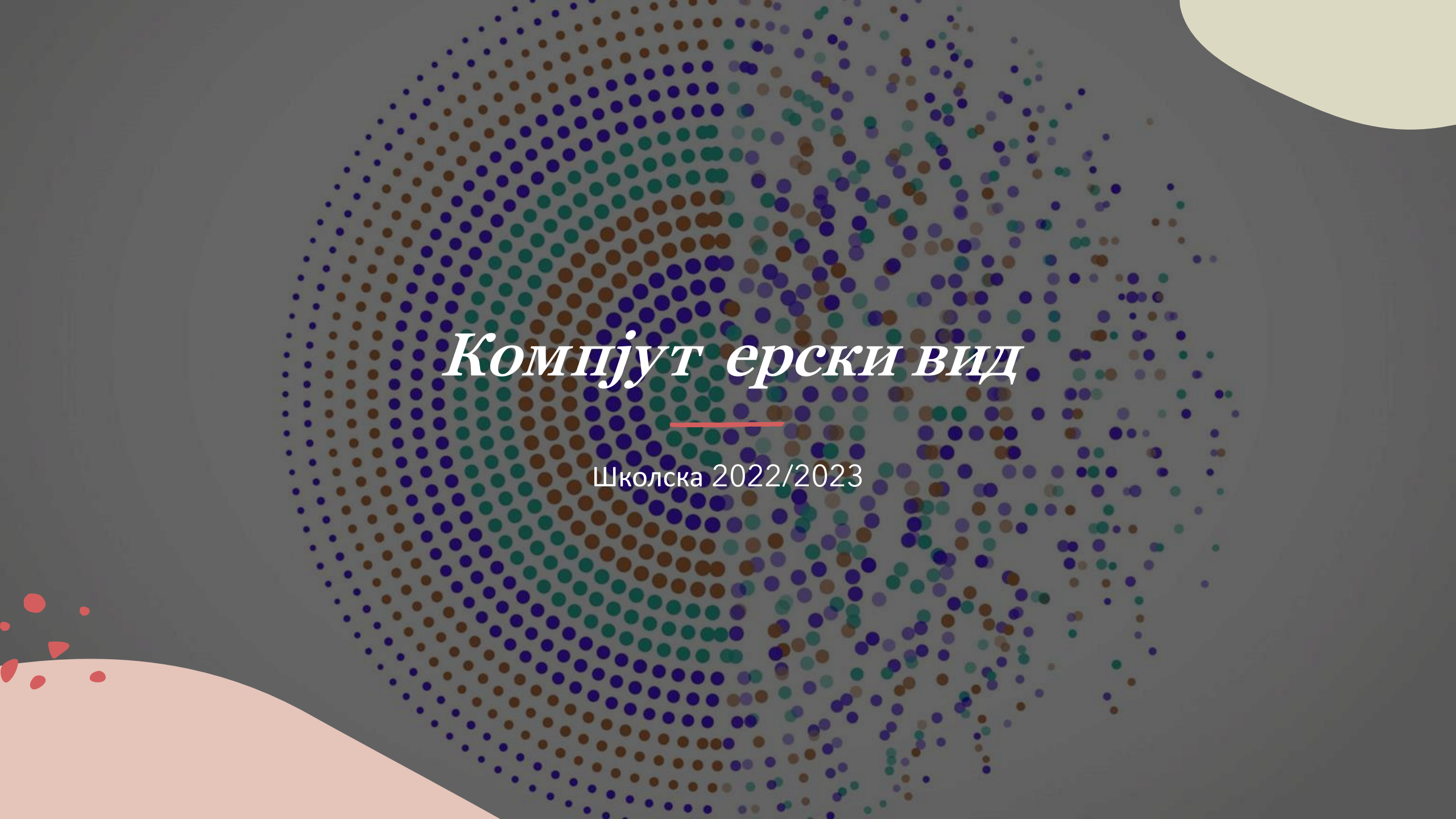




Компјутерски вид

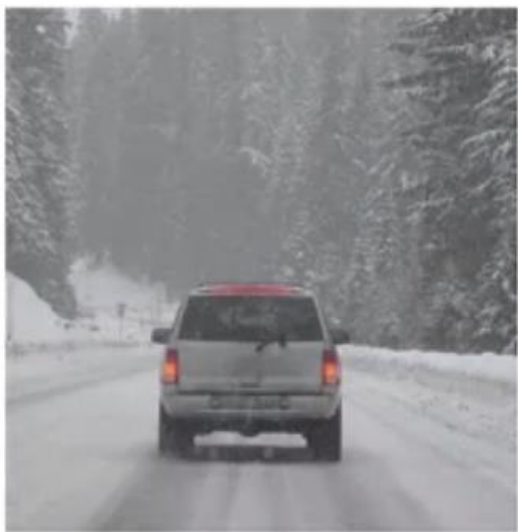
Школска 2022/2023

План рада

1. Детекција
 - YOLO алгоритам

Класификација – локализација – детекција

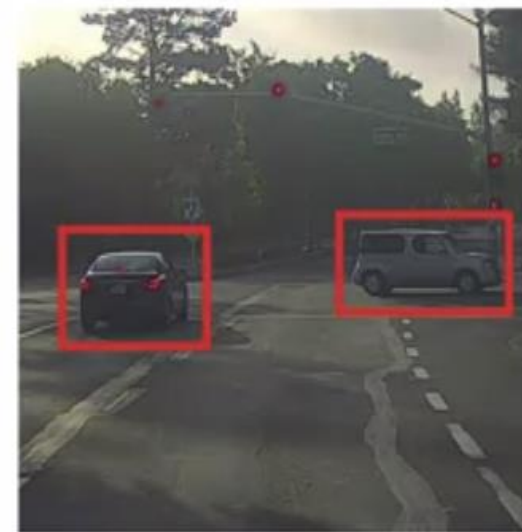
Image classification



Classification with localization



Detection

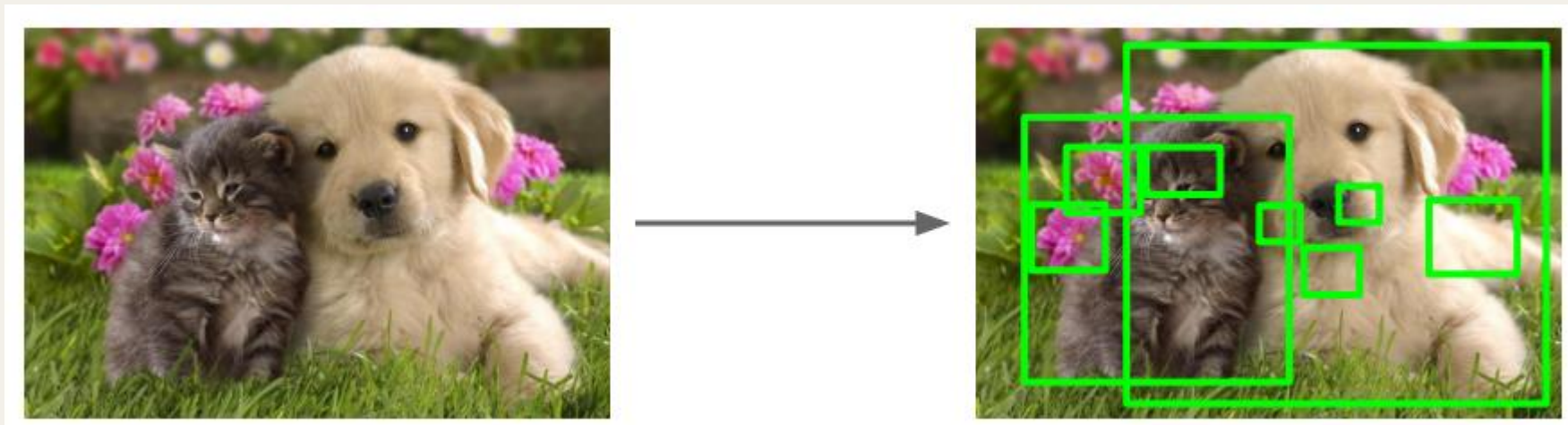


Детекција – локализација више објеката

- Две групе алгоритама
 - Region-based detection (R-CNN, Fast R-CNN, Faster R-CNN)
 - Detections without proposals (YOLO)

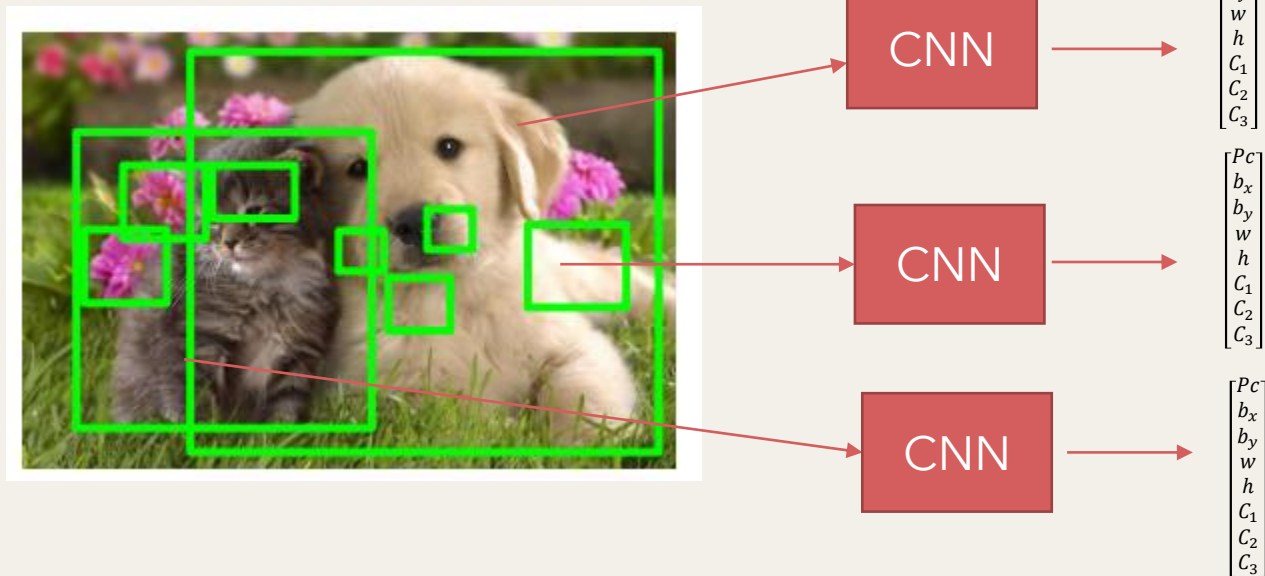
Детекција – локализација више објеката

- Интуиција: R-CNN
 - Пронаћи регионе који вероватно садрже објекте (класичне методе обраде слике)



Детекција – локализација више објеката

- Интуиција: R-CNN
 - Пронаћи регионе који вероватно садрже објекте (класичне методе обраде слике)



