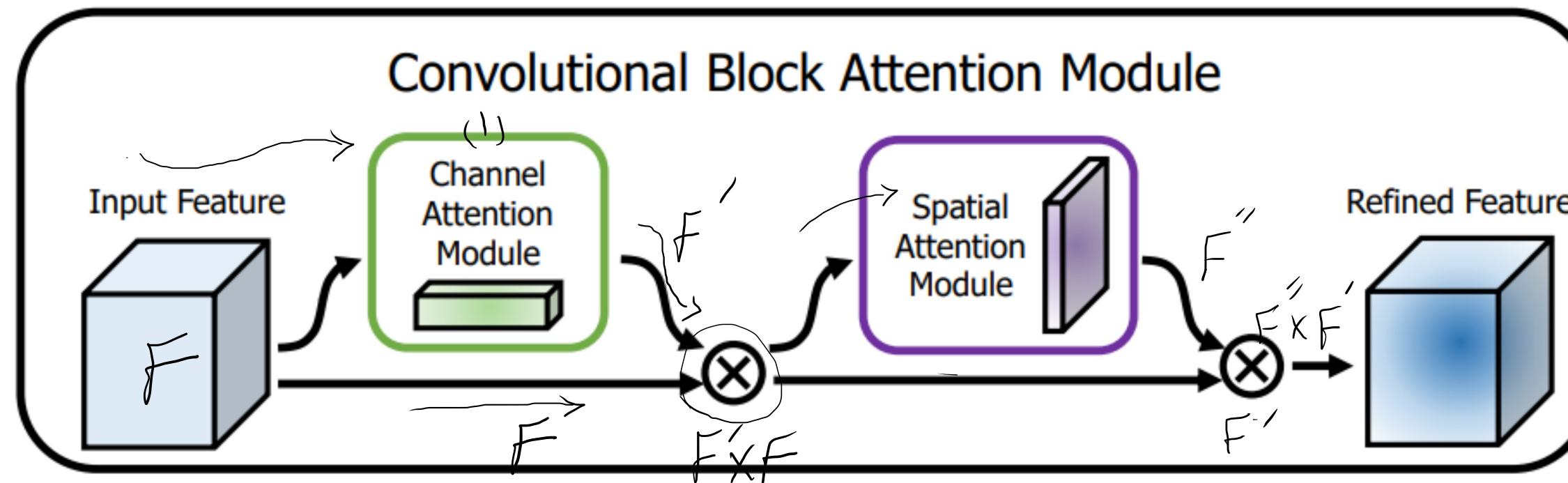


# CBAM معماری



# پایپ لاین CBAM



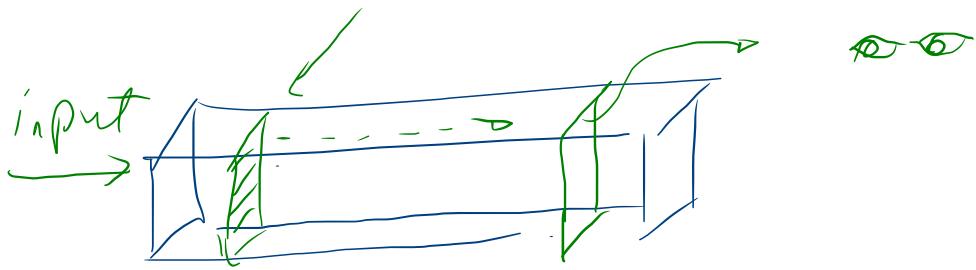
$$Mc(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)))$$



کدوم کanal  
(لبه، بافت، رنگ، ...)  
مهم تره؟

کجاهاي تصوير مهم تره؟  
(نواحي مكانی)

$$Ms(F') = \sigma(\text{Conv2D}7 \times 7([\text{Avg}; \text{Max}]))$$



امکانیز توجه CBAM



کمک می کند مدل بفهمد  
ویژگی "موهای سگ" یا "چشم‌ها" مهم‌تر از "چمن"  
هست.

