

OBJECT DETECTION

COMPUTER VISION
SPRING, 2023



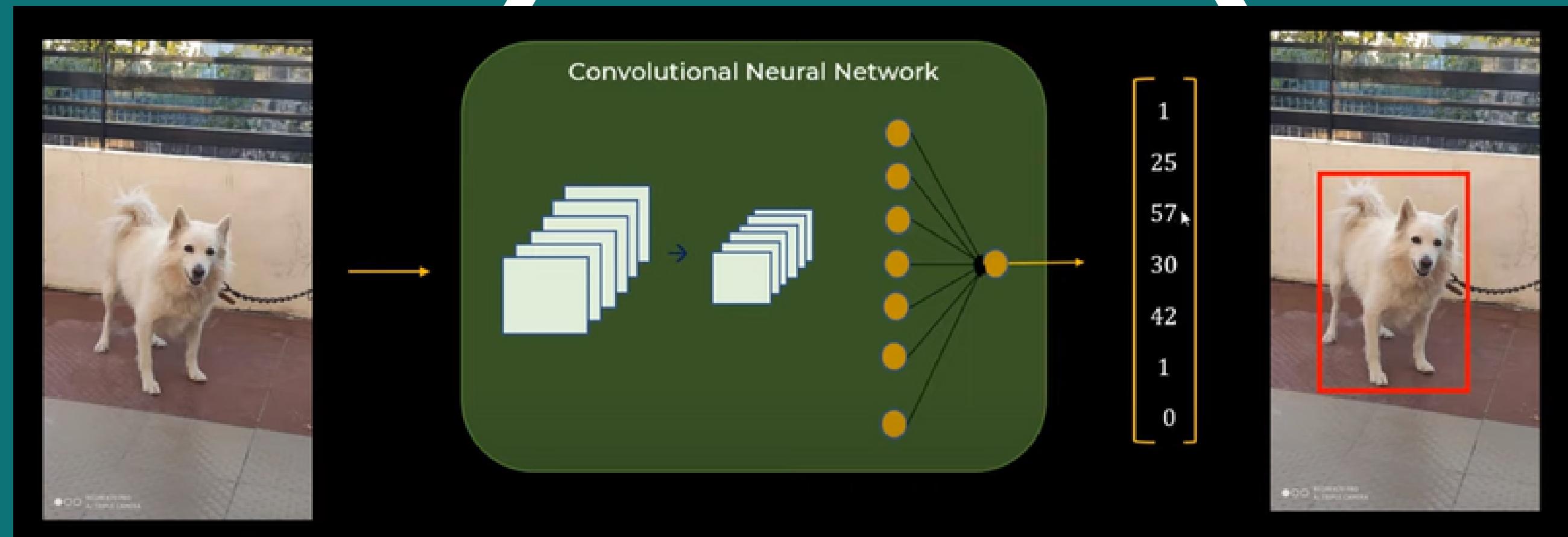


I have a
Plan

- **Single Pass**
- **Sliding window**
- **Region proposal**
- **SSD**
- **Transformer**

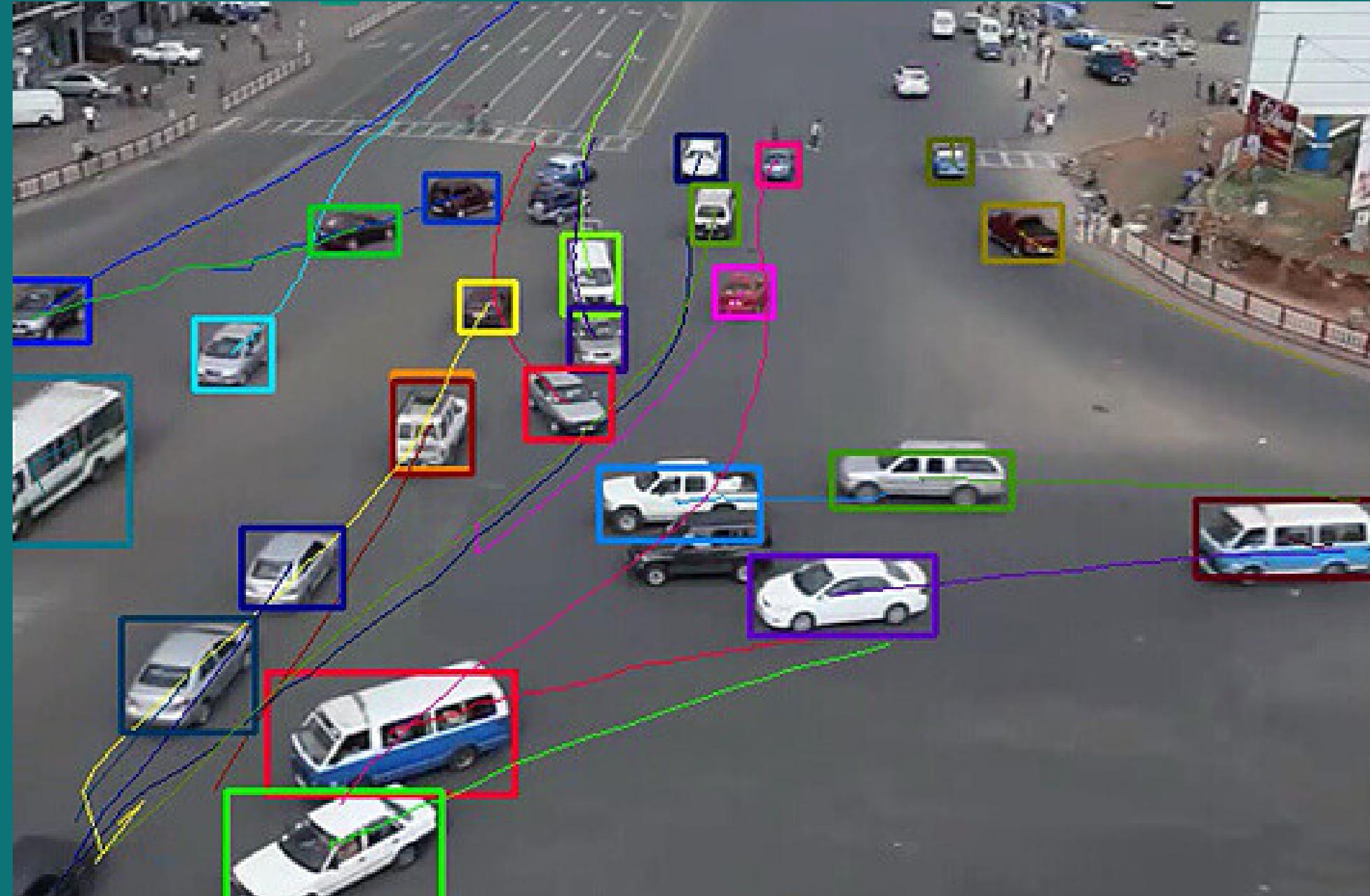
Single Pass

Works well for
single object
detection



What to do ?

