Pedestrian Detection

A introduction to YOLO and its implementation

Jenny Rong & Xuweiyi Chen

Outline

Some backgrounds about YOLO

- Object Detection

- Sliding Window Convolution-ally

- Bounding Box

- Intersection Over Union

- Non-max Suppression

- Anchor Boxes

- YOLO algorithm and YOLOv3 algorithm

Object Detection

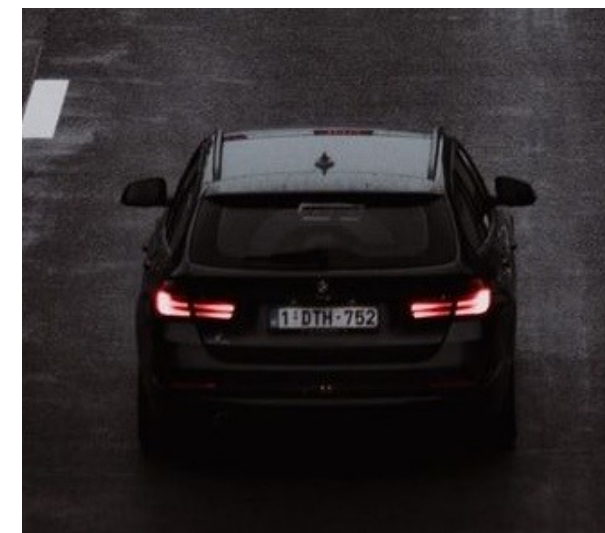## What are localization and detection?

- Image Classification

  - One object; Only identify the object

- Classification with Localization

  - One object; Identity the object and its position in the image

- **Detection**

  - Multiple objects; Identity the objects and their positions in the image

Object Detection

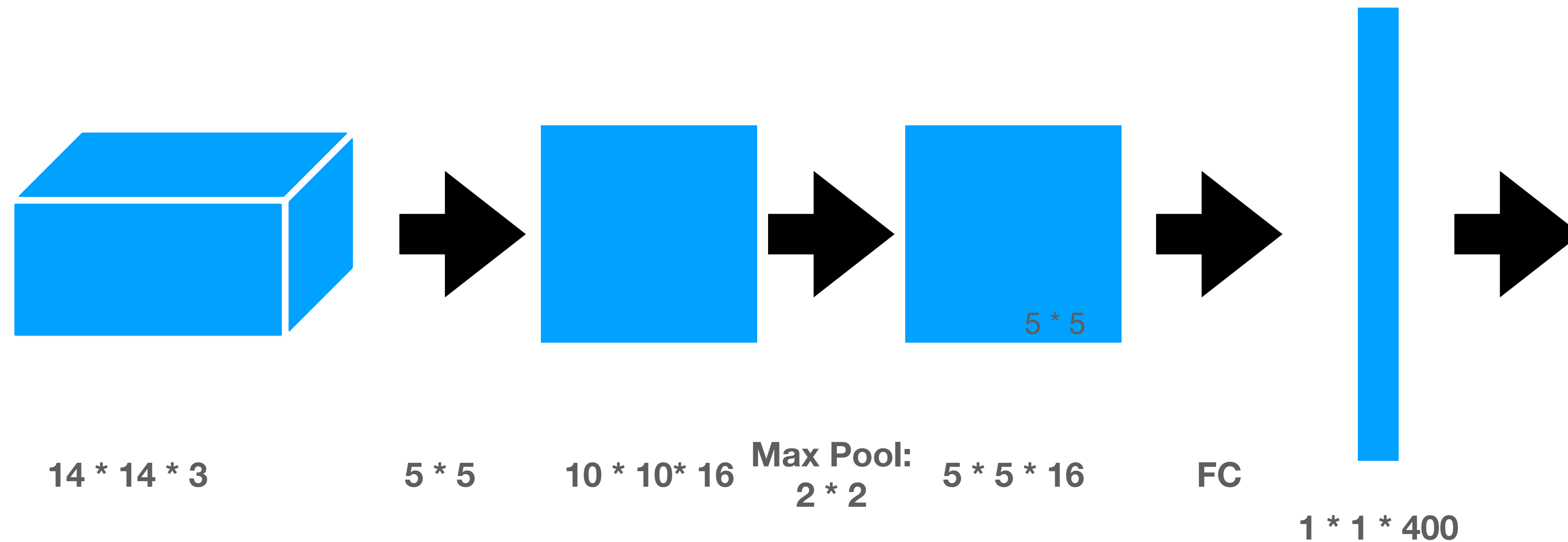## Sliding Window Detection
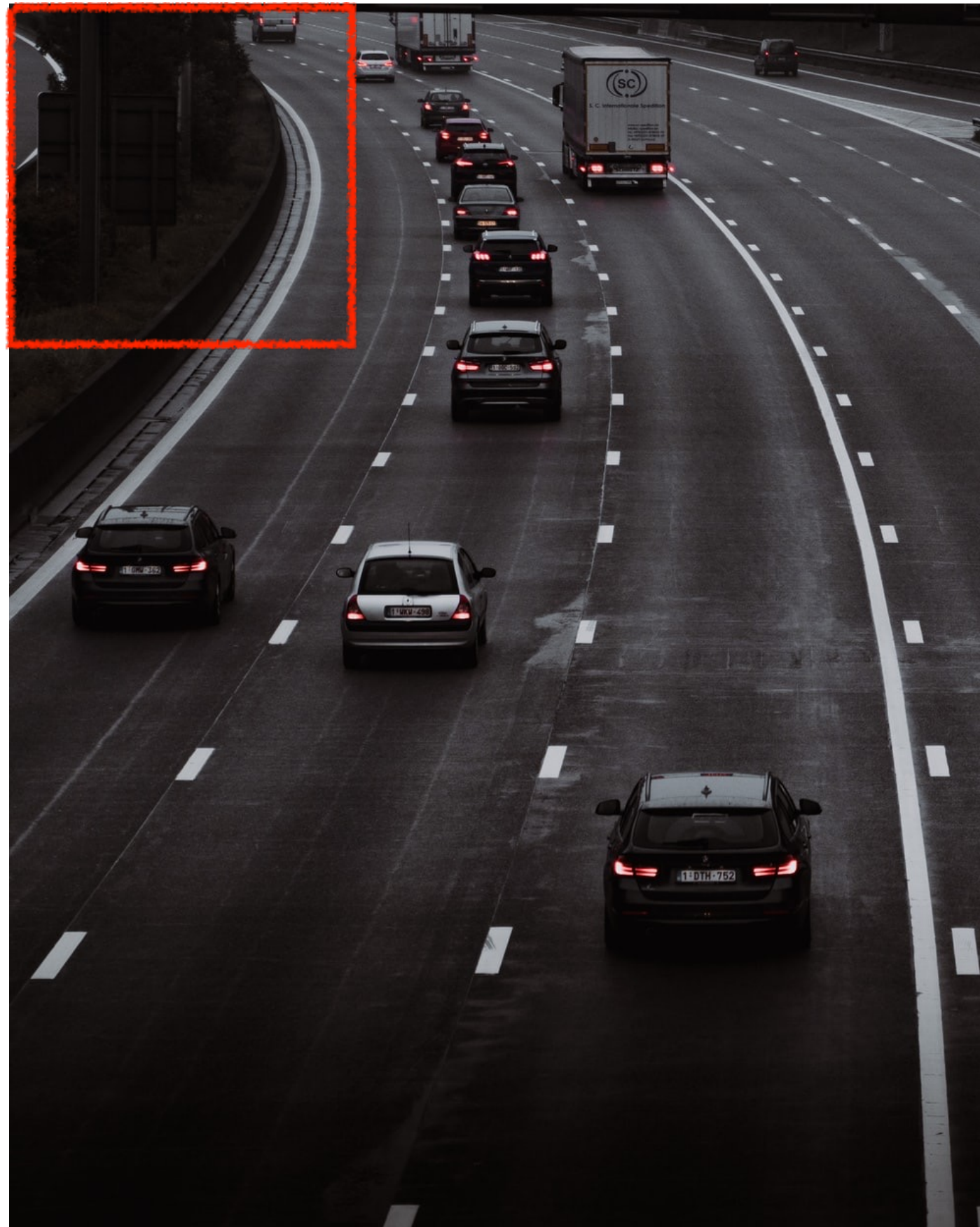
Crop Out:

Feed the red region into previous Conv net

Some Disadvantages of this algorithm:

1. This is very computational expensive.
2. Reducing the coarse granularity may hurt performance.

Sliding window into Convolutional Layers

Turning FC layer into Convolutional Layers

14 * 14 * 3   5 * 5   10 * 10* 16   **Max Pool:**   5 * 5 * 16   FC
                                                    2 * 2
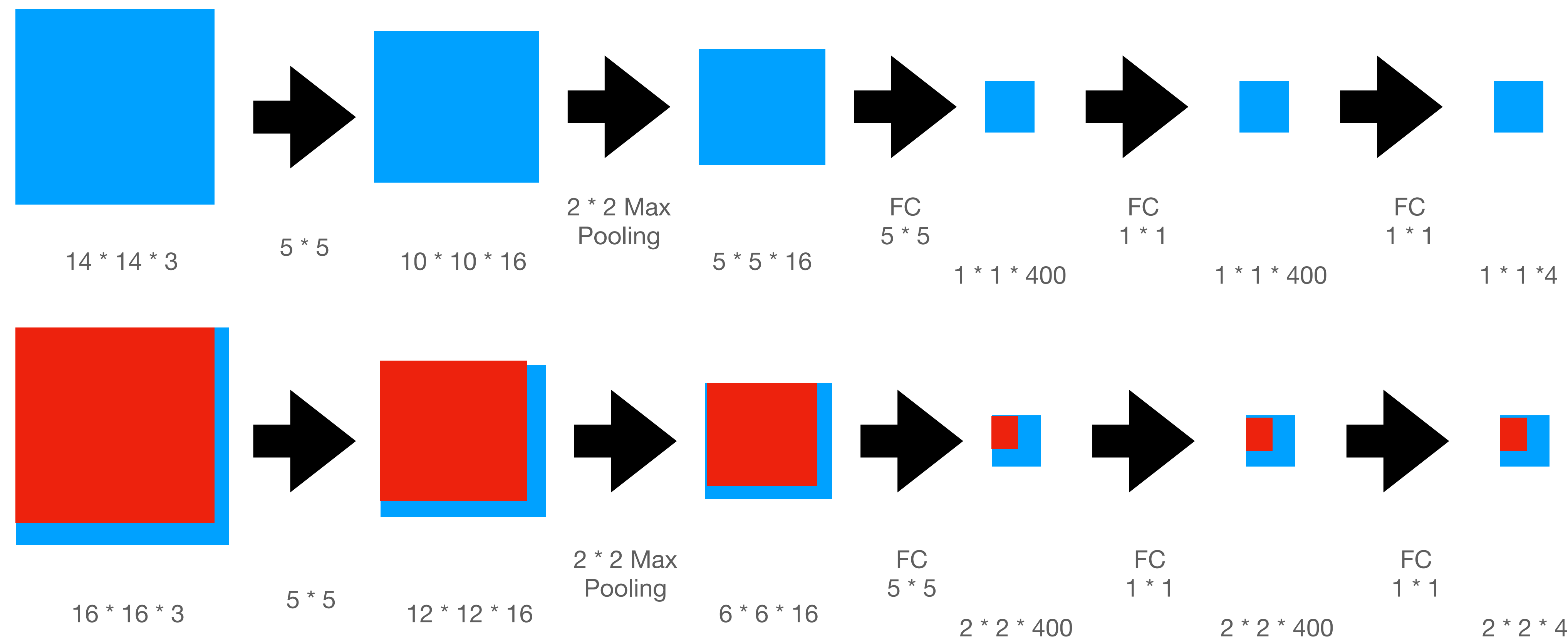
5 * 5 (label inside third box)

