

# OBJECT DETECTION

COMPUTER VISION



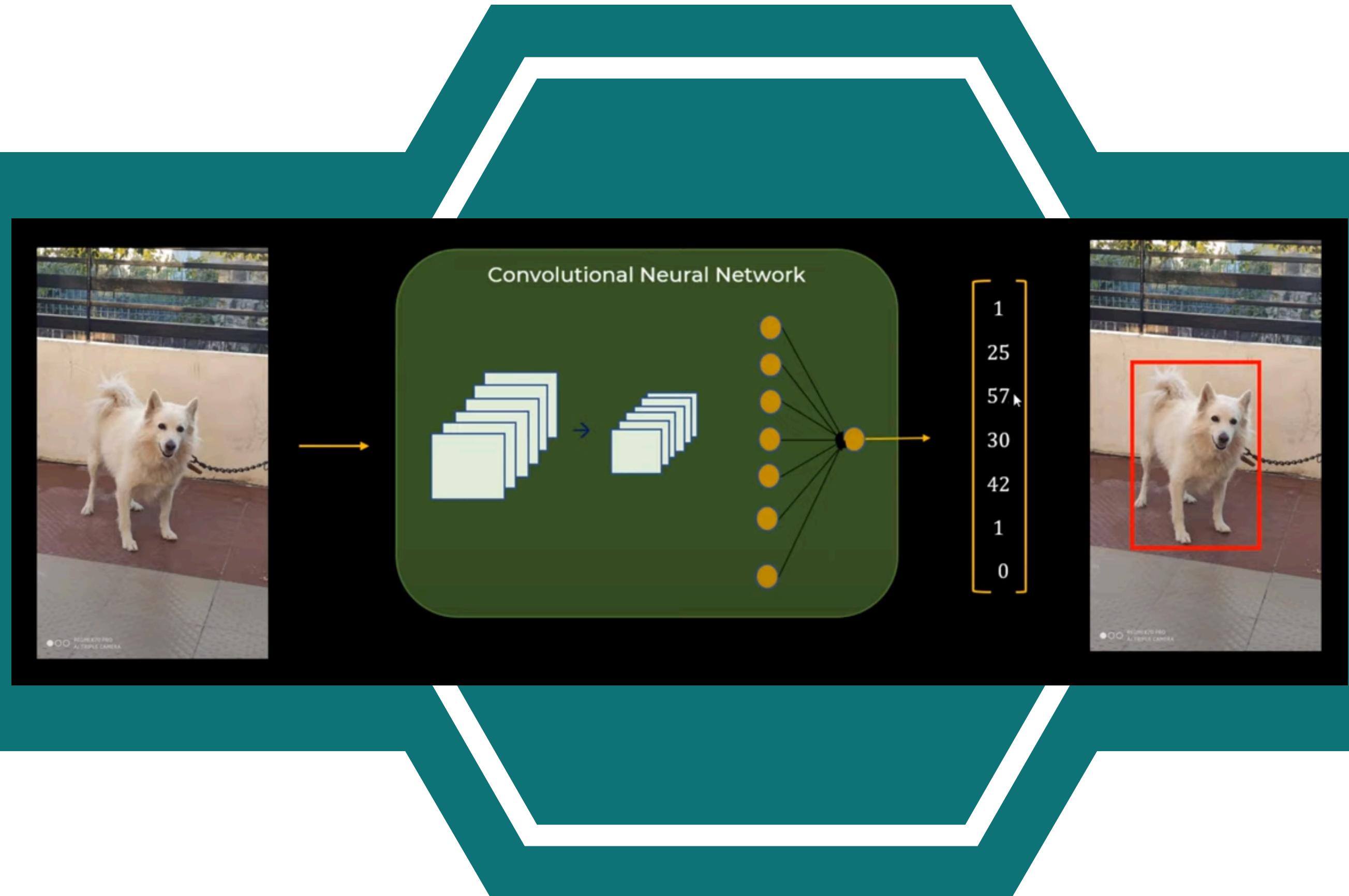


I have a  
**Plan**

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

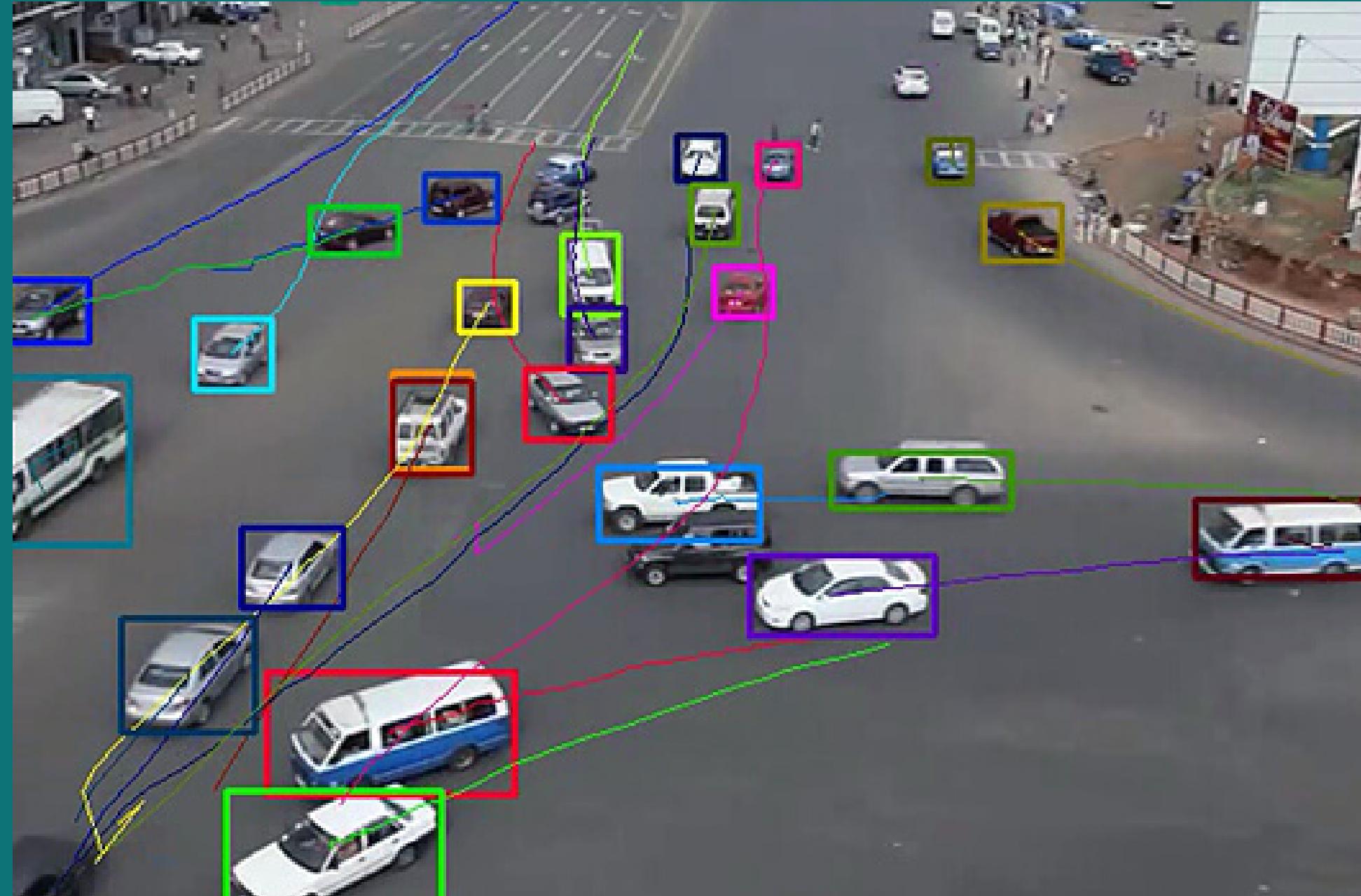
# Single Pass

Works well for  
single object  
detection



# What to do ?

In the case of many objects...