In the case of many objects...



Traditional Techniques

Intuitive but limited...

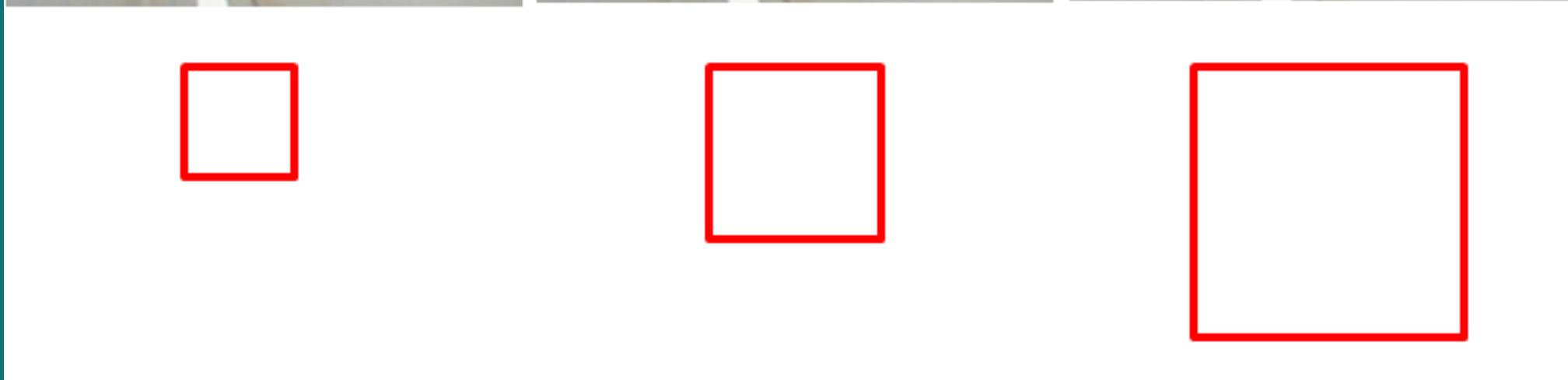
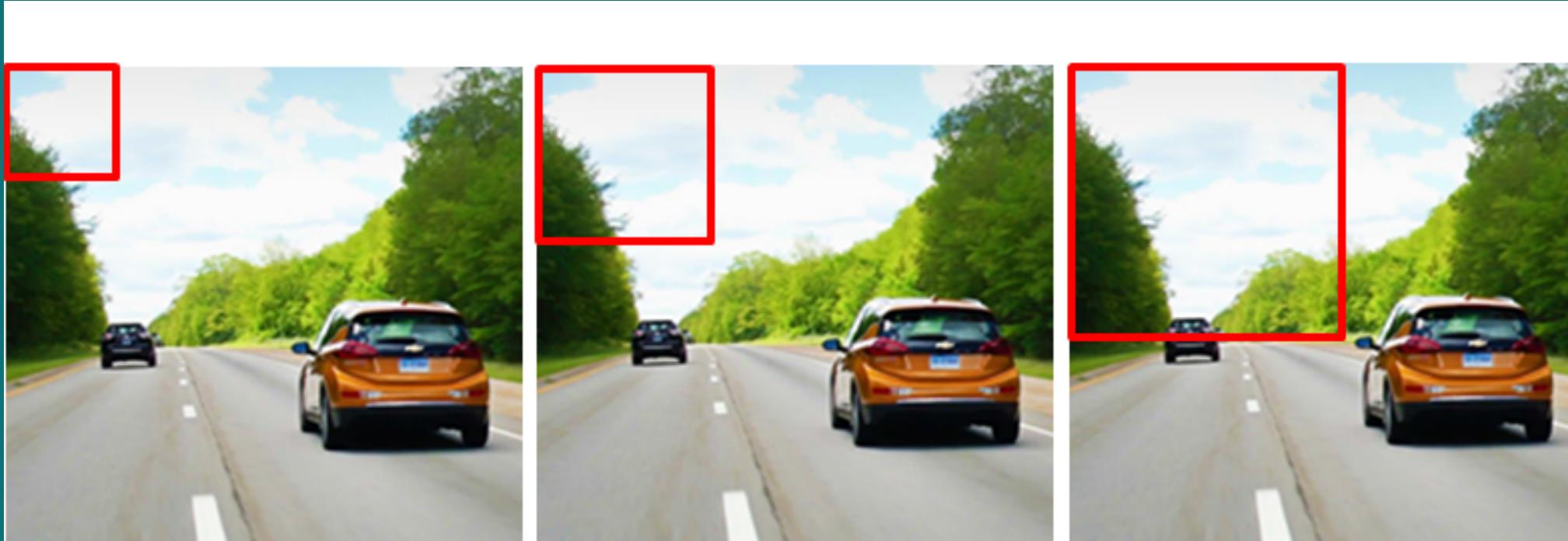