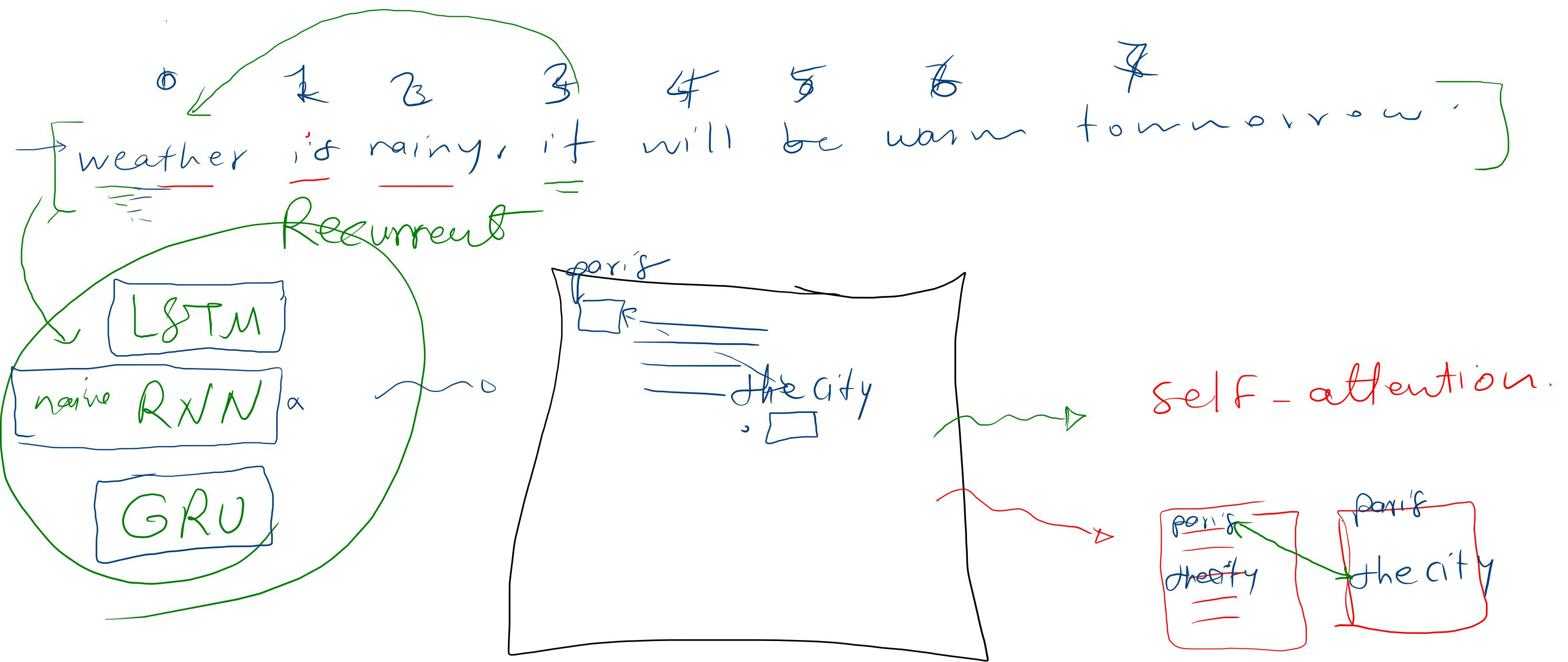
کمک می کند که مدل بفهمد  
کجا تصویر باید تمرکز کند (مثلاً ناحیه‌ای که سگ  
هست).

self - attention!

مکانیزم توجه به خود

self-attention mechanism

X



self-attention.

# نگاهی کوتاه به مقاله

## Attention Is All You Need

Ashish Vaswani\*

Google Brain

[avaswani@google.com](mailto:avaswani@google.com)

Noam Shazeer\*

Google Brain

[noam@google.com](mailto:noam@google.com)

Niki Parmar\*

Google Research

[nikip@google.com](mailto:nikip@google.com)

Jakob Uszkoreit\*

Google Research

[usz@google.com](mailto:usz@google.com)

Llion Jones\*

Google Research

[llion@google.com](mailto:llion@google.com)

Aidan N. Gomez\* †

University of Toronto

[aidan@cs.toronto.edu](mailto:aidan@cs.toronto.edu)

