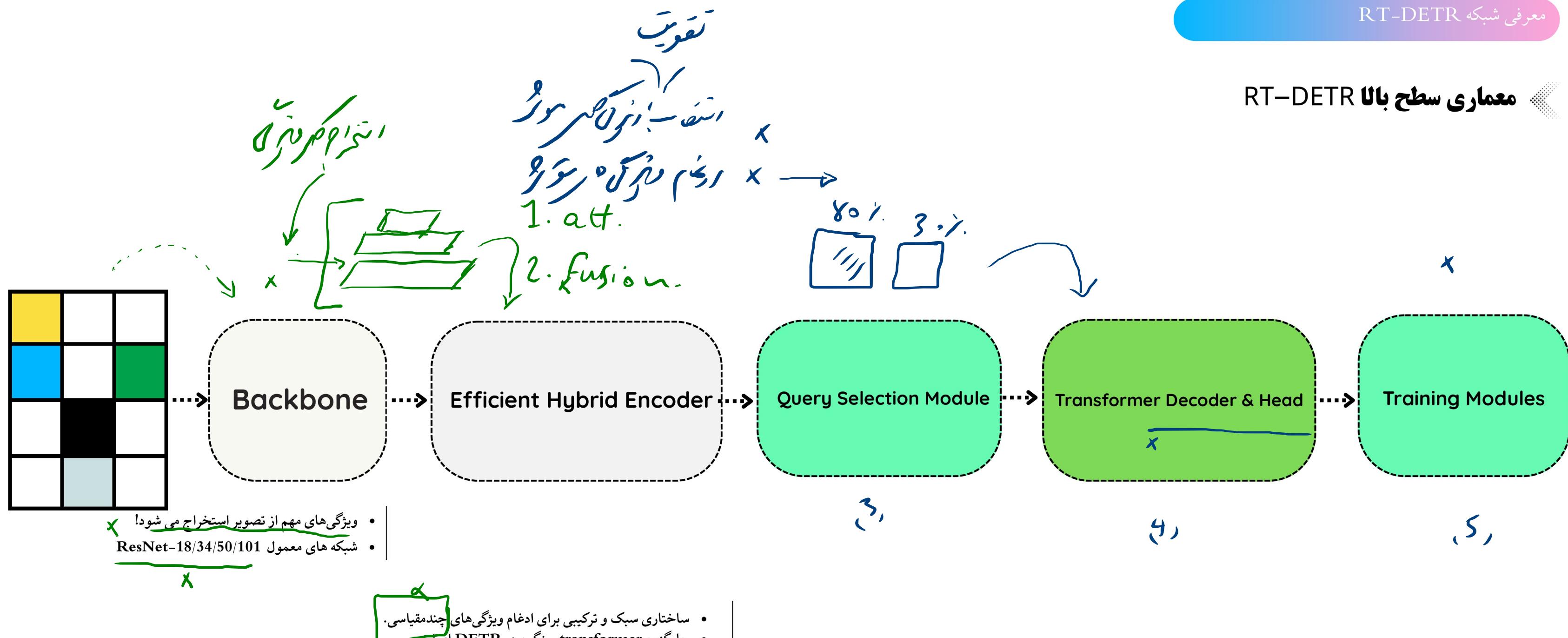
مناسب برای روی دستگاه‌های edge با استفاده از TensorRT یا DeepStream deployment.