# Traditional Techniques

Intuitive but limited...



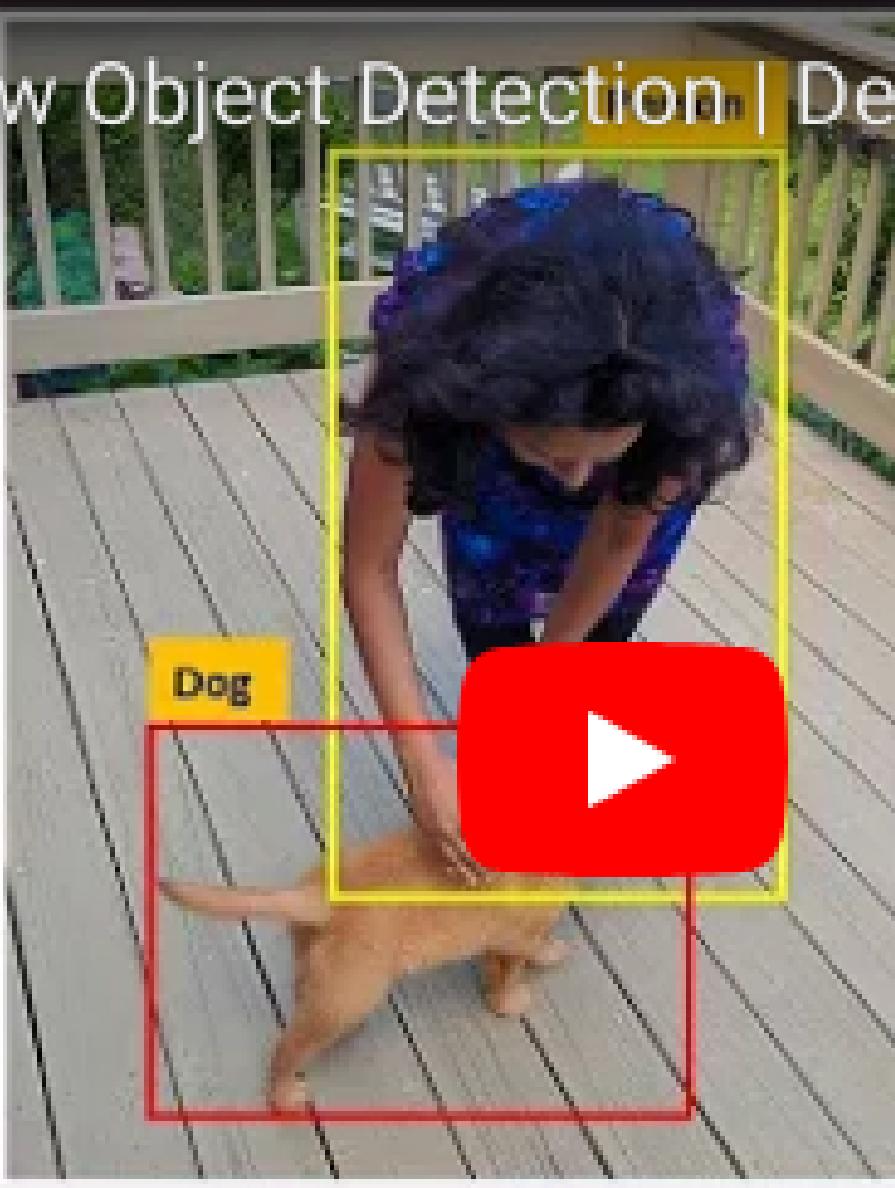
**Sliding  
window**



**Region  
proposal**



# Dhaval Patel



A YouTube video thumbnail titled "Sliding Window Object Detection | Deep Learning Tutorial". It shows a woman sitting on a wooden deck with a small dog. A yellow bounding box highlights the dog, and a red bounding box highlights the woman's head. A yellow play button icon is overlaid on the video frame. The thumbnail includes the "CODE BASICS" logo, a share button, and a timestamp of "30".

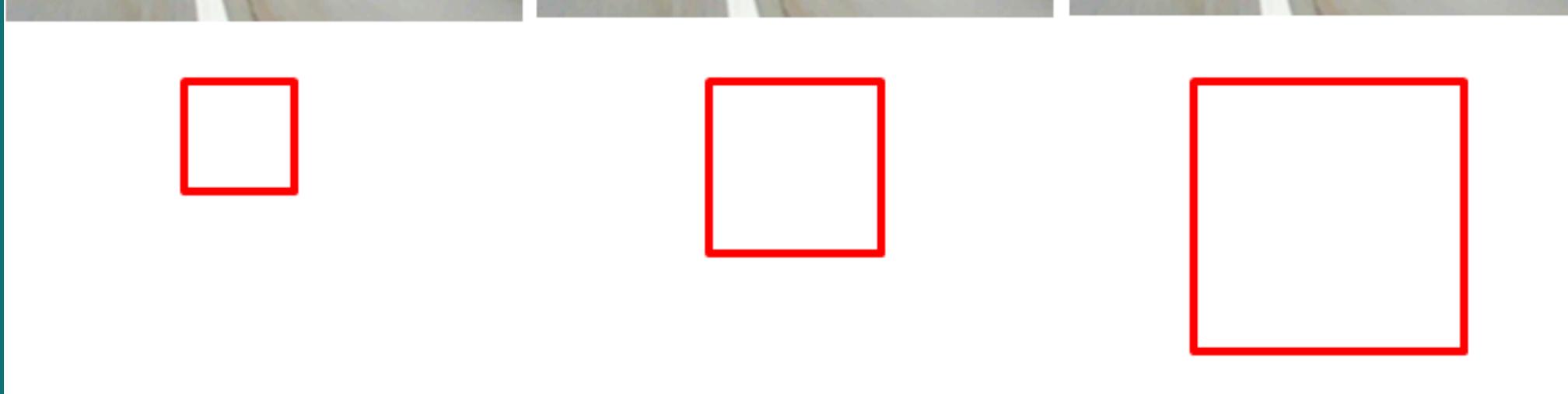
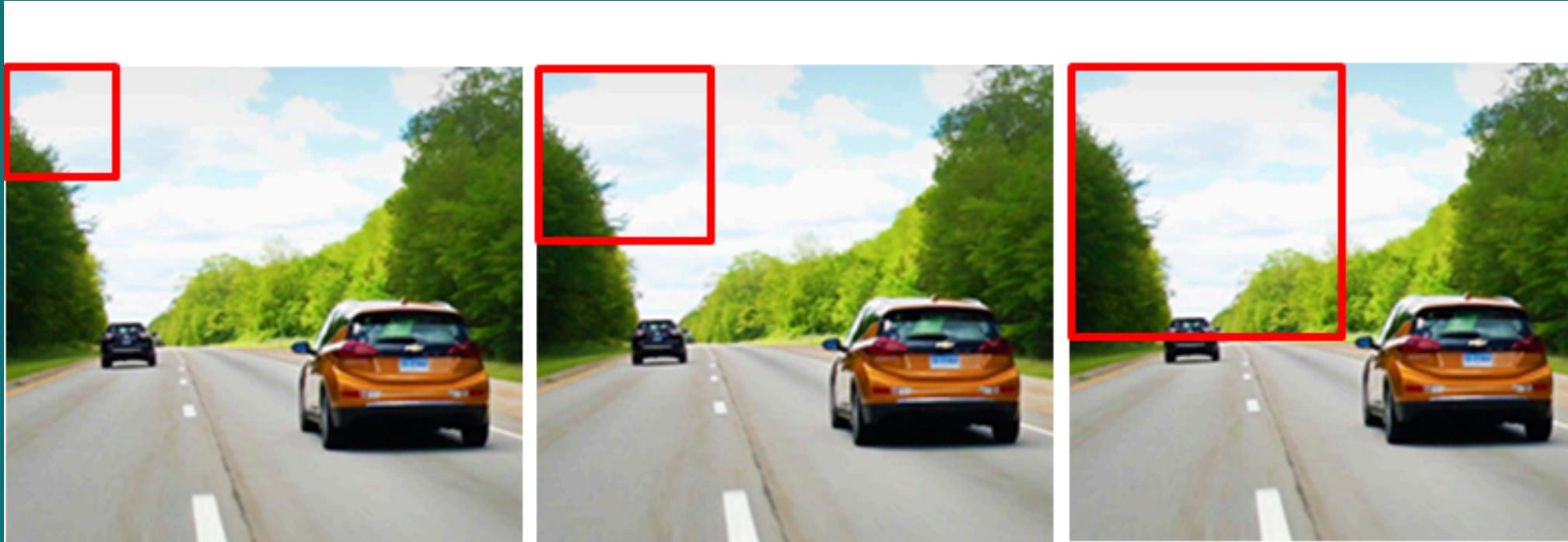
Sliding Window Object Detection | Deep Learning Tutori... Share

