

موضع

InternImage

Semantic Segmentation

**InternImage: Exploring Large-Scale Vision Foundation Models with
Deformable Convolutions**

Wenhai Wang^{1*}, Jifeng Dai^{2,1*}, Zhe Chen^{3,1*}, Zhenhang Huang^{1*}, Zhiqi Li^{3,1*}, Xizhou Zhu^{4*},
Xiaowei Hu¹, Tong Lu³, Lewei Lu⁴, Hongsheng Li⁵, Xiaogang Wang^{4,5}, Yu Qiao^{1✉}

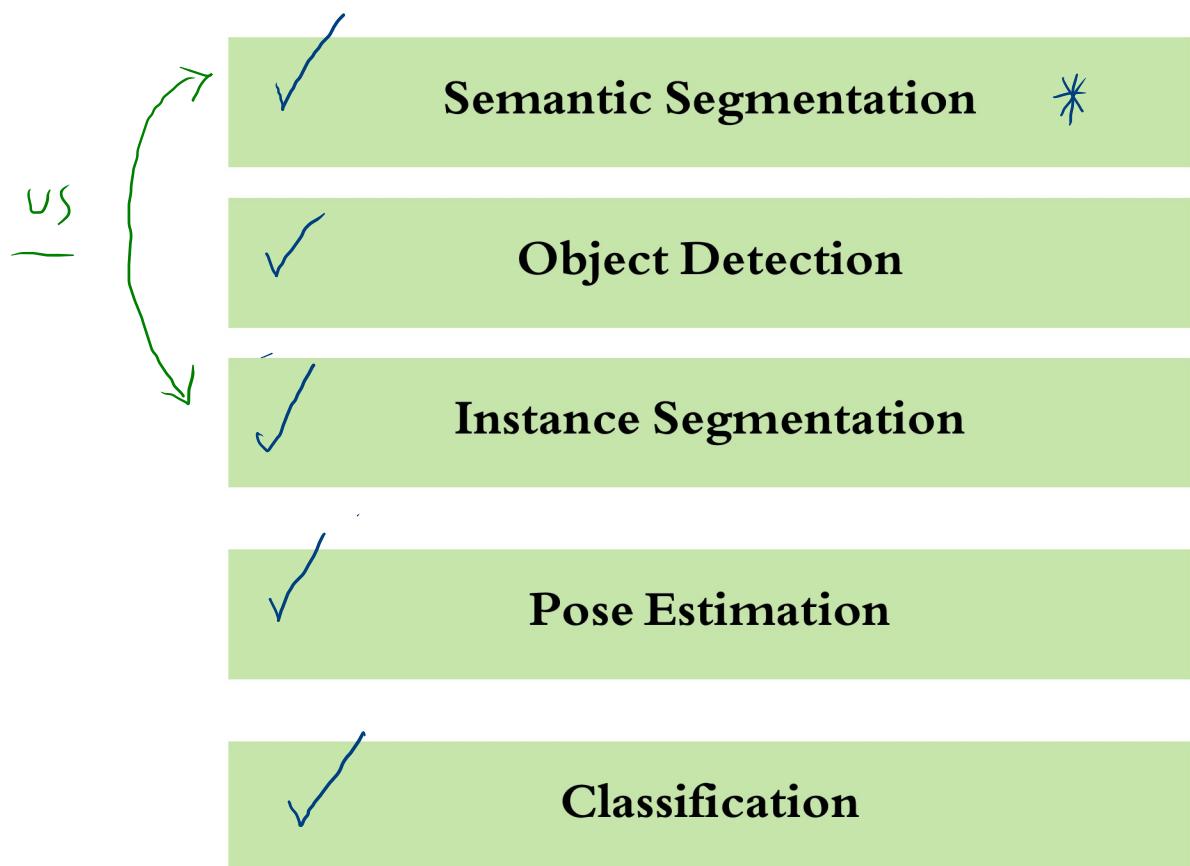
¹Shanghai AI Laboratory ²Tsinghua University

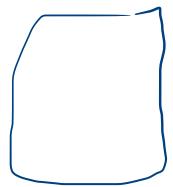
³Nanjing University ⁴SenseTime Research ⁵The Chinese University of Hong Kong

<https://github.com/OpenGVLab/InternImage>

internimage

کاربردها

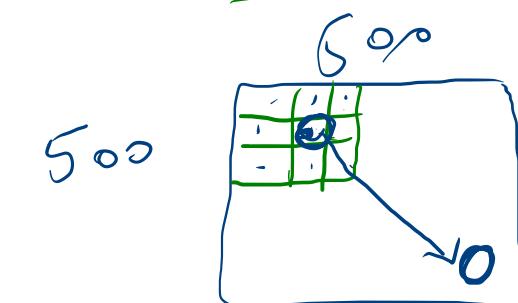
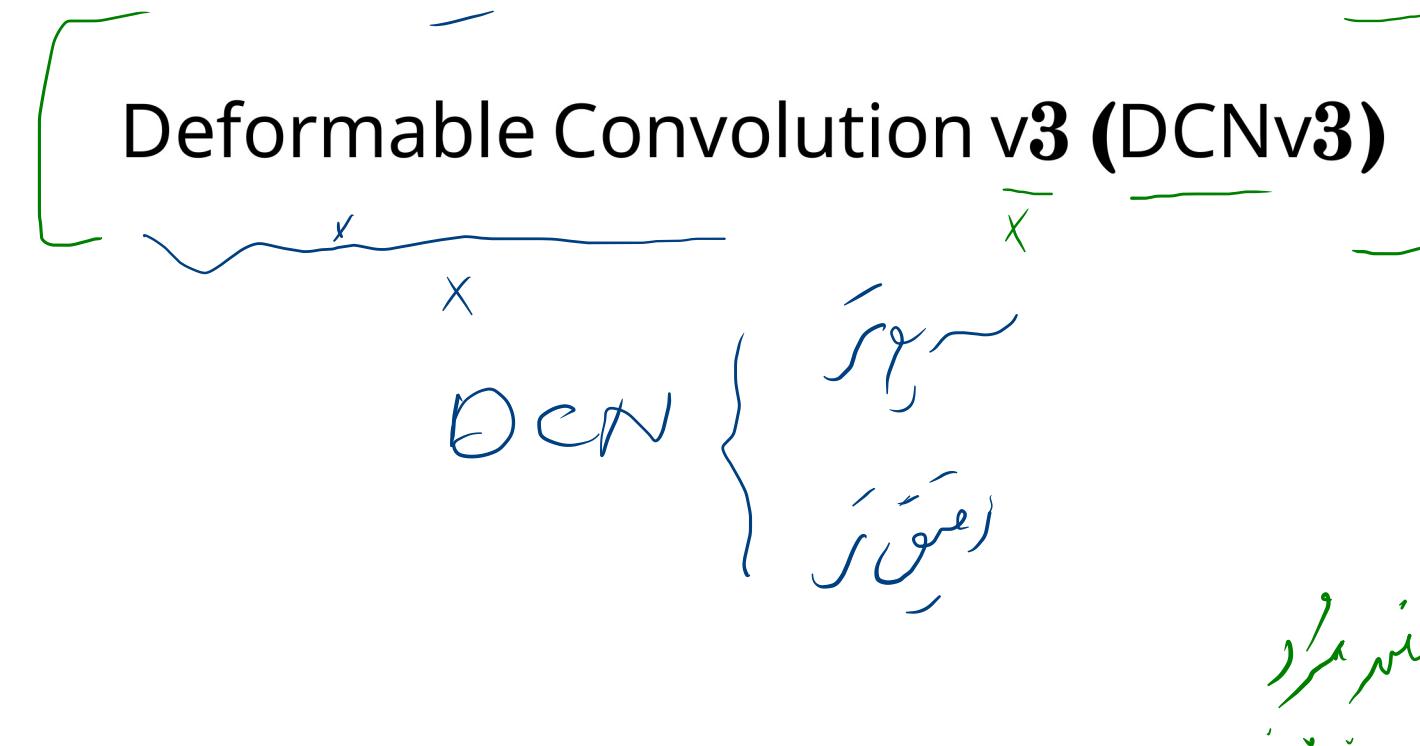