Lukasz Kaiser\*

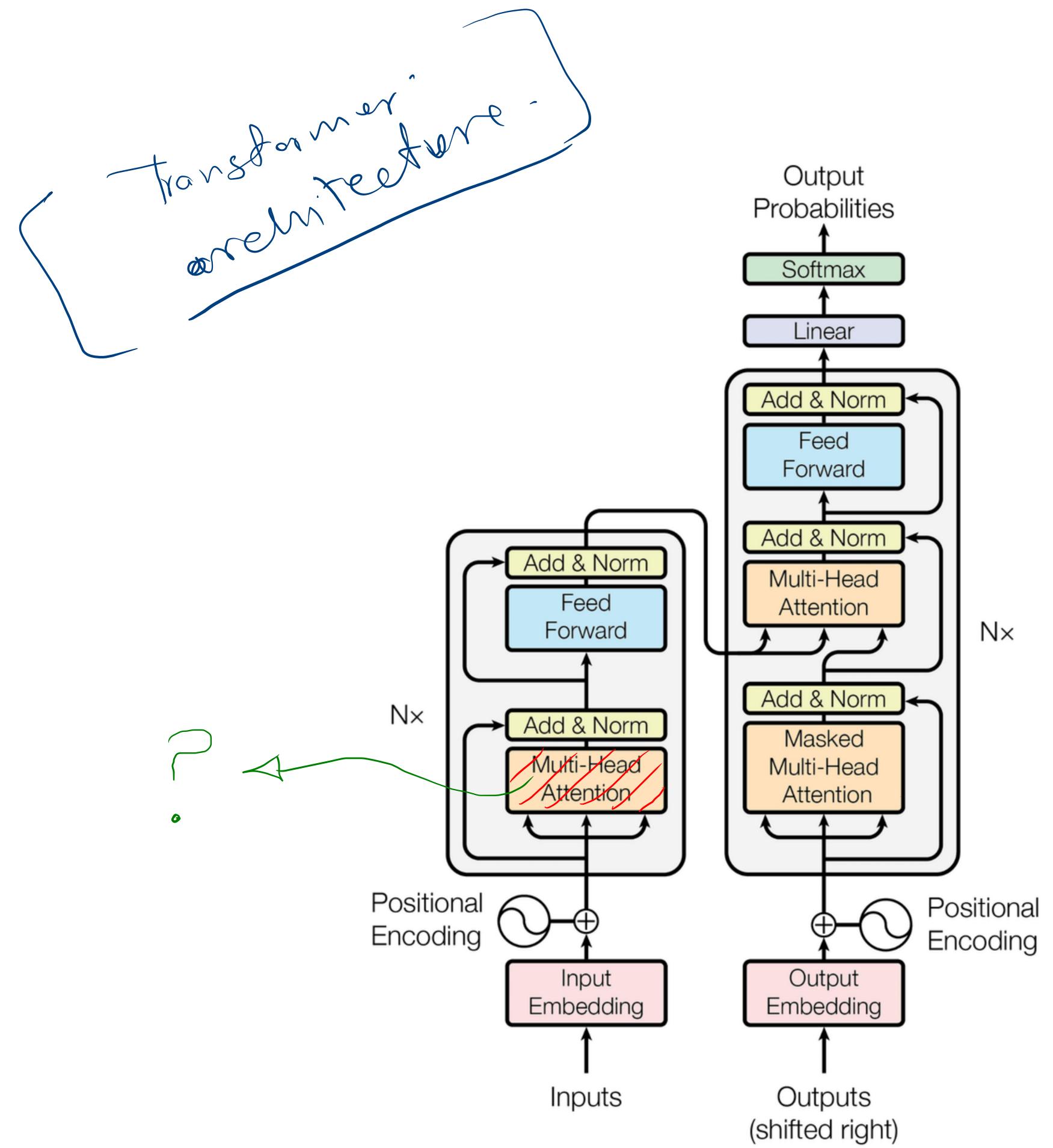
Google Brain

[lukaszkaiser@google.com](mailto:lukaszkaiser@google.com)

Illia Polosukhin\* ‡

[illia.polosukhin@gmail.com](mailto:illia.polosukhin@gmail.com)