✓ قابل انتقال به ONNX/TensorRT ✗

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

✓ مازوئار و منعطف ✗

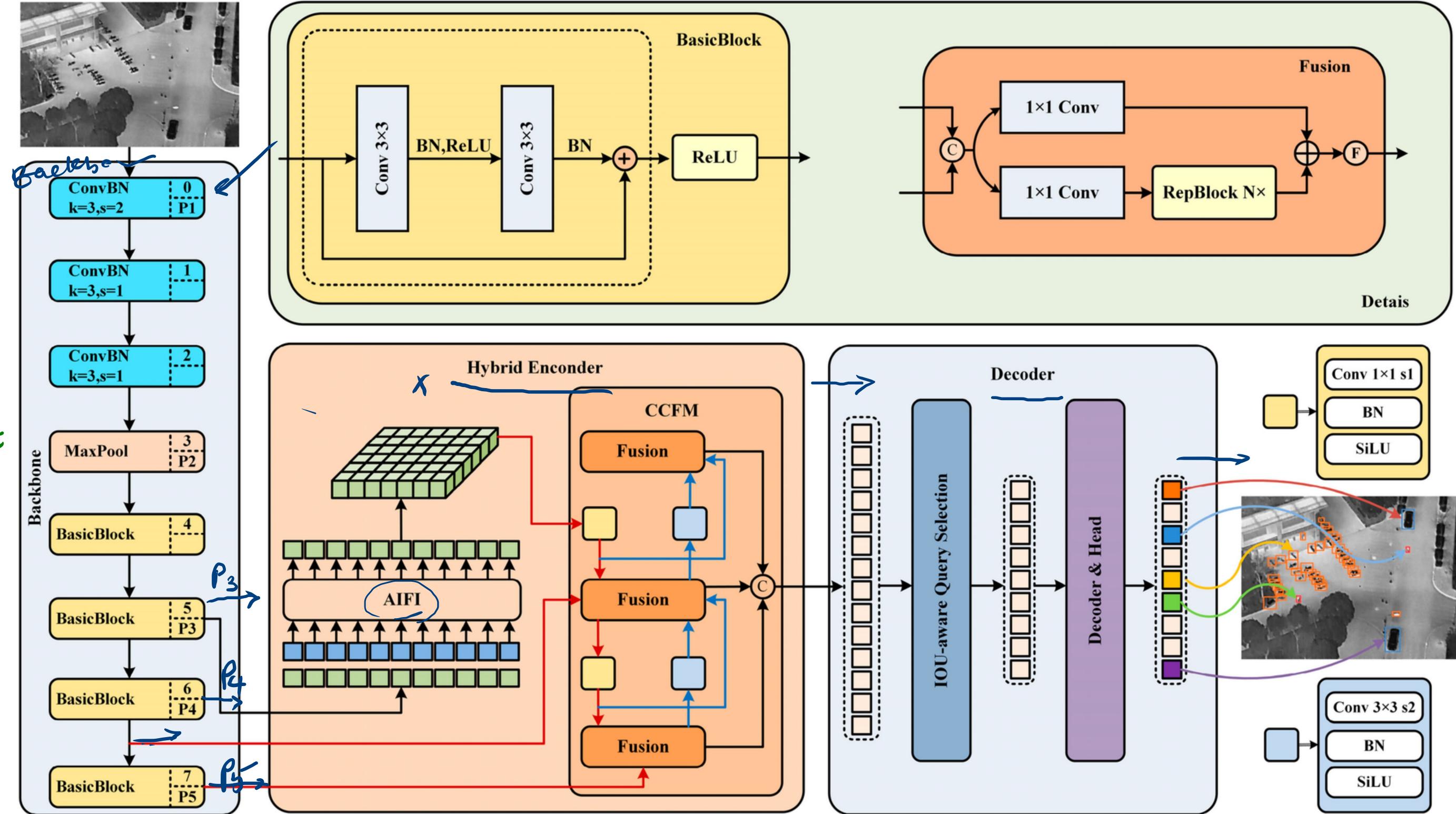
معماری سطح بالا

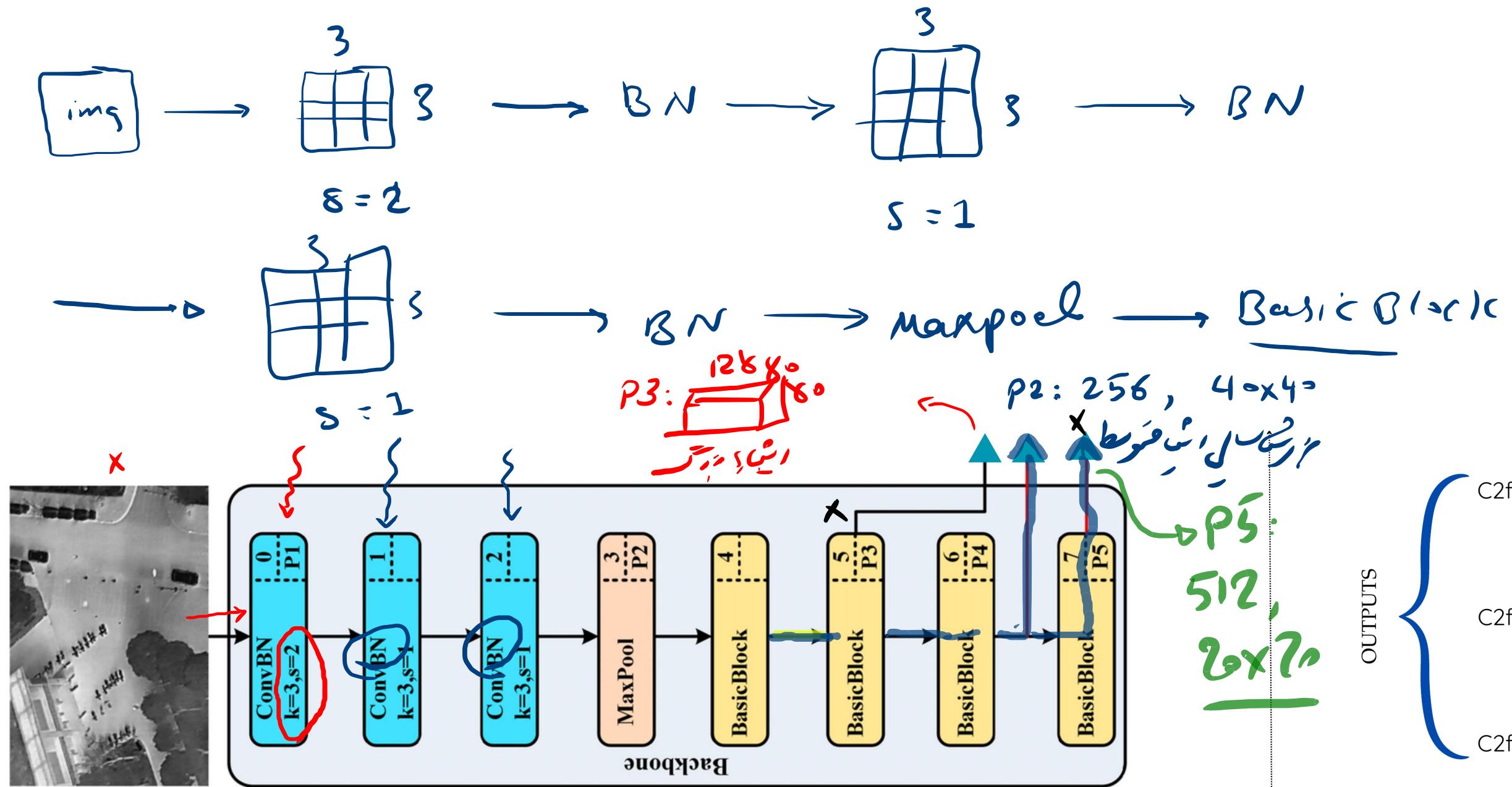


- با انتخاب k top- k ویژگی ها از خروجی Encoder، Object Query ها را تولید می کند.

