



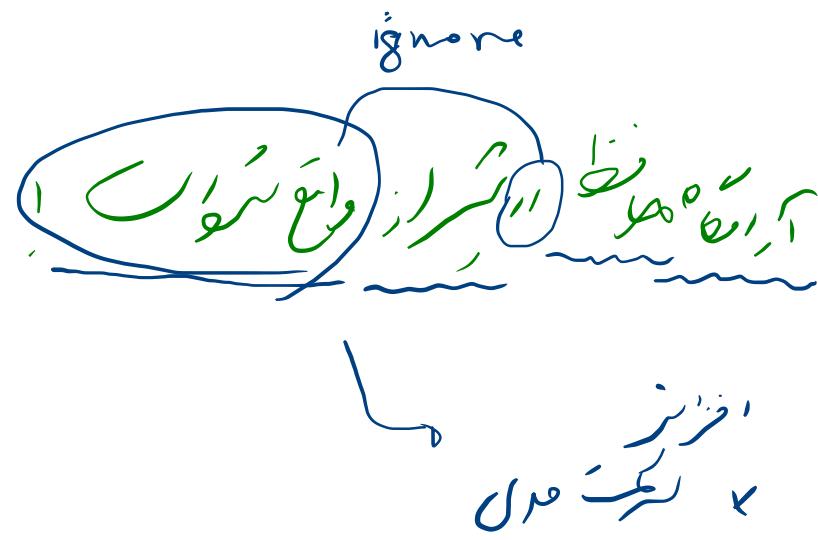
MLP
CN
—

مکانیزم توجه

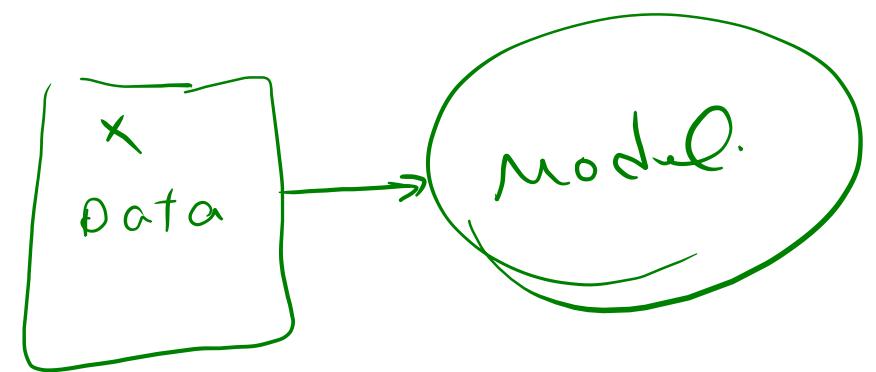
Attention Mechanism

اولین جمله‌ی کوتاهی که با ترکیب واژه‌های زیر چه جمله‌ای به ذهنتان می‌رسد را بگویید!





حافظ
آرامگاه
شیراز



داریوش
تخت جمشید

سیر - روز
داریوش تخت جمشید را ساخت.

آرامگاه حافظ در شیراز قرار دارد.
دایریوش

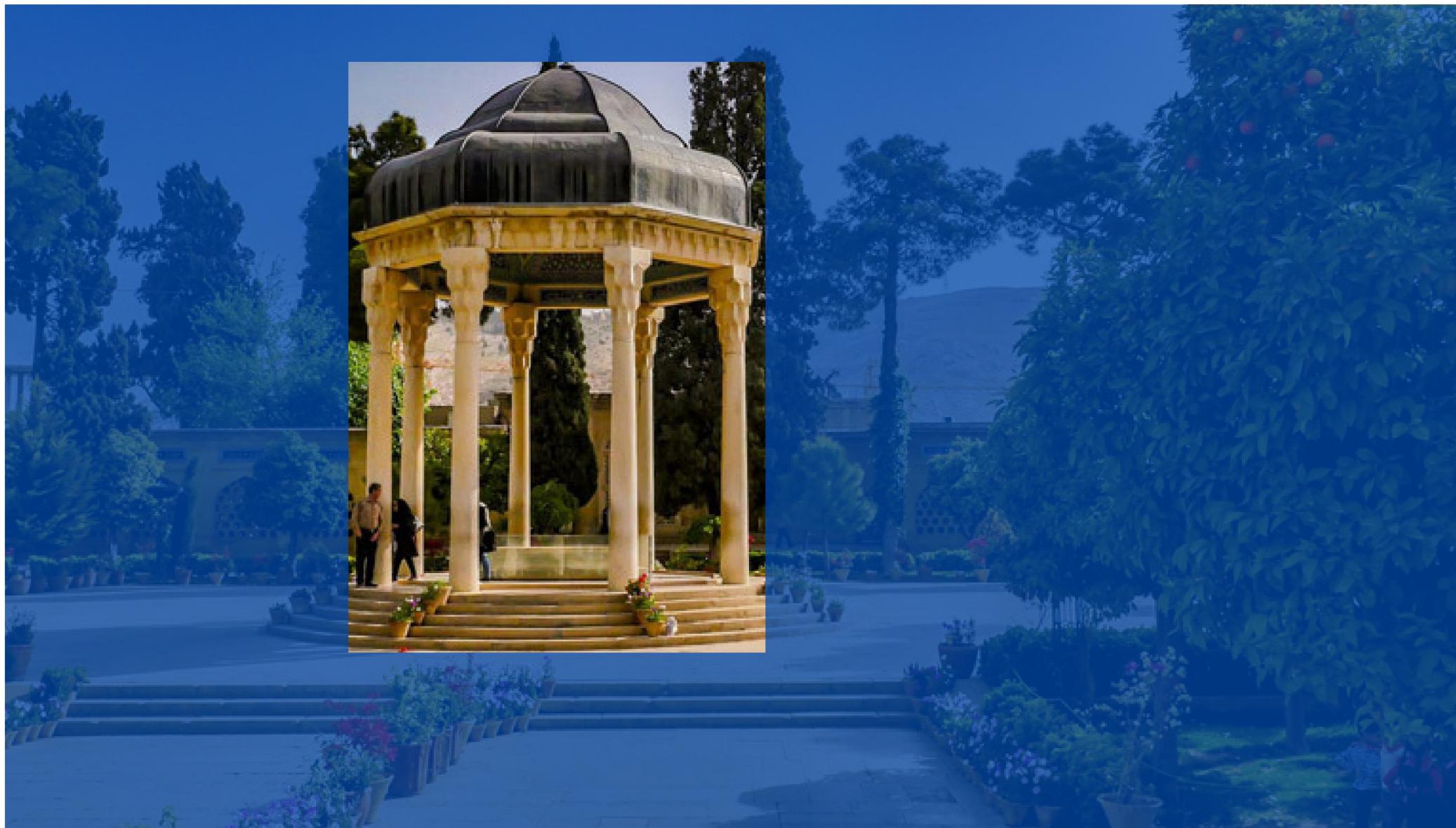
۱ ثانیه به این تصویر نگاه کنید!