self attention



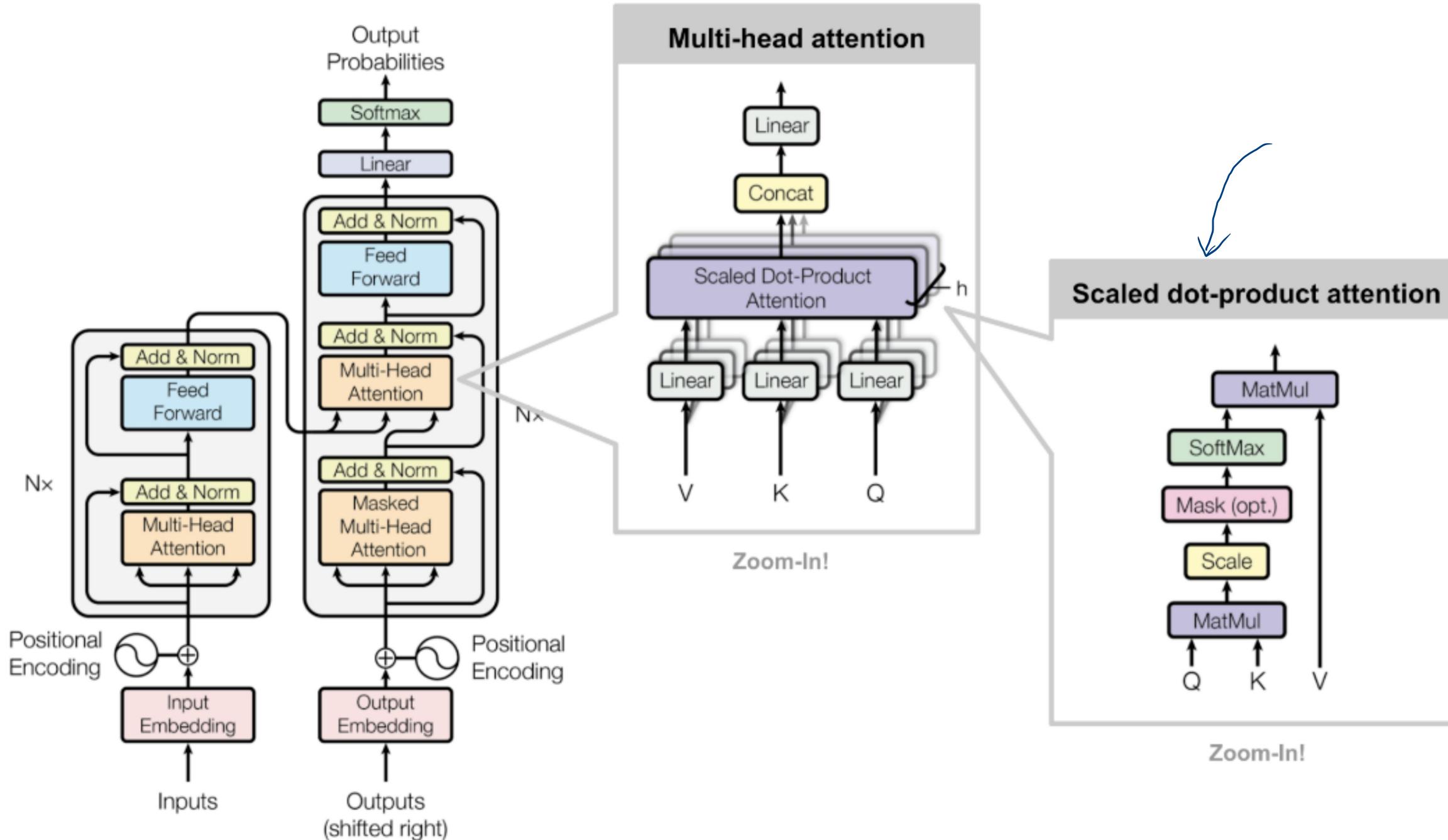