# Sliding Window Object Detection

Watch on YouTube

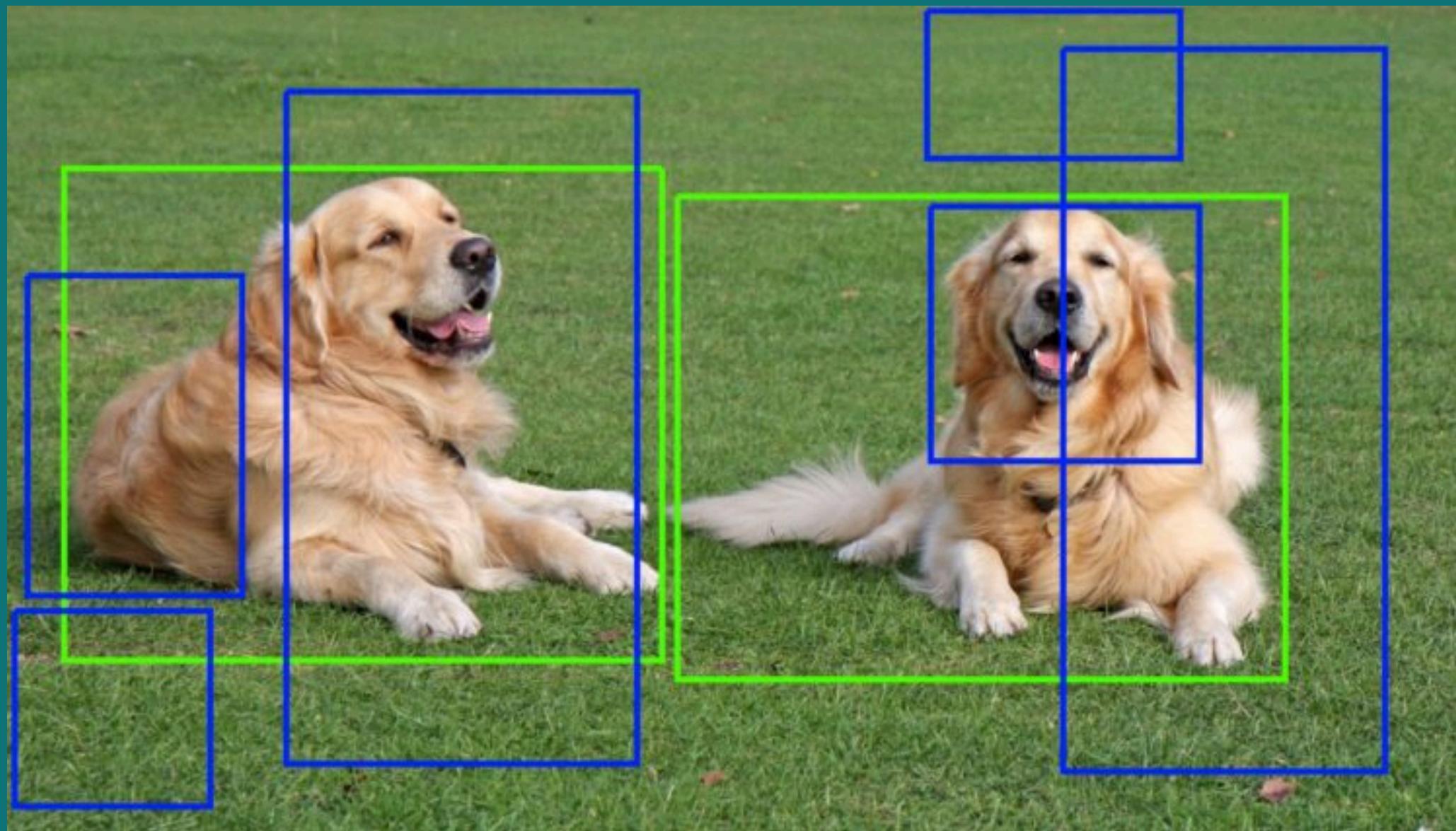
30

# Sliding Window



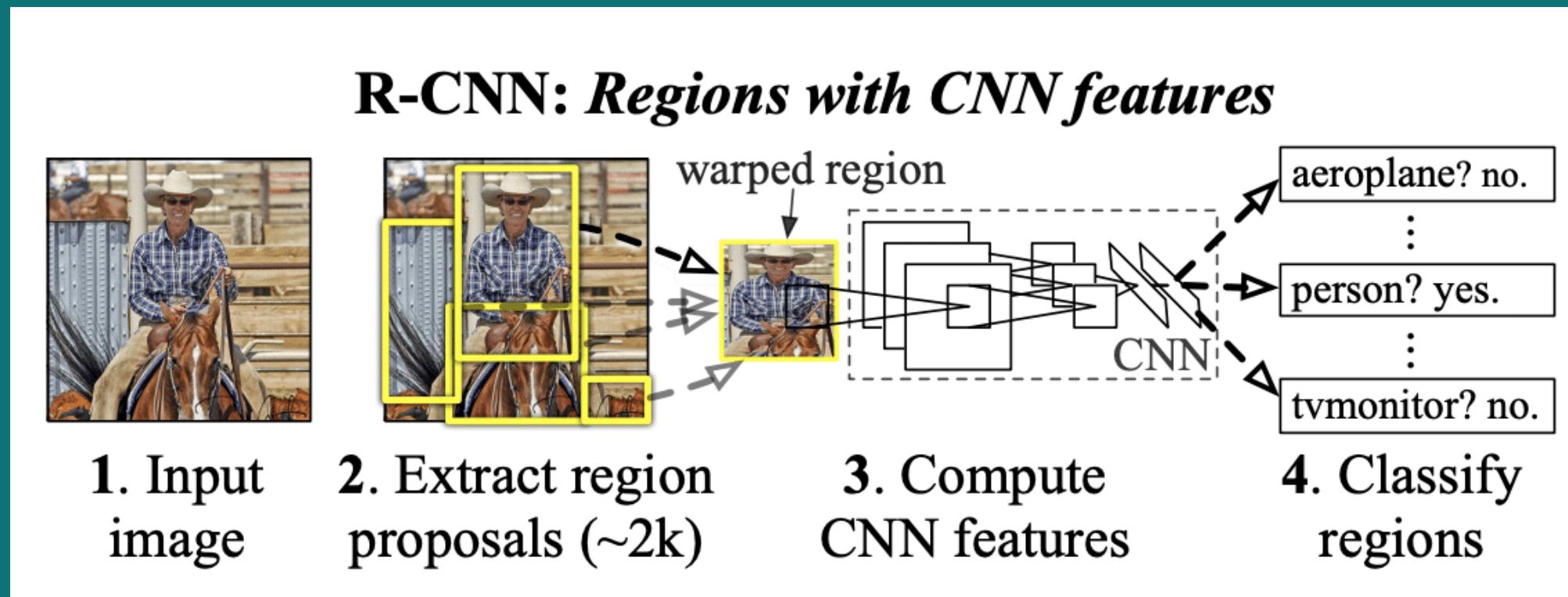
Guess the scale... !