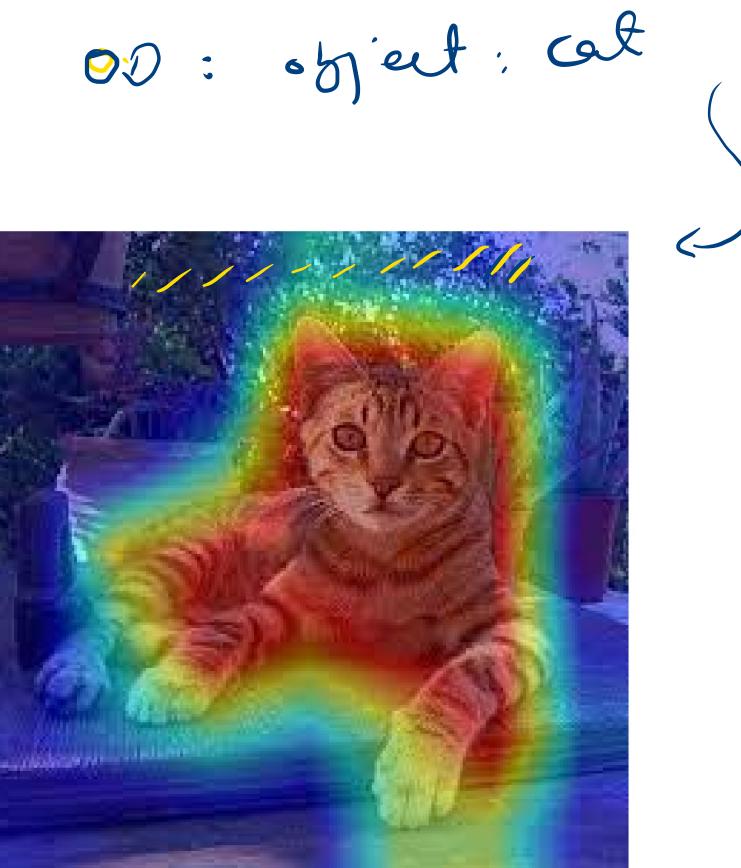


کجا تصویر را بیشتر نگاه کردید؟



مکانیزم توجه چیست ؟

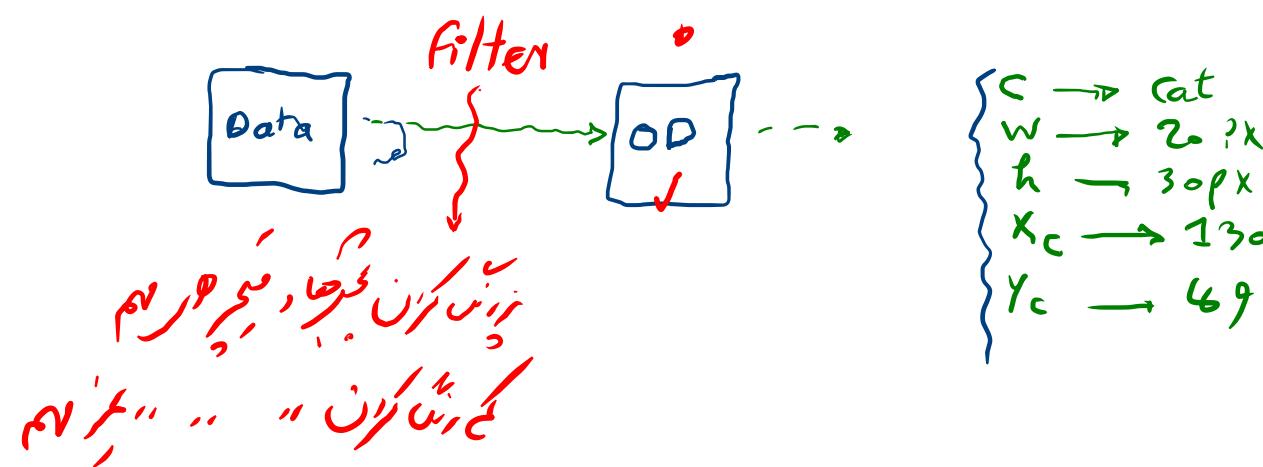




Featuremap
Image.

"توجه" مکانیزمی است که به مدل کمک می‌کند تا روی بخش‌های مهم تصویر (در بینایی کامپیوتر) تمرکز کند، به جای آن که تمام تصویر را به یک میزان پردازش کند.