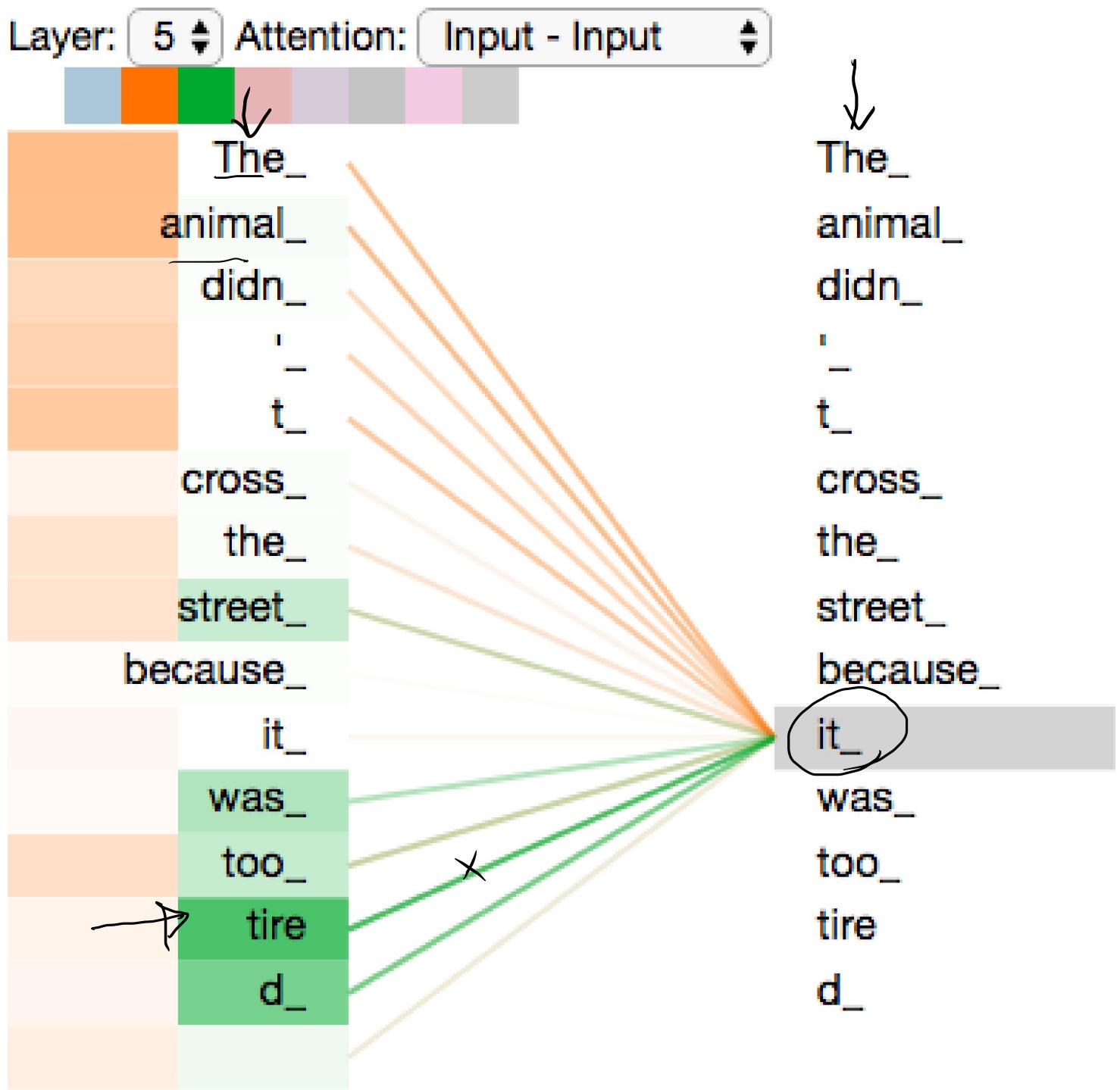
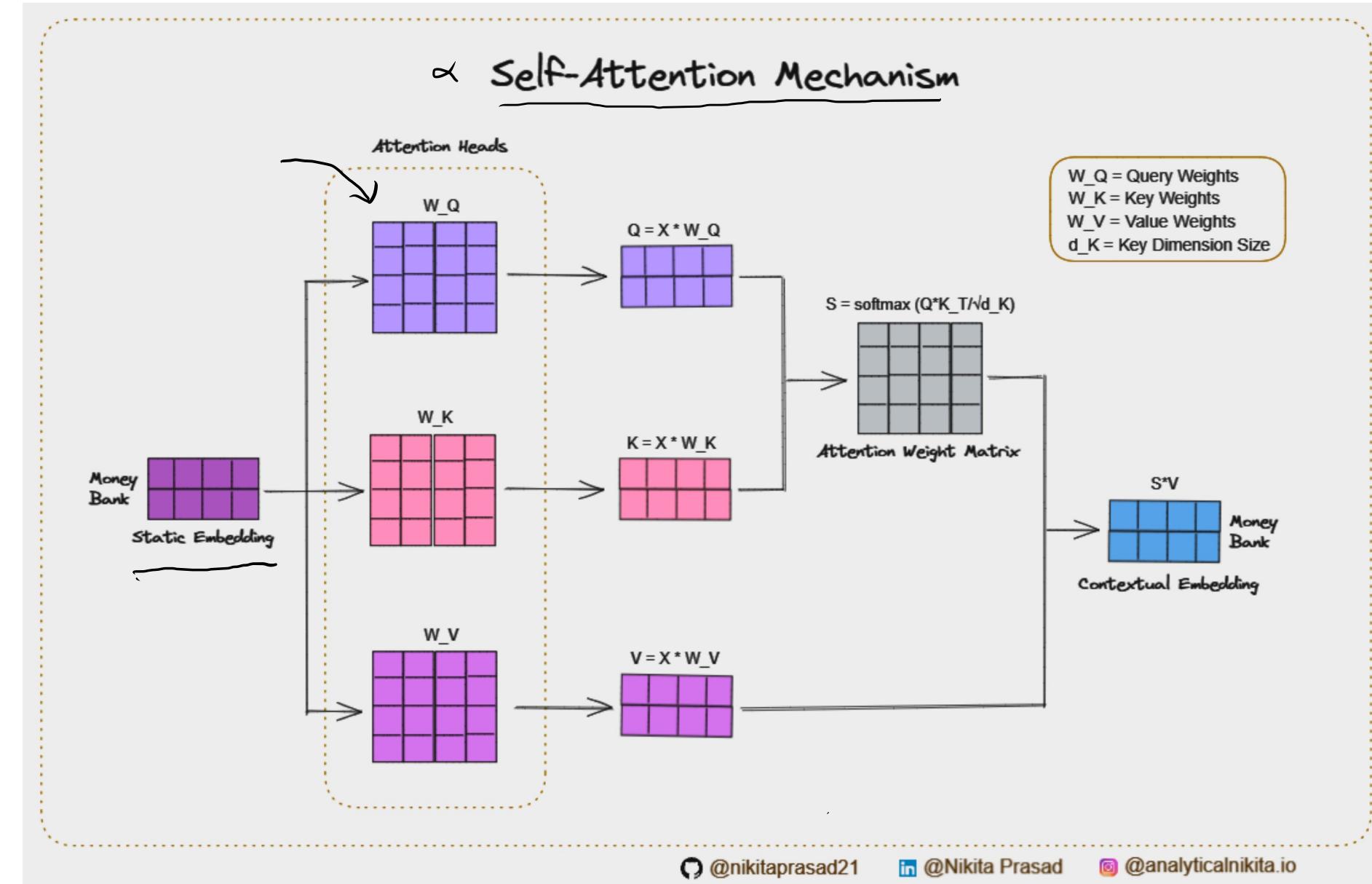