**Sliding
window**



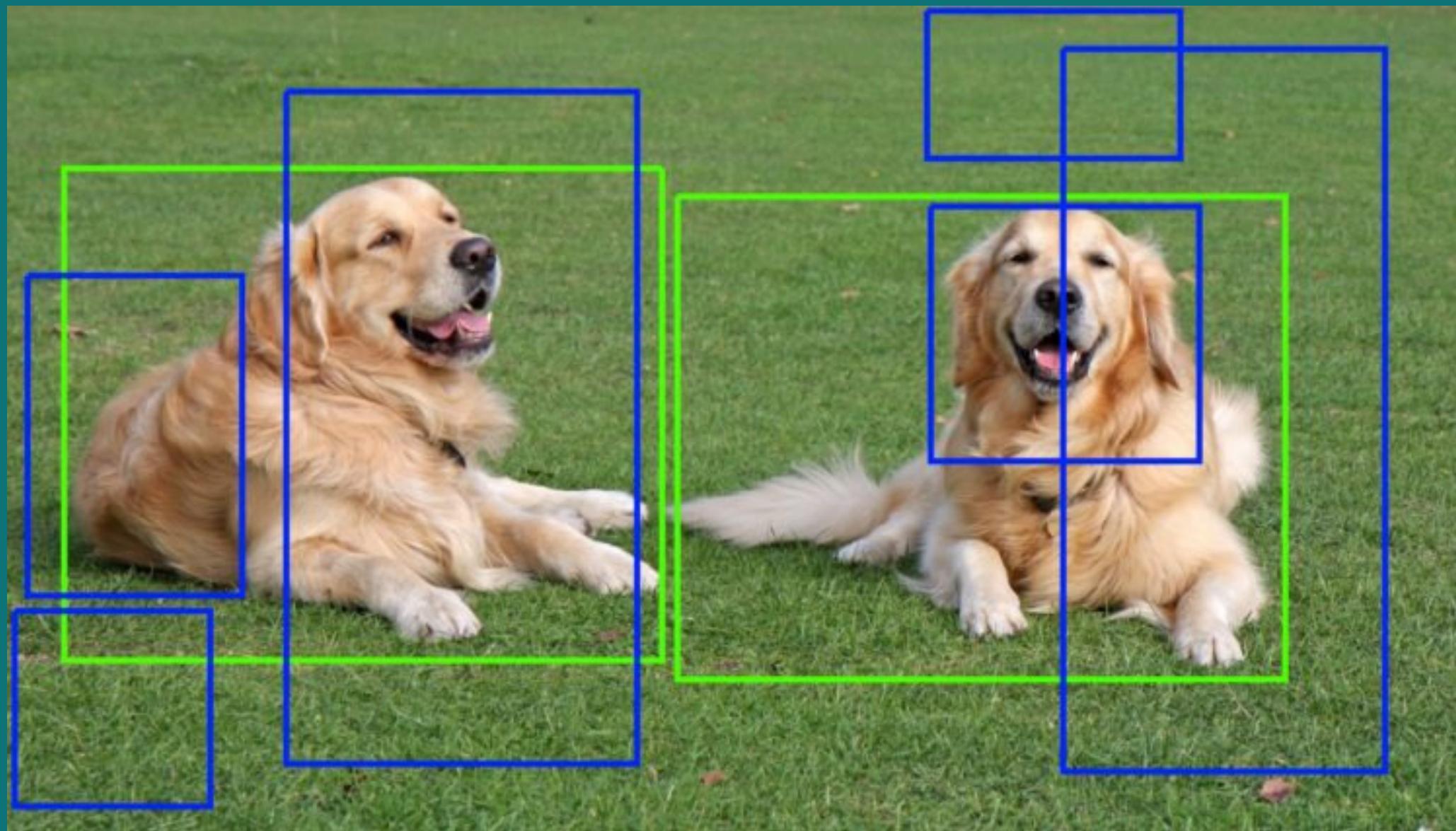
**Region
proposal**

Sliding Window



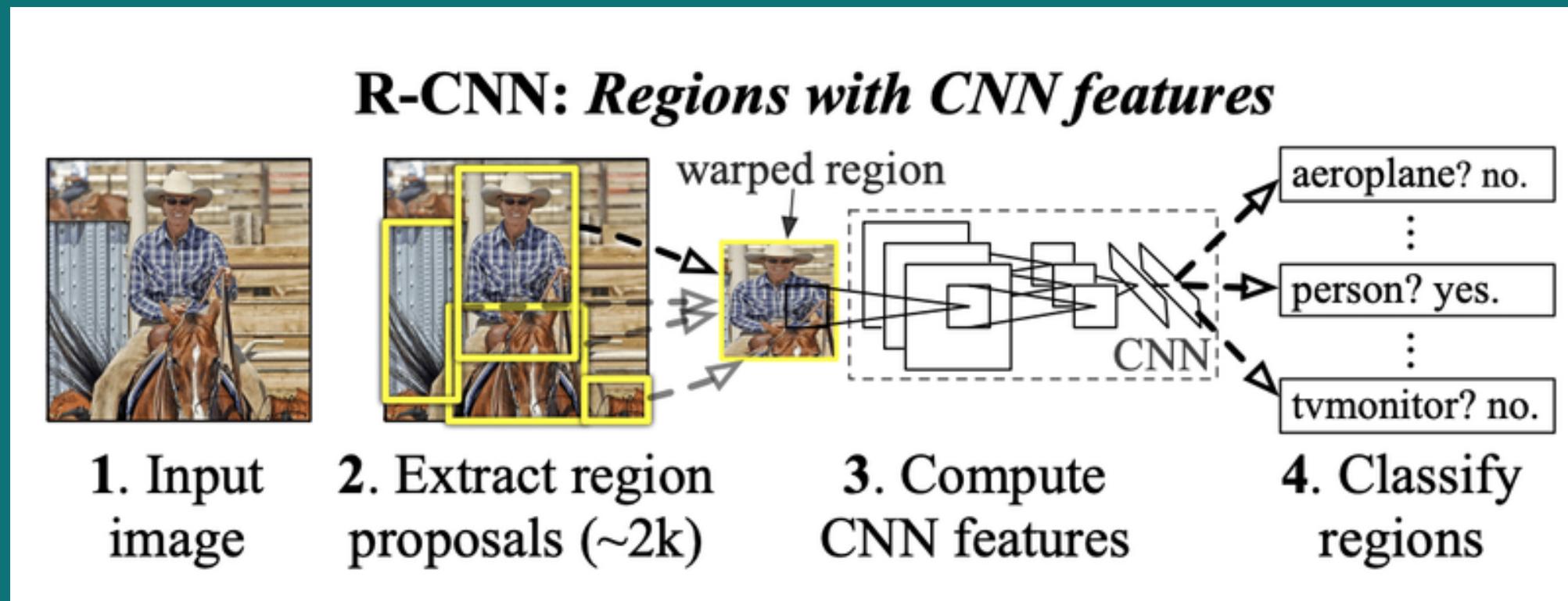
Guess the scale... !