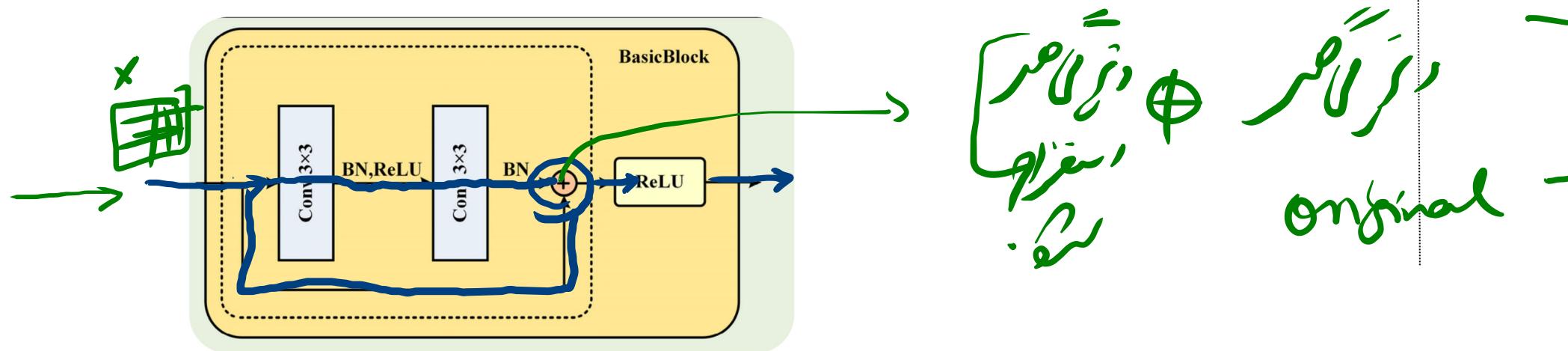
Self-Attention + FFN •

معماری شبکه

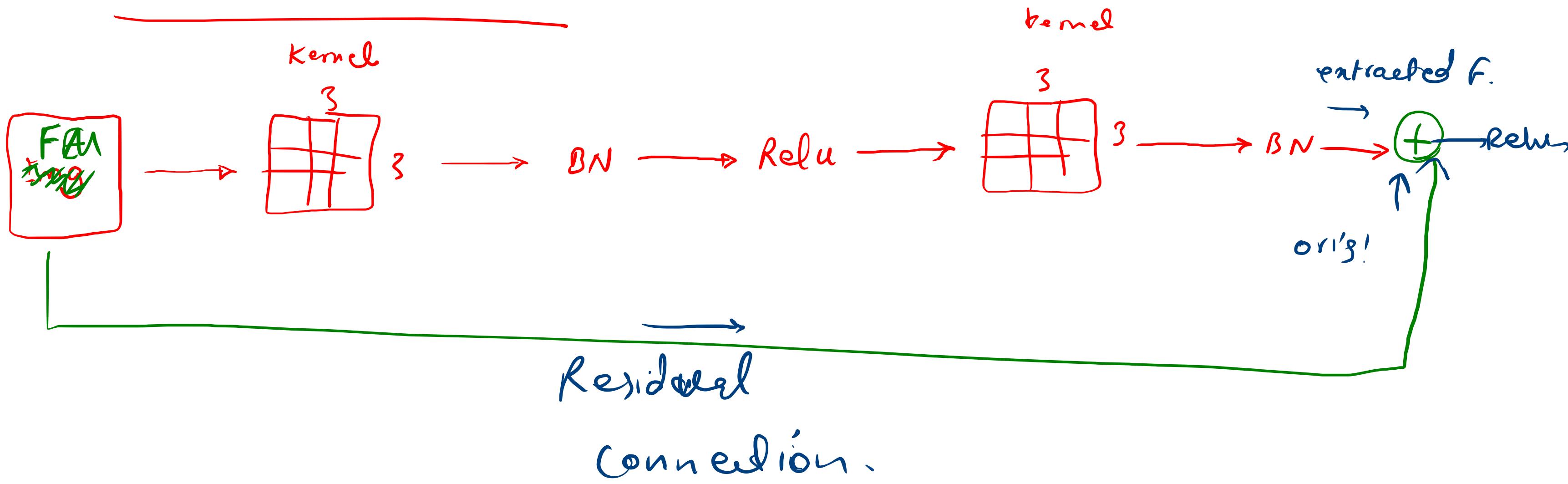


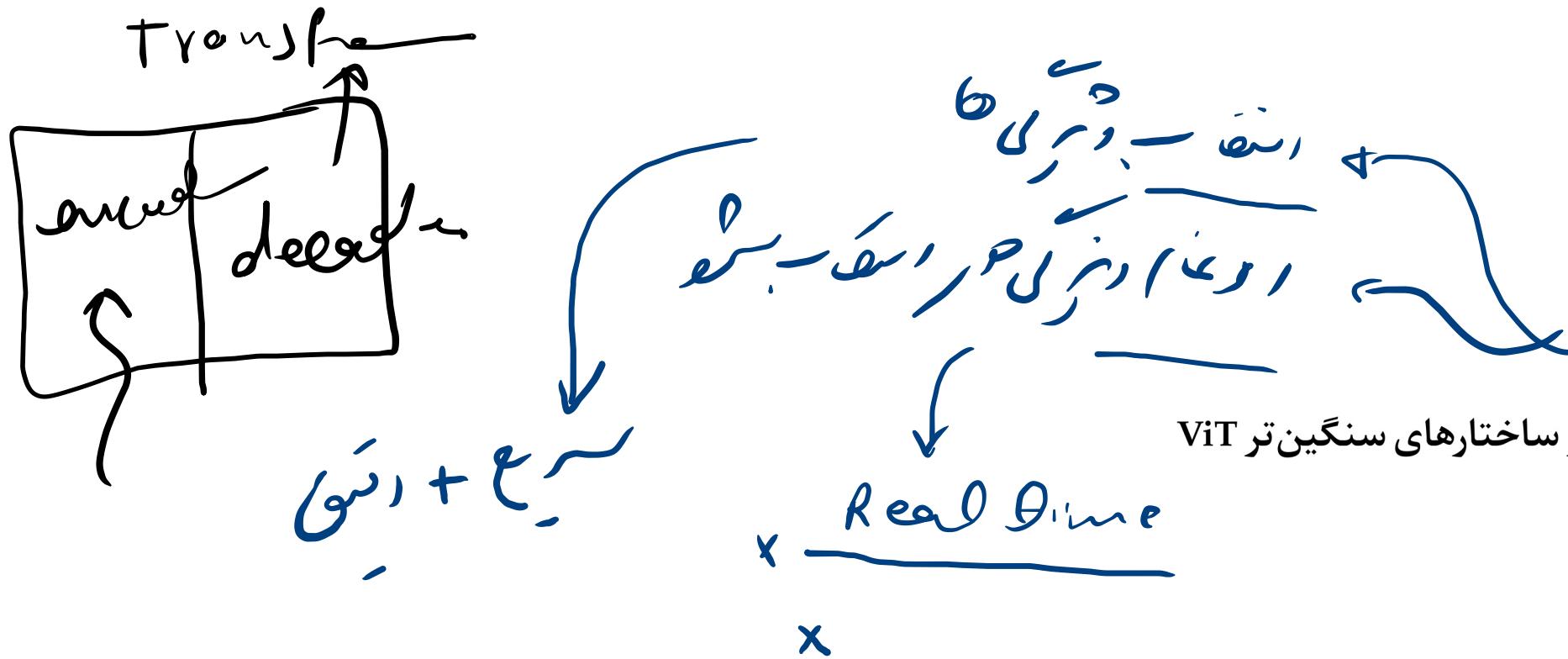
Backbone

Head	Feature Map Size	Channels
C2f (x2)	80x80	128
C2f (x3)	40x40	256
C2f (x4)	20x20	512