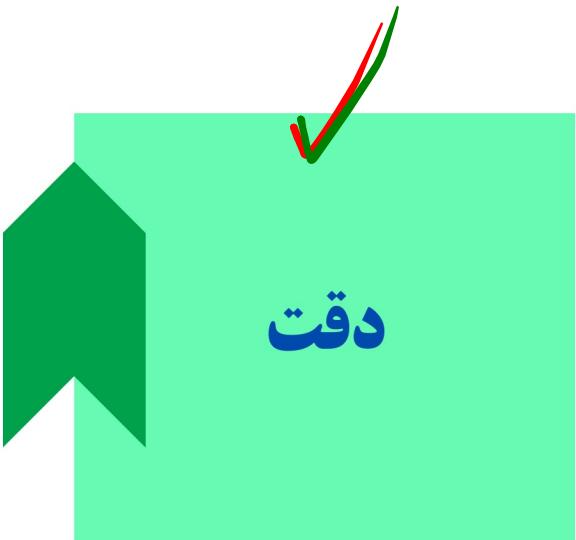
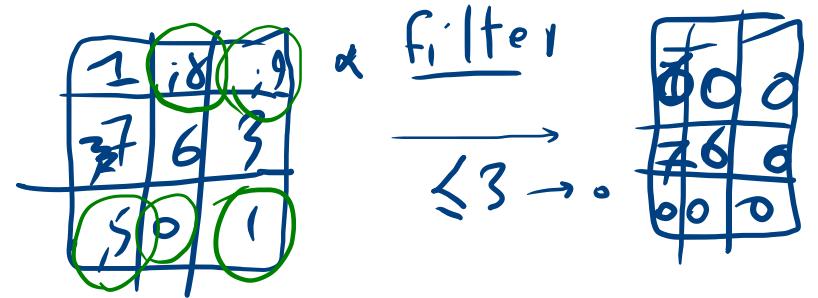
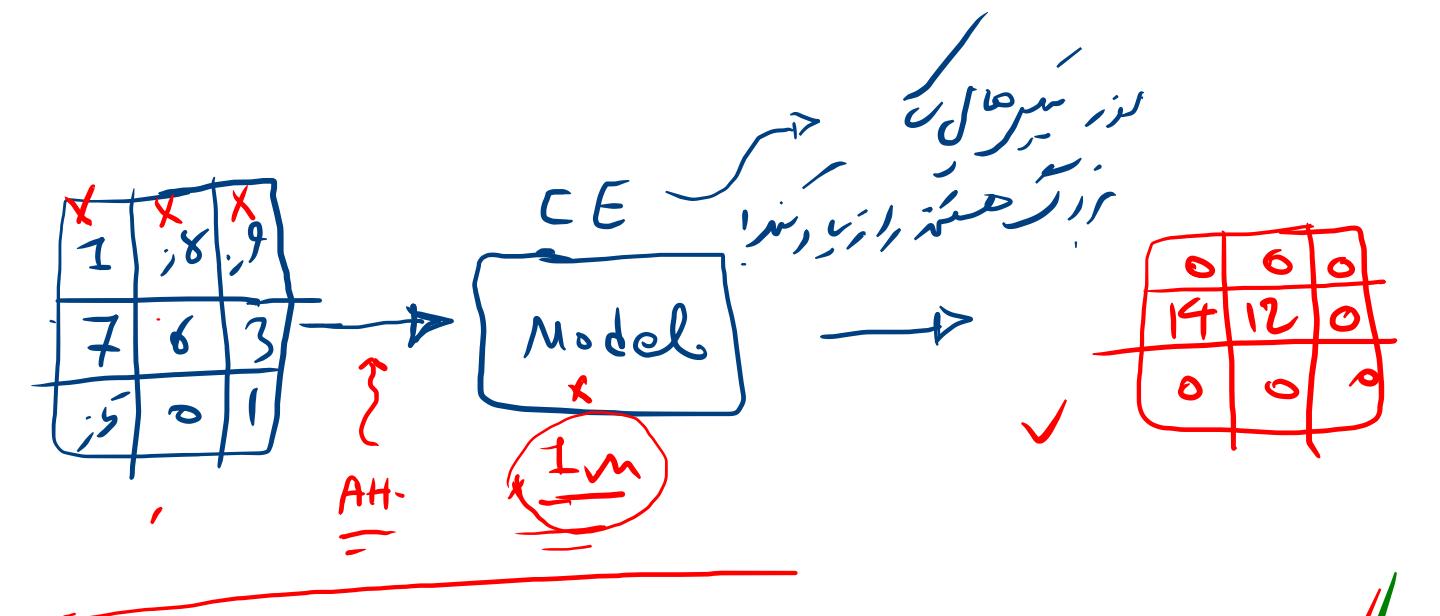
هدف اصلی مکانیزم توجه، تمرکز روی مناطق مهم تصویر که اطلاعات بیشتری دارند (مثل چهره‌ها، اشیاء، لبه‌ها، و غیره) و کاهش اهمیت بخش‌هایی که کمتر مفید هستند (مثل پس زمینه صاف).



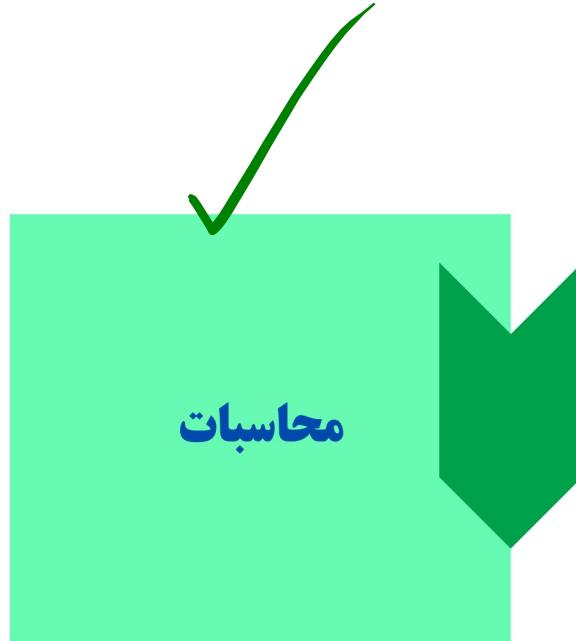
$$\left\{ \begin{array}{l} c \rightarrow \text{cat} \\ w \rightarrow 2 \times ?x \\ h \rightarrow 30px \\ x_c \rightarrow 130 \\ y_c \rightarrow 49 \end{array} \right.$$

مکانیزم توجه : چرا ؟

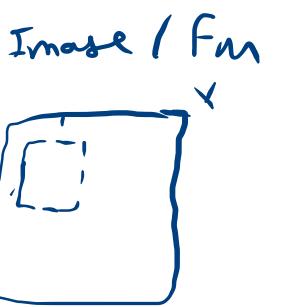
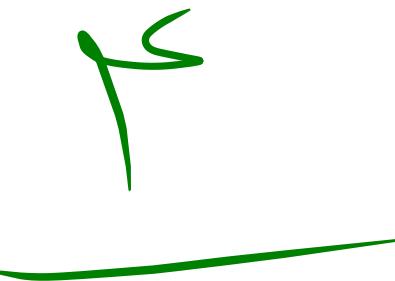
✓



باعث افزایش دقت در کارهایی مثل تشخیص اشیاء،
طبقه‌بندی تصاویر و تشخیص چهره می‌شود.