Sliding Window

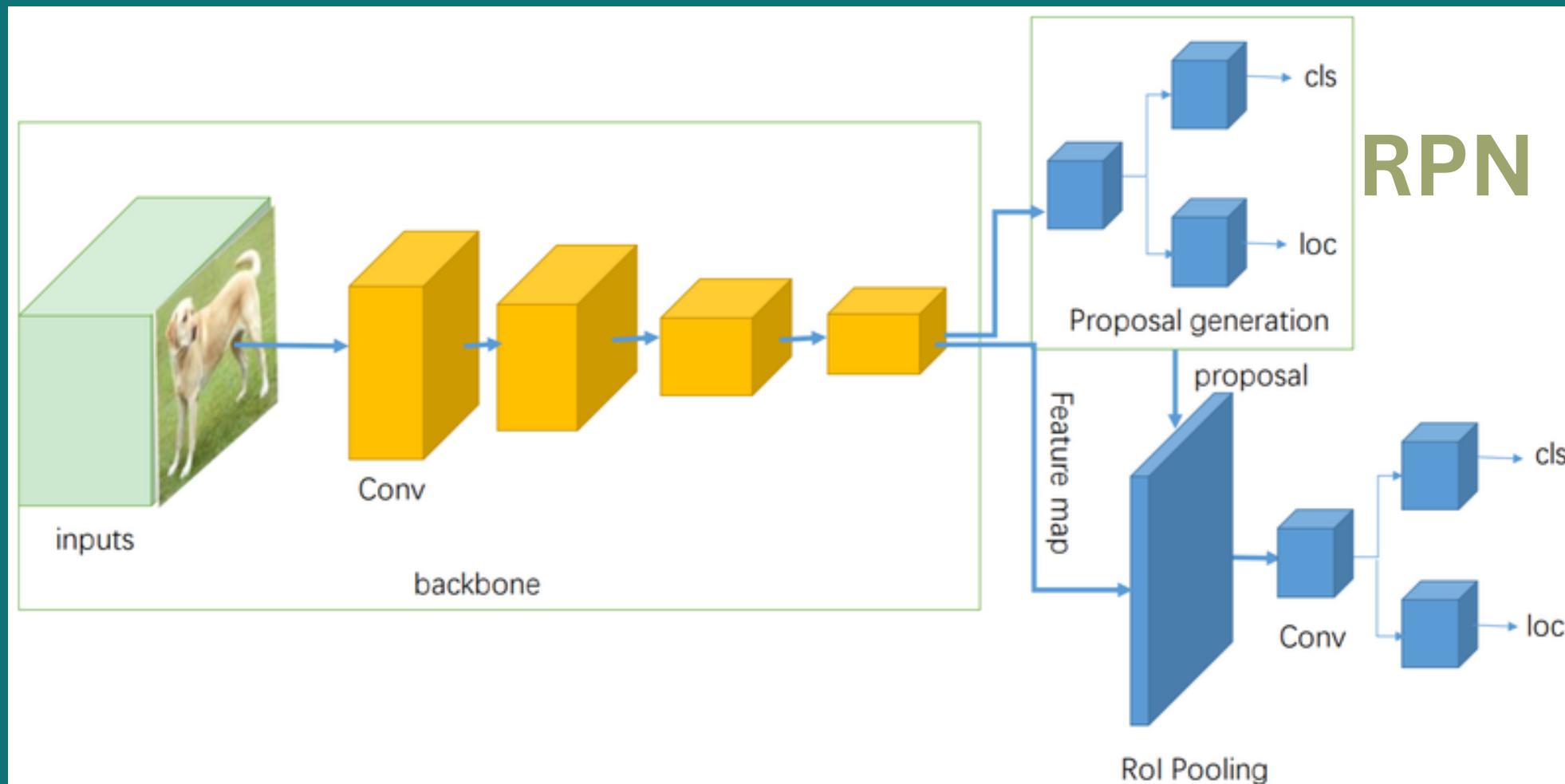


Too many predictions to make... !

Regions Proposal



R-CNN



Two-Stage Detector

Region Proposal
&
Region Classification / refinement



Recap !



The image shows a YouTube thumbnail for a tutorial titled "Sliding Window Object Detection | Deep Learning Tutorial 30 (Tensorflow, Keras & Python)". The thumbnail features a video frame where a person is bending over a small brown dog on a wooden deck. Two bounding boxes are overlaid on the image: a yellow one labeled "Person" and a red one labeled "Dog". A play button icon is positioned in the center of the red box. The thumbnail has a dark background with large white and yellow text for the title and number. A "Watch on YouTube" link is at the bottom left, and a "Share" button is at the top right. To the right of the thumbnail is a circular portrait of a man with glasses and a hoodie, standing against a red and black background.

Sliding Window Object Detection | Deep Learning Tutorial 30 (Tensorflow, Keras & Python)

Person

Dog

30

Watch on YouTube

Share

Modern Techniques

Fast and... faster !