نحوه انتقال دهنده

inference time -

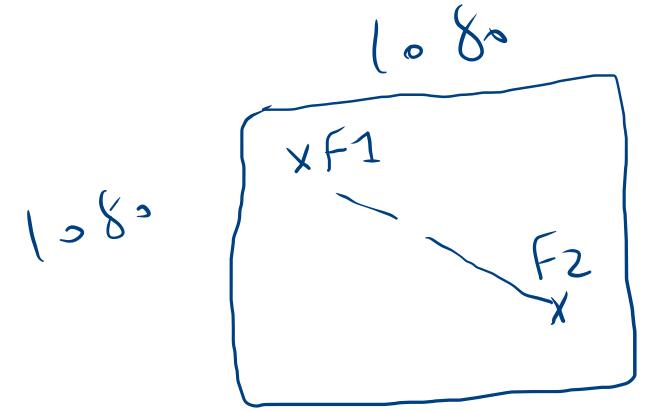
performance met.



- دارای میدان دید پویا (long-range dependencies) است.
- از نظر محاسباتی و مصرف حافظه بهینه است.

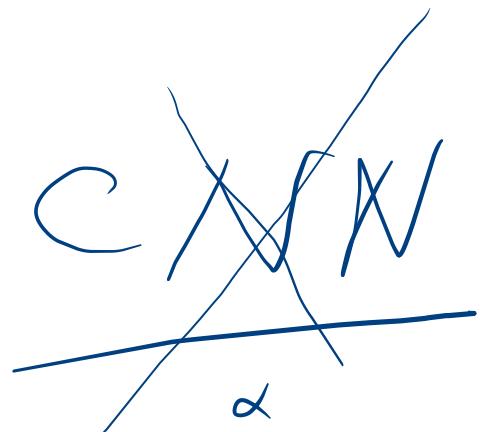
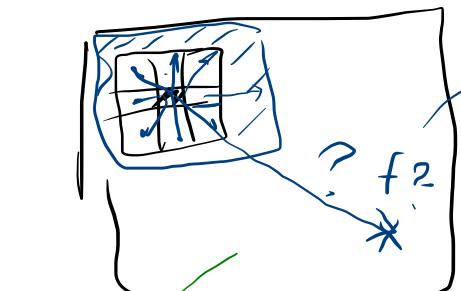
نوآوری
intern image

Long Range Dependencies -



issue: resource consumption -

no interleaving
kernel



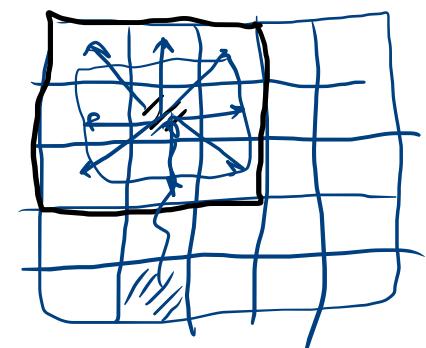
vs VIT

parallel
DCN

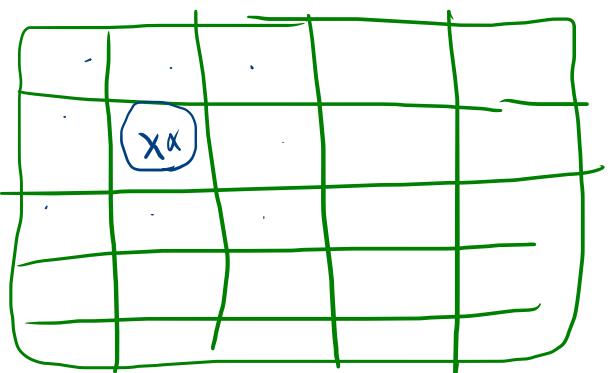
OCN

—

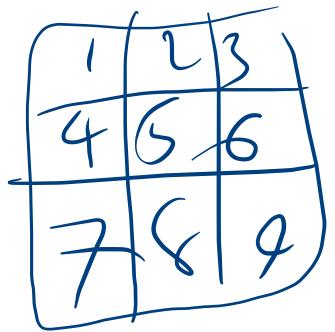
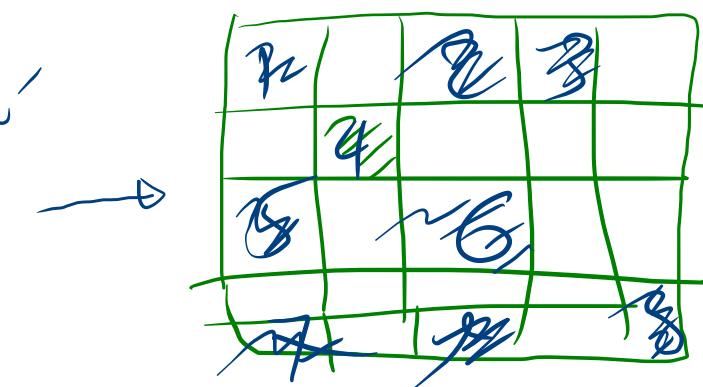
Trad. conv.