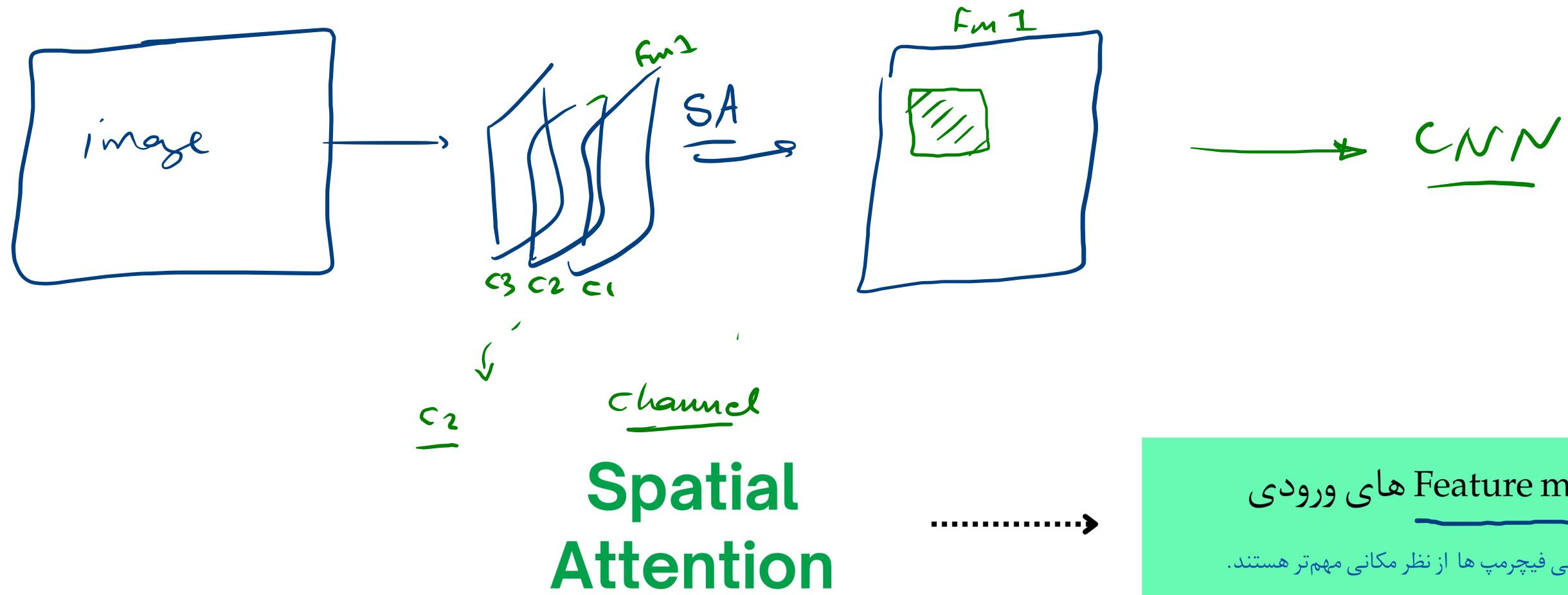


باعث کاهش محاسبات بی‌مورد در بخش‌های غیرضروری
تصویر می‌شود.