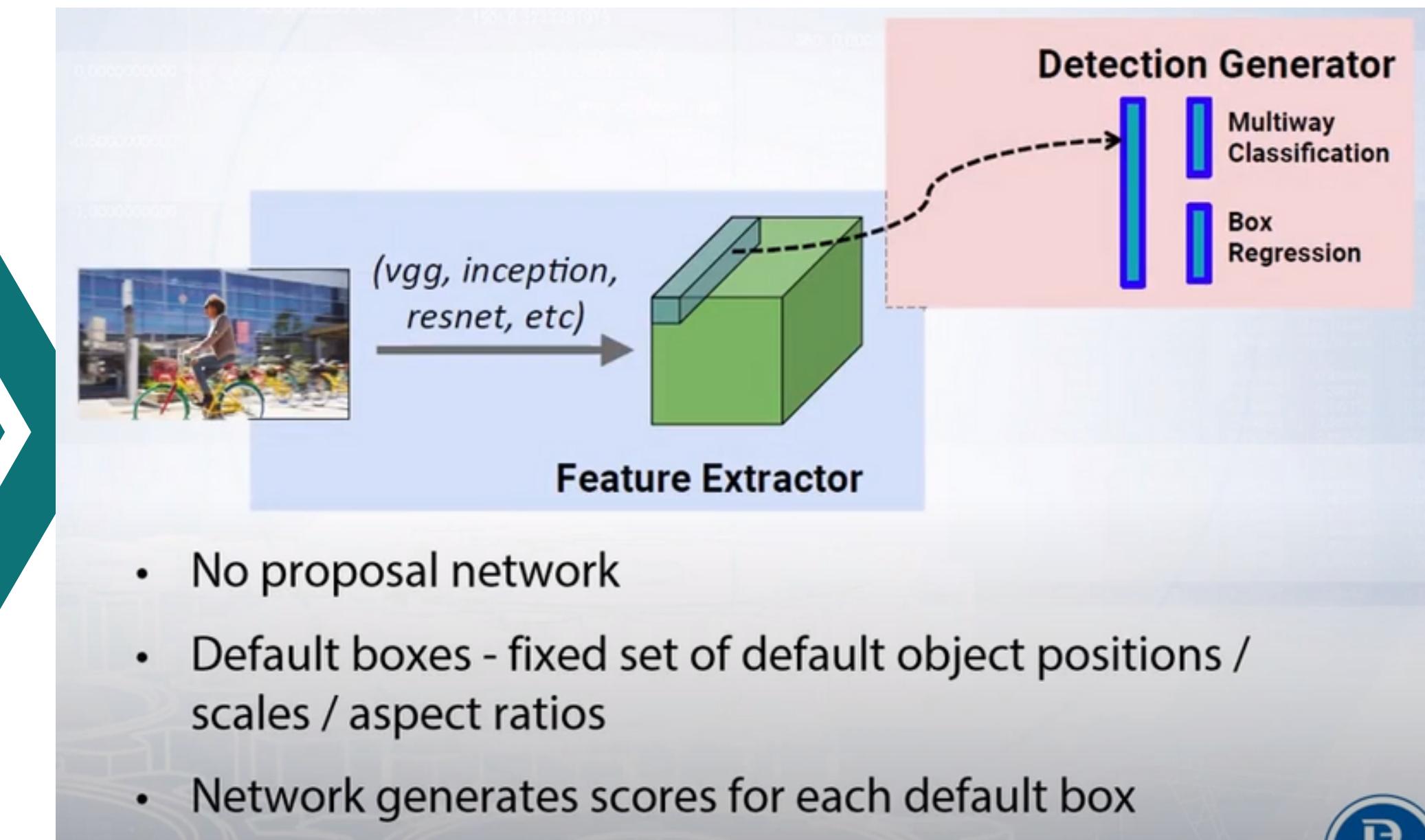


SSD



Transformer

Single Shot Detector



Single Shot Detector

- Separate input into 3×3 cells.
- For each cell, we want to produce

$$y = \begin{bmatrix} p \\ c_1 \\ c_2 \\ x \\ y \\ w \\ h \end{bmatrix} = \begin{bmatrix} 0 \text{ if "no object", 1 if "object"} \\ 1 \text{ if "pedestrian", 0 "otherwise"} \\ 1 \text{ if "car", 0 "otherwise"} \\ \text{x-position} \\ \text{y-position} \\ \text{width} \\ \text{height} \end{bmatrix}.$$





To be clear !

A YouTube thumbnail for a tutorial on the YOLO algorithm. The thumbnail features a dark background with the title "What is YOLO Algorithm?" in large yellow letters. A white circle containing the number "31" is positioned to the left of the title. In the center, there is a small video frame showing a person interacting with a dog, with two bounding boxes drawn around them: a red one for the dog and a yellow one for the person. A red YouTube play button icon is overlaid on the bottom right of the video frame. The top of the thumbnail has the text "What is YOLO algorithm? | Deep Learning Tutorial 31 (Tensorflow, Keras & Python)" and the bottom has a "Watch on YouTube" button.

What is
YOLO
Algorithm?

Watch on YouTube

31

What is YOLO algorithm? | Deep Learning Tutorial 31 (Tensorflow, Keras & Python)