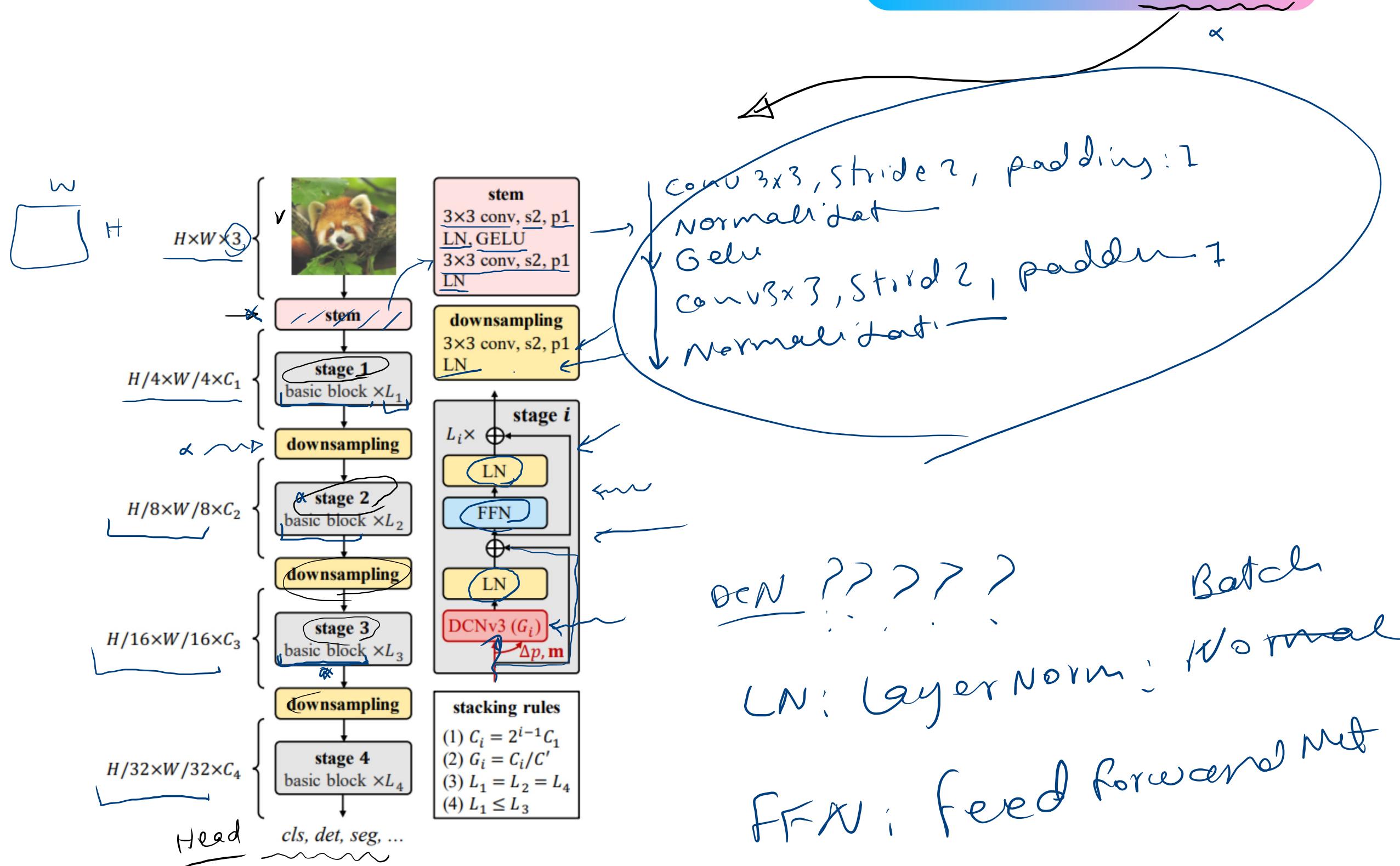
DCN



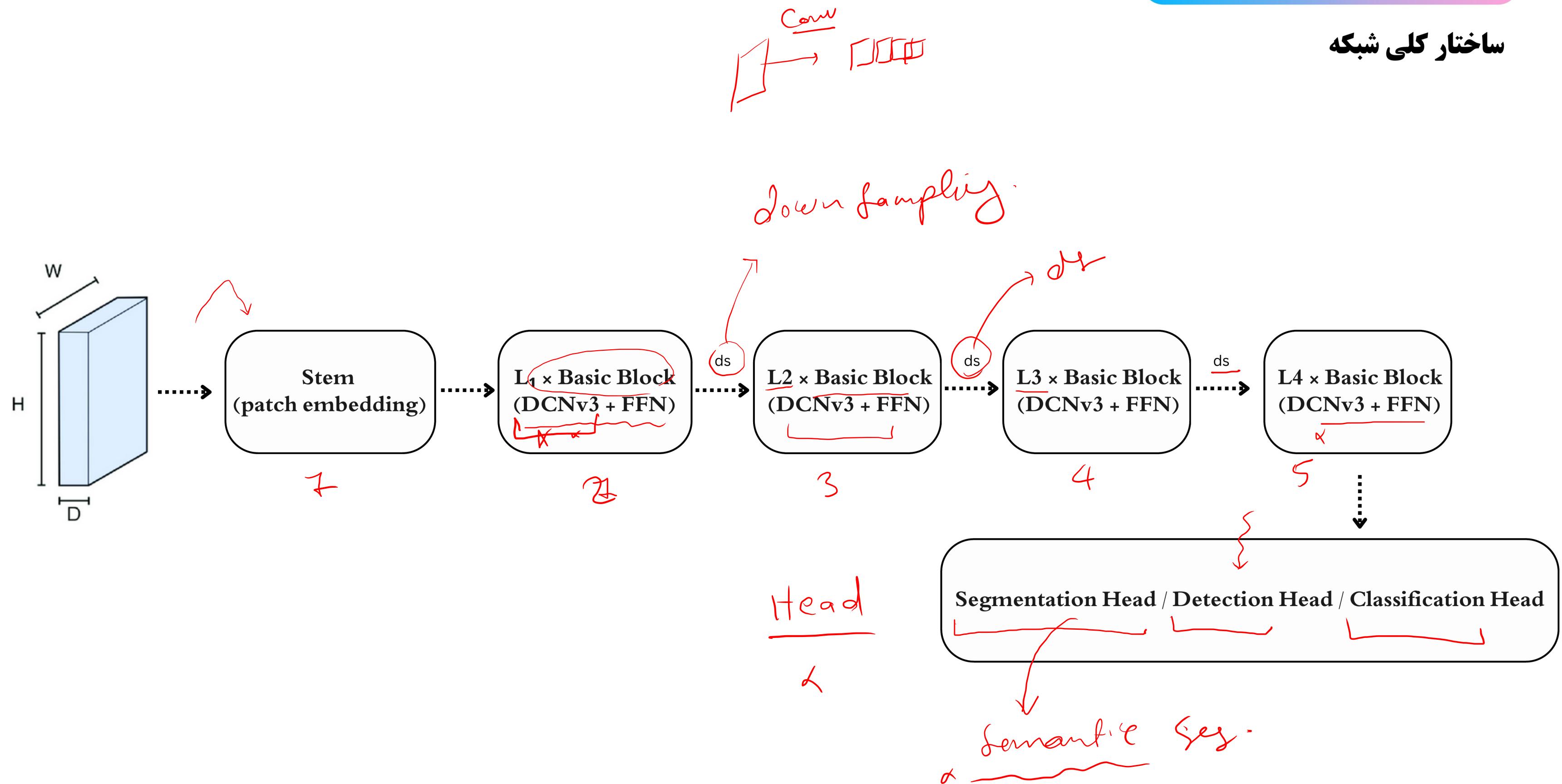
العنصر المختار
العنصر المختار
العنصر المختار