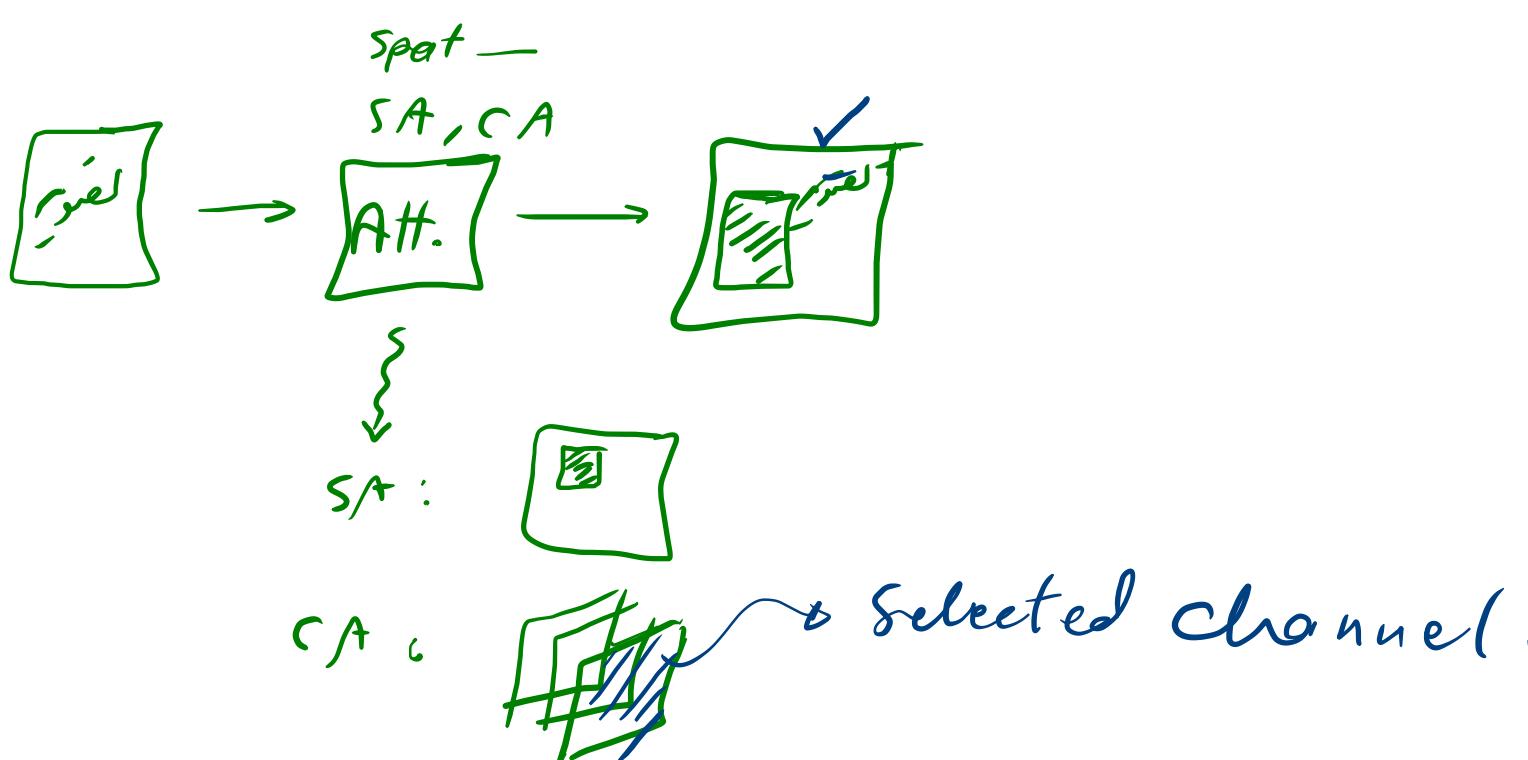
مکانیزم توجه فضایی

Spatial Attention

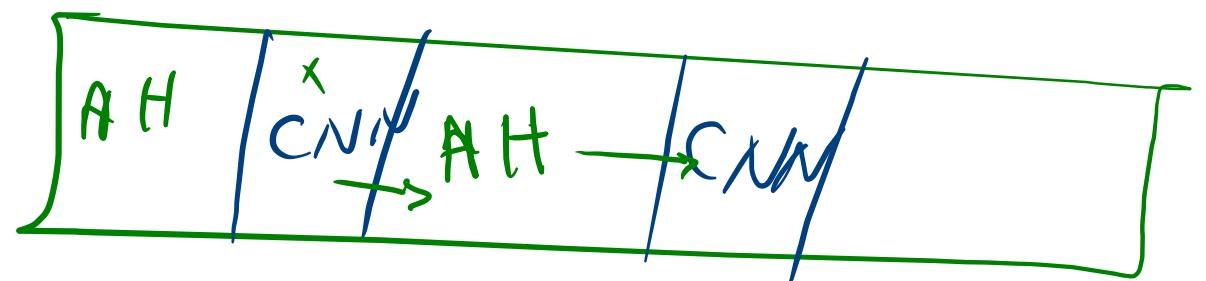


جاهای مهم Feature map های ورودی

مدل یاد می‌گیرد که کدام نواحی فیچرمپ‌ها از نظر مکانی مهم‌تر هستند.



مکانیزم توجه فضایی (Spatial Attention)

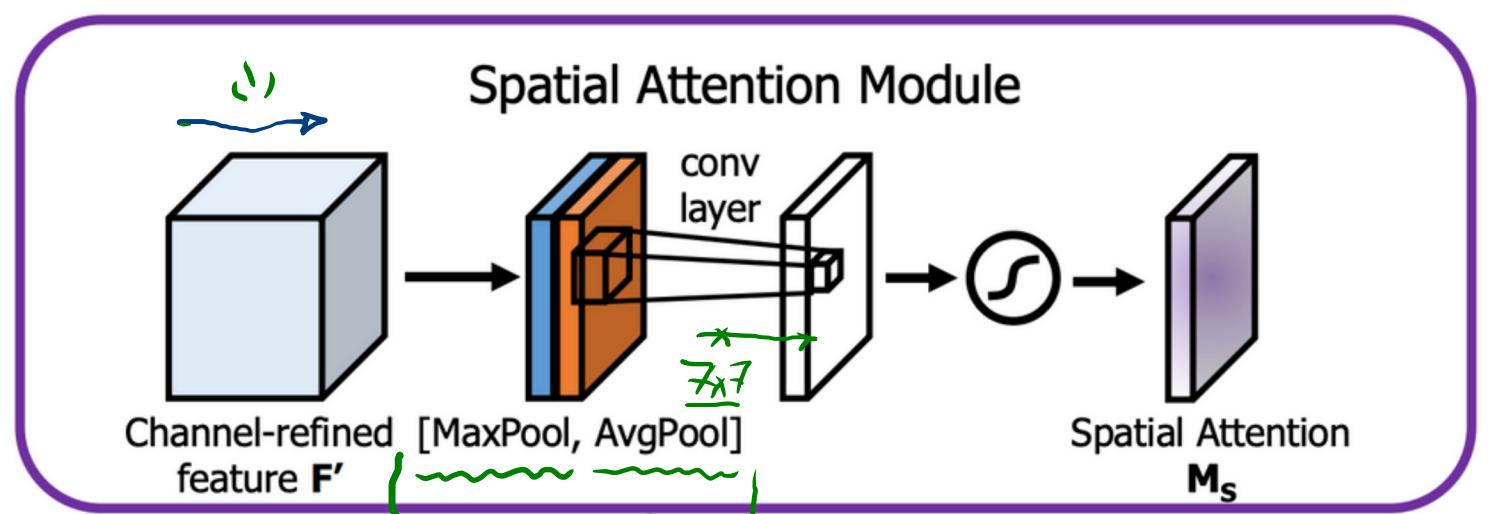


$$\rightarrow M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)]))$$

۱. ورودی فیچرmap از لایه قبل با ابعاد مثلاً خروجی یک لایه کانولوشنی) که ابعادی به صورت $[B, C, H, W]$

\times

- B: Batch size .
- C: تعداد کانال‌ها .
- H: ارتفاع تصویر .
- W: عرض تصویر .