# Sliding Window

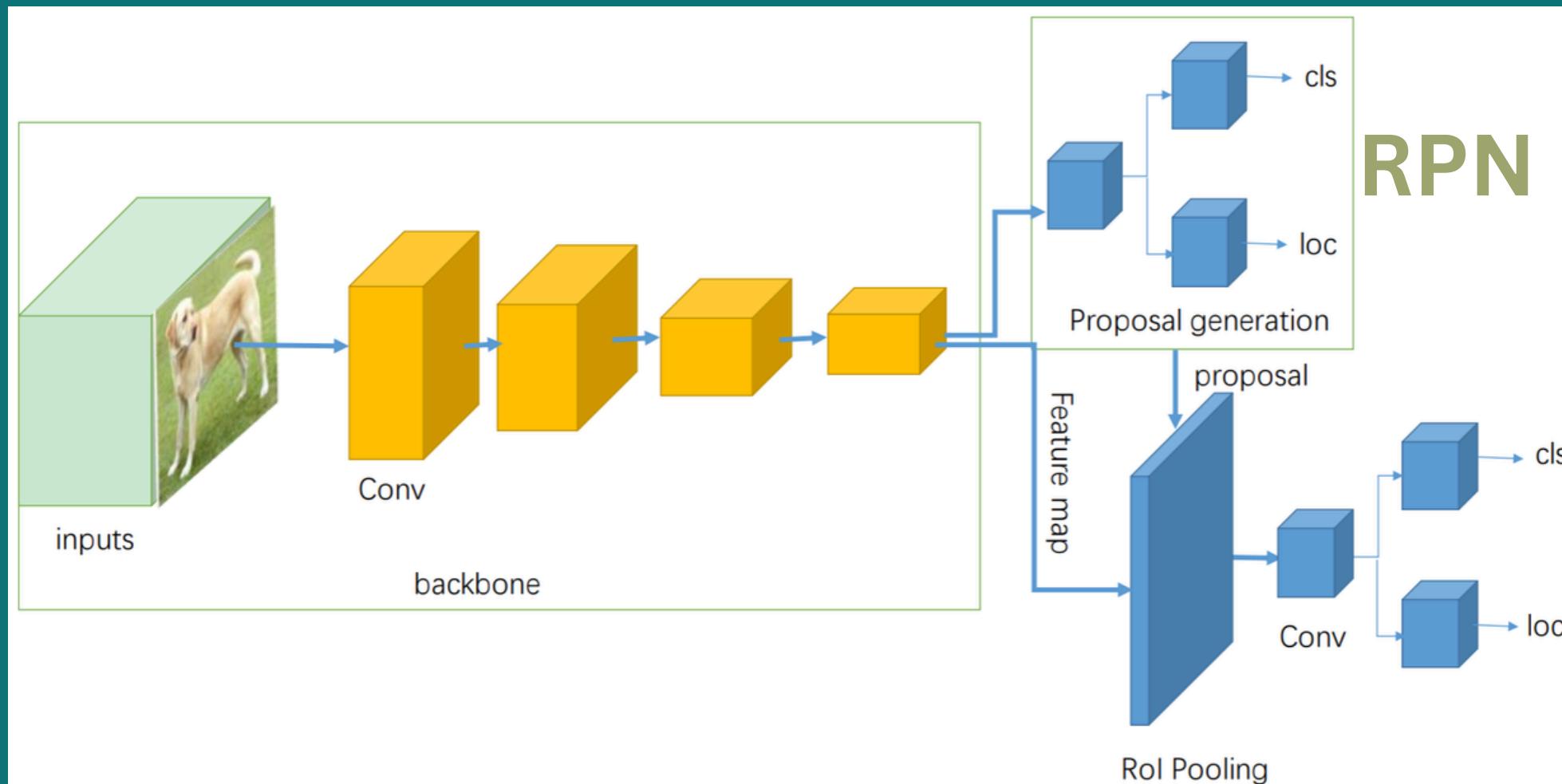


Too many predictions to make... !

# Regions Proposal



# R-CNN

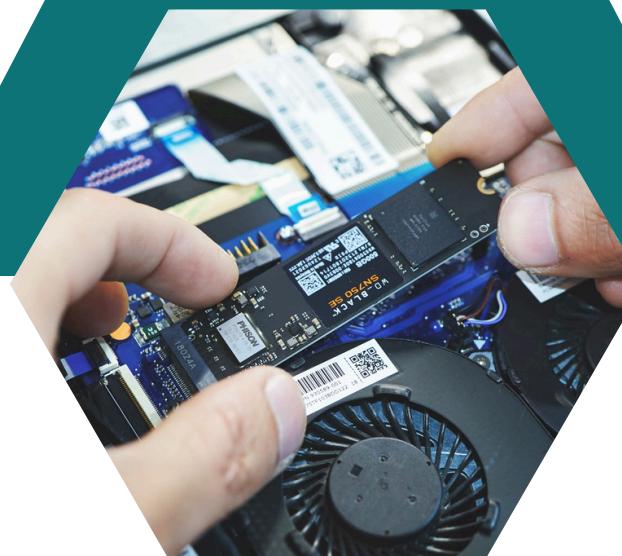


# Two-Stage Detector

Region Proposal  
&  
Region Classification / refinement

# Modern Techniques

Fast and... faster !