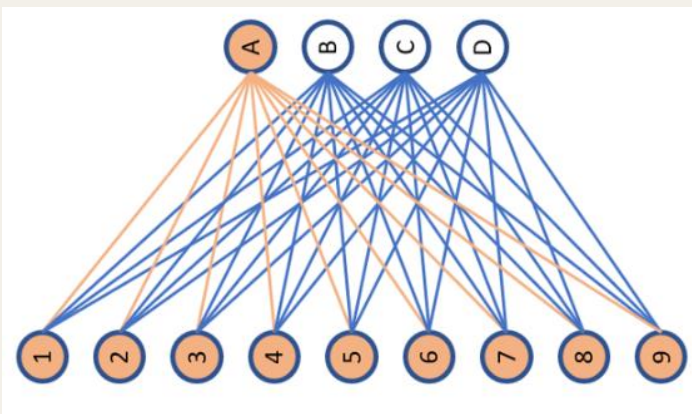
Детекција – локализација више објеката

- Интуиција: R-CNN
 - Пронаћи регионе који вероватно садрже објекте (класичне методе обраде слике)
- Проблеми:
 - Споро тренирање (неколико хиљада потенцијалних региона за једну слику)
 - Спора примена тренираних модела спора (око 40 секунди по слици)

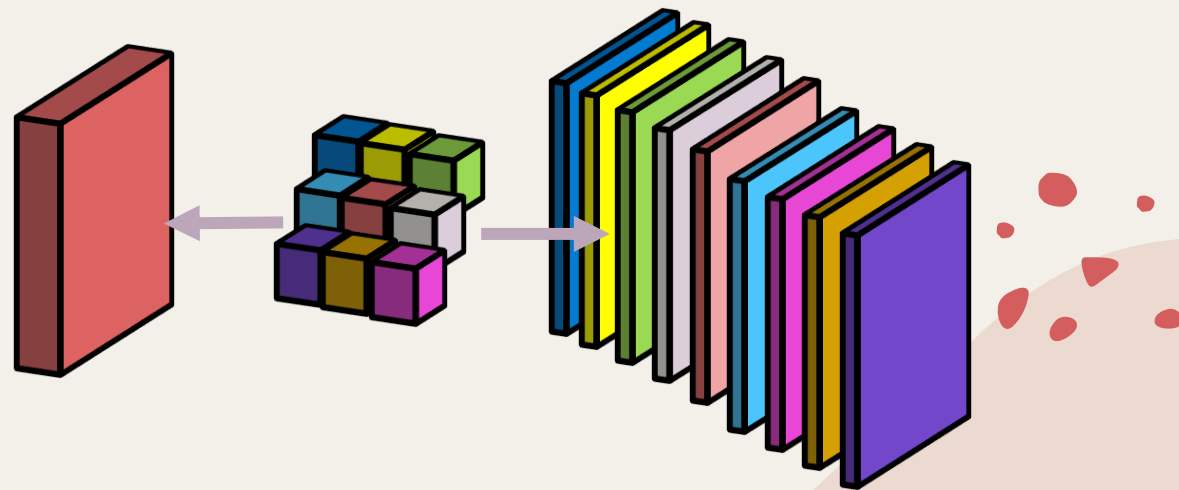
Детекција – локализација више објеката

- Конволуциони слојеви раде конволуцију улазног тензора са филтерима.
 - Не зависе од димензије улазног тензора (докле год је димензија улазног тензора већа од филтера)
- Потпуно повезани слојеви у мрежама захтевају улаз **фиксних** димензија

Потпуно-повезани слој



Конволуциони слој



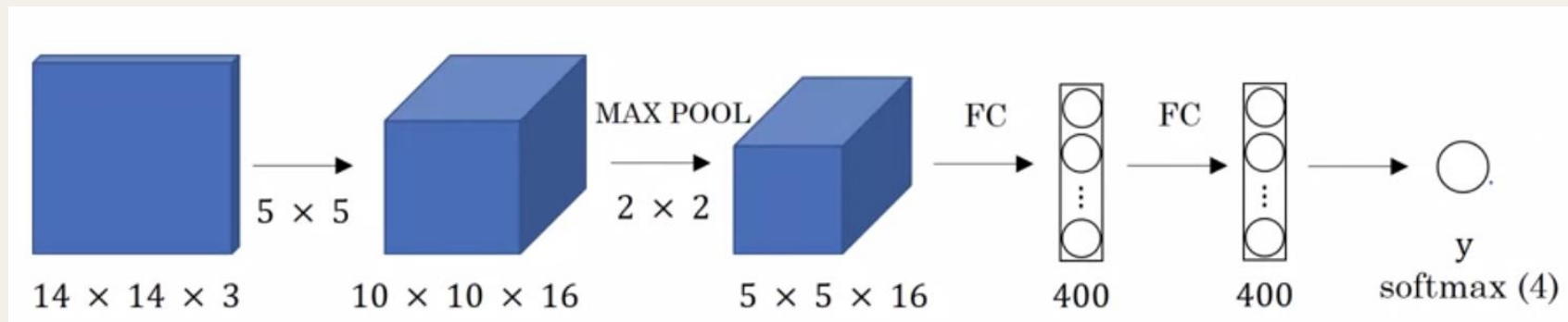
Детекција – локализација више објеката

- Конволуциони слојеви раде конволуцију улазног тензора са филтерима.
 - Не зависе од димензије улазног тензора (докле год је димензија улазног тензора ваћа од филтера)
- Потпуно повезани слојеви у мрежама захтевају улаз **фиксних** димензија
- **Идеја: Имплементирати потпуно повезане слојеве као конволуције**

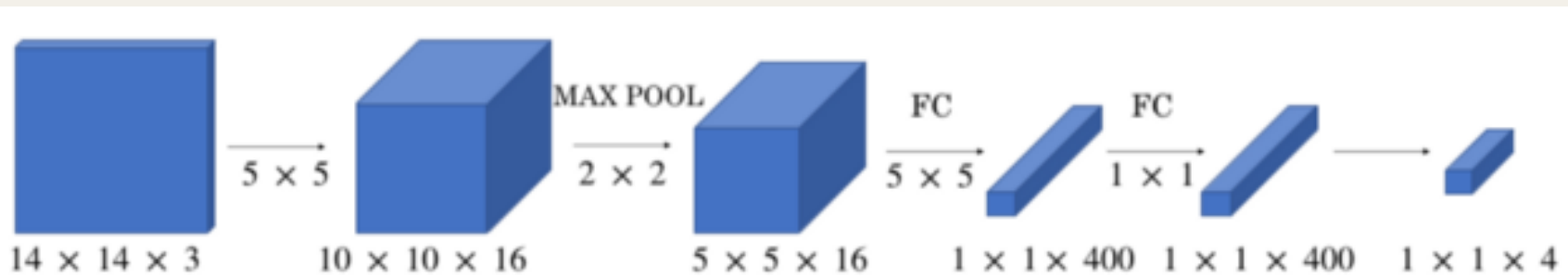
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична
архитектура



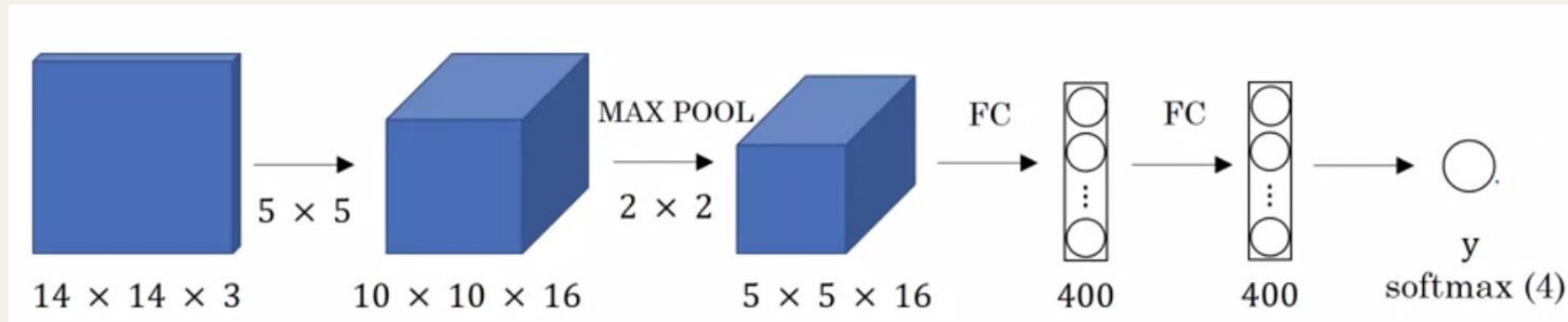
Конволуциона
архитектура



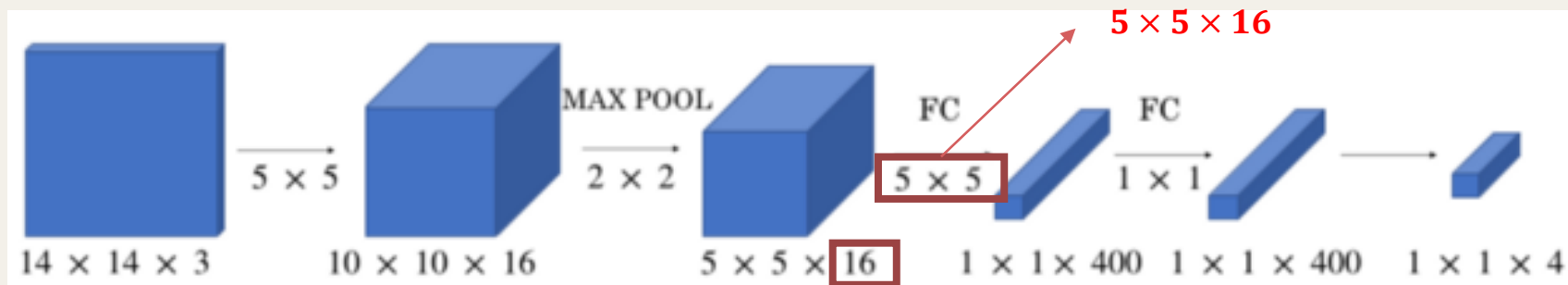
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична архитектура