1 * 1 * 400

400 different 5 *5 * 16 filter we can also achieve the same influence

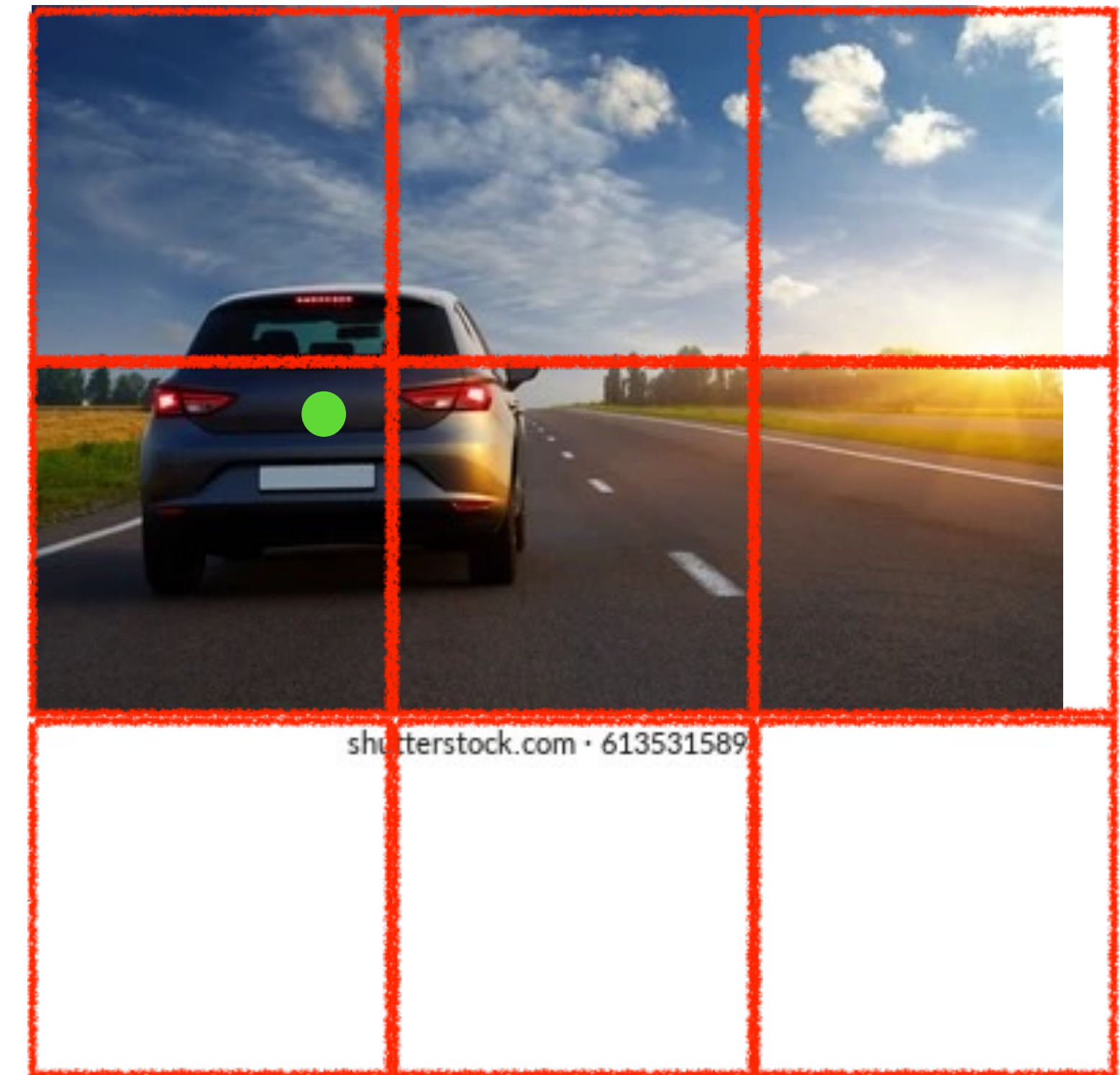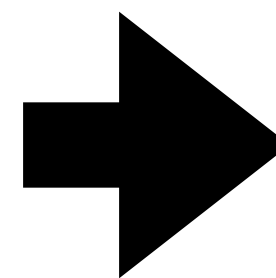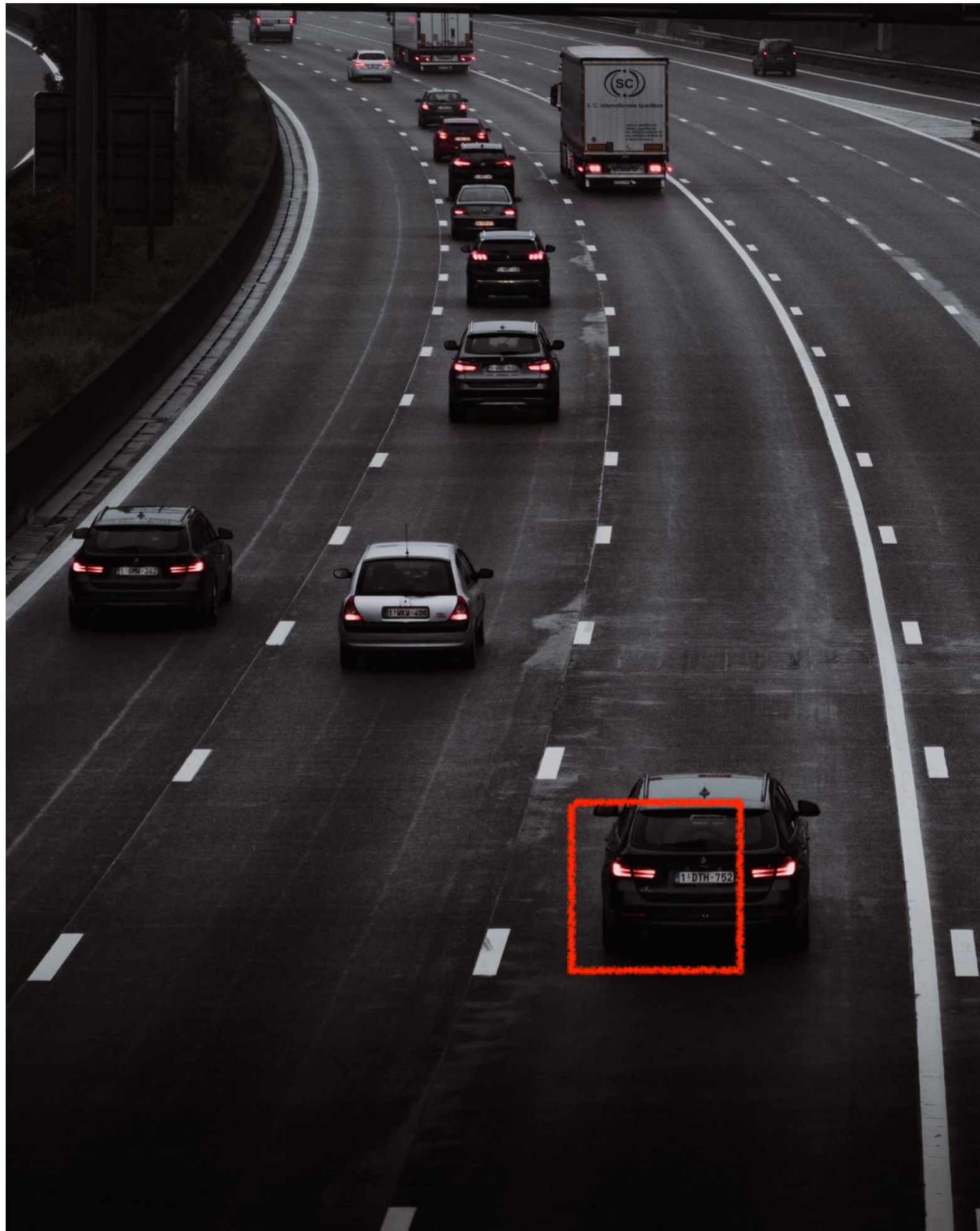This shows how we can take these fully connected layers and implement them using convolutional layers.

Result of Turning FC layer into Convolutional Layers

We do not have to input each section of image into networks

14 * 14 * 3   5 * 5   10 * 10 * 16   2 * 2 Max Pooling   5 * 5 * 16   FC 5 * 5   1 * 1 * 400   FC 1 * 1   1 * 1 * 400   FC 1 * 1   1 * 1 *4