Semantic seg: internimage

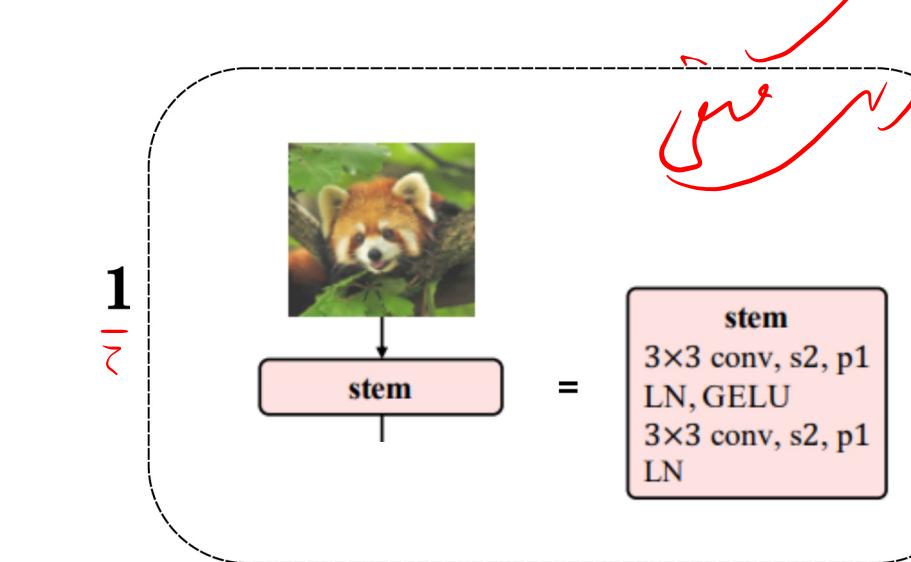


ساختار کلی شبکه



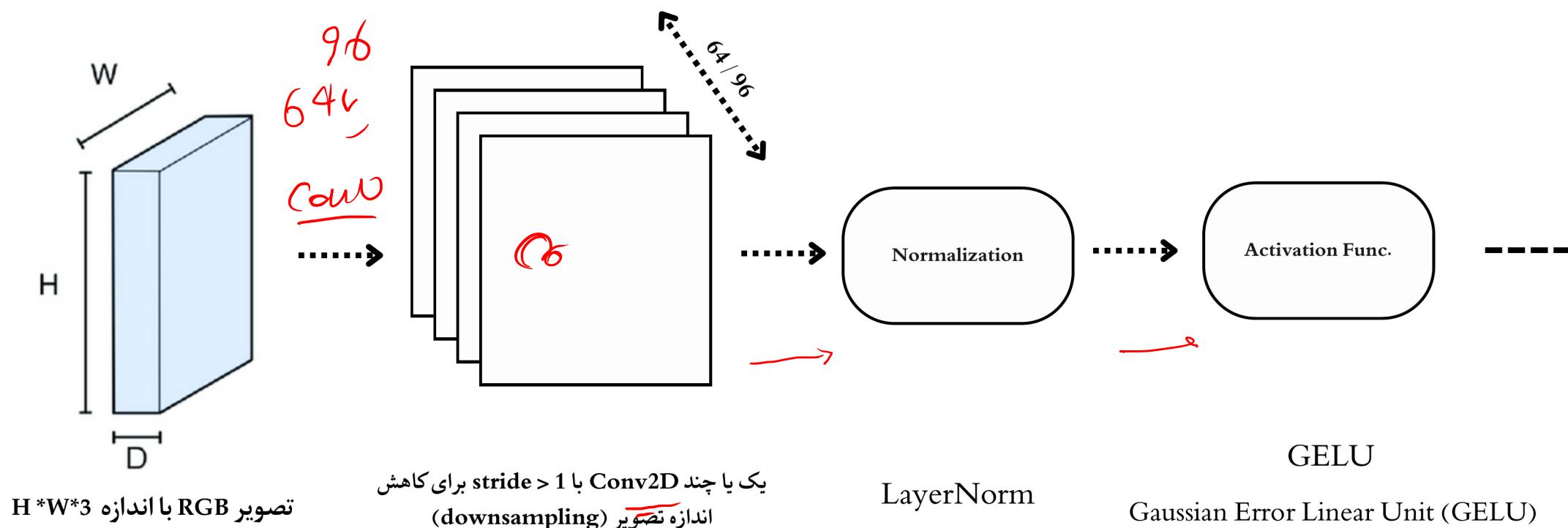
$$\text{ReLU} \rightarrow X < 0 \rightarrow X = 0$$

$$\text{GELU} \rightarrow \text{GELU}(\text{Input})$$



~~STEM LAYER (PATCH EMBEDDING)~~

تبدیل تصویر ورودی به نقشه ویژگی اولیه برای ورود به بلوک‌های اصلی شبکه

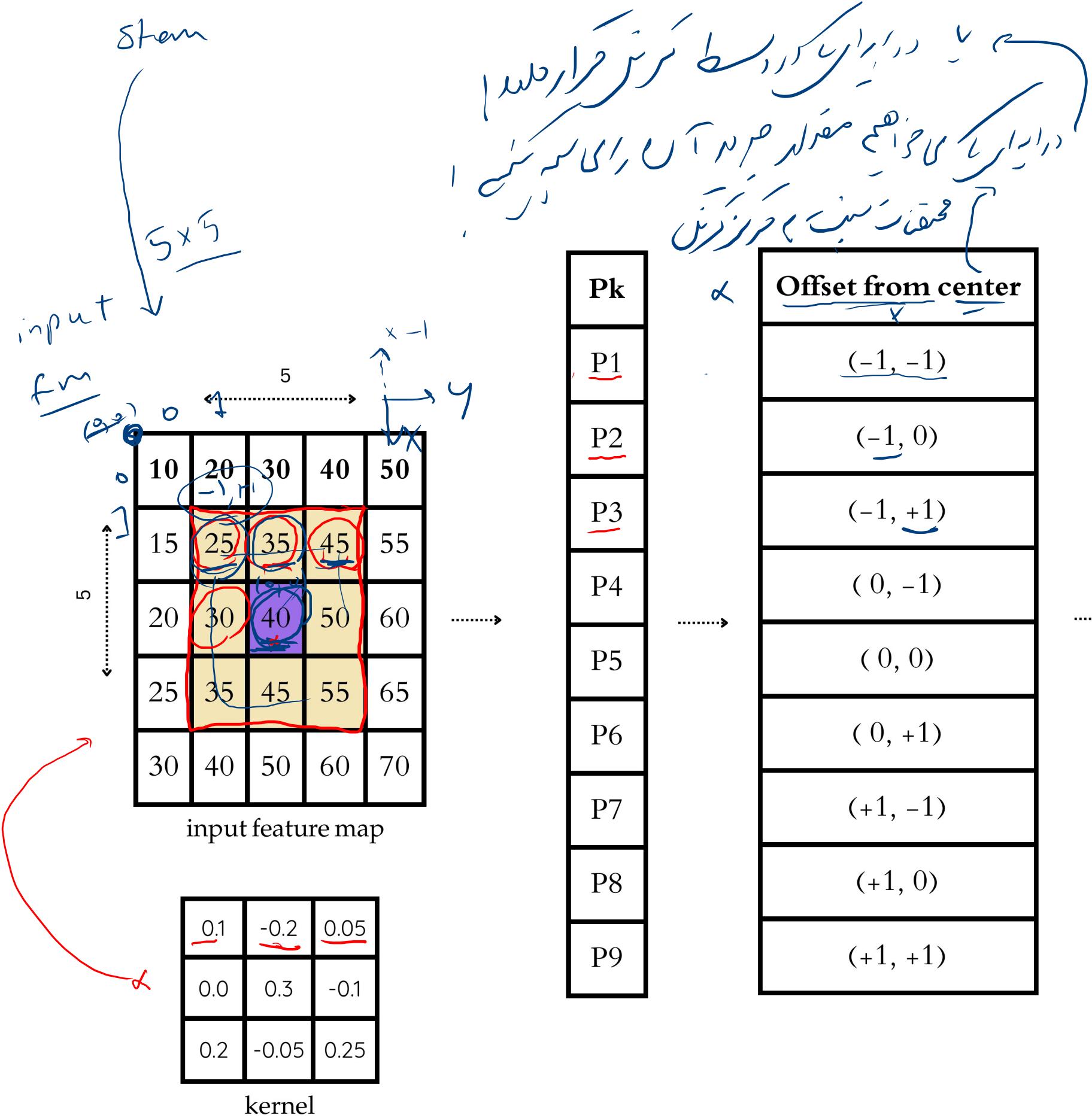


$$GELU(x) = 0.5 \cdot x \cdot \left(1 + \tanh \left(\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right) \right)$$