Basic Block

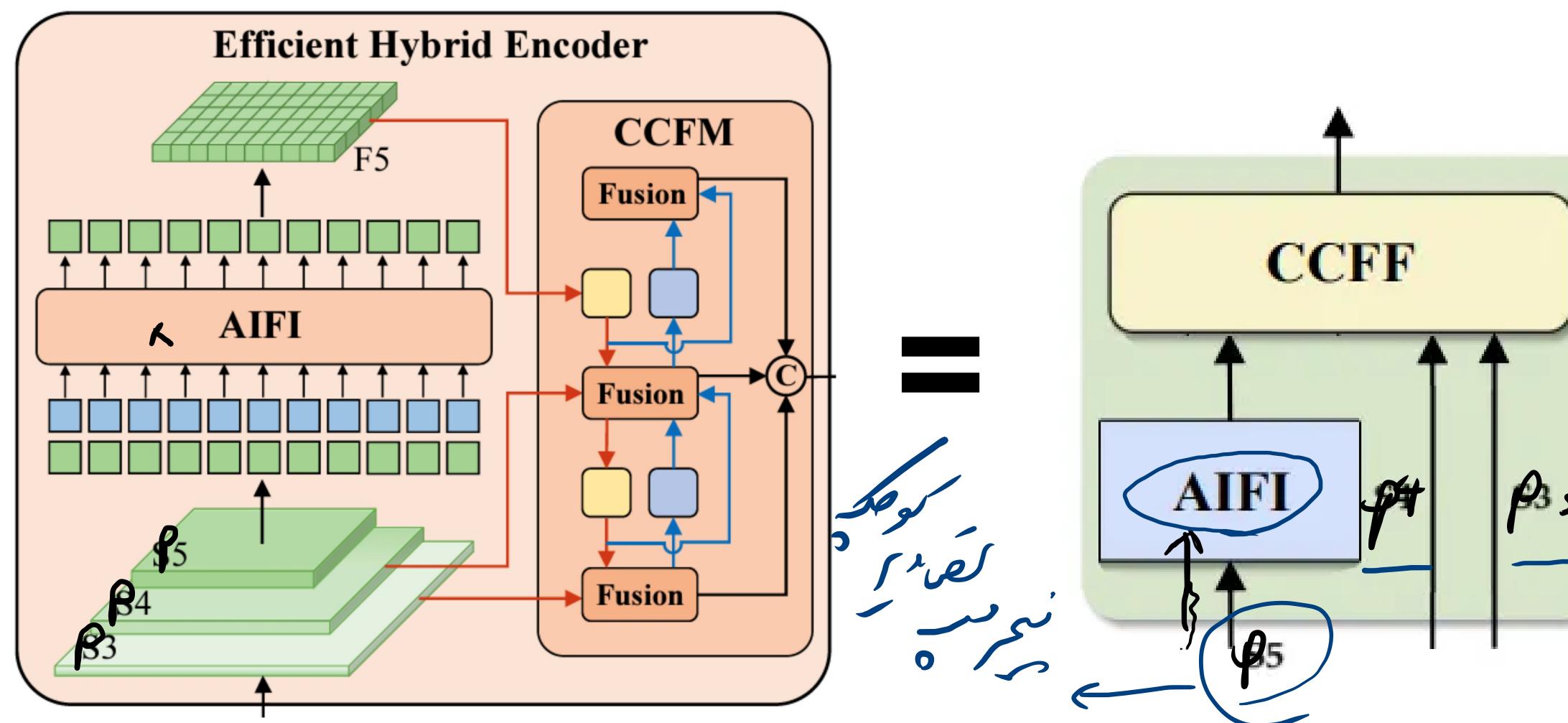


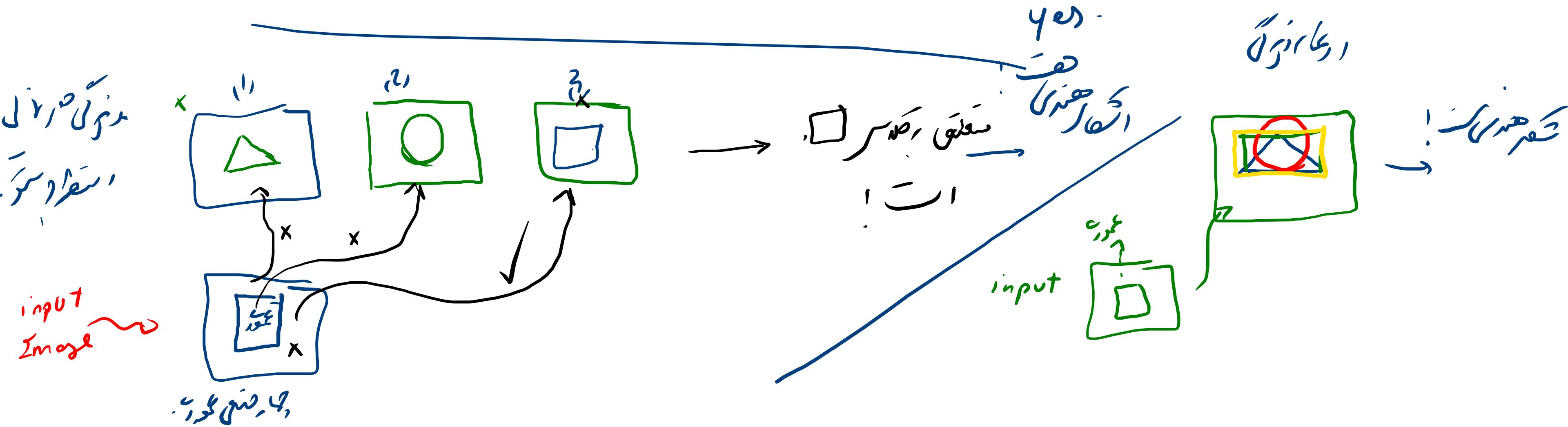
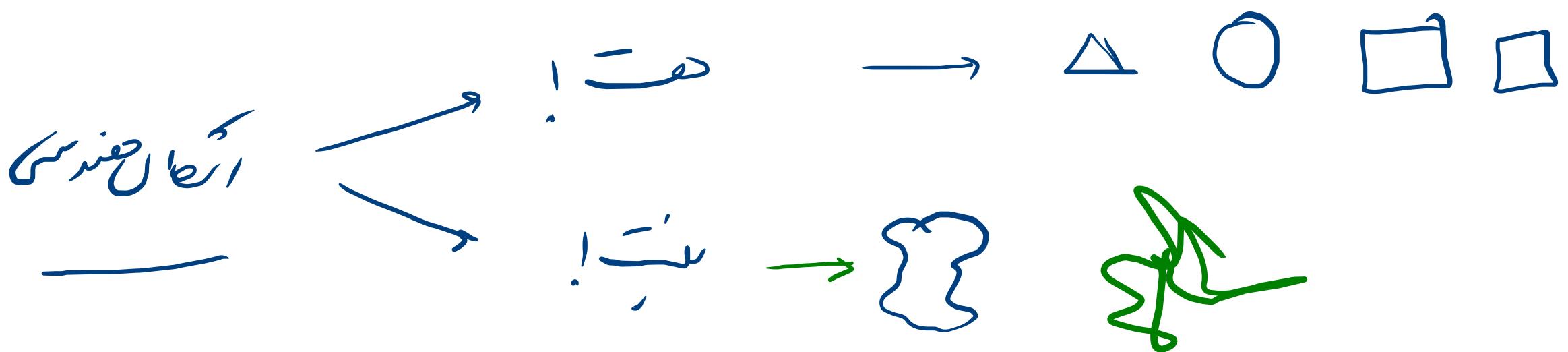


RT-DETR شبکه معماری

- Encoder ← Efficient Hybrid Encoder

 - پردازش عمیق تر خروجی Backbone : ویژگی های چندمقیاسی (53, 54, 55)