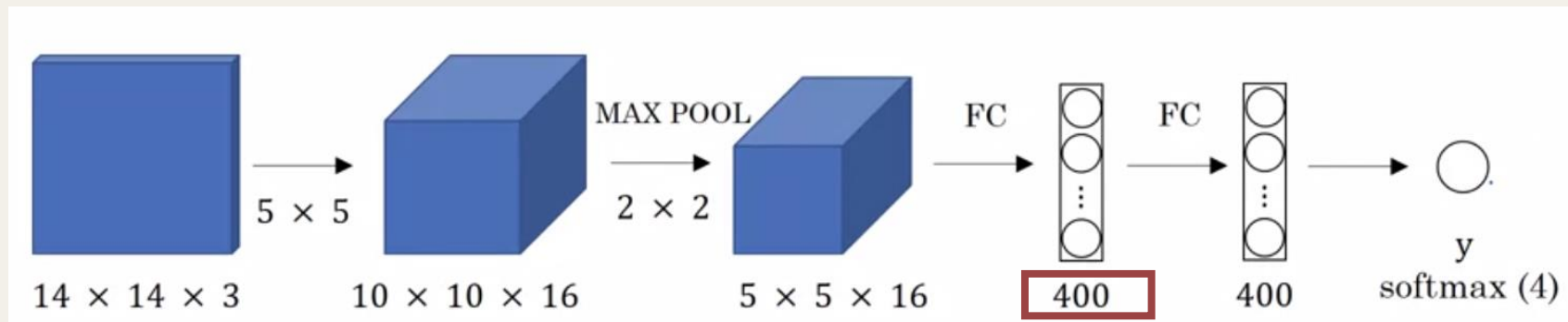
Конволуциона архитектура



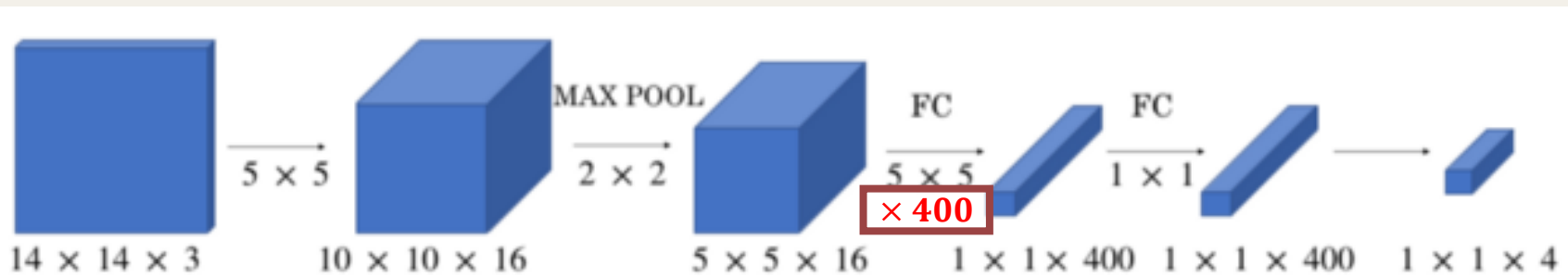
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична архитектура



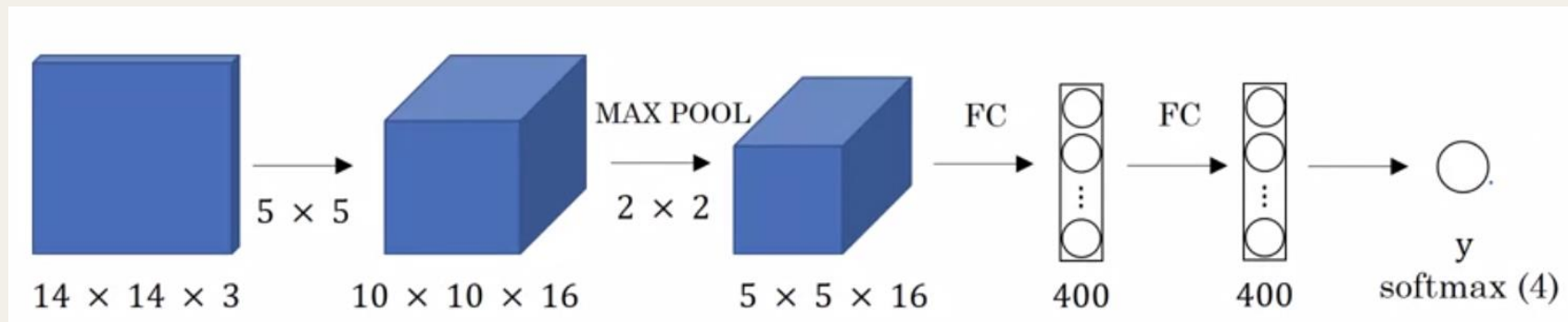
Конволуциона архитектура



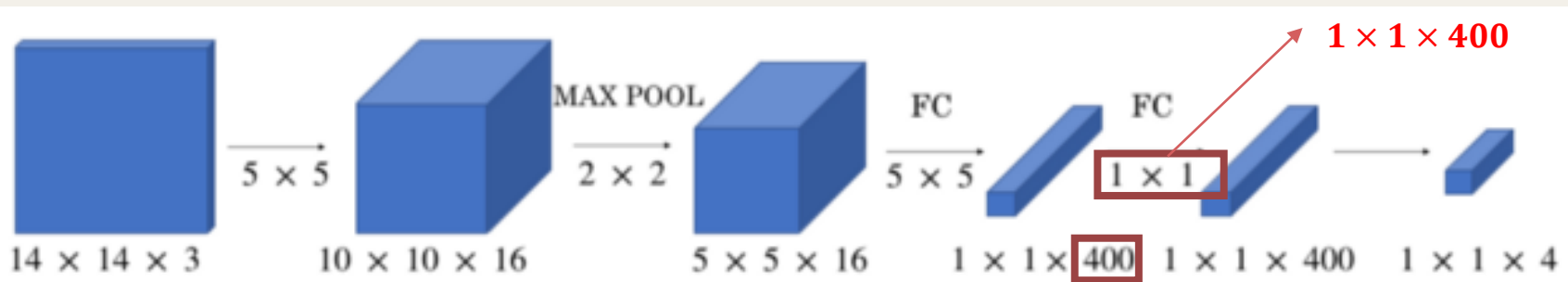
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична архитектура



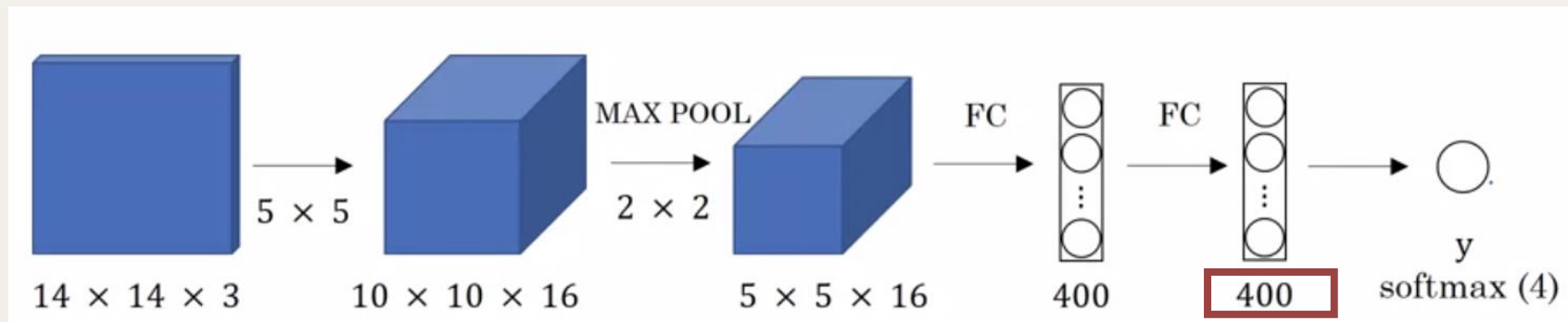
Конволуциона архитектура



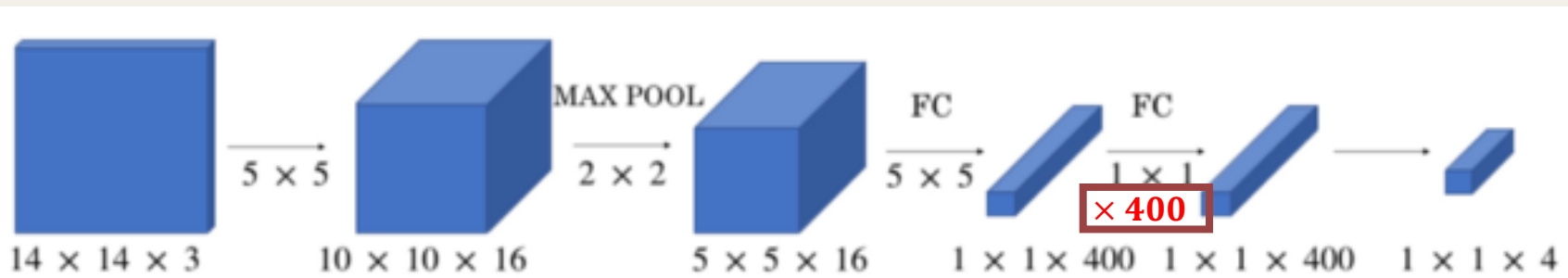
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична архитектура