STAGE 1

Basic Block (DCNv3 + FFN)



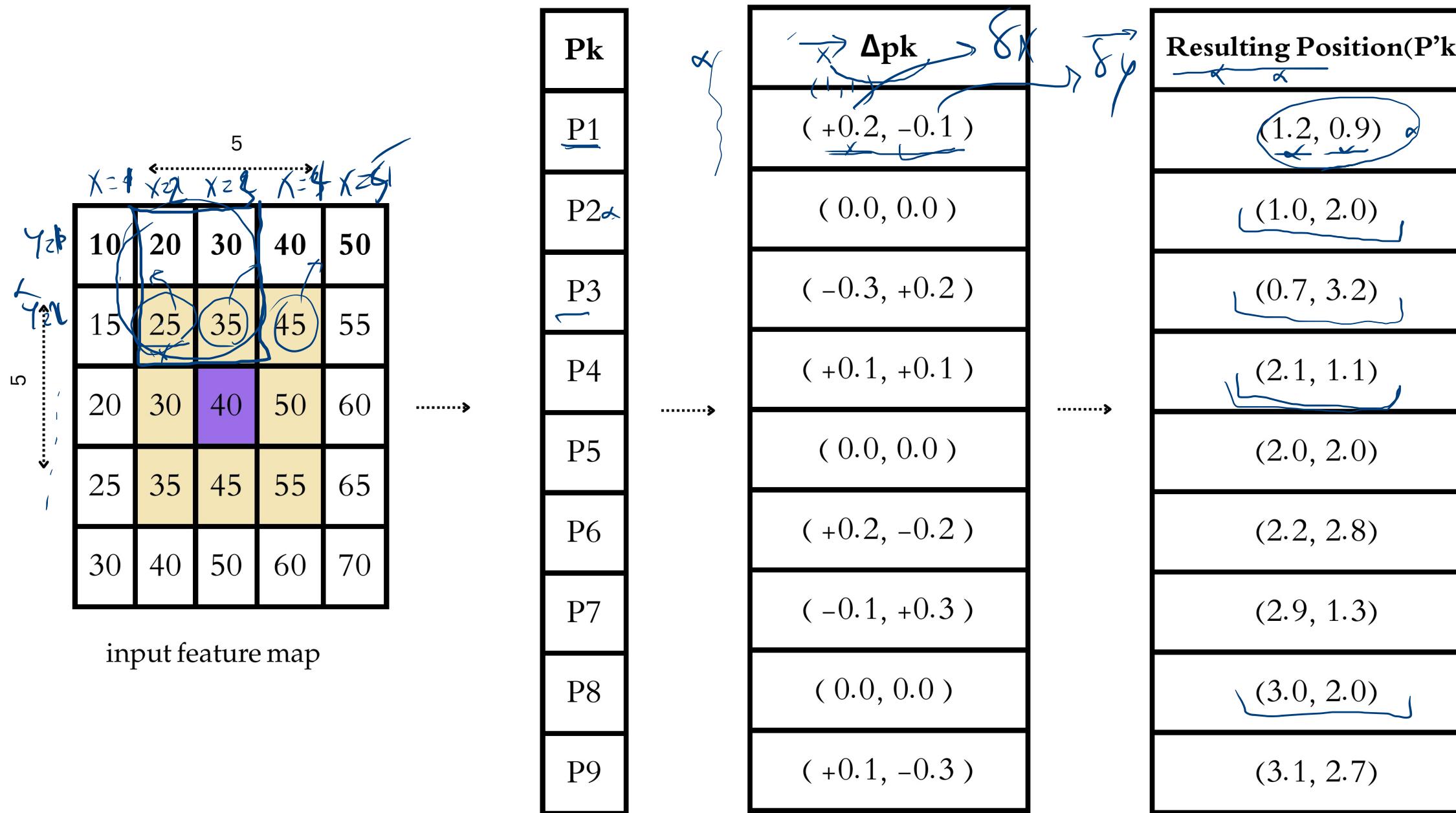
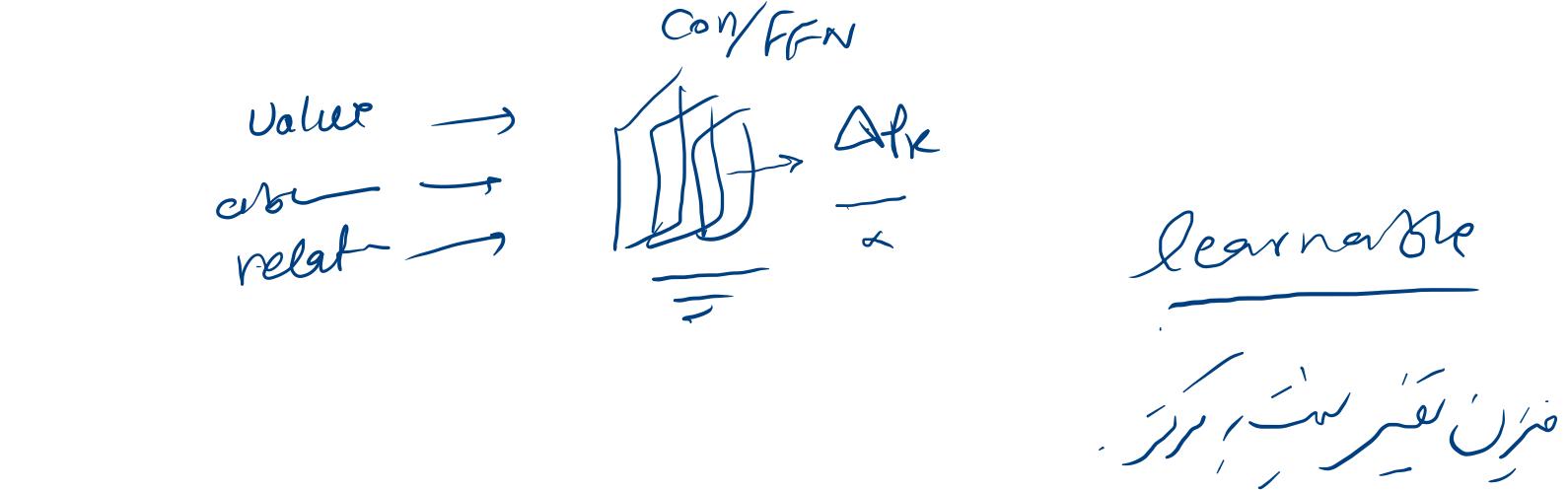