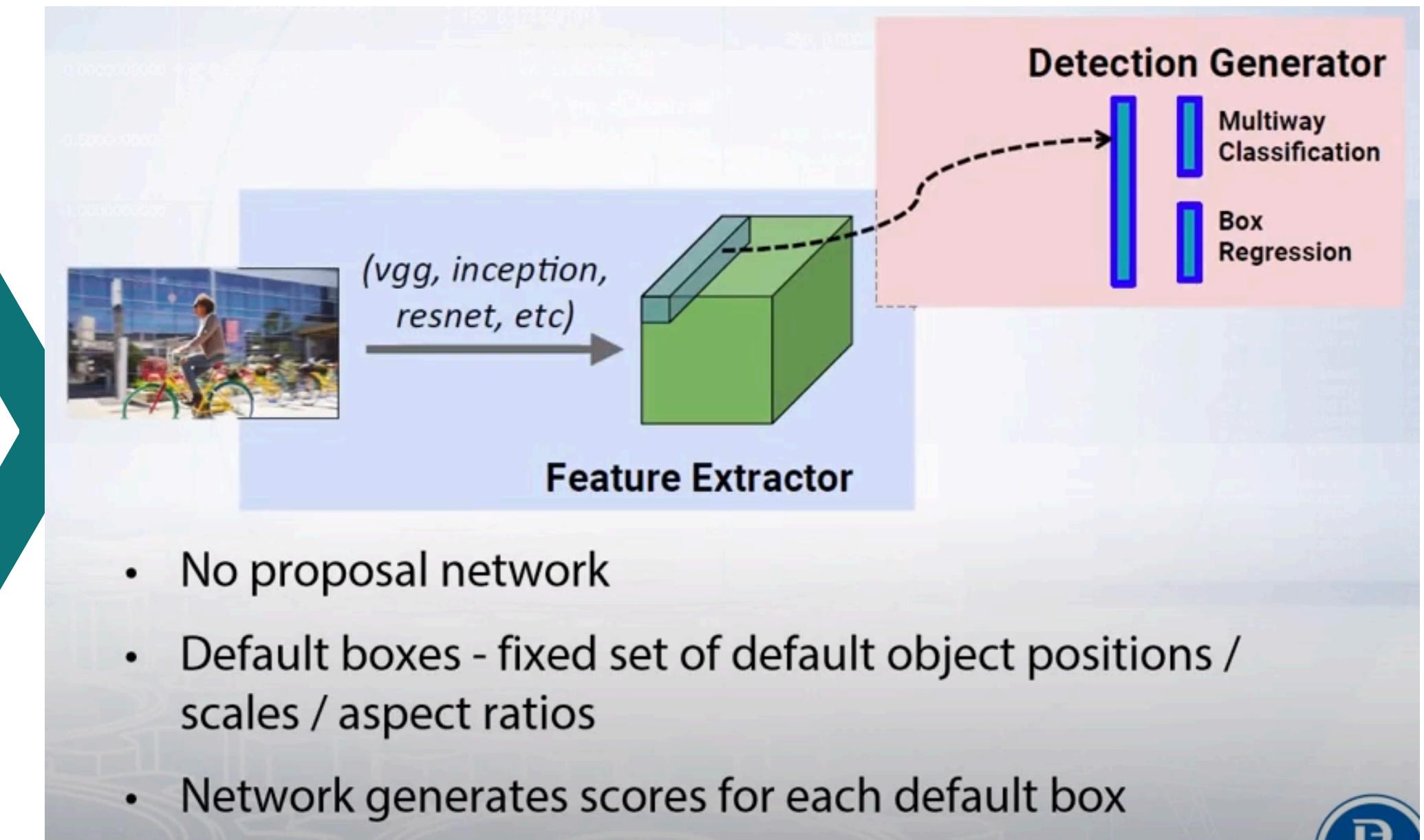


**SSD**



**Transformer**

# Single Shot Detector



# Single Shot Detector

- Separate input into  $3 \times 3$  cells.
- For each cell, we want to produce

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





# Dhaval Patel

The image shows a YouTube thumbnail for a video titled "What is YOLO algorithm? | Deep Learning Tutorial 31". The thumbnail features a man wearing glasses and a blue hoodie, standing in front of a red background. The video frame itself shows a person's arm reaching towards a small dog, with two bounding boxes overlaid: a red one around the dog and a yellow one around the person's arm. A large play button icon is in the bottom right corner of the thumbnail frame. The overall design includes a purple header bar at the top and a teal footer bar at the bottom.

What is YOLO Algorithm?

Watch on YouTube

What is YOLO algorithm? | Deep Learning Tutorial 31 (Te... Share

31

# YOLO v8

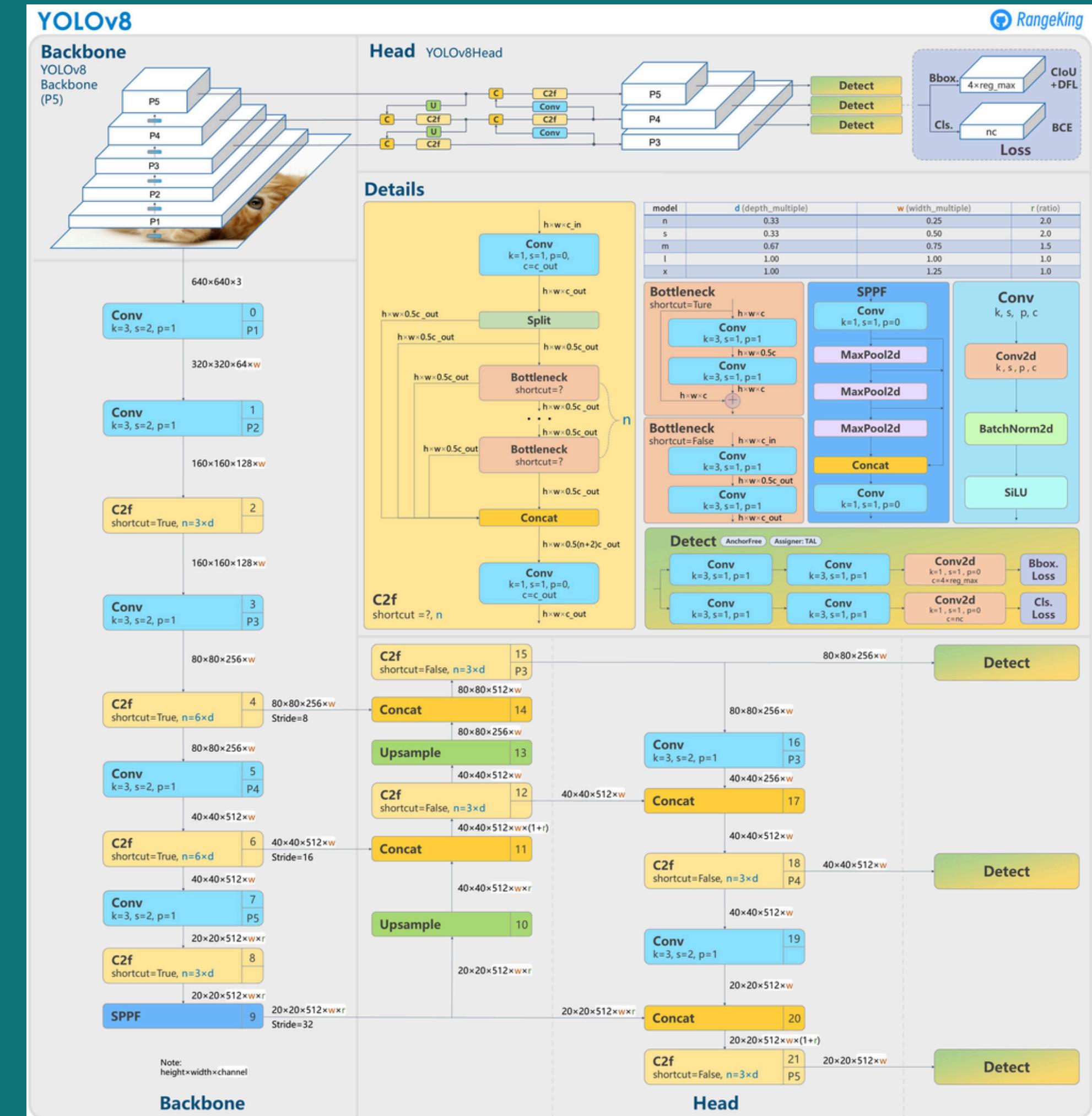
The state of the art solution



# YOLO V8

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

Source: [blog.roboflow.com/whats-new-in-yolov8/](https://blog.roboflow.com/whats-new-in-yolov8/)



# Modern Techniques

Fast and... faster !