<https://www.mdpi.com/2076-3417/12/8/3846>

۲. اعمال روی کانال‌ها به صورت channel wise

۳. اعمال روی کانال‌ها به صورت max pooling

۴. feature map دو concatenate .

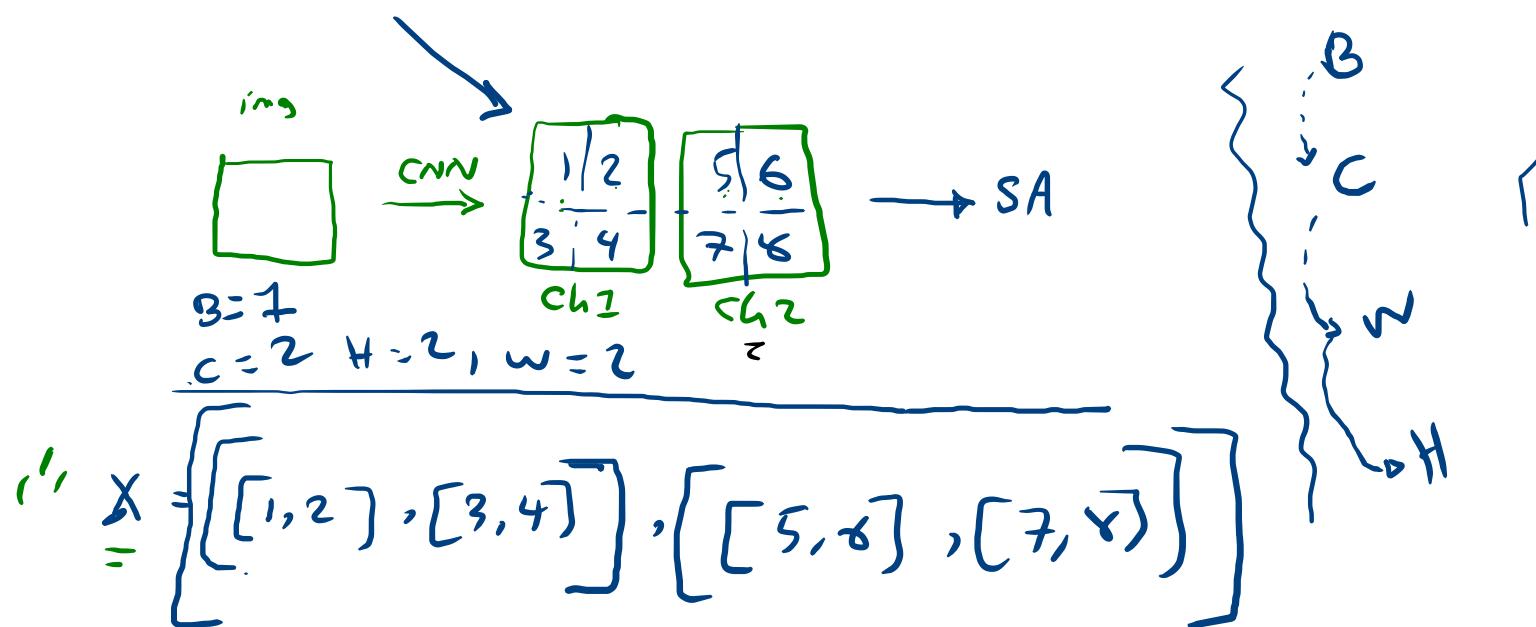
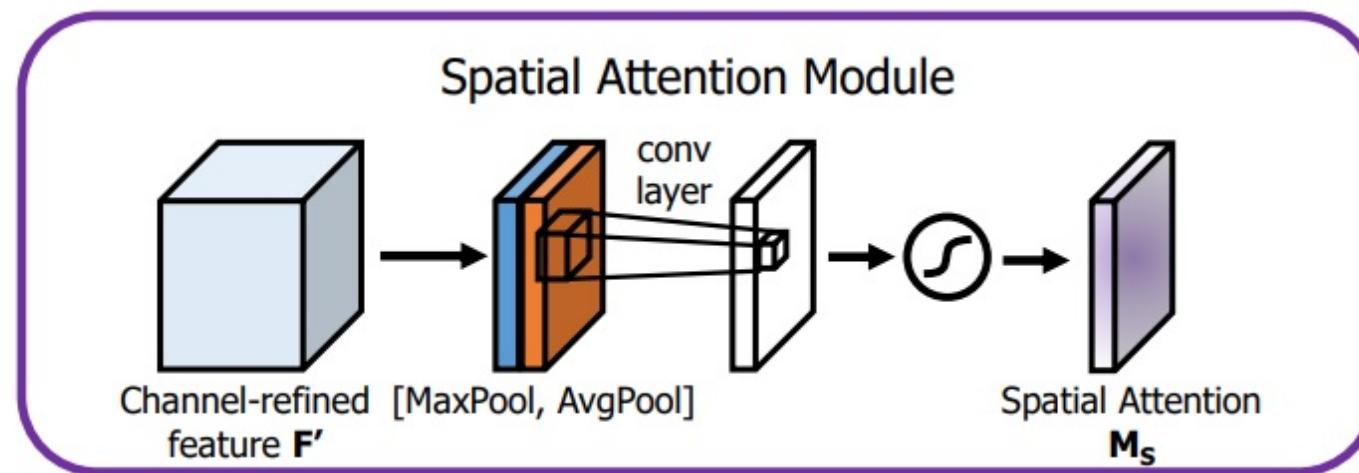
۵. اعمال کانولوشن با کرنل 7×7 روی آن

۶. اعمال سیگموید روی خروجی گام قبل

۷. ضرب کردن feature map در attention map اولیه

{ ۰, ۱ }

لسا لسا



6. attention-map:

Sigmoid(conv)

$$\begin{bmatrix} * & 9 & * & 7 \\ * & 3 & * & 9 \\ 9 & * & 7 & * \\ 3 & * & 9 & * \end{bmatrix}$$