DCNv3

Deformable Convolution v3

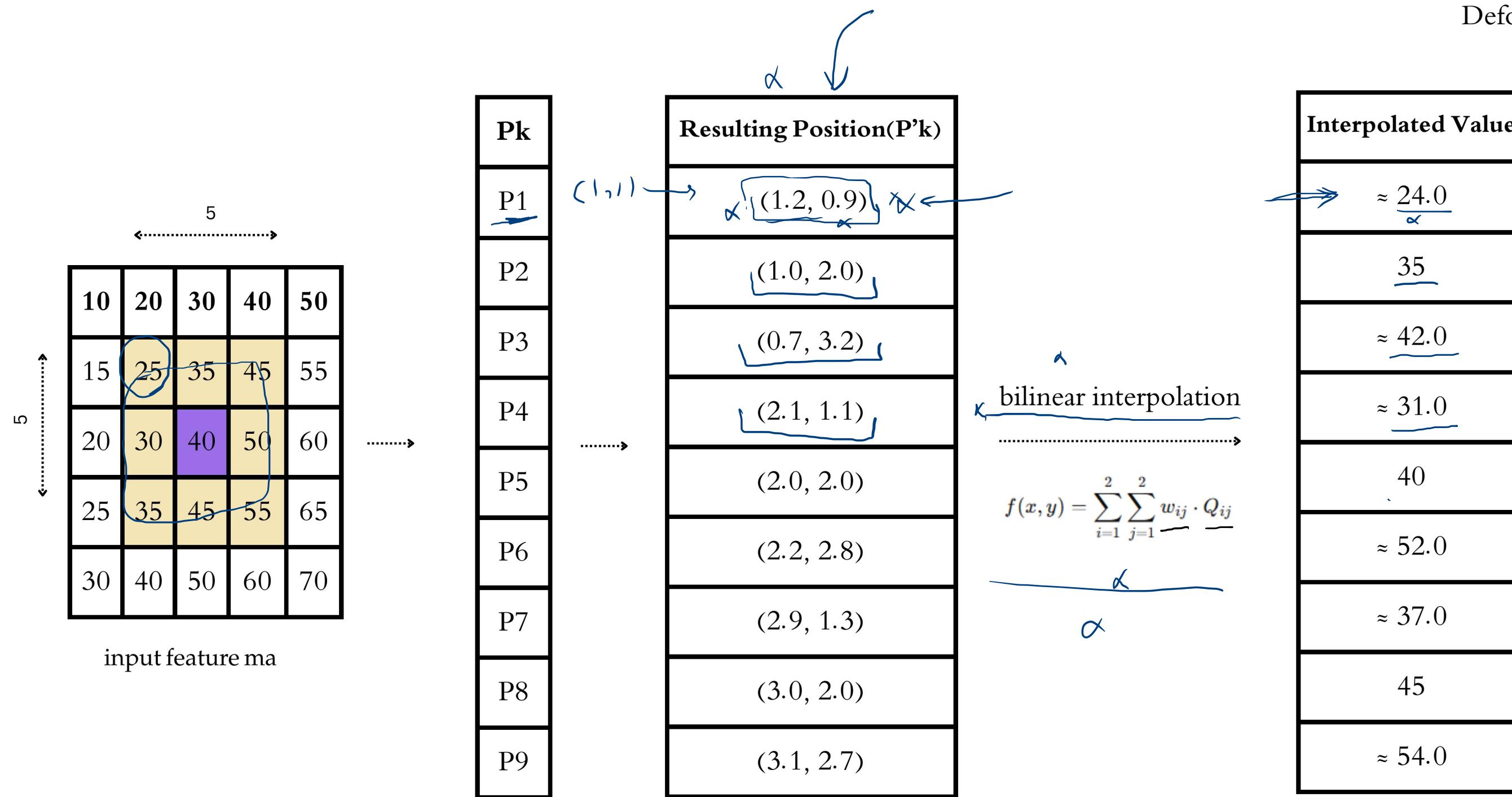
DCNv3

Deformable Convolution v3

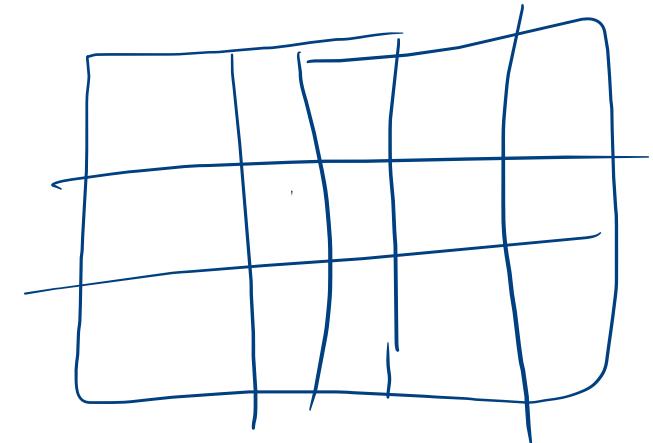
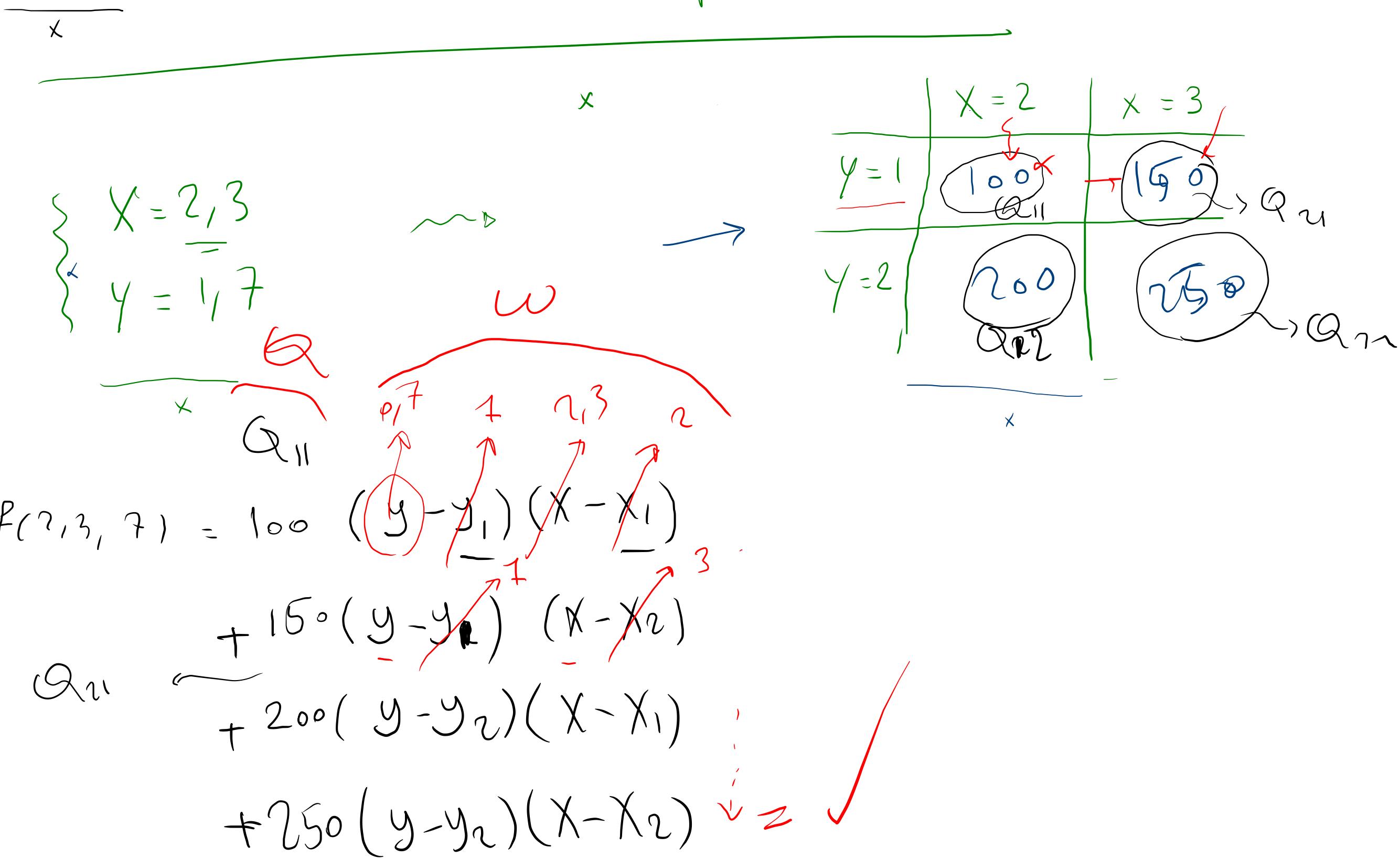