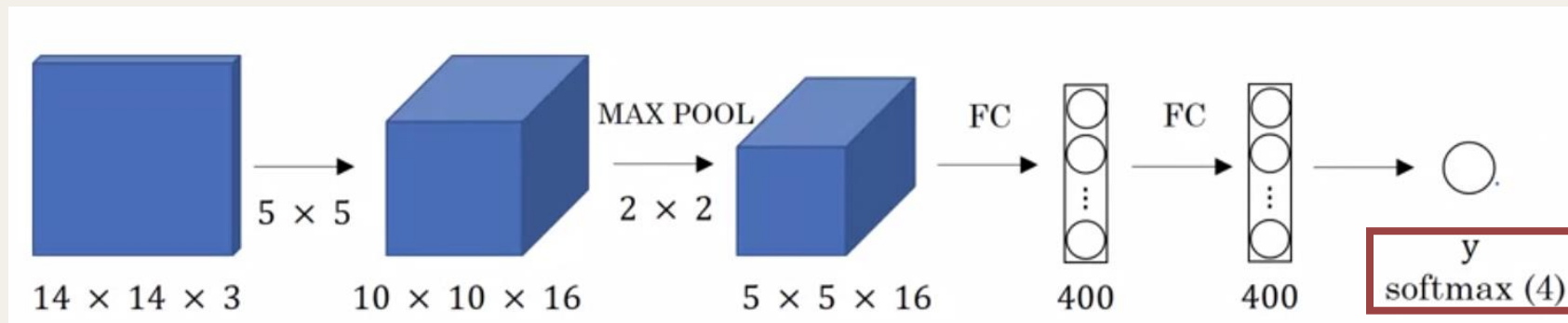
Конволуциона архитектура



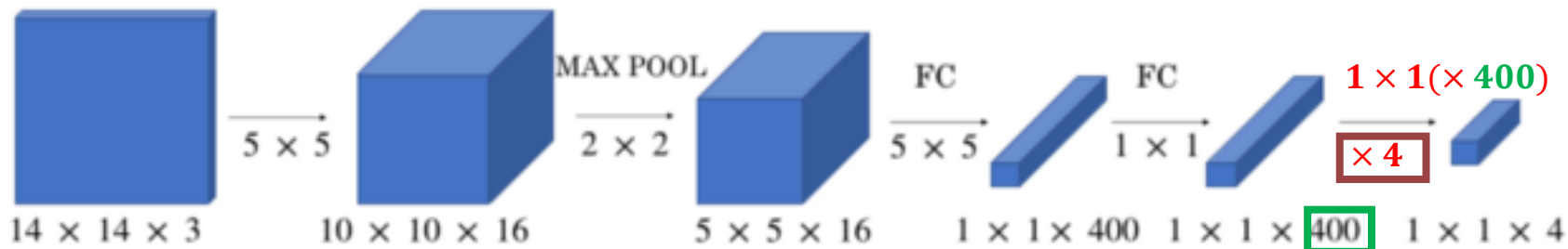
Детекција – локализација више објеката

- Идеја: Имплементирати потпуно повезане слојеве као конволуције

Класична архитектура

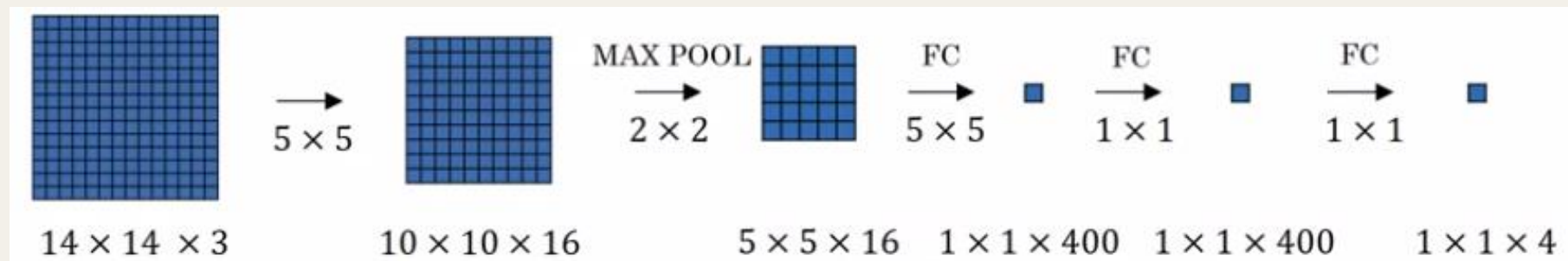


Конволуциона архитектура



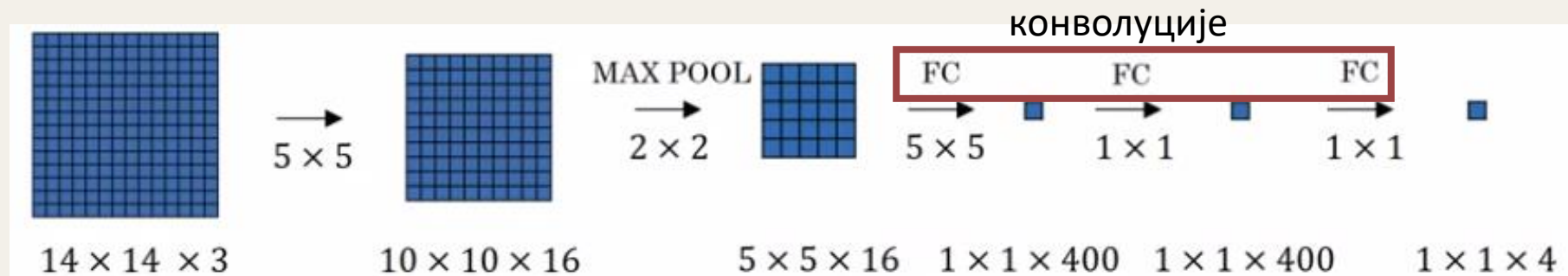
Детекција – локализација више објеката

- Клизајући прозори као конволуција



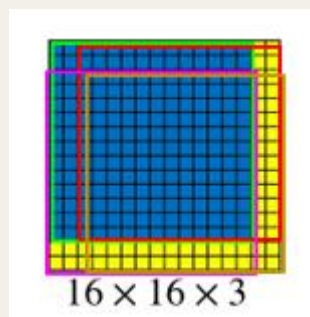
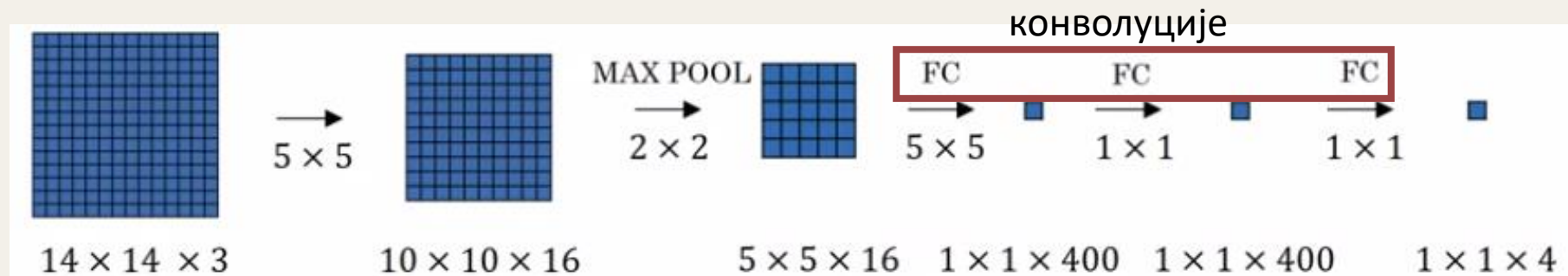
Детекција – локализација више објеката

- Клизајући прозори као конволуција



Детекција – локализација више објеката

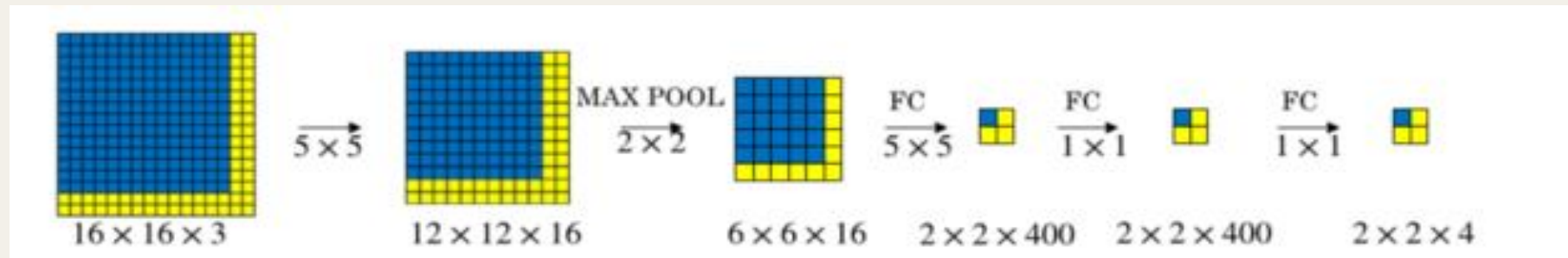
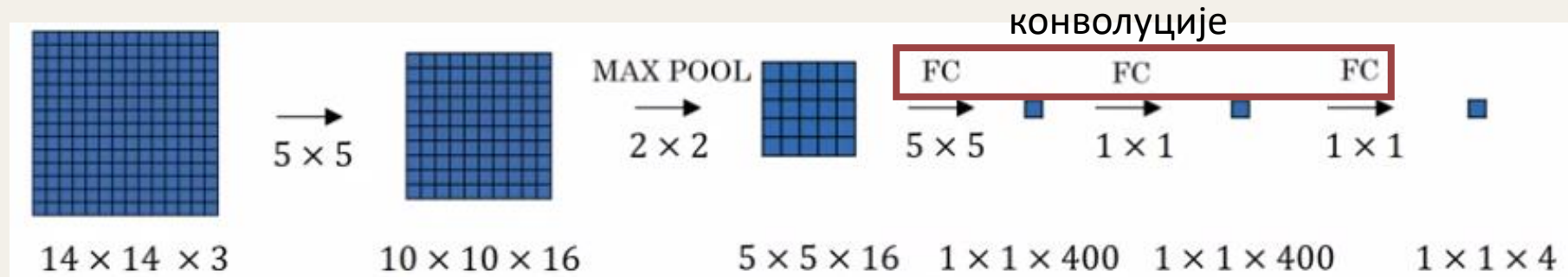
- Клизајући прозори као конволуција



4 клизајућа прозора димензије 14×14
Доста заједничких пиксела (дуплираних рачунања у конволуцији)

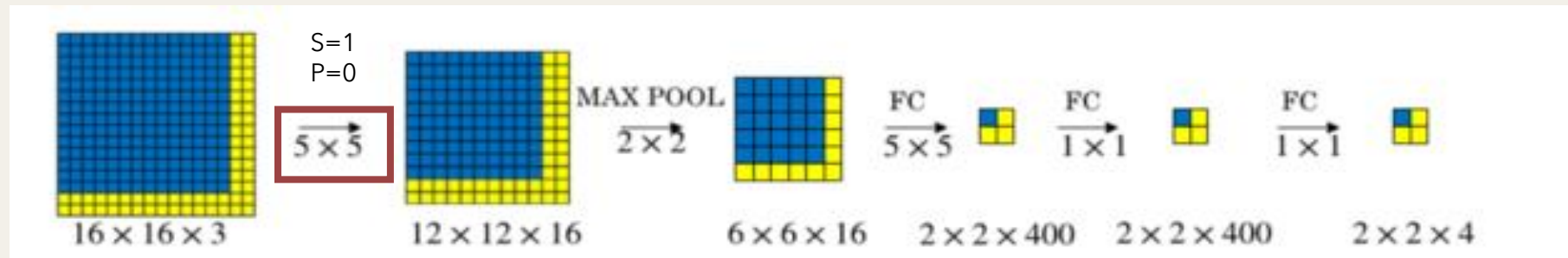
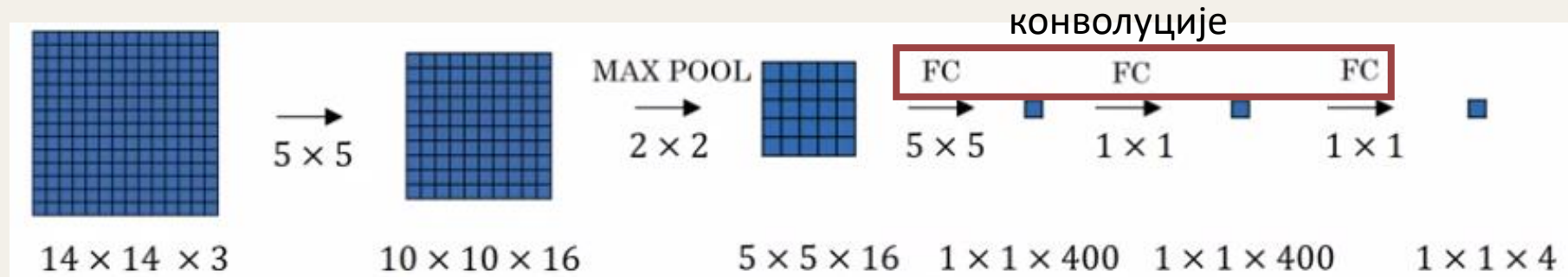
Детекција – локализација више објеката