16 * 16 * 3   5 * 5   12 * 12 * 16   2 * 2 Max Pooling   6 * 6 * 16   FC 5 * 5   2 * 2 * 400   FC 1 * 1   2 * 2 * 400   FC 1 * 1   2 * 2 * 4
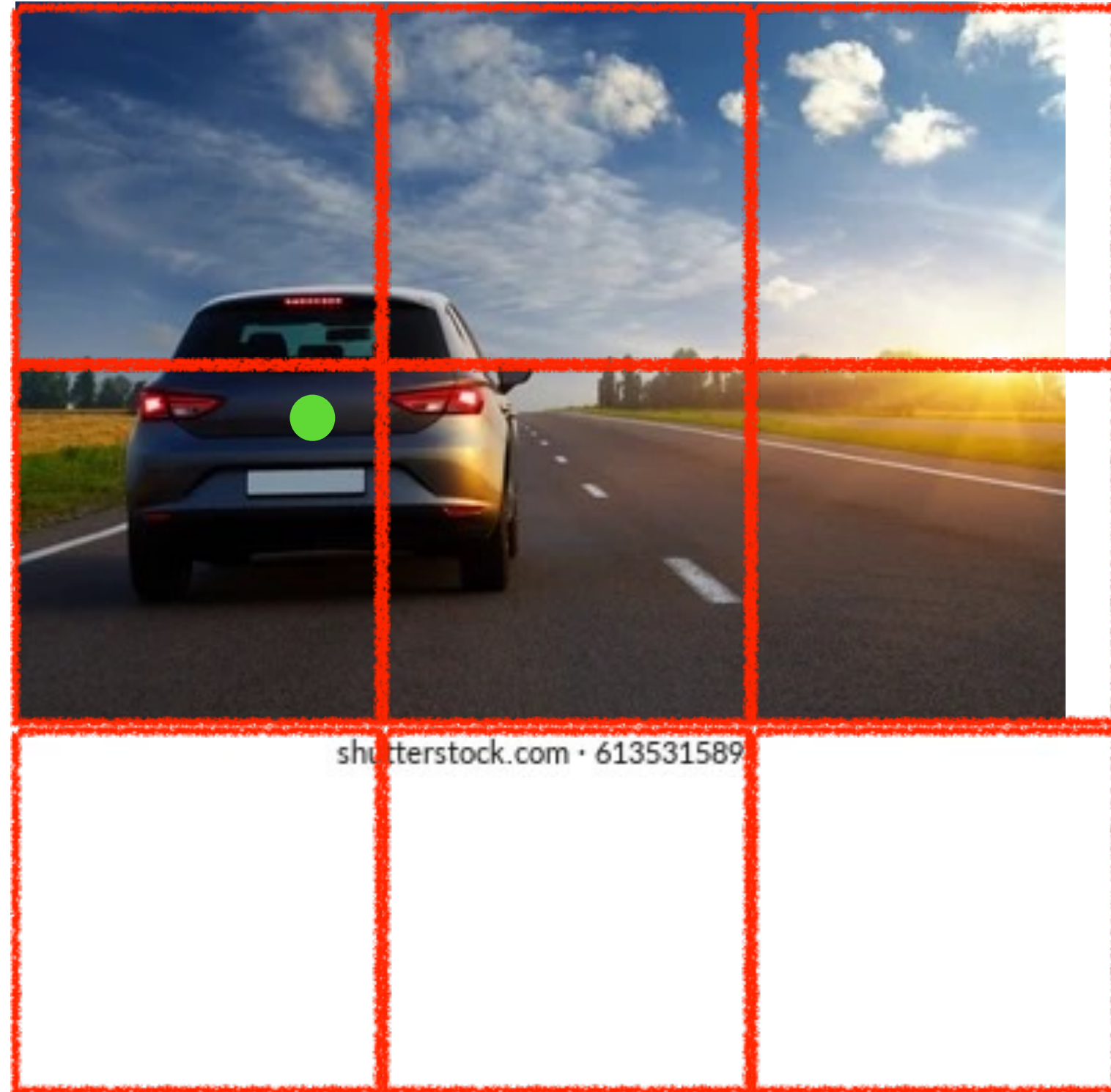
Bounding Box Prediction

Enhance bounding box prediction

Consider the training vectors and its dimension

Labels for training for each grid cell:

[P1,
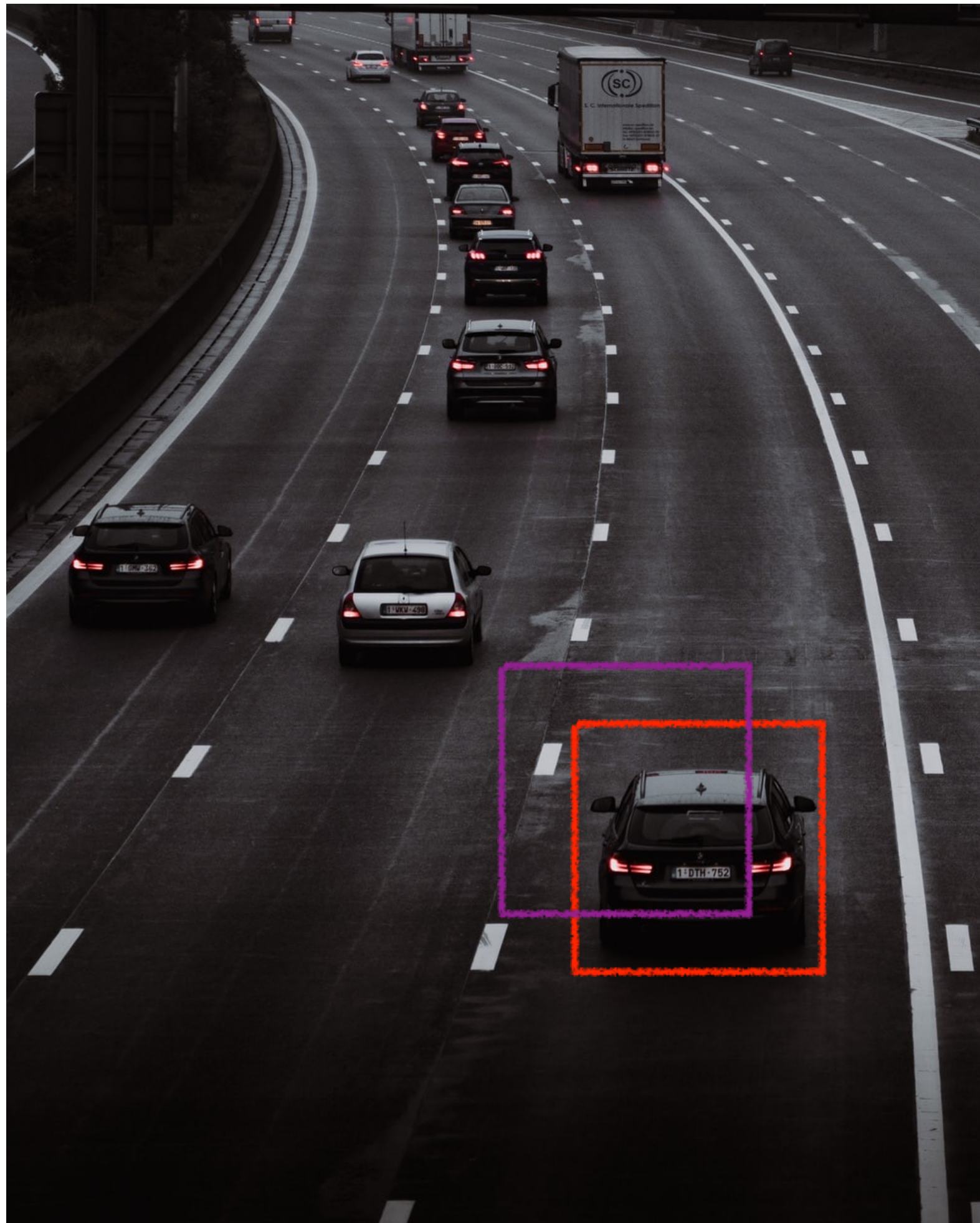Bx,
By,
Bh,
Bw,
O1,
O2,
O3]

Since we are using 3 * 3 grid, we should have 3 * 3 * 8 vectors

Consider the training vectors and its dimension

Problem: What if there are more than one object in one grid?

IOU

How Intersection over union works?

Intersection over union (IOU) is
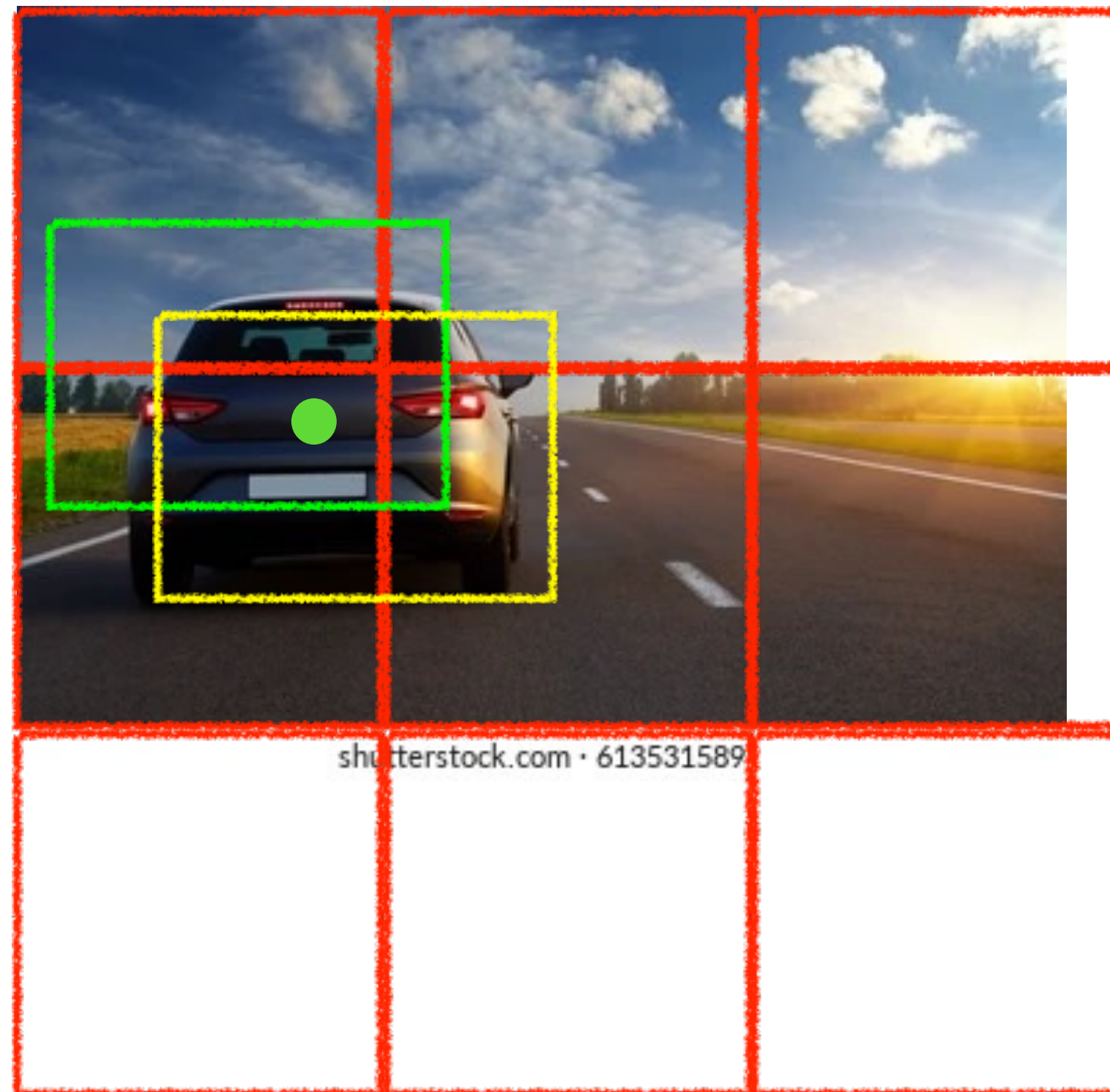
The intersection
between red and purple
boxes
/
The Union of purple
and red boxes