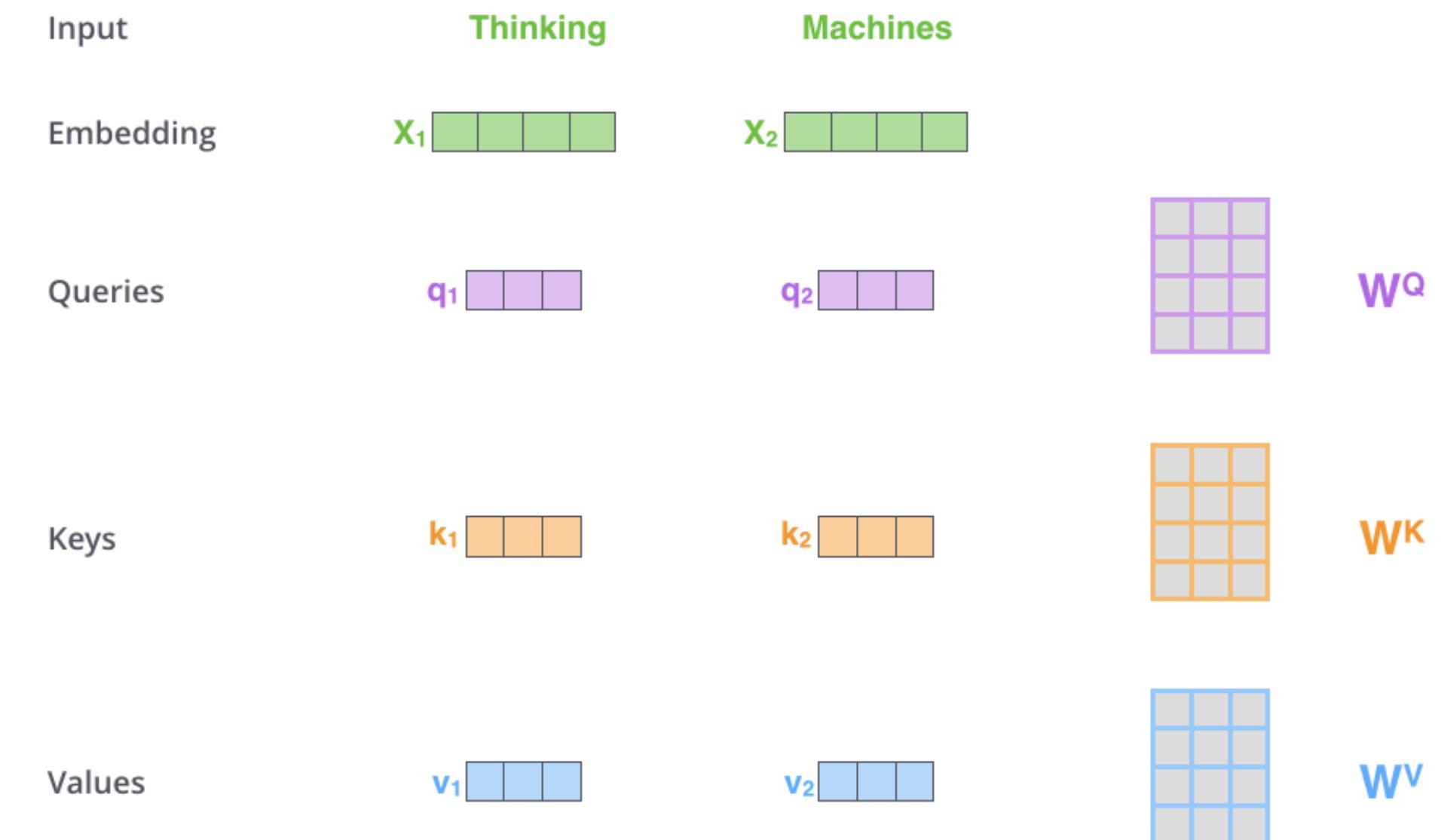
**SSD**



**Transformer**

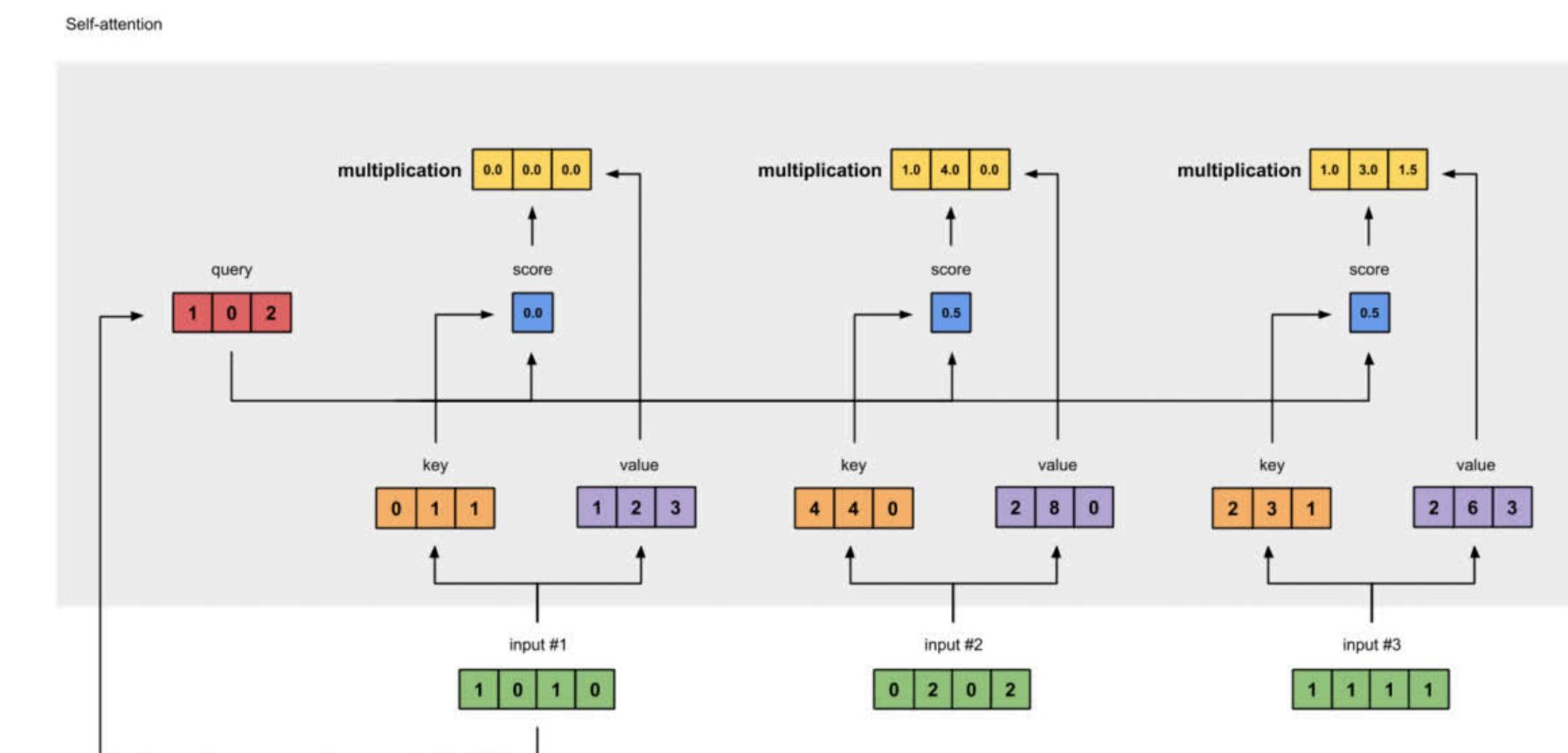
# Transformer

## Self-attention (attention head)



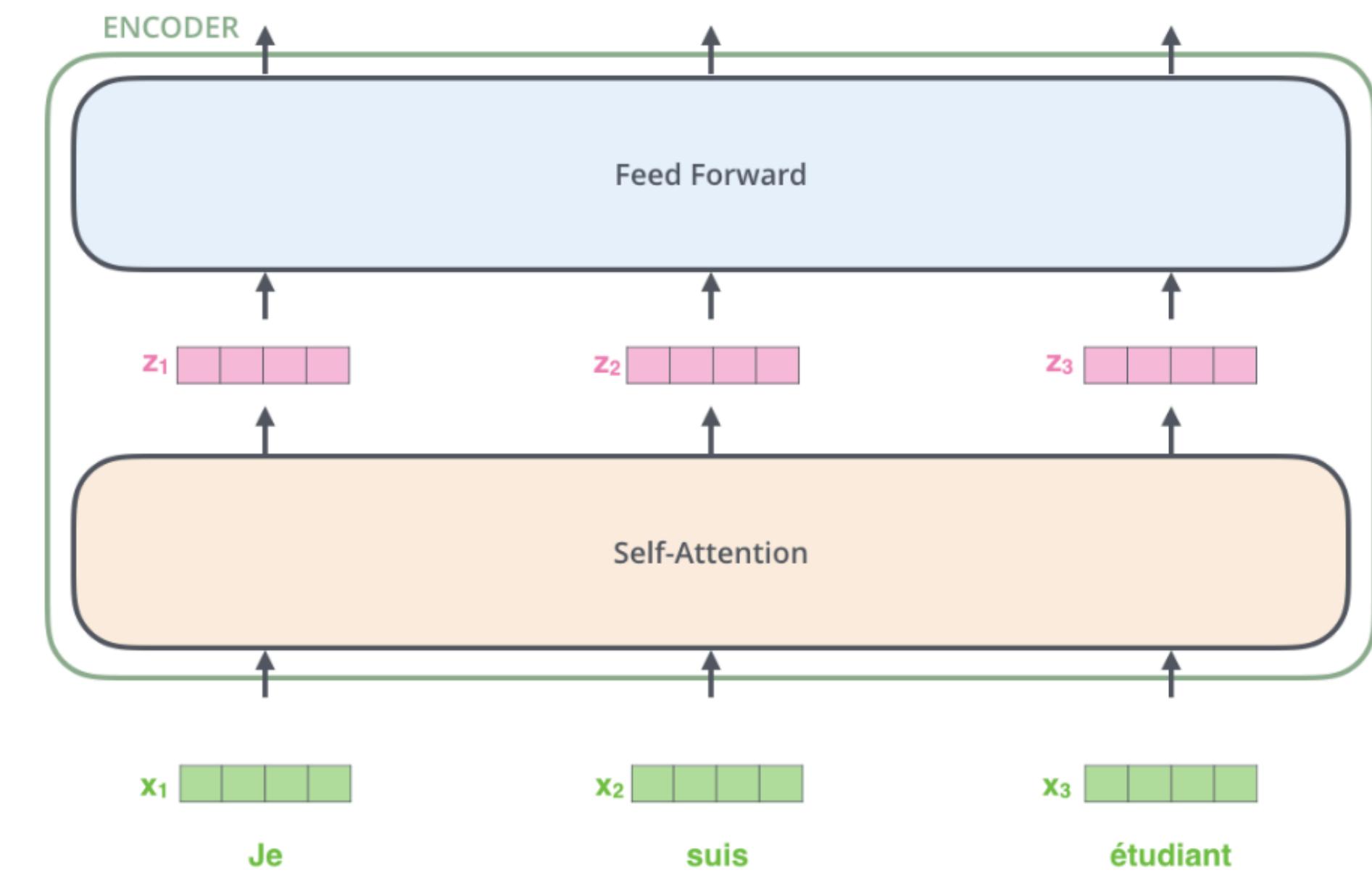
# Transformer

Self-attention  
(attention head)