7: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} * \begin{bmatrix} 9 & 7 \\ 3 & 9 \end{bmatrix} = \begin{bmatrix} 9 & 7 \\ 3 & 9 \end{bmatrix}$

att.map

ch1-modified

2, avg-pooling = $\left[\left[\frac{1+5}{2}, \frac{2+6}{2} \right], \left[\frac{3+7}{2}, \frac{4+8}{2} \right] \right] = \left[[3, 4], [5, 6] \right]$

3, max-pooling: $\left[[\max(1, 5), \max(2, 6)], [\max(3, 7), \max(4, 8)] \right] \Rightarrow \left[[5, 6], [7, 8] \right]$

4) concat = $\left[[3, 4], [5, 6], [5, 6], [7, 8] \right]$

5: conv = conv2D($k=7$, concat)

مکانیزم توجه کانالی

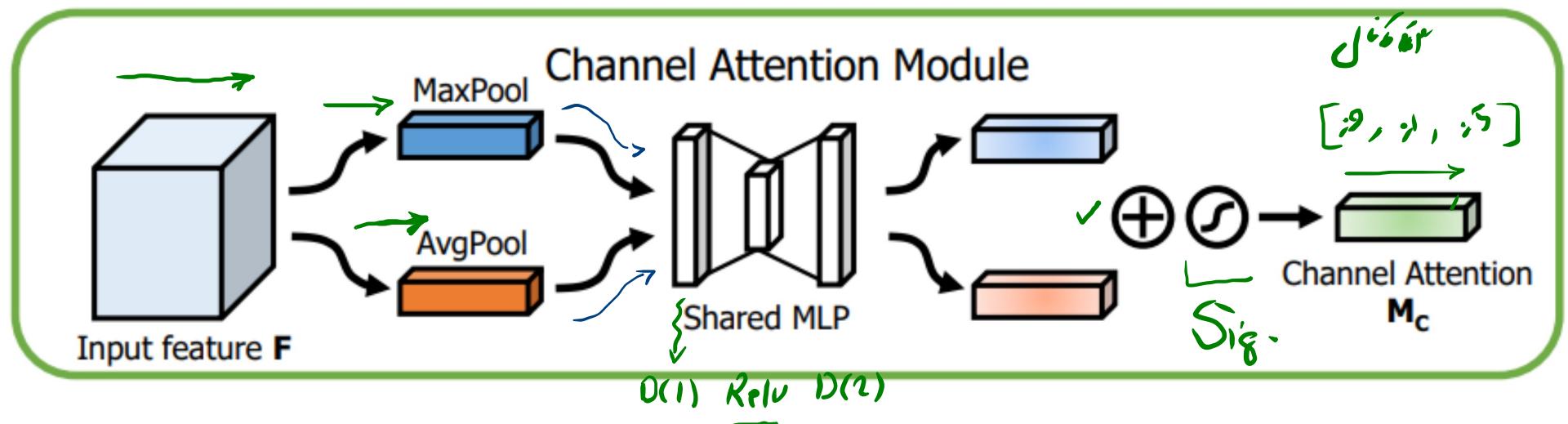
Channel Attention

Channel Attention

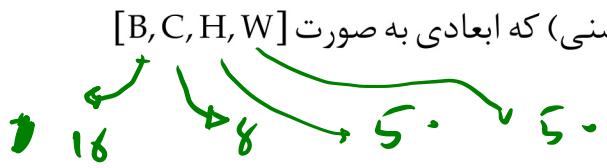


به جای تمرکز بر مکان‌های خاص در نظریه f_m (مثل spatial attention)، مدل یاد می‌گیرد کدام کانال‌ها (ویژگی‌ها) از اهمیت بیشتری برخوردارند.

$$Mc(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F)))$$



<https://www.mdpi.com/2076-3417/12/8/3846>

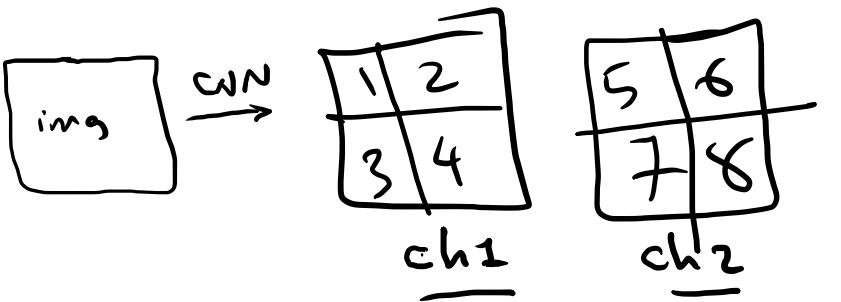
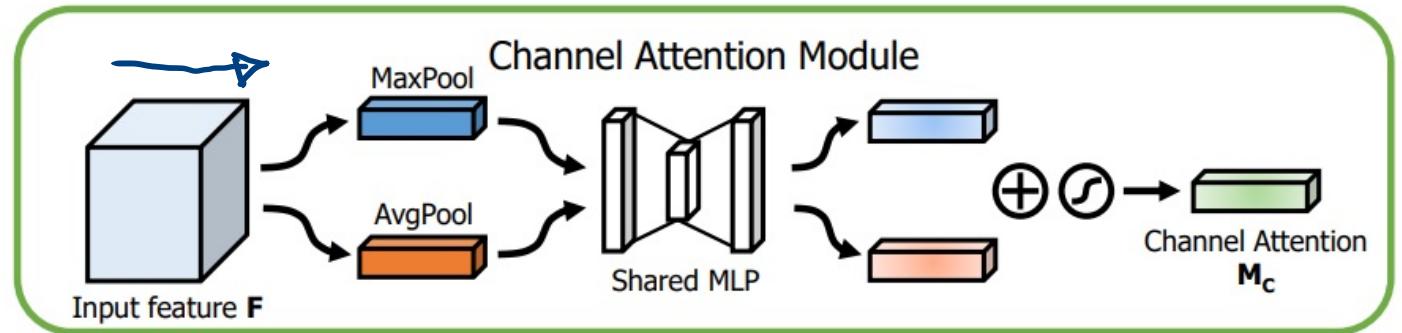


✓

1. ورودی فیچرمپ از لایه قبل با ابعاد مثلاً خروجی یک لایه کانولوشنی) که ابعادی به صورت $[B, C, H, W]$ است!
2. B: Batch size .
3. C: تعداد کانال‌ها .
4. H: ارتفاع تصویر .
5. W: عرض تصویر .