- Клизајући прозори као конволуција



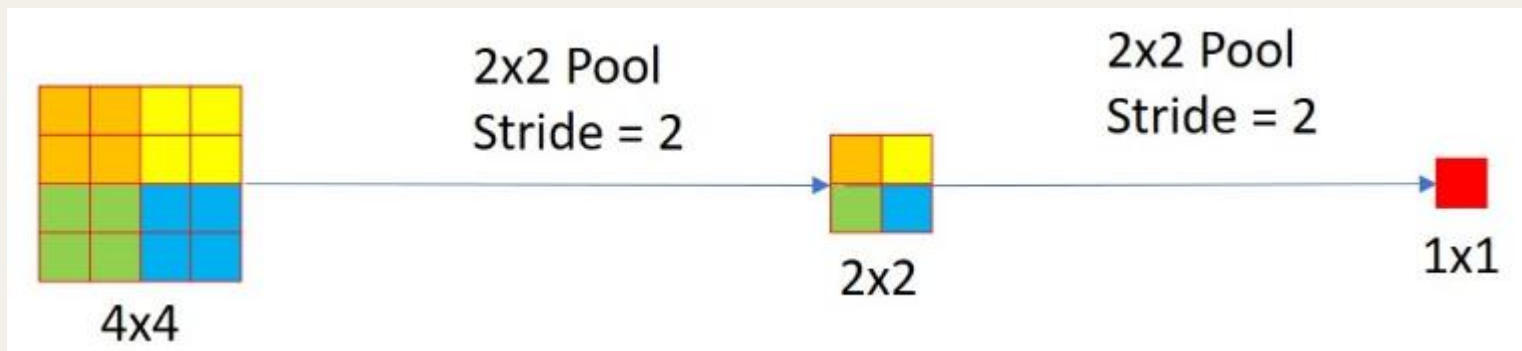
Детекција – локализација више објеката

- Клизајући прозори као конволуција



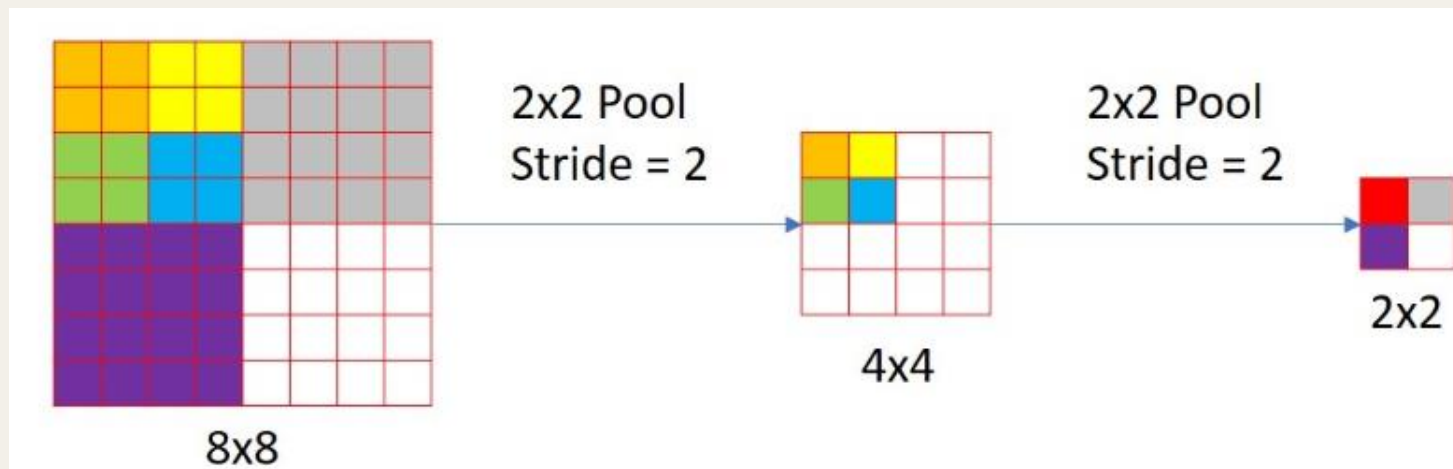
Дет екција – локализација више објекат а

- Чему одговара једна вредност у активационој мапи (енг. receptive field)



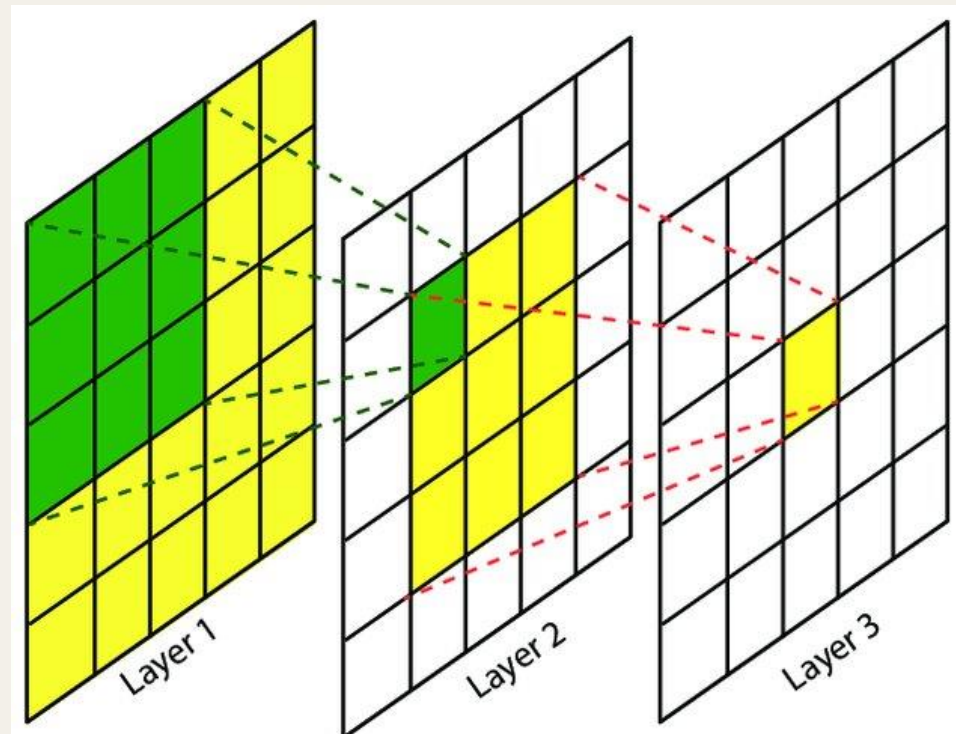
Дет екција – локализација више објекат а

- Чему одговара једна вредност у активационој мапи (енг. receptive field)



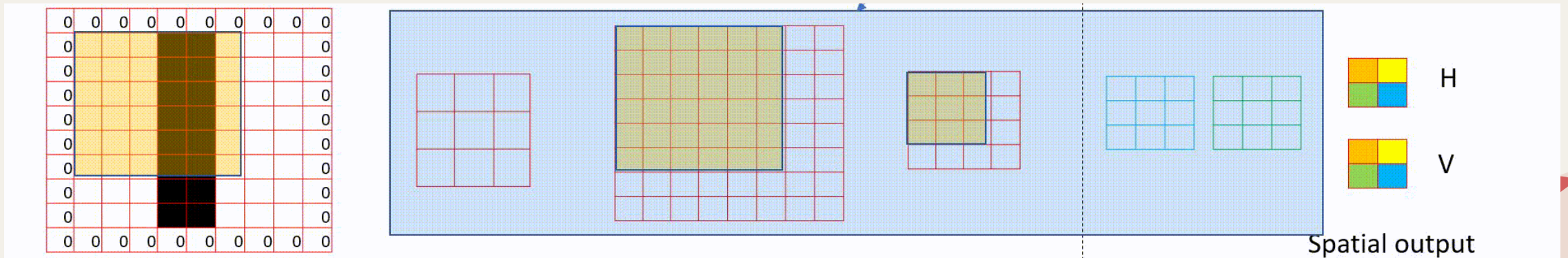
Детекција – локализација више објеката

- Чему одговара једна вредност у активационој мапи (енг. receptive field)



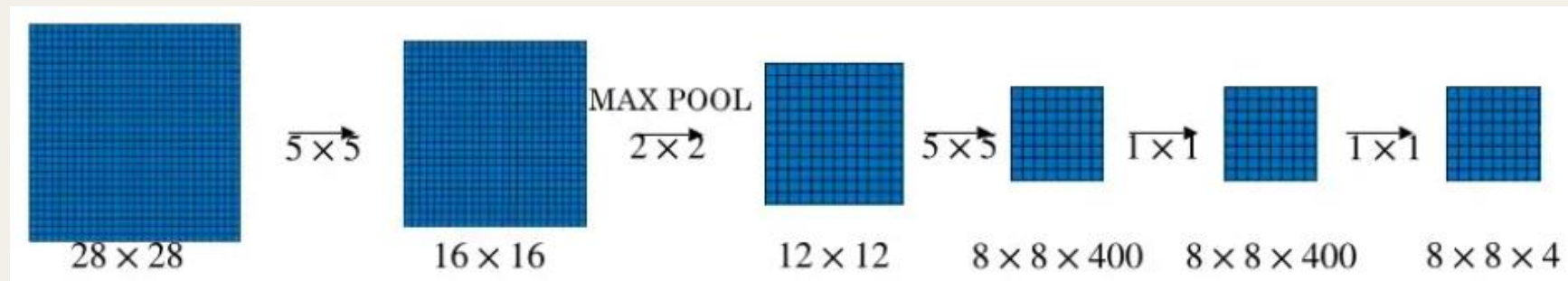
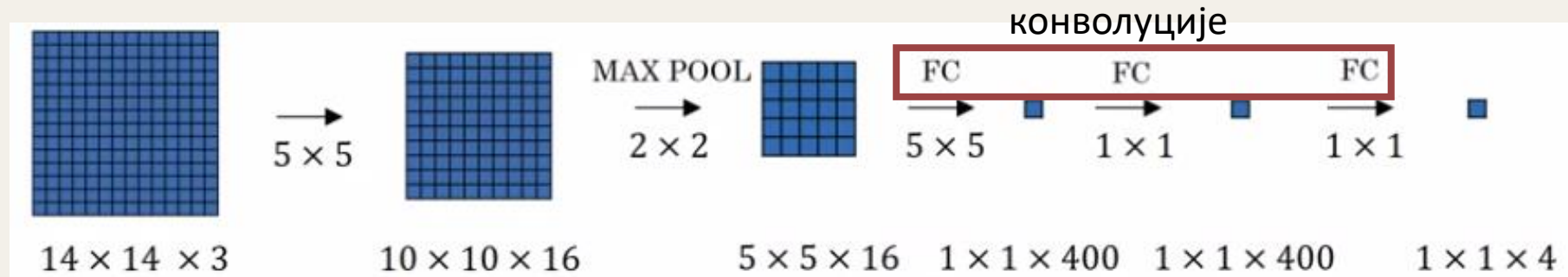
Детекција – локализација више објеката

- Чему одговара једна вредност у активационој мапи (енг. receptive field)
 - Вредност у крајњем тензору (излазу из мреже) еквивалентна је као да је сваки исечак оригиналне слике примењен независно
 - Конволуцијом се све те вредности добијају у једном проласку кроз мрежу



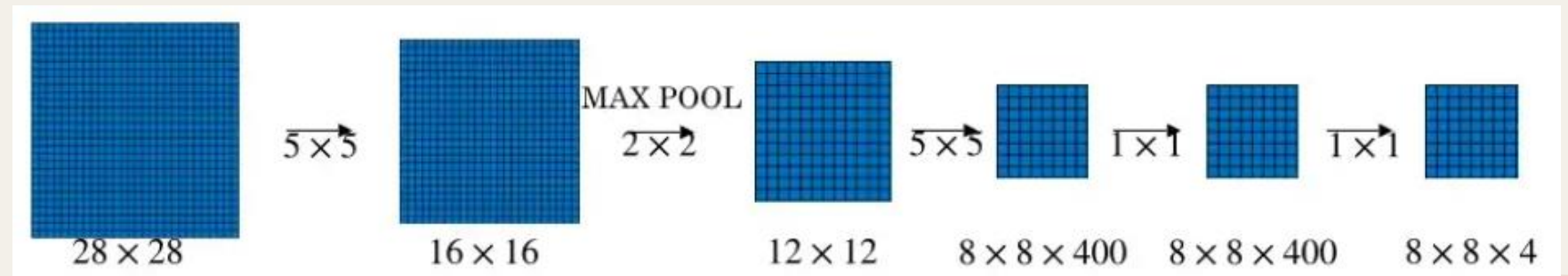
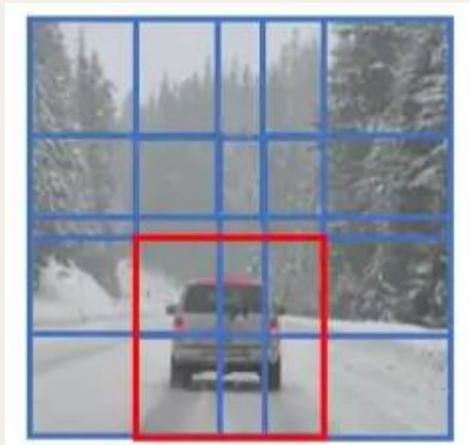
Детекција – локализација више објеката