 - پردازش ویژگی های چندمقیاسی به صورت کارآمد و سریع، بدون استفاده مستقیم
 - ارتباط بین پیکسل هایش را بهبود می دهد
 - ویژگی های مهم را تقویت می کند!

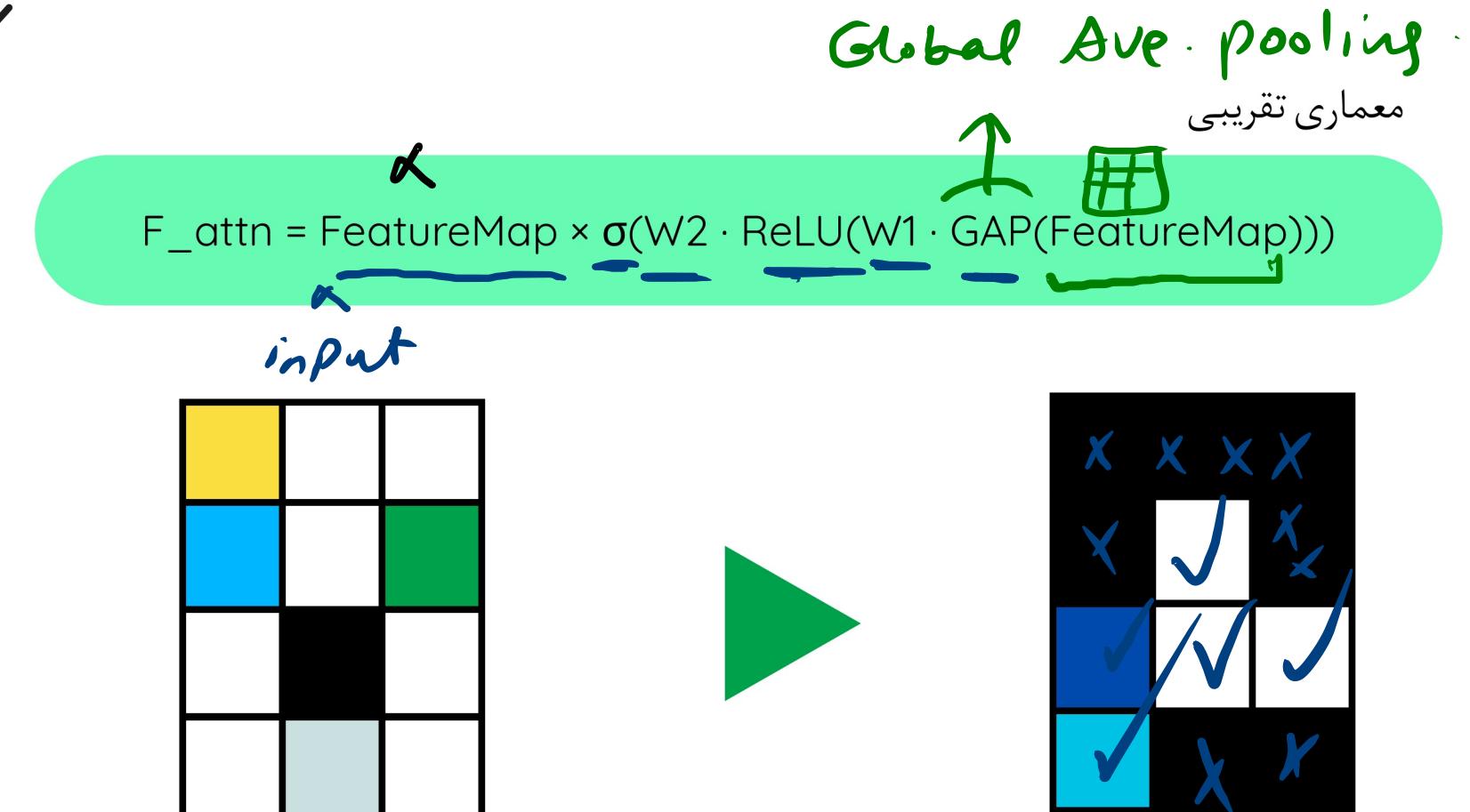
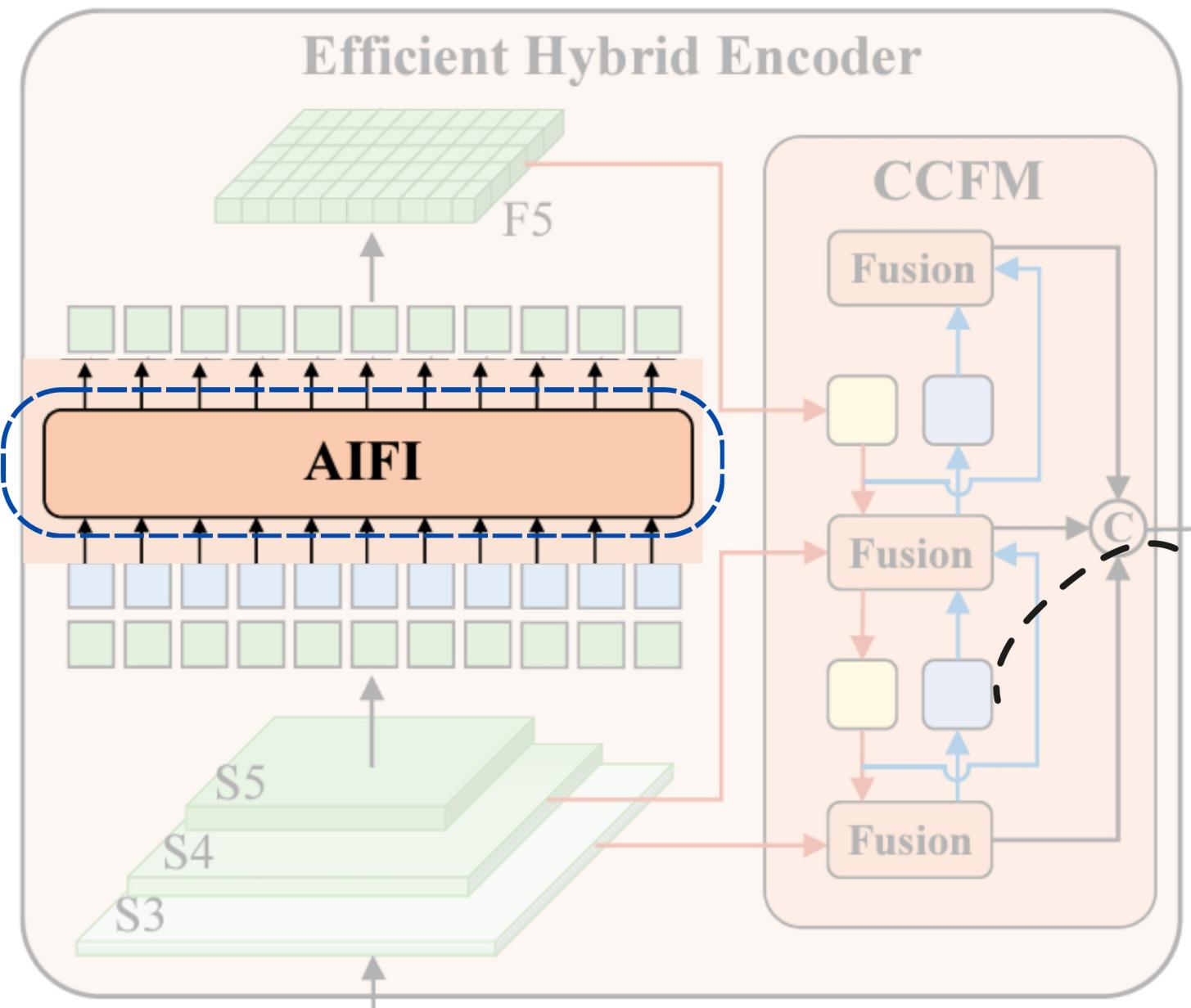