Fig. 17. The full model architecture of the transformer. (Image source: Fig 1 & 2 in  
Vaswani et al. 2017)

? - /





S: I love tea.

10,000 words

$$E = [0, 1, 0, 1]$$

love: [1, 0, 1, 0]

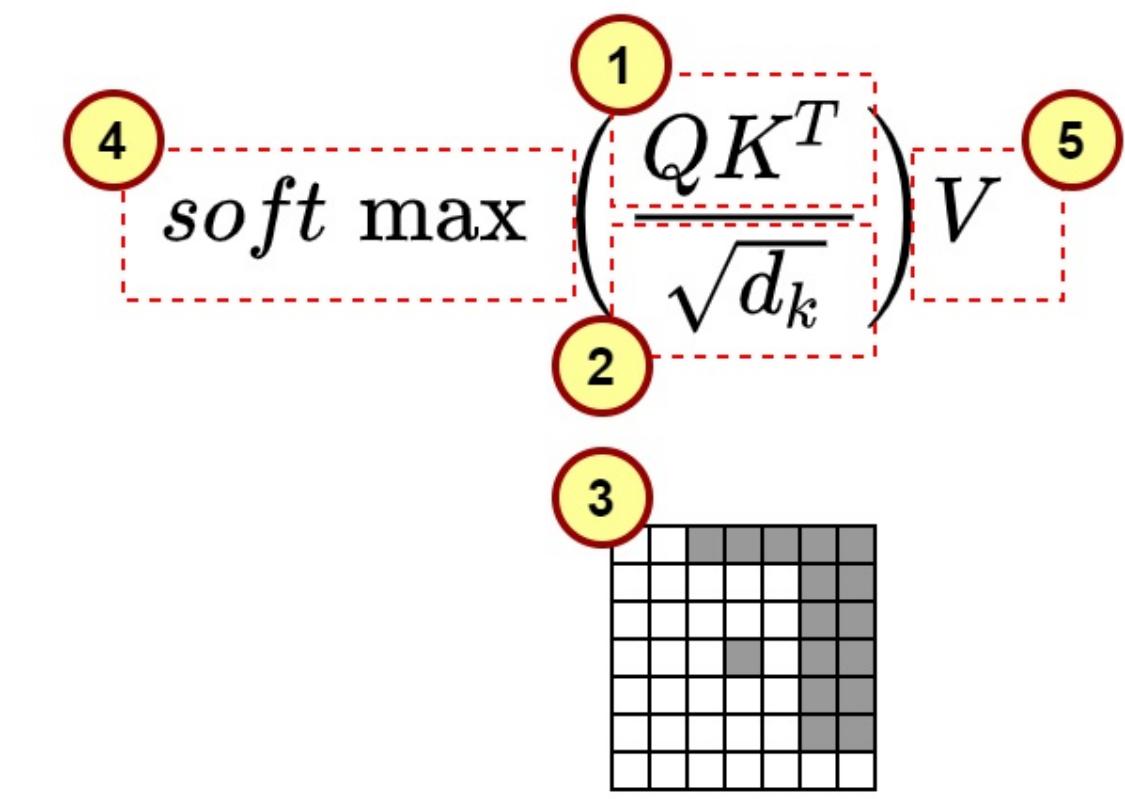
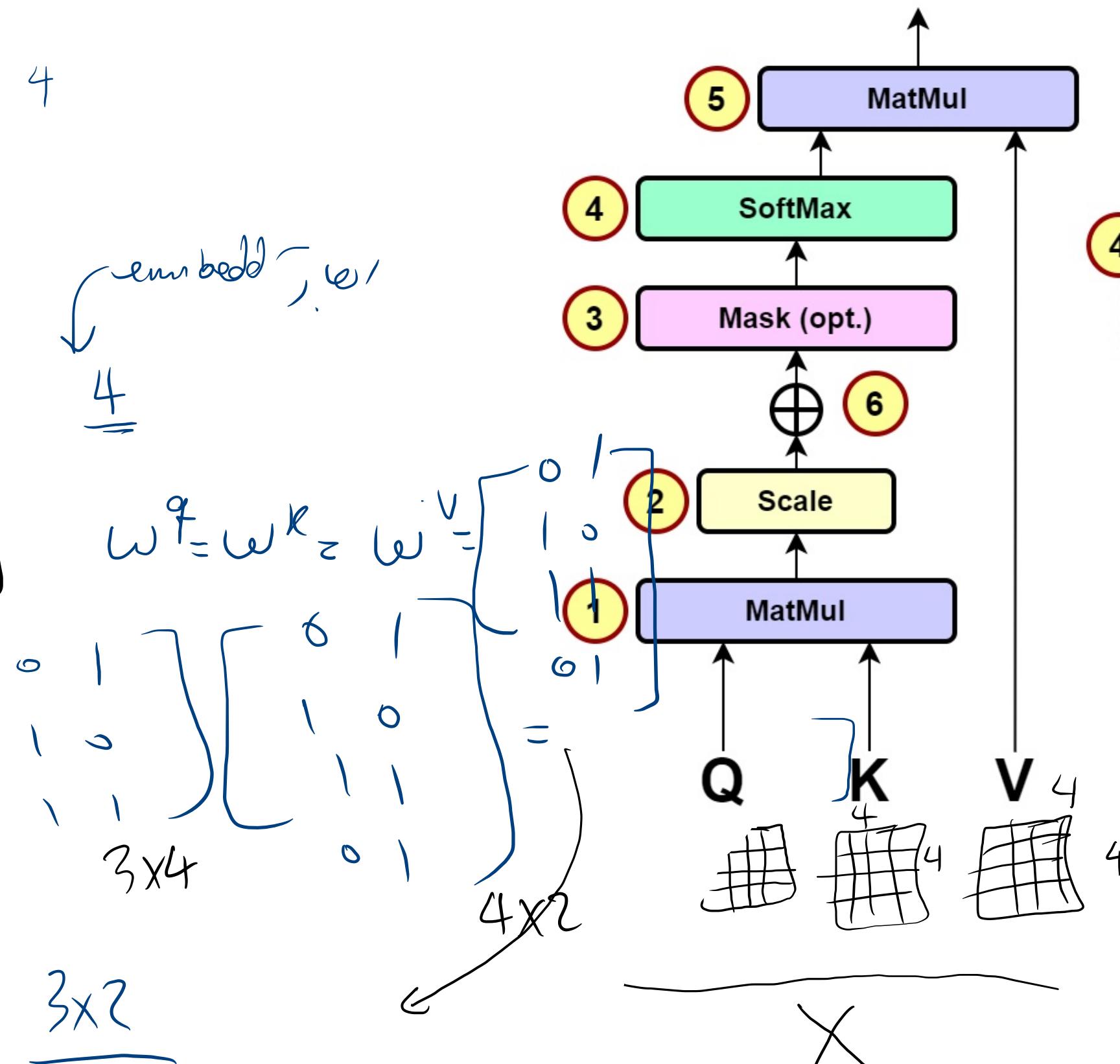
tea: [1, 1, 1, 1]