- Клизајући прозори као конволуција



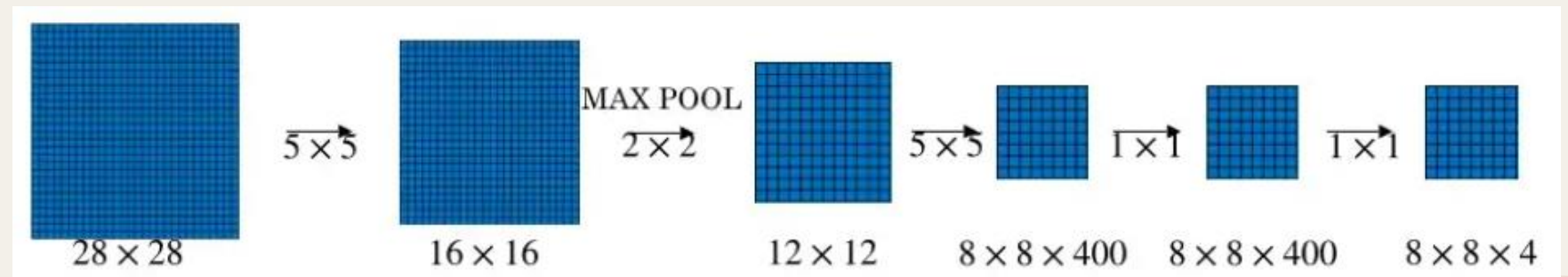
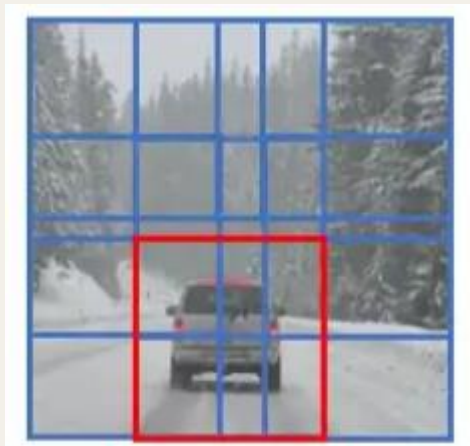
Детекција – локализација више објеката

- Клизајући прозори као конволуција



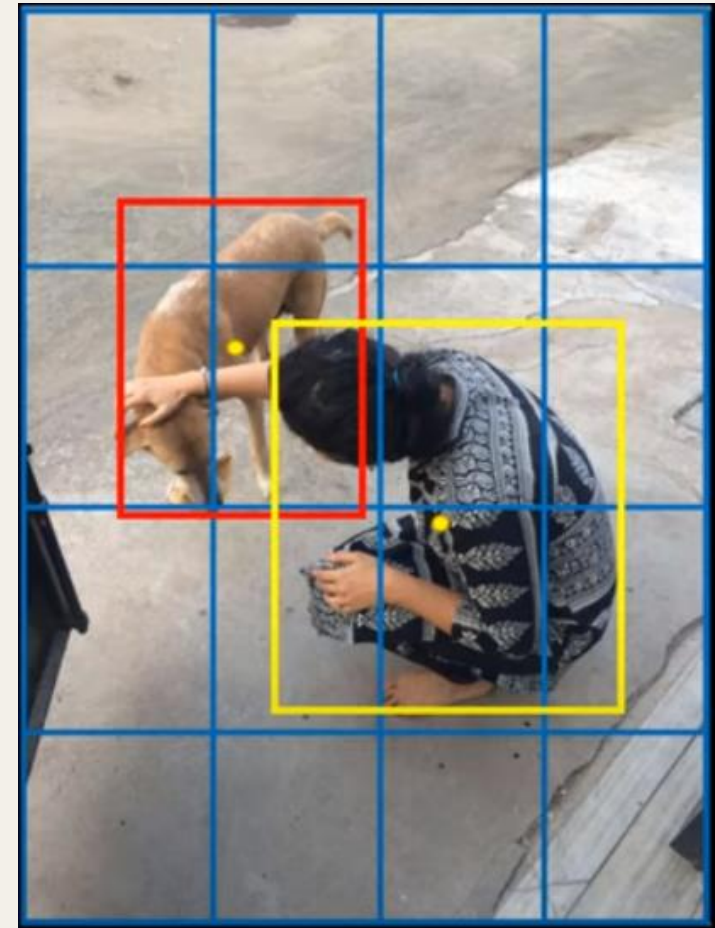
Детекција – локализација више објеката

- Питања:
 - Шта ако објекат не стаје у квадратни прозор?
 - Шта ако је објекат већи од одабраног прозора?



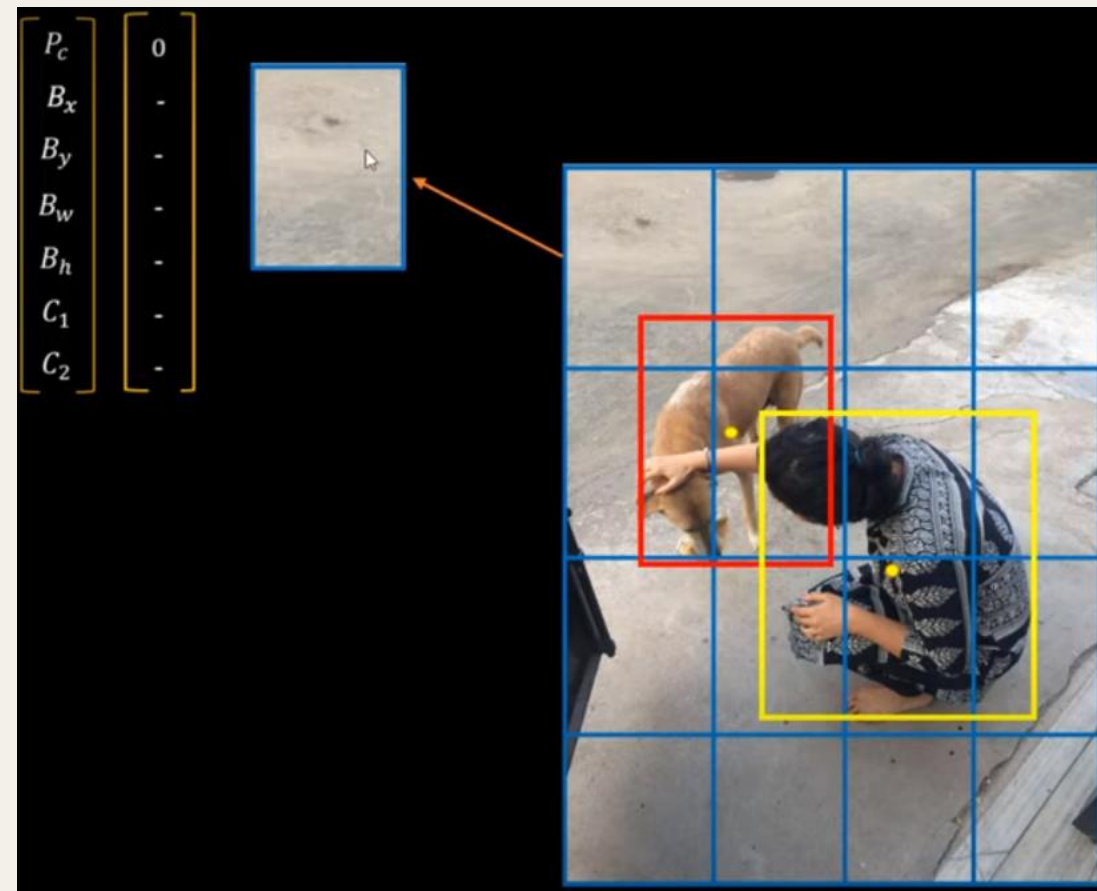
Детекција – локализација више објеката

- Улазну слику поделити на одређен број прозора (нпр. 7×7)