Convention if IoU > 0.5

Non-max suppression

Detect one object only once

Suppose P(yellow) = 0.9 and P(Green) = 0.6.

We will select P(yellow) based on non-max suppression.

Details of the algorithm:
1. Scan all output prediction
2. Discard all boxes with p < 0.6
3. If there are remaining boxes, we pick the highest P, and consider as a prediction
4. Discard any remaining box with IoU > 0.5 with the box we predicted in the previous step
5. Go back to step 3.

Question: What if the box we decide to discard contains different categories of object?

Anchor Box

Overlapping Objects
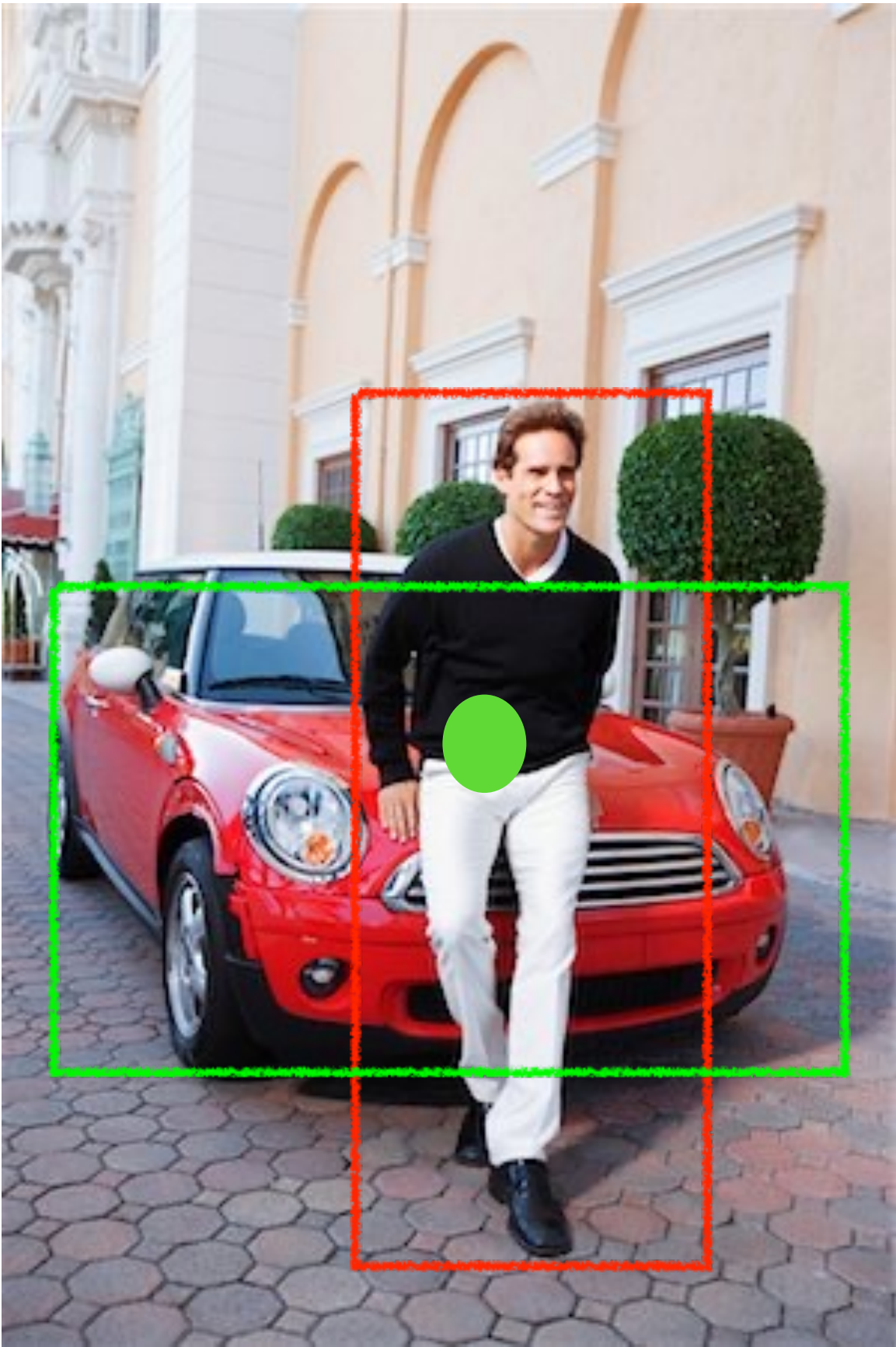
Let us construct two anchor box:

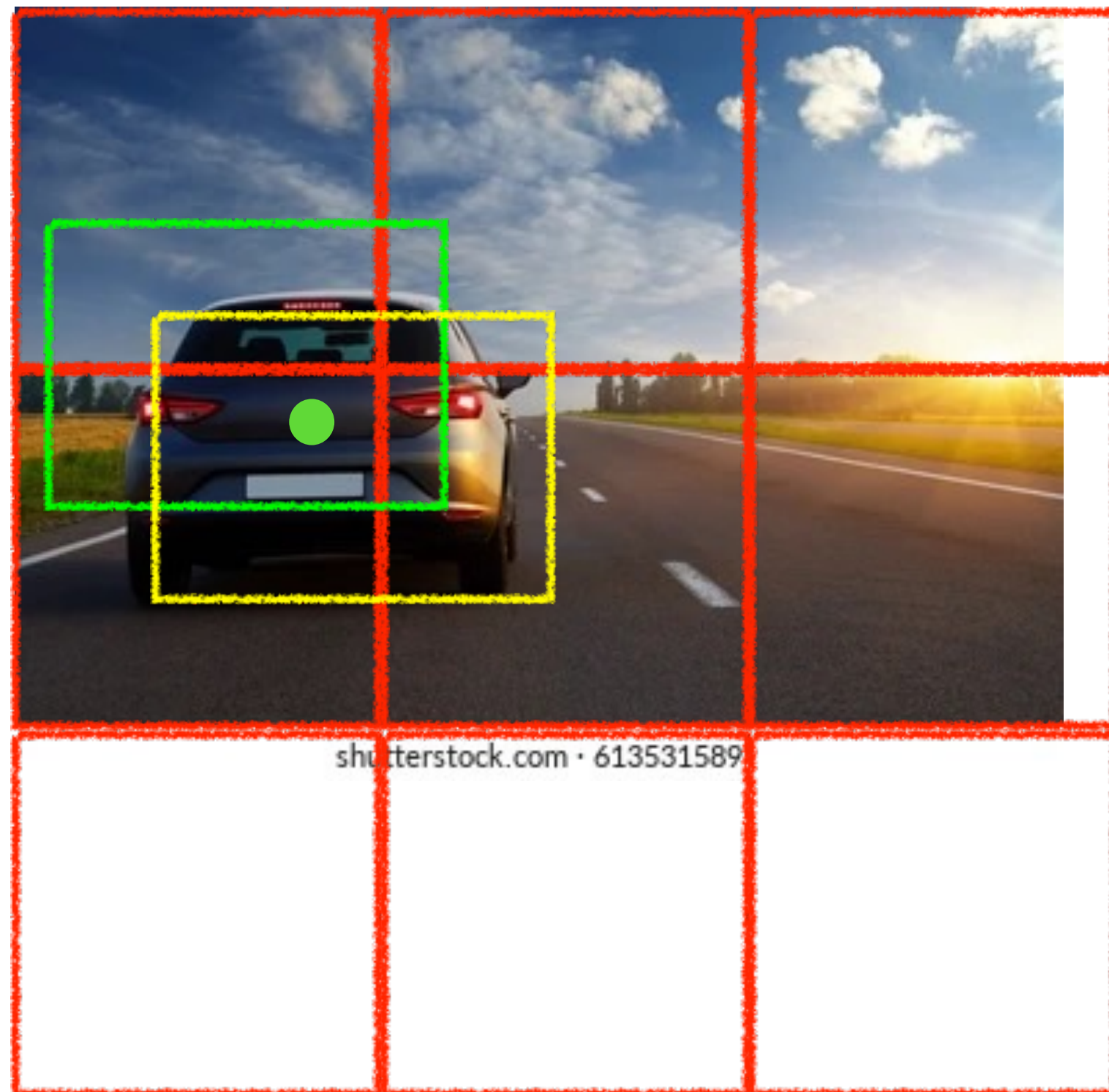**Anchor Box 1**

Anchor Box 2

[P1, Bx, By, Bh, Bw, O1, O2, O3]

[P1, Bx, By, Bh, Bw, O1, O2, O3, P1', Bx', By', Bh', Bw', O1', O2', O3']

Finally, YOLO

We will discuss the algorithm of YOLO

[P1,
Bx,
By,
Bh,
Bw,
O1,
O2,
O3,
P1',
Bx',
By',
Bh',
Bw',
O1',
O2',
O3']

And then run Non-compression Algorithm

After all the process, we can output the result.

Framework of YOLO

Visualization of how YOLO works

YOLO v3 network Architecture

Demo & Results

Reference

- Redmon et al., 2015, You Only Look Once: Unified real-time object detection

- Redmon et al., 2018, YOLOv3: An Incremental Improvement

Thank you!