$X = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

$$Q = X \cdot w^q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$K = X \cdot w^K =$$

$$V = X \cdot w^V =$$



Probability

P = 0.6 / 11.9 = 0.05

$S: I \text{ love tea.}$

$$X = \text{embedding}(S)$$

$$Q = \underline{\underline{X}} \cdot \underline{w^q} = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} \quad 3 \times 2$$

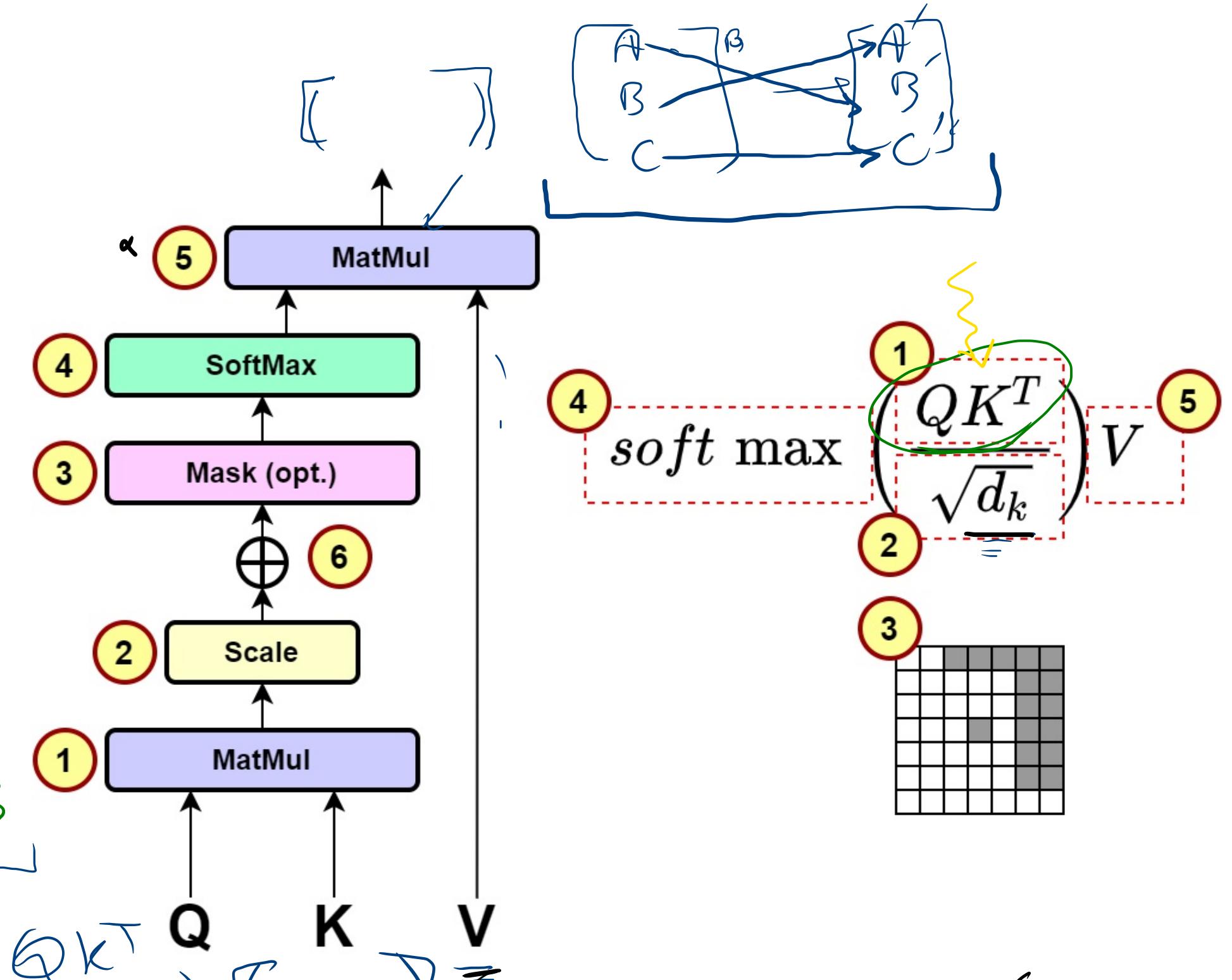
$$K = \underline{\underline{X}} \cdot \underline{w^K} = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} \quad 3 \times 2$$