Detr → self-attention

RT-Detr → GAP (Global attention projection)

- حفظ سرعت بالا در عین افزایش دقت
- استخراج نواحی مهم



8	7	3
1	1	0
1	2	1

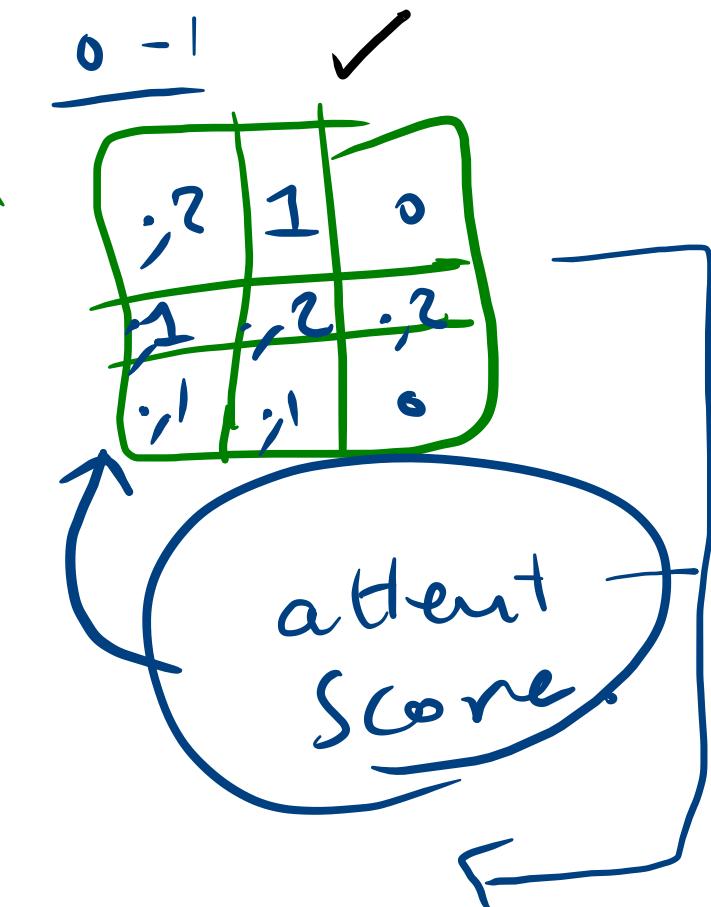
GAP
جزویت

2	1	-1
1	2	2
1	1	0

ReLU

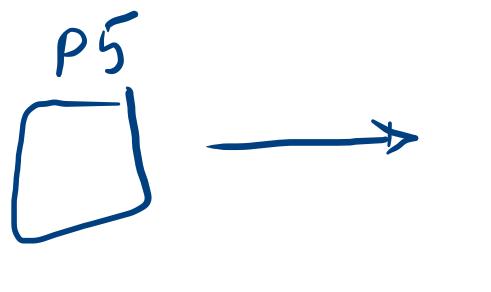
2	1	0
1	2	2
1	1	0

softmax
کسری



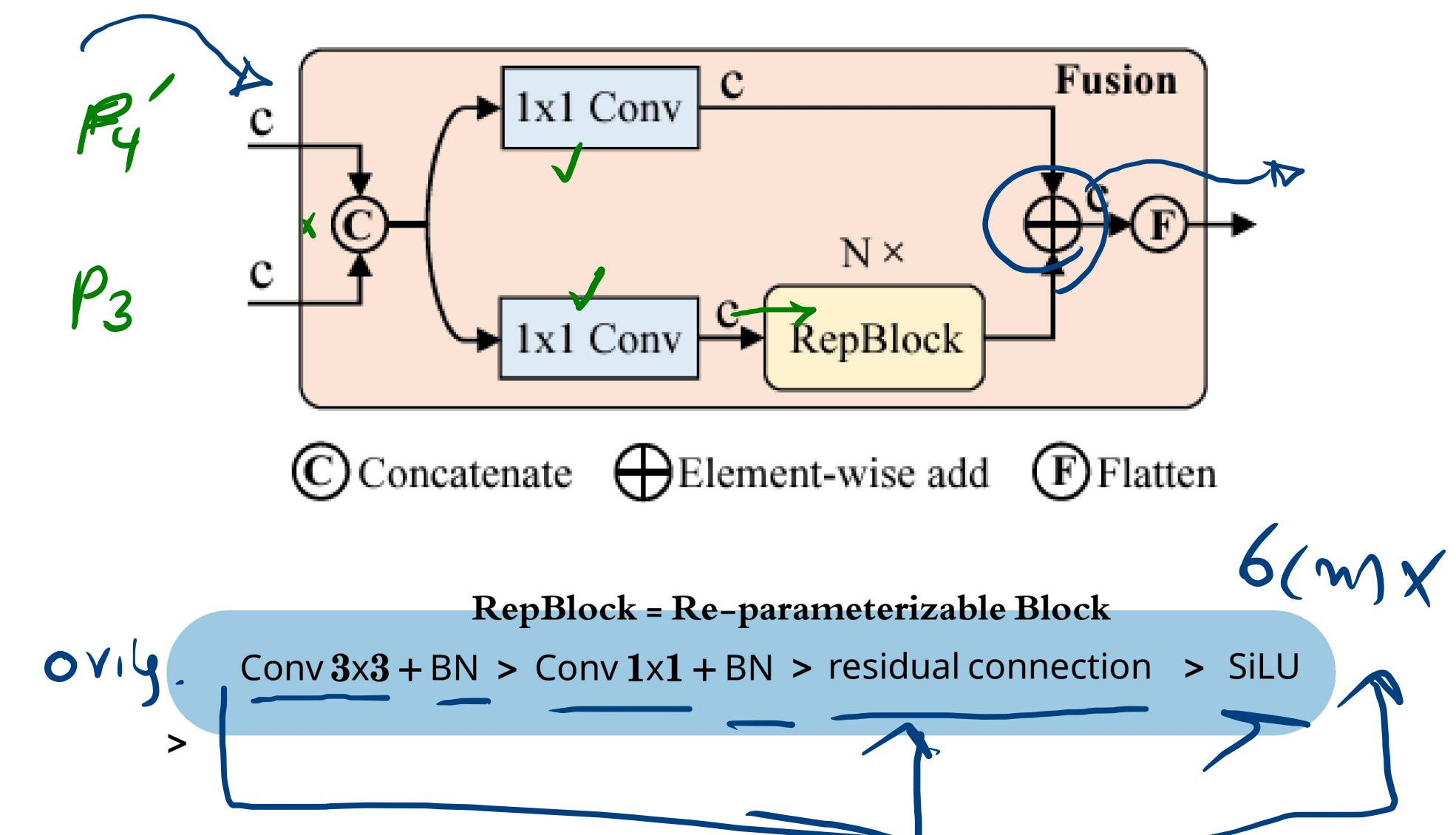
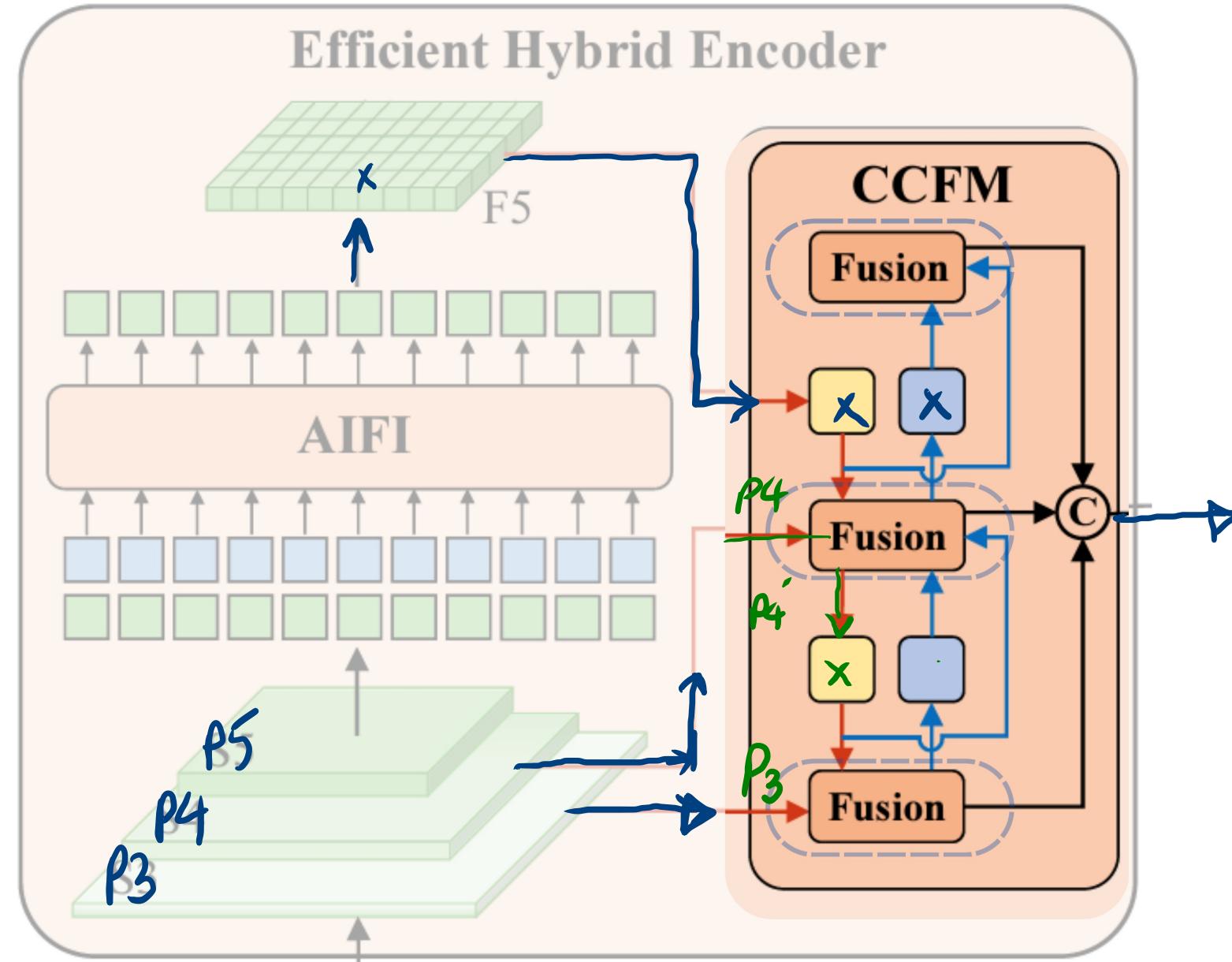
FM