DCNv3

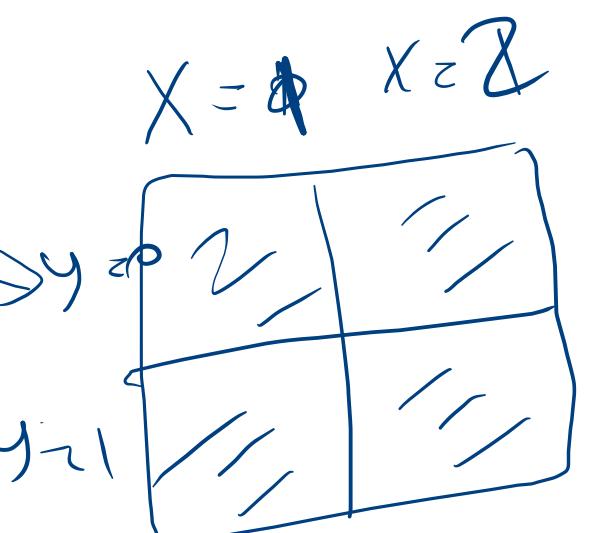
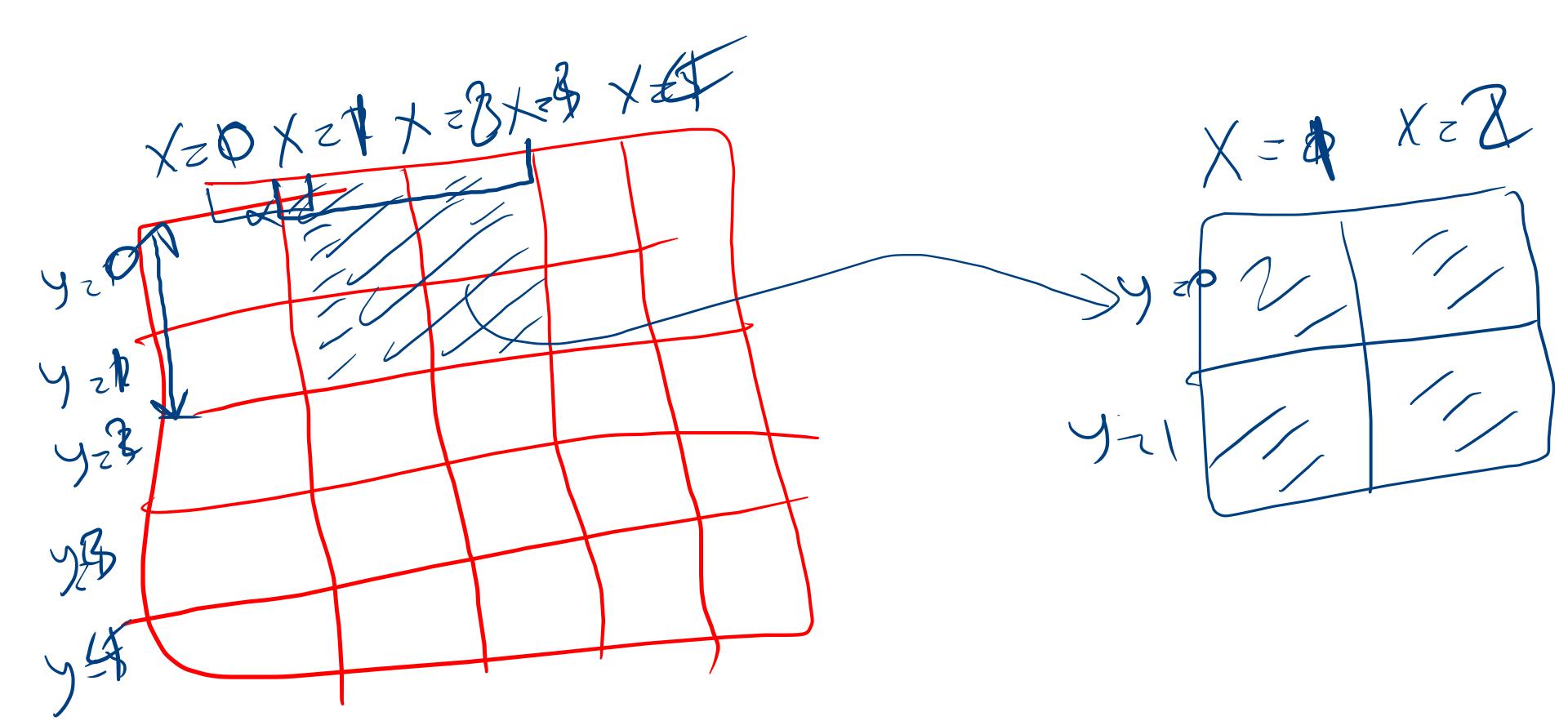
Deformable Convolution v3



Bilinear Interpolation.