Share

YOLO v8

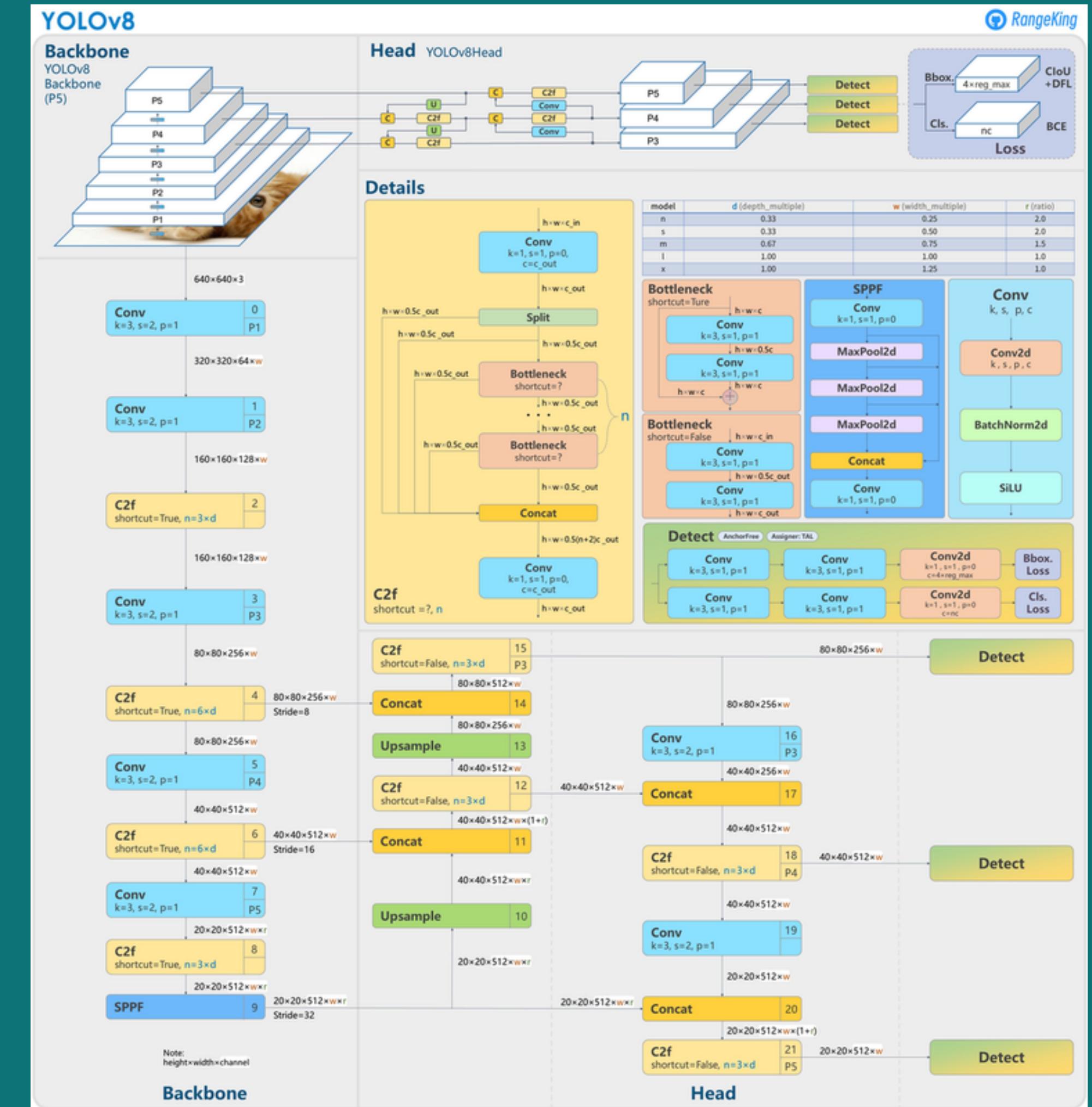
The state of the art solution



YOLO V8

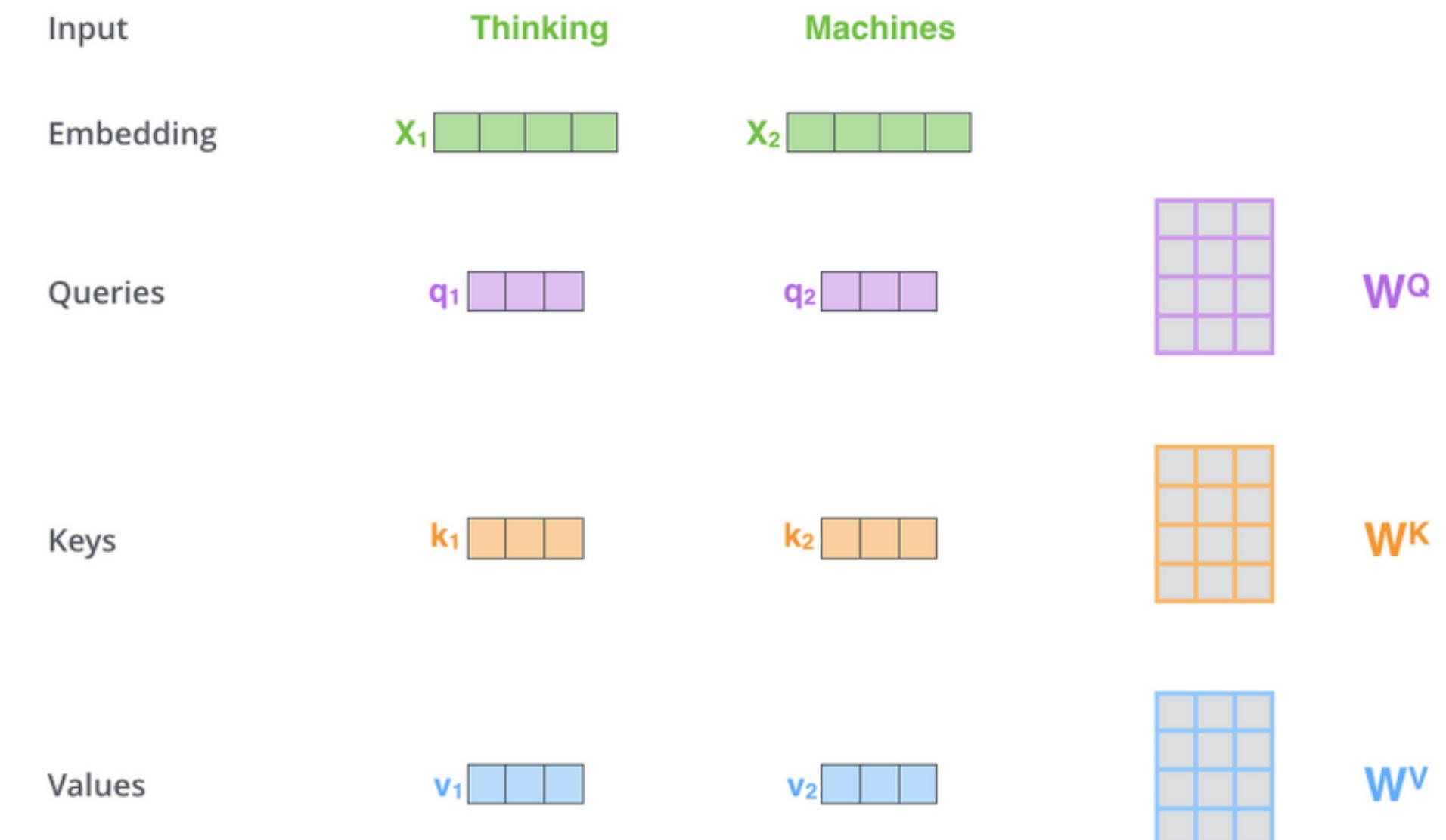
- SSD
- Data augmentation
- Anchor free
- Predict center of objects
- Classify objects
- Can do segmentation