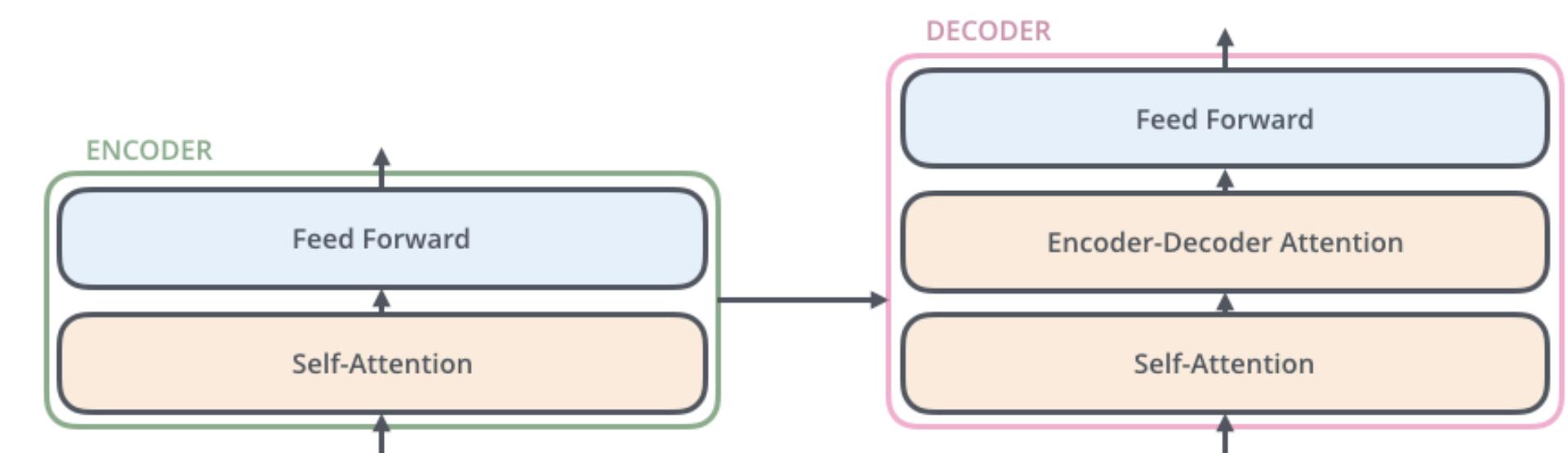
# Transformer

Encoder

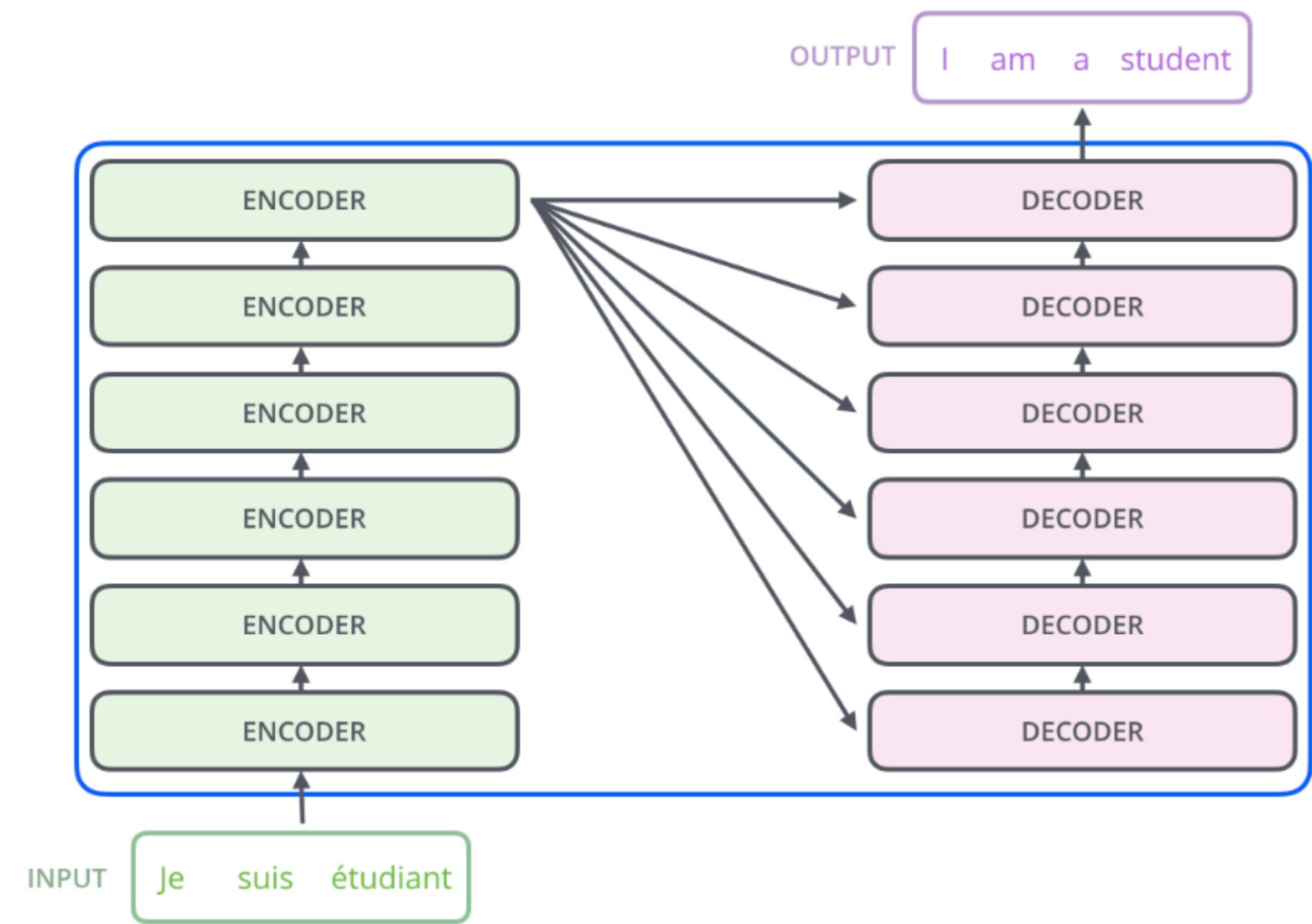


# Transformer

Encoder / decoder



# Transformer



# Transformer

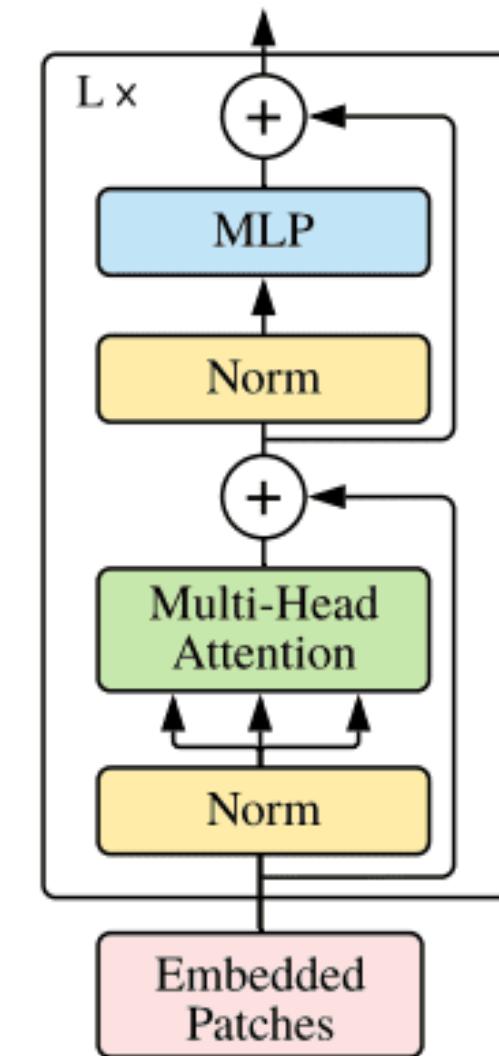
For images ? (ViT)



# Transformer

For images ?

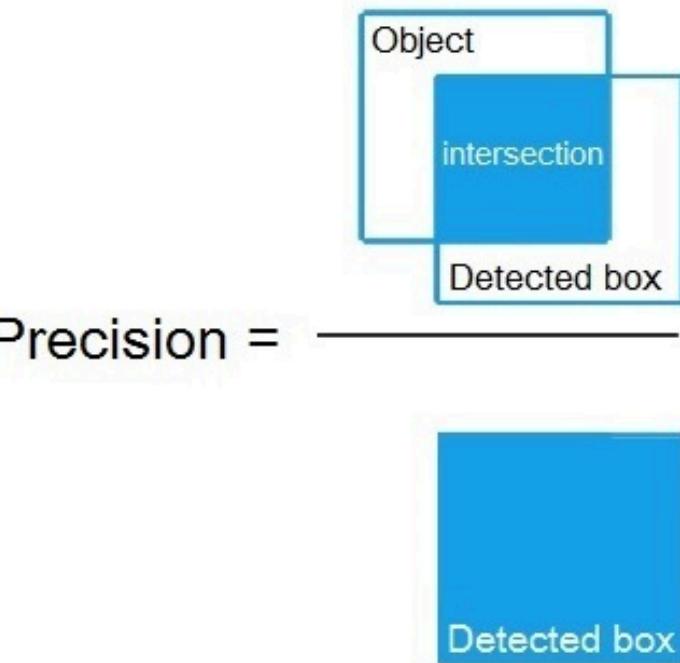