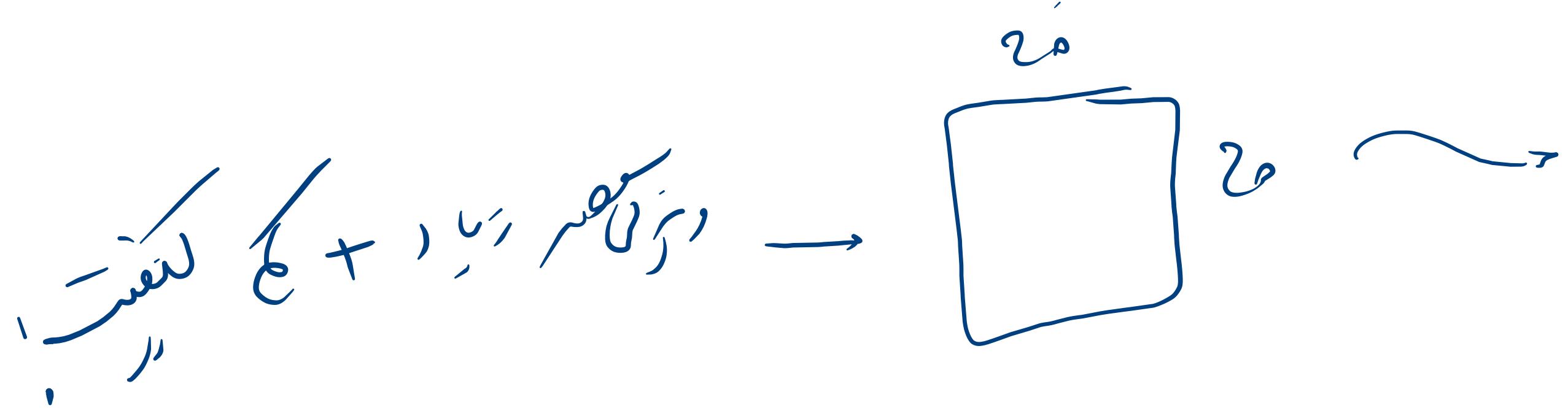
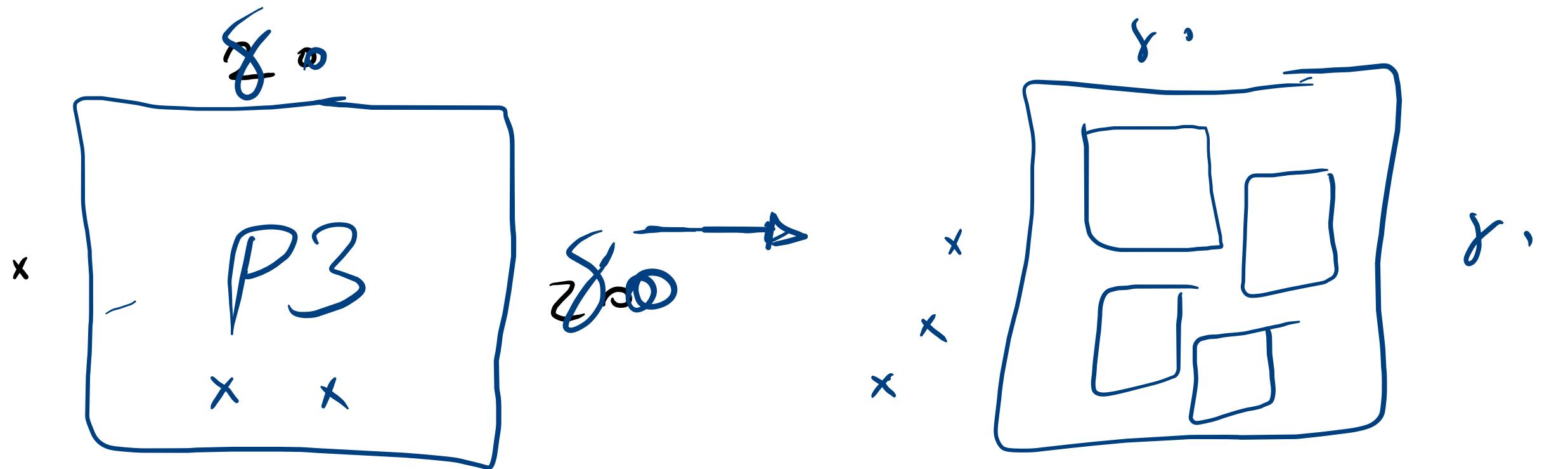
$$\begin{matrix} 8 & 7 & 3 \\ \hline 1 & 1 & 0 \\ 1 & 2 & 1 \end{matrix} \times \begin{matrix} .2 & .1 & 0 \\ .1 & .2 & .1 \\ .1 & .1 & 0 \end{matrix} = D^x \begin{matrix} 1.6 & .21 & 0 \\ .21 & .2 & 0 \\ .21 & .2 & 0 \end{matrix}$$



Cross-Scale Cross-Feature Fusion Module

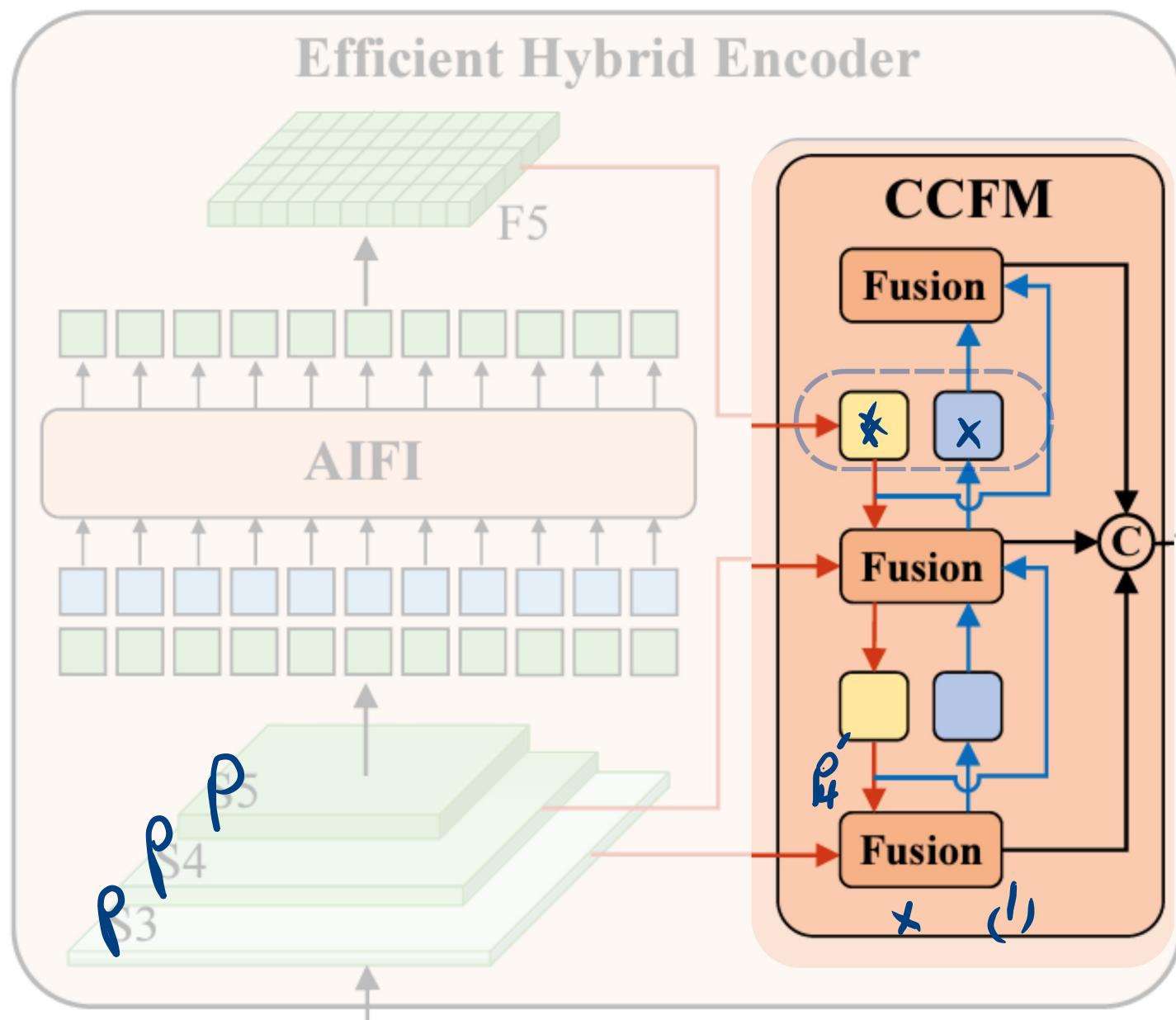
- ترکیب هوشمندانهٔ خروجی‌های چندمقیاسی (P3, P4, P5) پس از پردازش در AIFI
- اطلاعات اشیای کوچک (P3)، متوسط (P4)، و بزرگ (P5) در هم ادغام شوند.