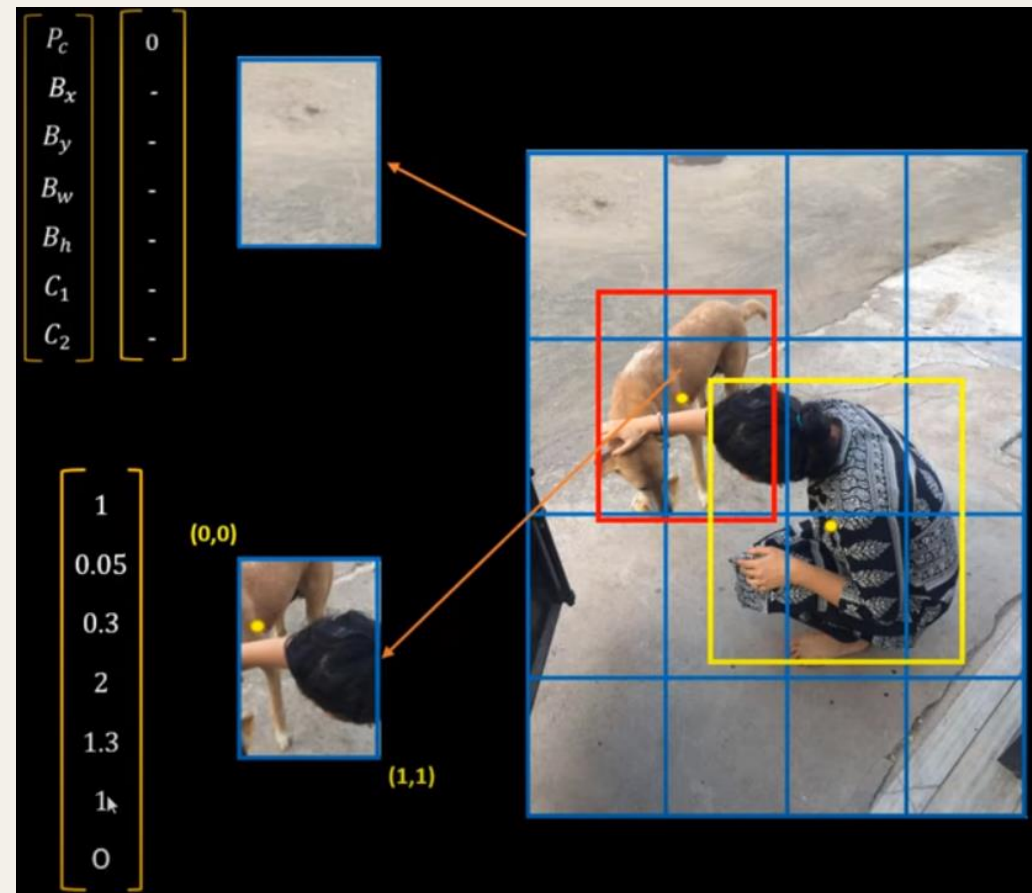
Детекција – локализација више објеката

- Улазну слику поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију



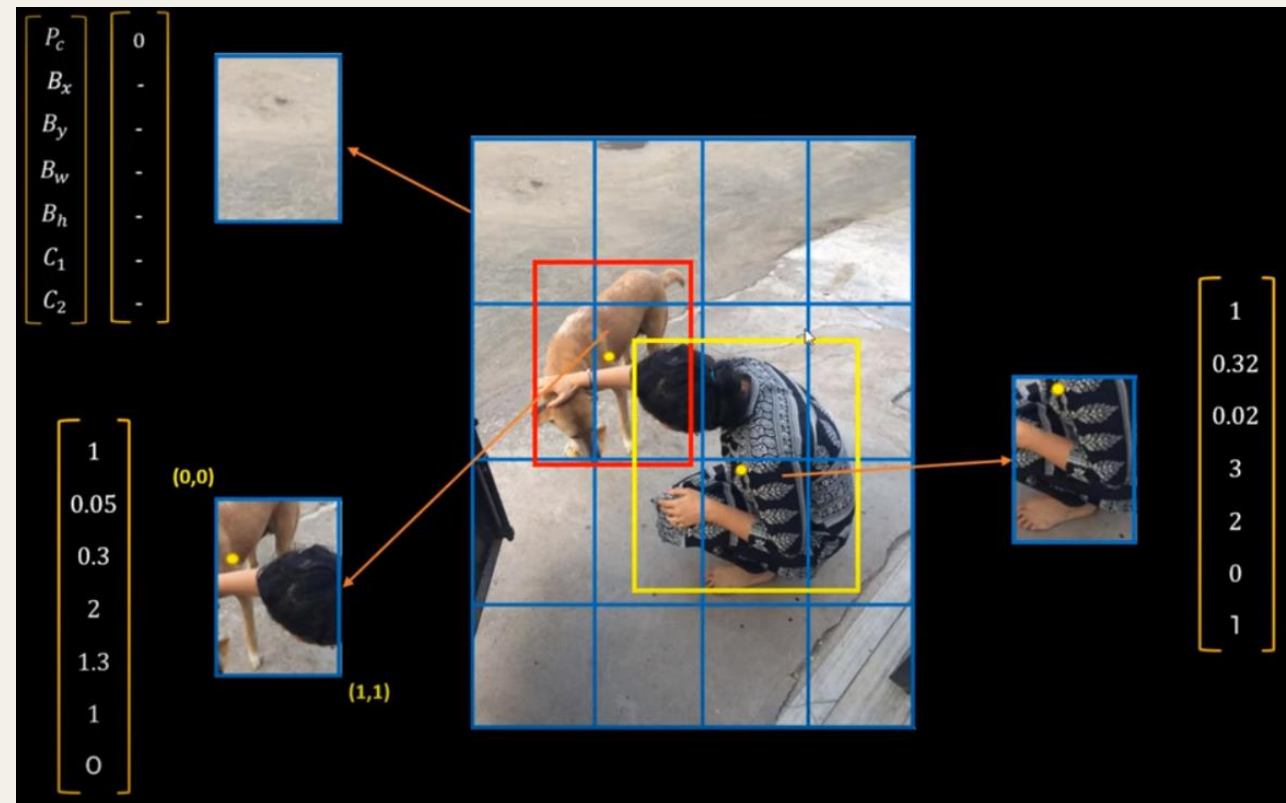
Детекција – локализација више објеката

- Улазну слику поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију
 - Ако се објекат протеже кроз више прозора, доделити га оном у коме се налази његов центар

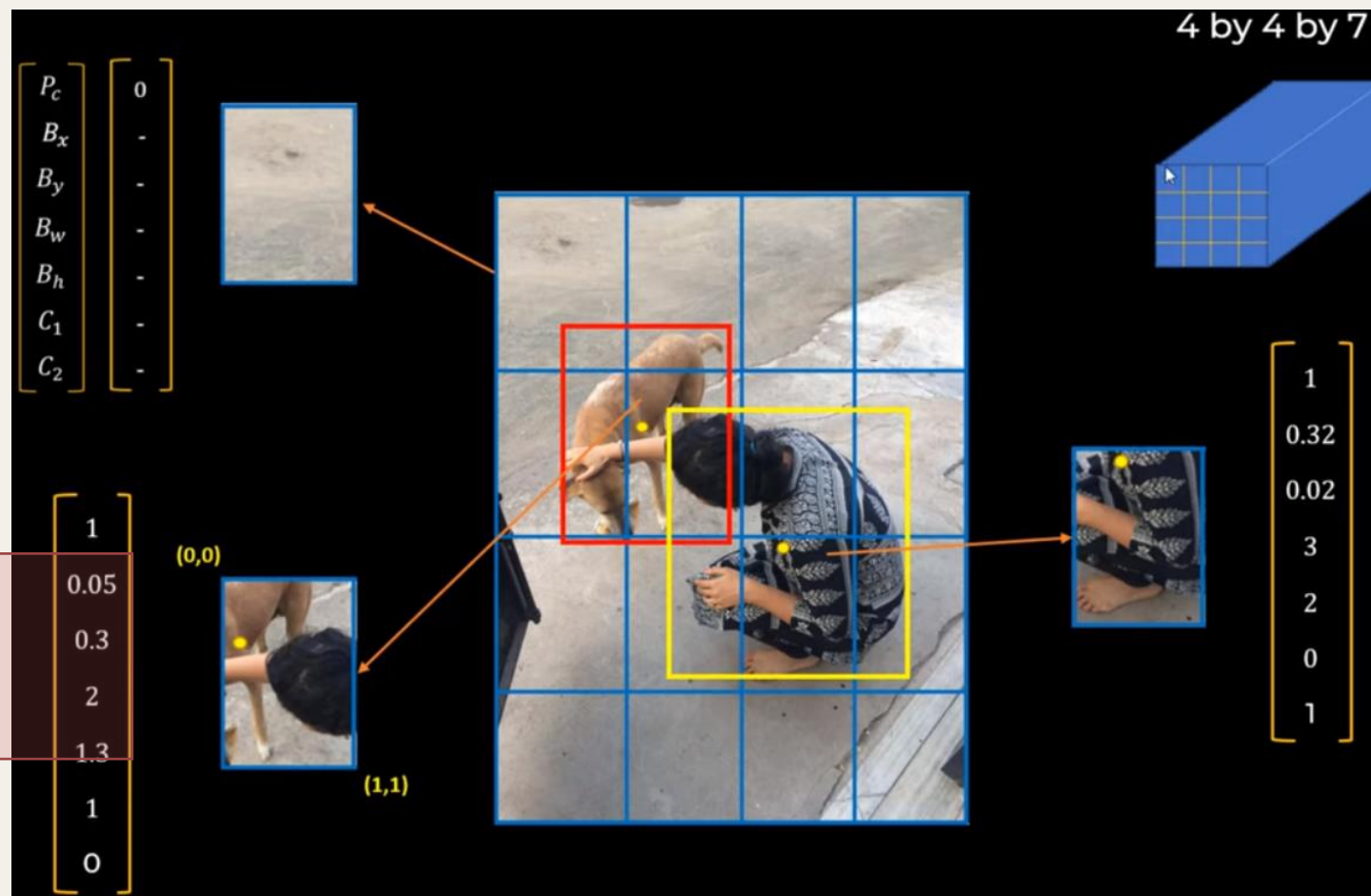


Детекција – локализација више објеката

- Улазну слику поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију
 - Ако се објекат протеже кроз више прозора, доделити га оном у коме се налази његов центар