۲. اعمال روی هر کanal (Global Pooling و Global Average Pooling)

1. ورود خروجی لایه‌ی قبل به لایه MLP (این لایه به صورت shared است!)
2. جمع خروجی‌های دو مسیر و اعمال sigmoid
3. ضرب کردن attention map در feature map اصلی



$H = 2$
 $W = 2$
 $C = 2$
 $B = 1$

$X = \left[\left[[1, 2], [3, 4] \right], \left[[5, 6], [7, 8] \right] \right]$

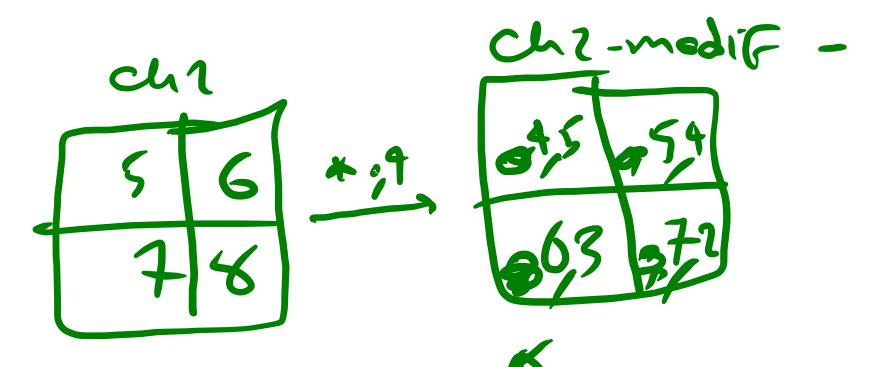
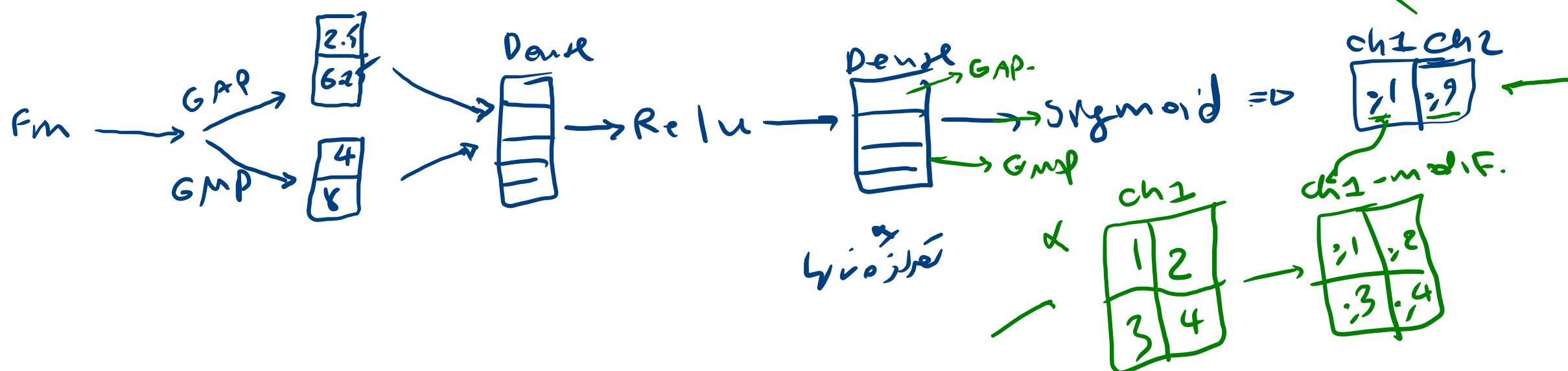
$GAP = \left[\left[\frac{1+2+3+4}{4} \right], \left[\frac{5+6+7+8}{4} \right] \right] = \left[[2.5], [6.25] \right]$

$GMP = \left[[4], [8] \right]$

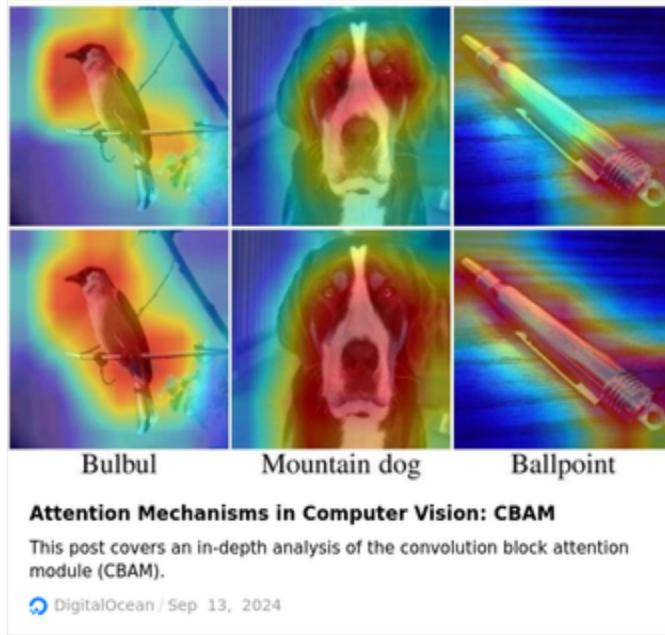
②

$GAP: [2.5], [6.25]$

$GMP: [4], [8]$



[https://arxiv.org/pdf/2111.07624](https://arxiv.org/pdf/2111.07624.pdf)



https://openaccess.thecvf.com/content_ECCV_2018/papers/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.pdf