$$V = \underline{\underline{X}} \cdot \underline{w^V} = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix} \quad 3 \times 2$$

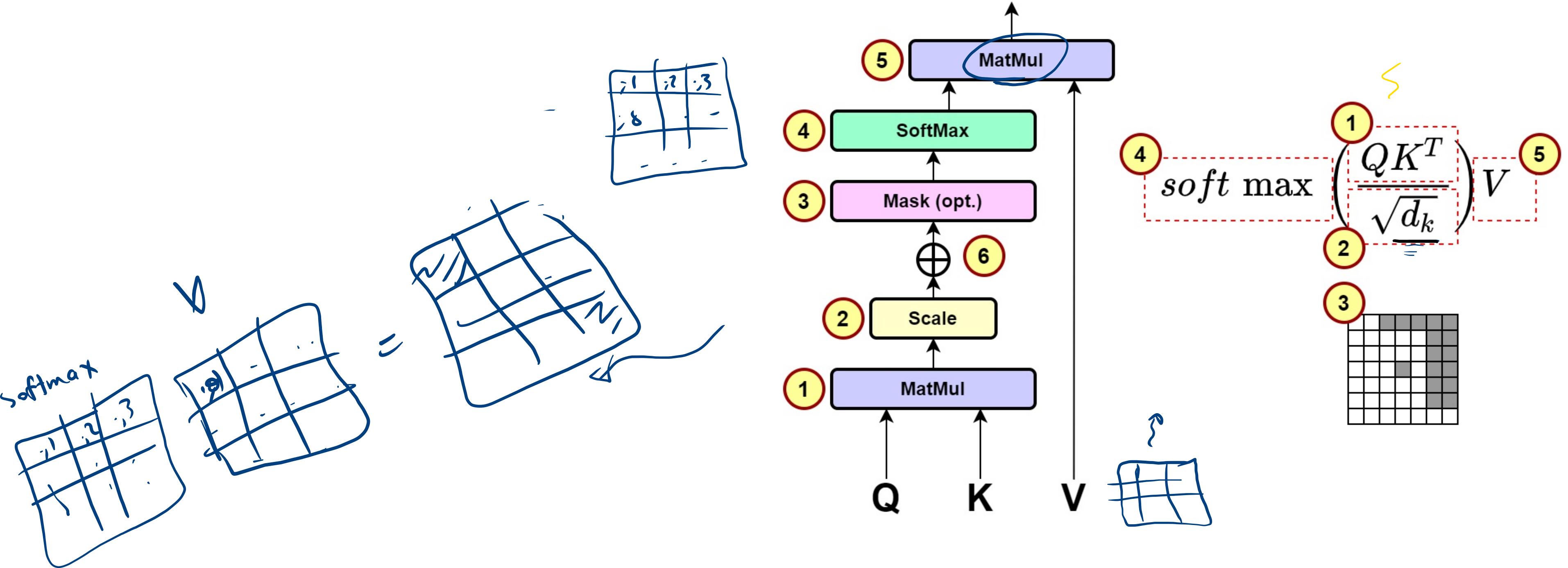
$$\cancel{QK^T} : \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}_{3 \times 2} \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}_{2 \times 3} = \begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}_{3 \times 3}$$

$$\left( \frac{QK^T}{\sqrt{2}} \right) \sim [0, 1] \rightarrow \text{softmax} \left( \frac{QK^T}{\sqrt{n}} \right) \in \mathbb{R}^{3 \times 3}$$

10,000 words



$$\rho = \sqrt{6} / 119 \cdot P$$



word prediction

x

[this is an online class about w details.]

is

label

This is an - - - -

This is an online - - -

this is an online class - - -

This is an

This is an online class.

- - -

this is an online class - - -



$$X = \begin{bmatrix} & 128 \\ 1000 & \end{bmatrix}$$

Text →

1000 word

$$\overline{\overline{1000 \times 128}}$$