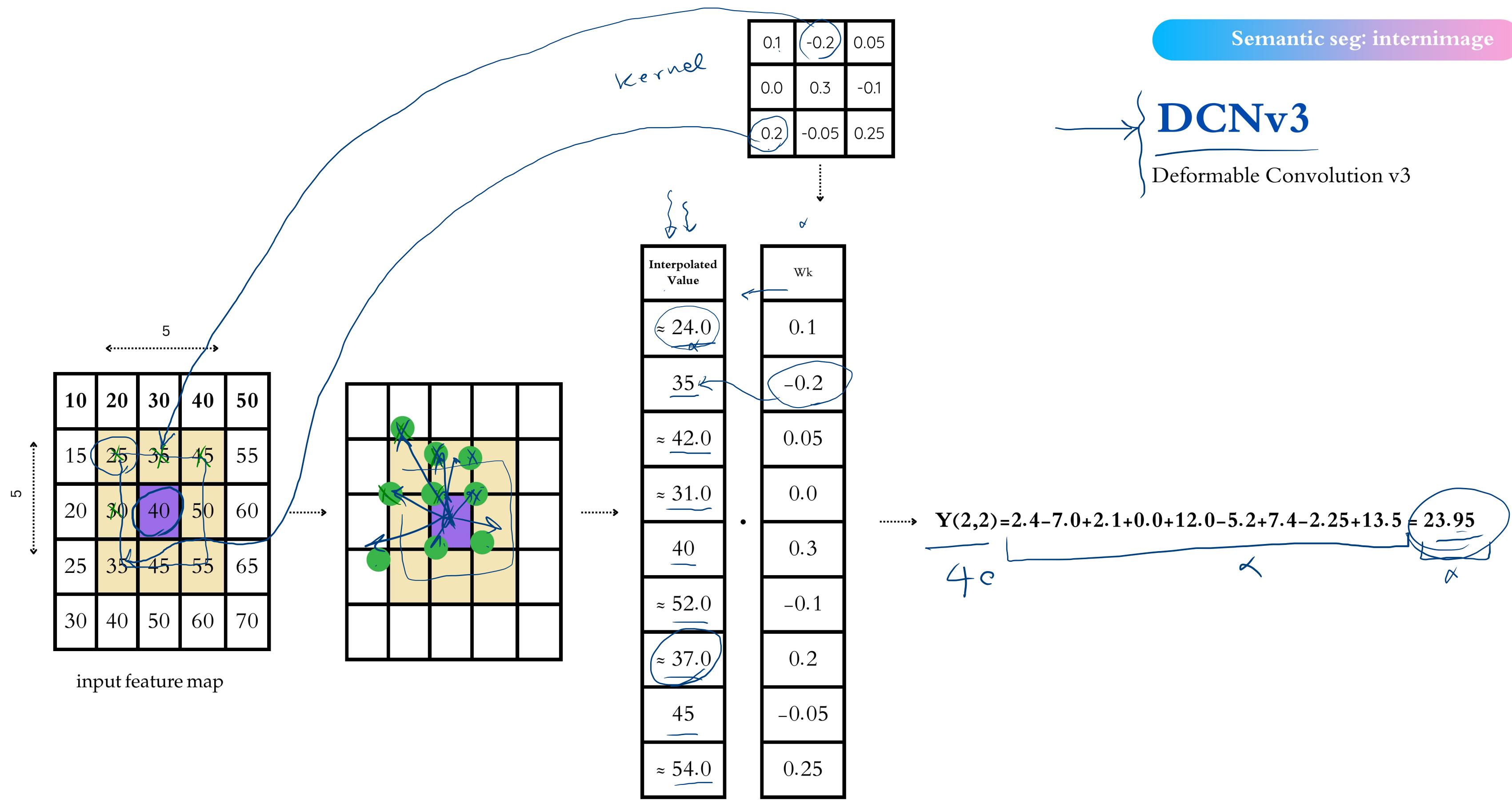
$$\sum w_{ij} Q_{ij}$$



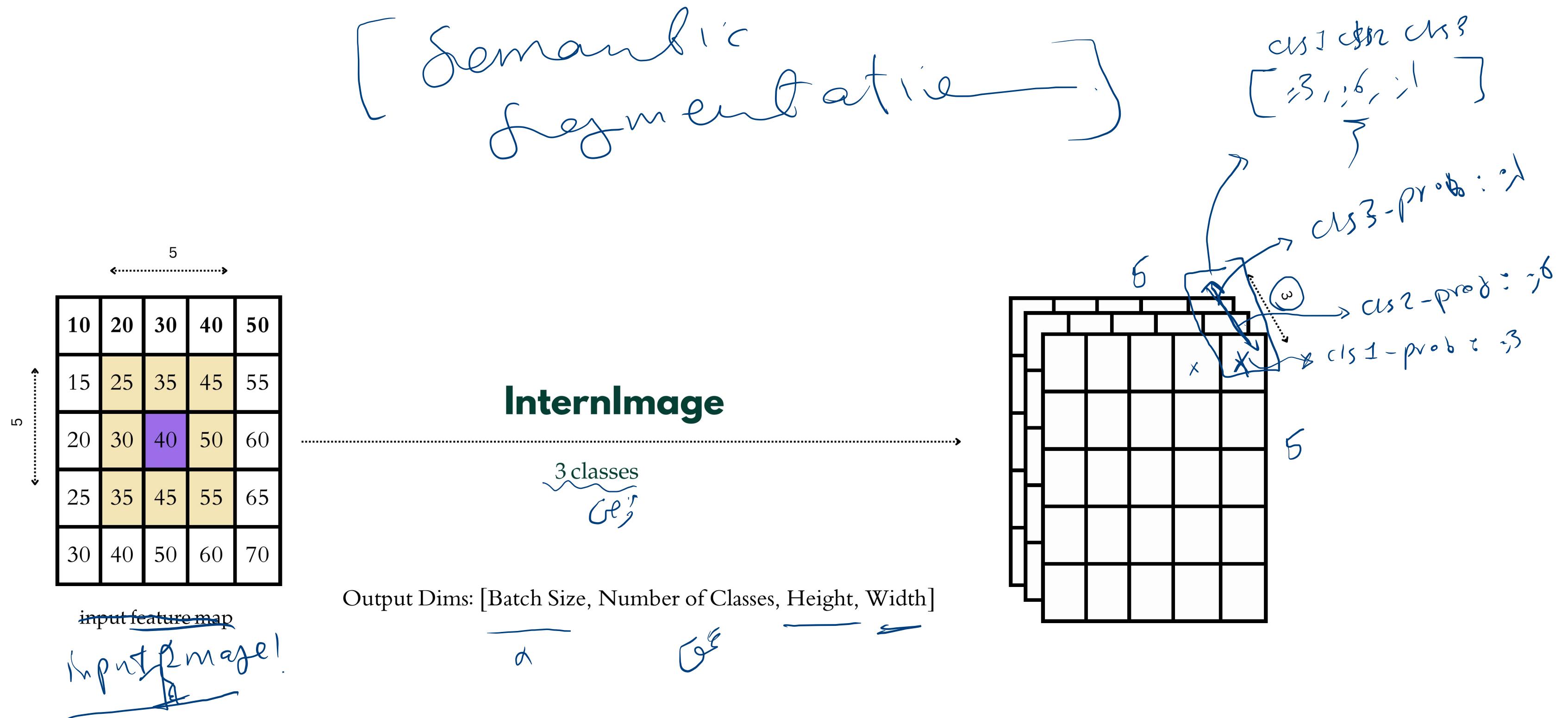
$$\underline{x = 1, 2}$$

$$\underline{y = 0, 1, 2}$$

Semantic seg: internimage



$$Y(p_0) = \sum_{k=1}^9 w_k \cdot X(\hat{p}_k)$$



بریم سراغ کدش!

InternImage