Source: blog.roboflow.com/whats-new-in-yolov8/



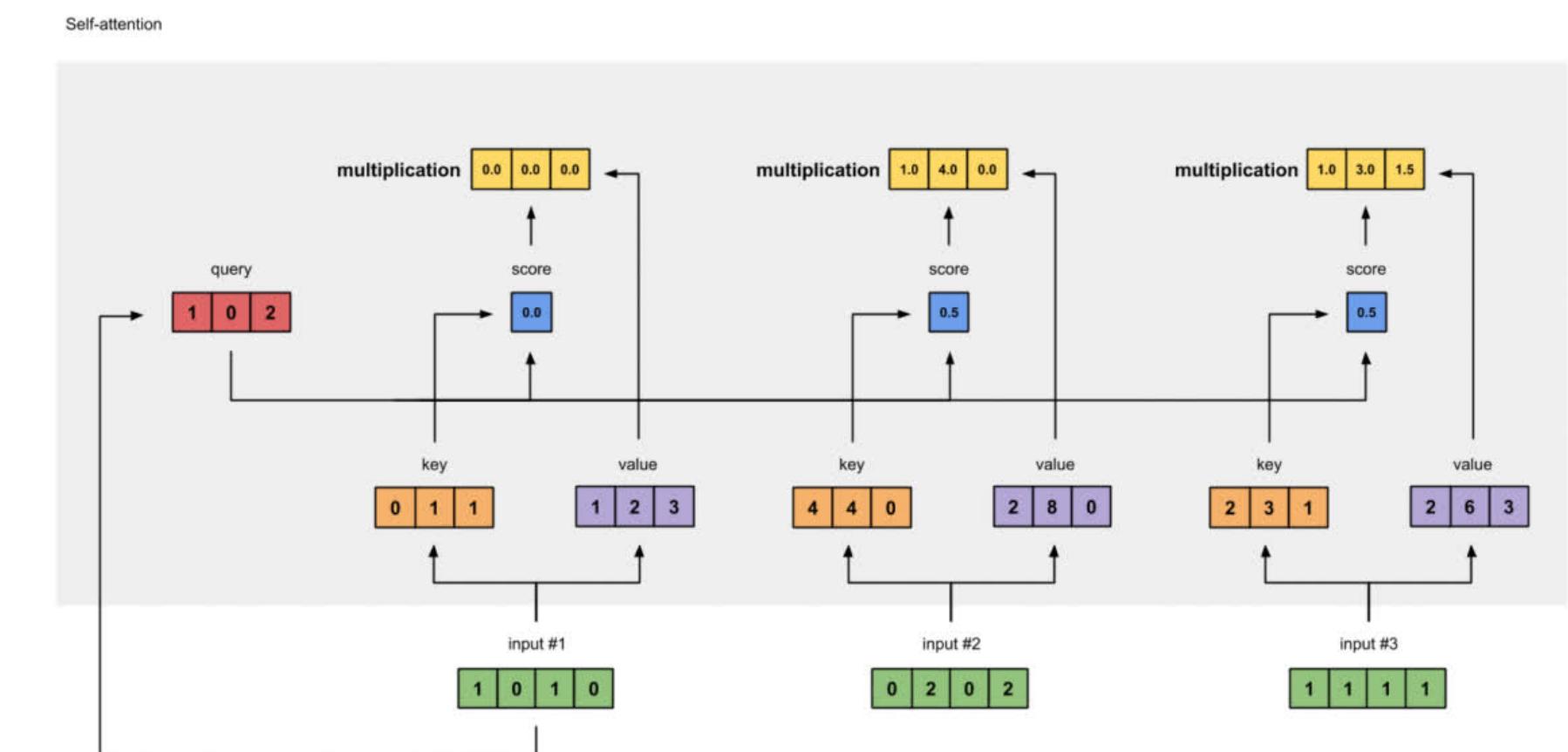
Transformer

Self-attention
(attention head)



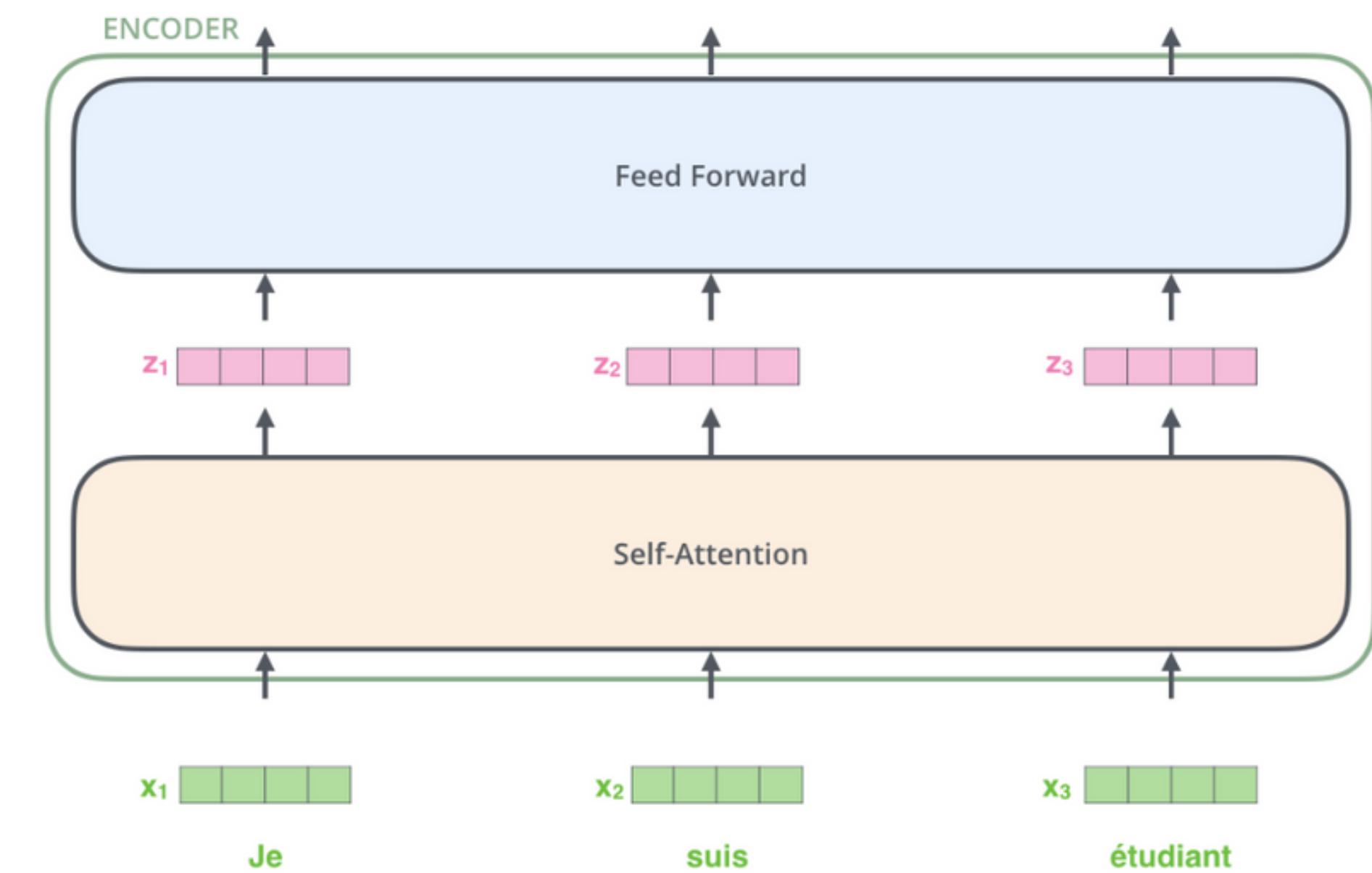
Transformer

Self-attention
(attention head)