**Transformer Encoder**



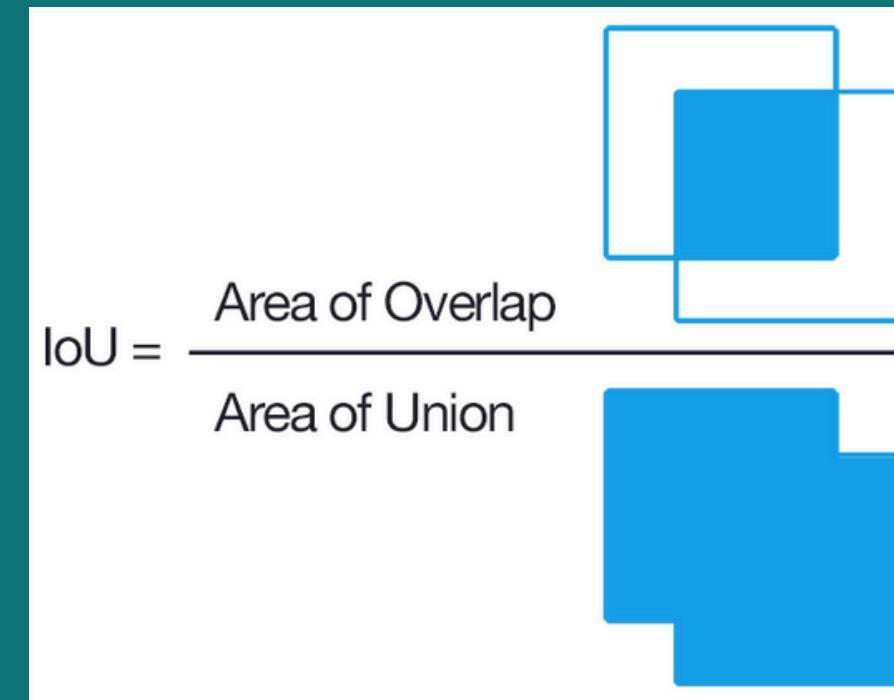
# Métrics

- Precision
- IoU
- mAP

Source: [v7labs.com](https://v7labs.com)



$$\text{Precision} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

*AP<sub>k</sub> = the AP of class k*  
*n = the number of classes*



**THANK  
YOU**