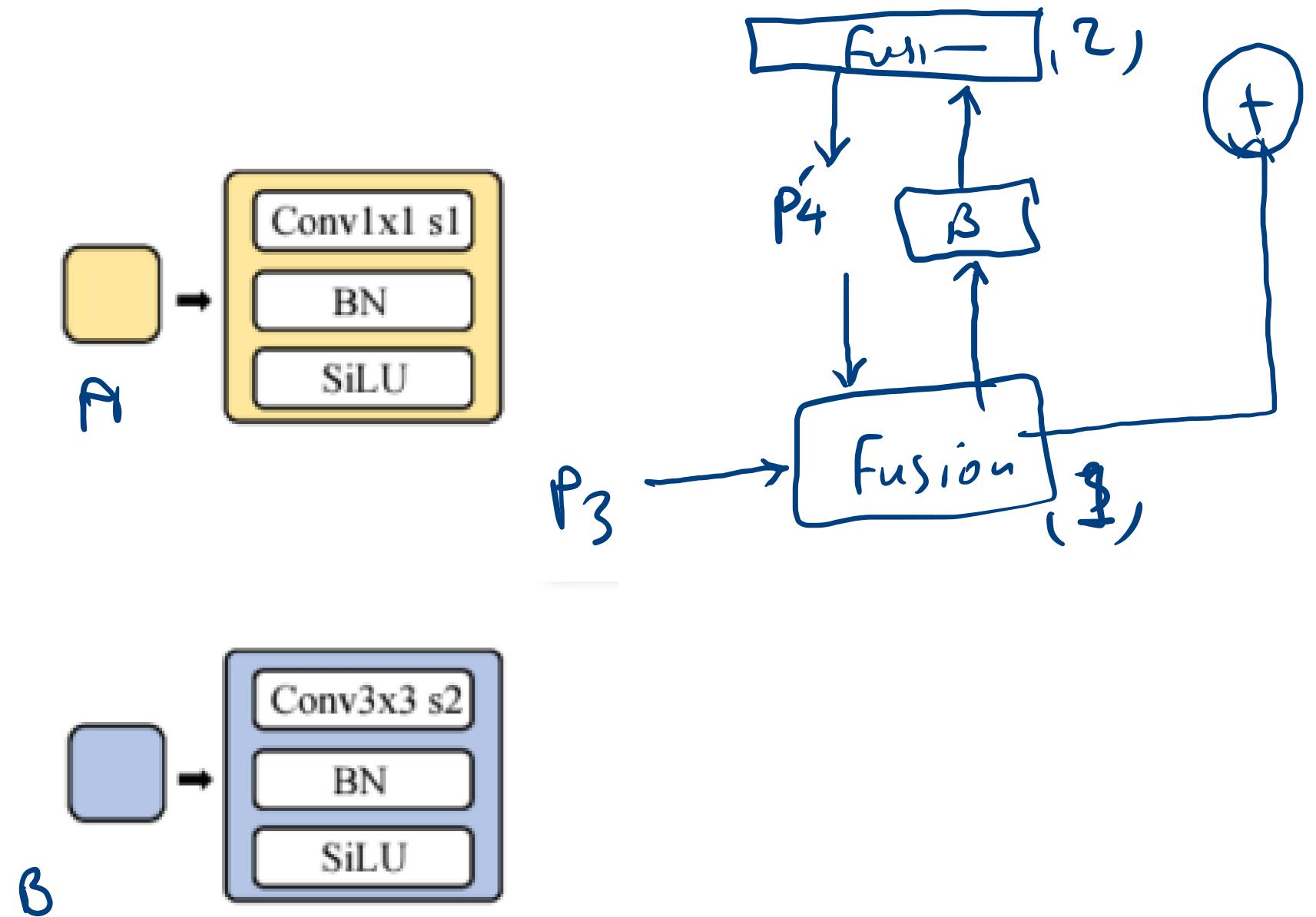


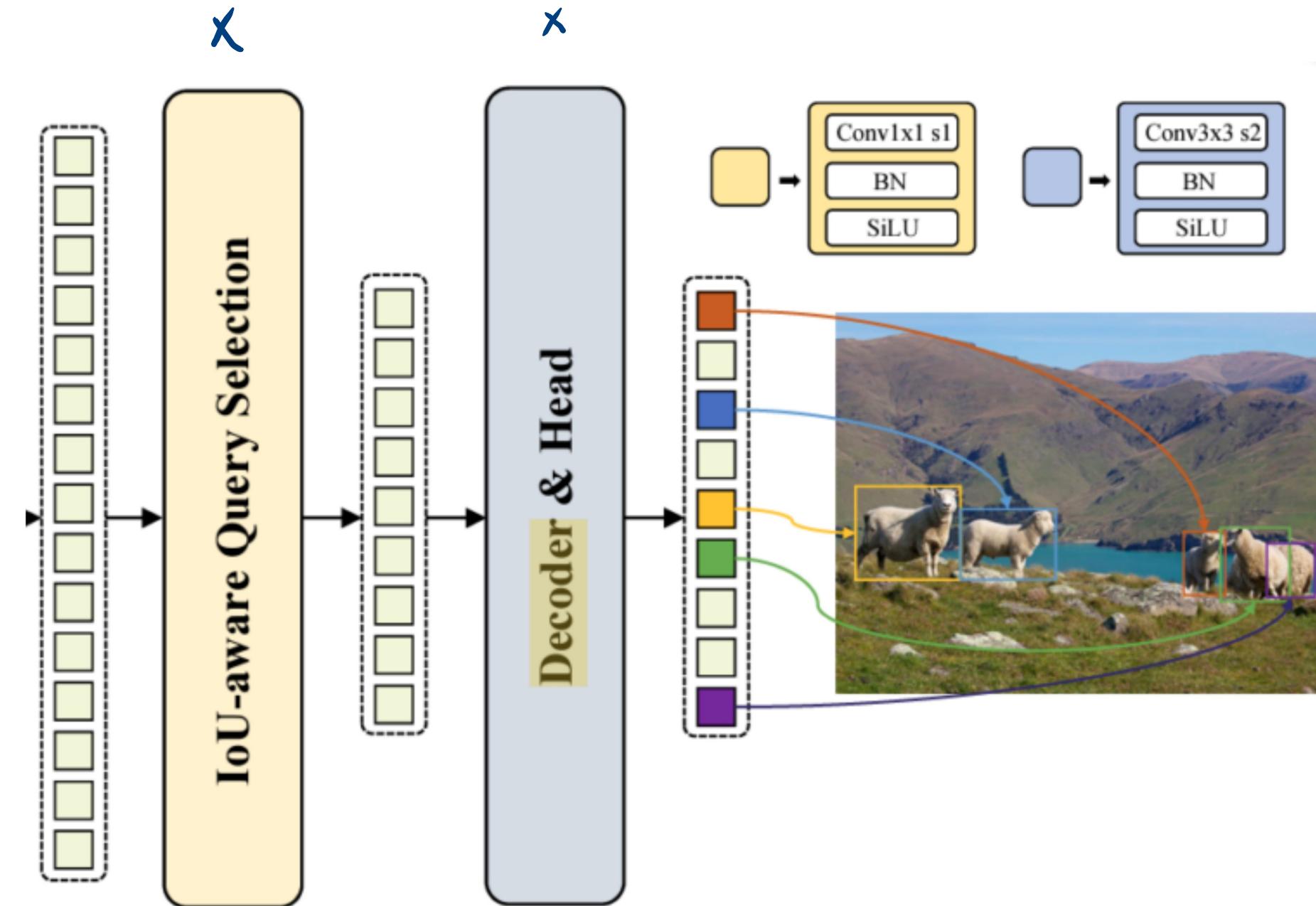
I

Y



Cross-Scale Cross-Feature Fusion Module





The end