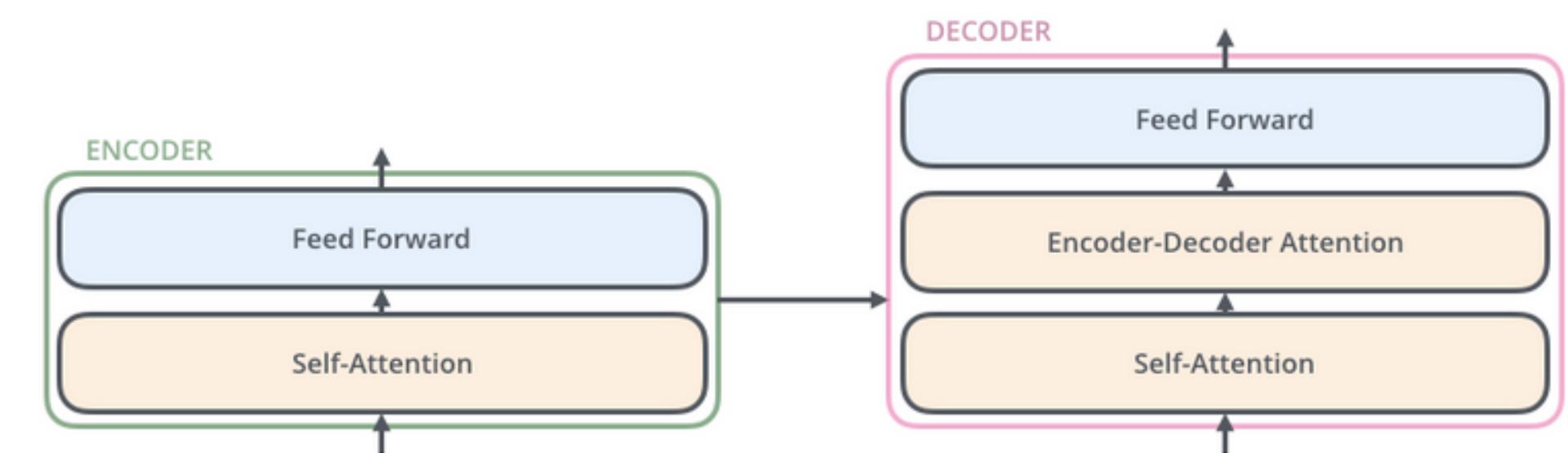
Transformer

Encoder

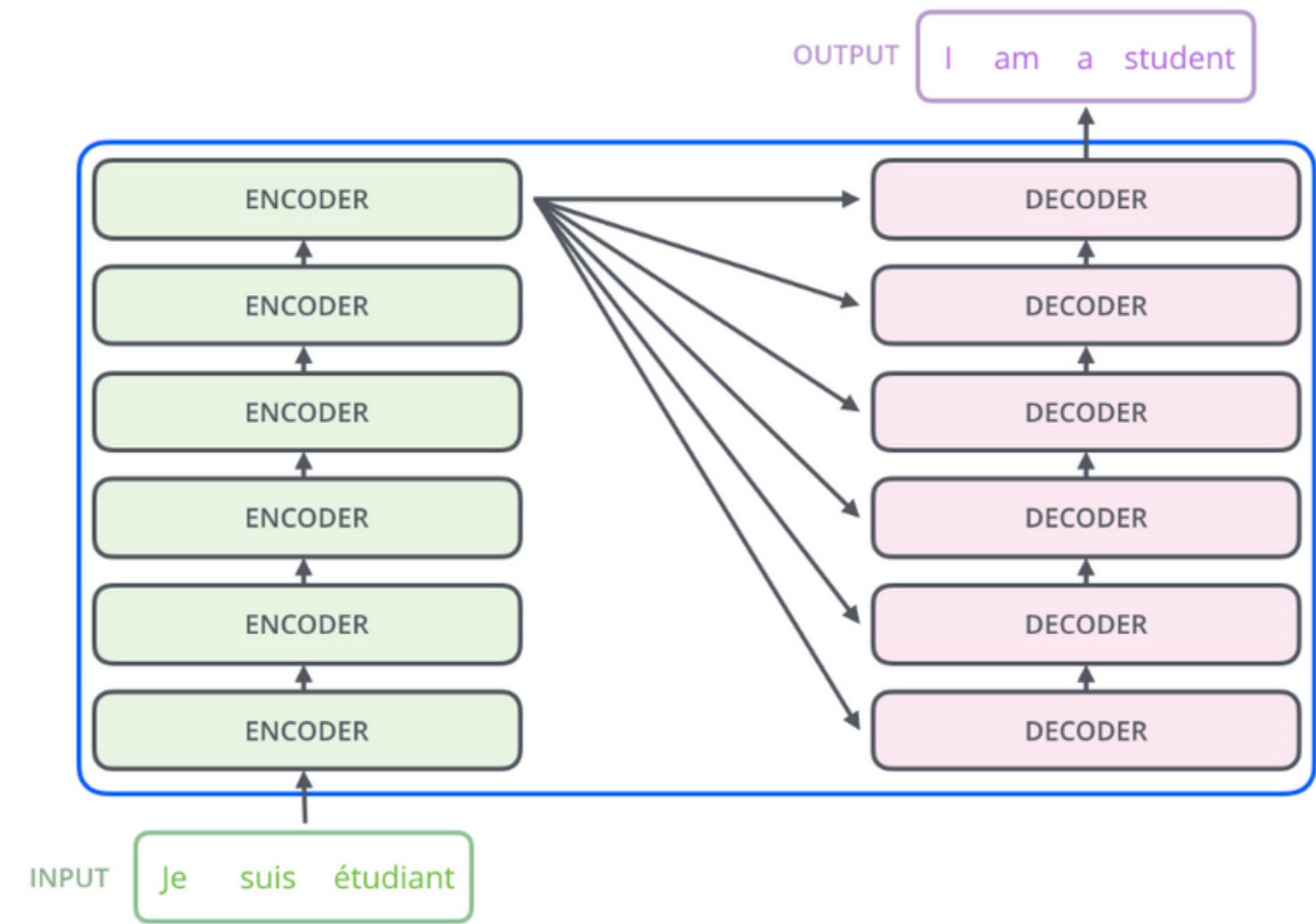


Transformer

Encoder / decoder



Transformer



Transformer

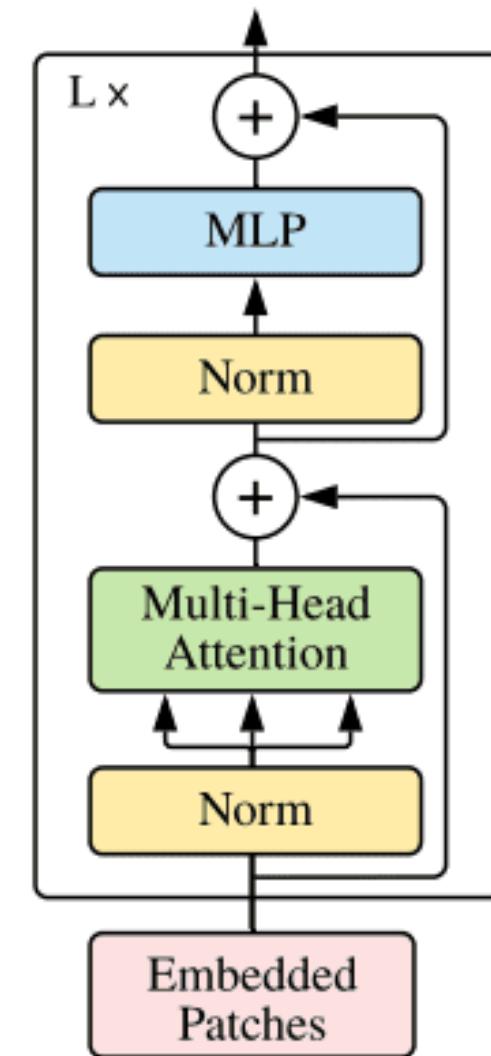
For images ? (ViT)



Transformer

For images ?

Transformer Encoder



Métrics

- Precision
- IoU
- mAP

Source: v7labs.com



**THANK
YOU**