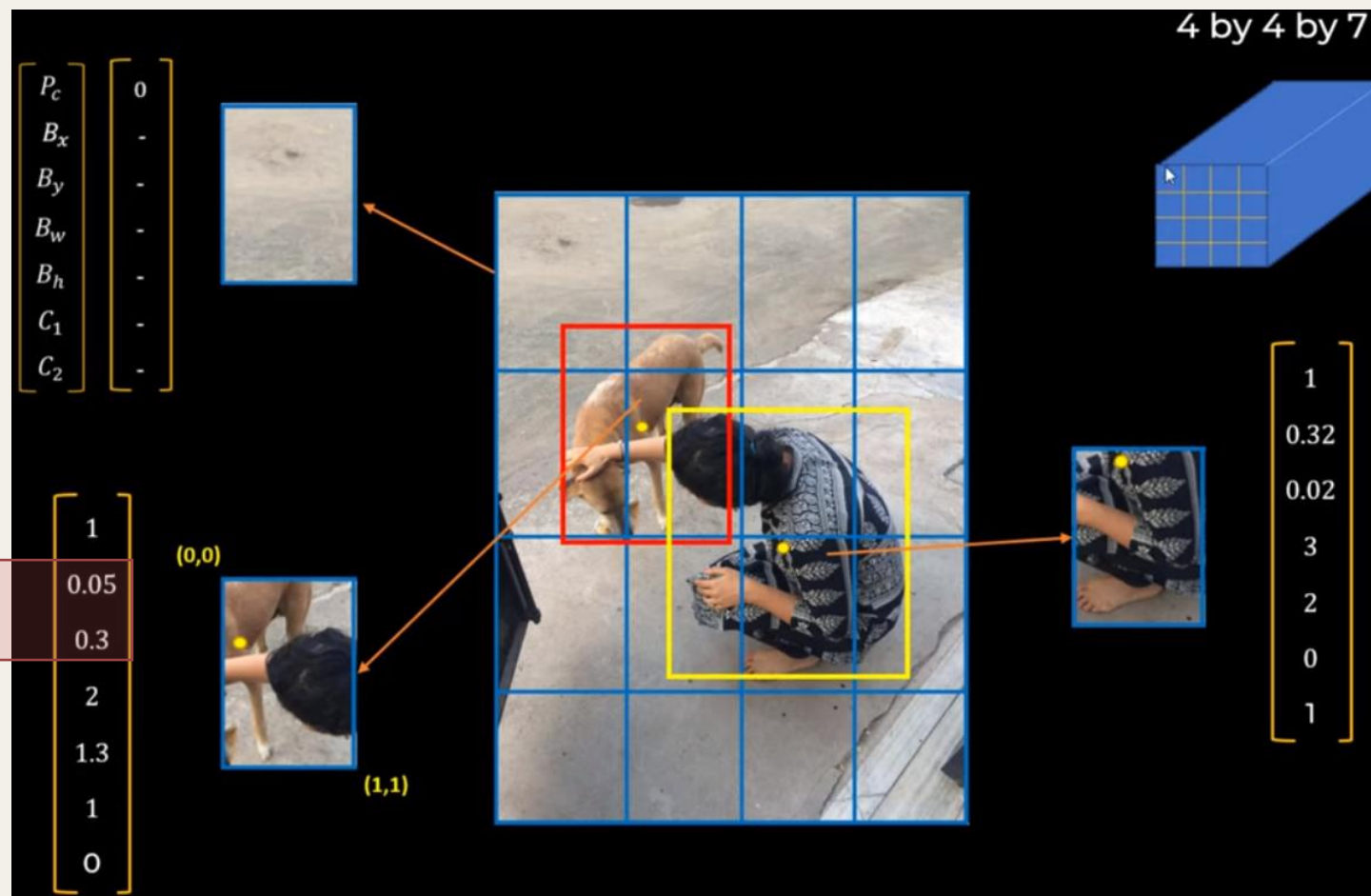


Детекција – локализација више објеката

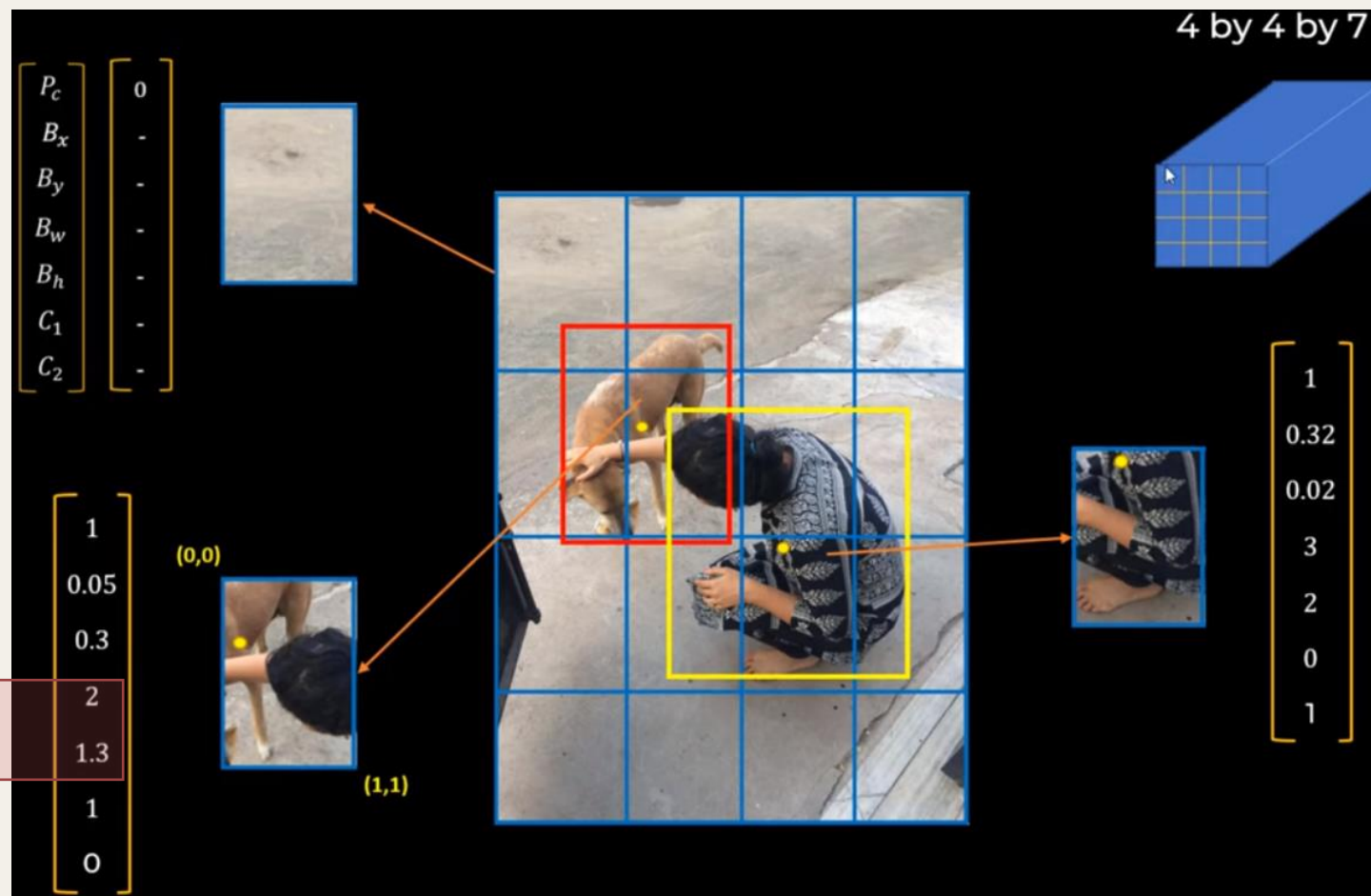


Релативно у
односу на
прозор

Детекција – локализација више објеката



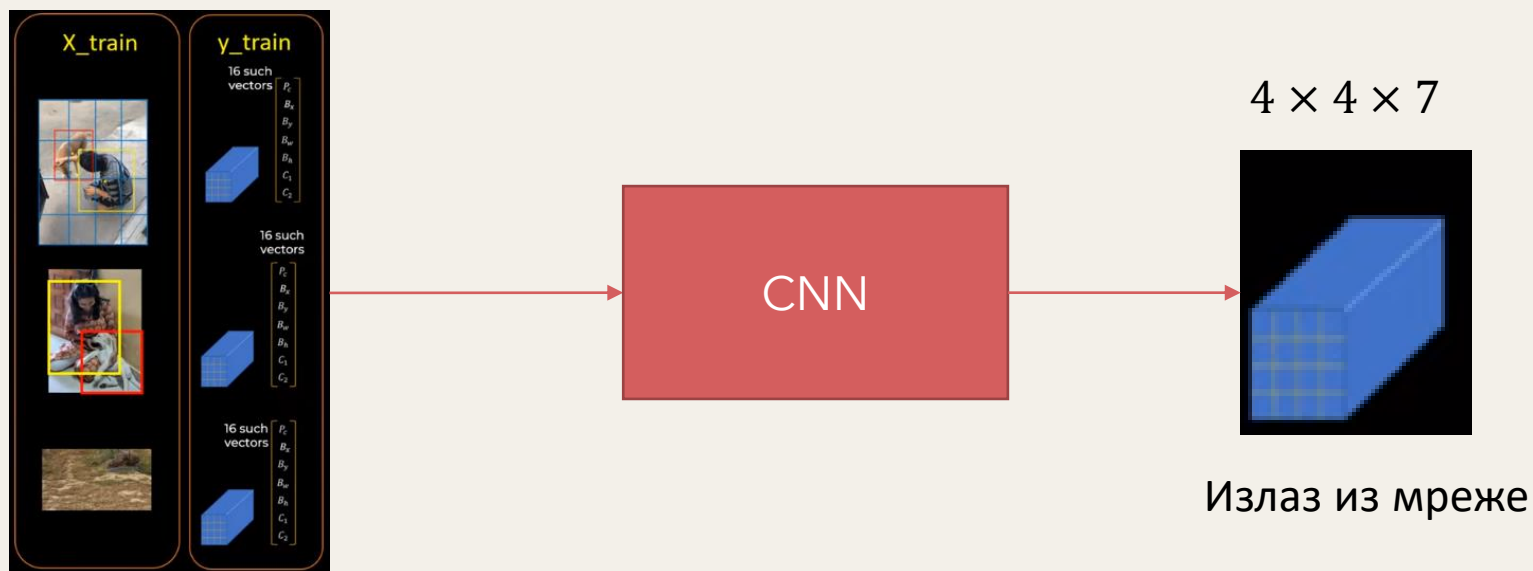
Детекција – локализација више објеката



Може да буде и веће од 1 ако се објекат протеже кроз више прозира

Детекција – локализација више објеката

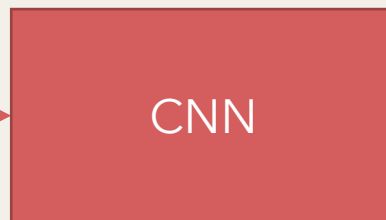
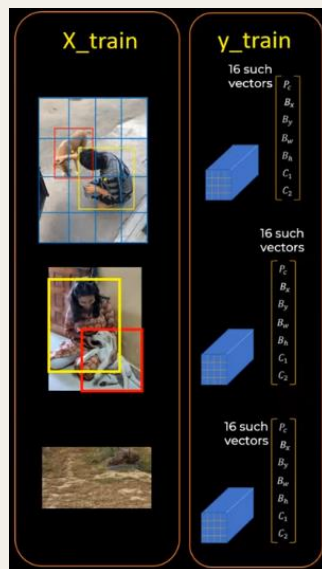
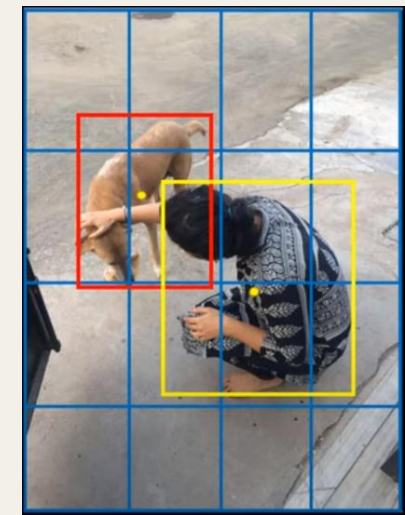
- Сliku поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију
- Тренирати мрежу



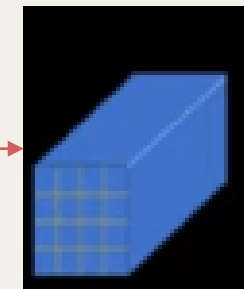
Детекција – локализација више објеката

- Сliku поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију
- Тренирати мрежу

Јер је слика подељена на 4×4 прозора (у пракси се користи већа гранулација, нпр. 19×19)



$4 \times 4 \times 7$

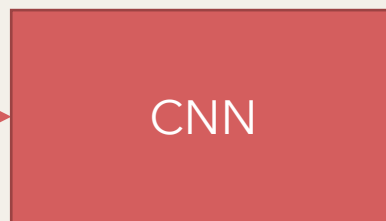
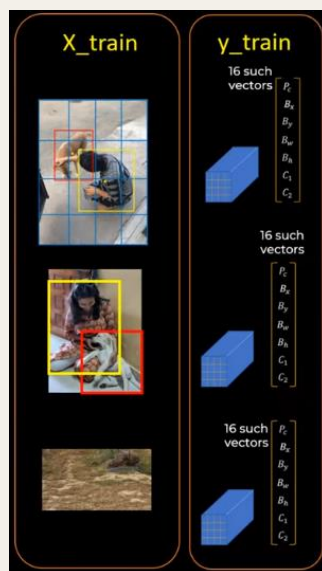


Излаз из мреже

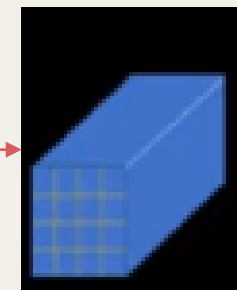
Детекција – локализација више објеката

- Сliku поделити на одређен број прозора (нпр. 7×7)
- Сваком прозору доделити анотацију
- Тренирати мрежу

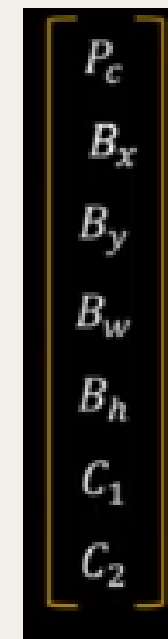
Анотација има 7 поља: једно за присуство објекта, четири за локацију и две за класу (редом)



$4 \times 4 \times 7$

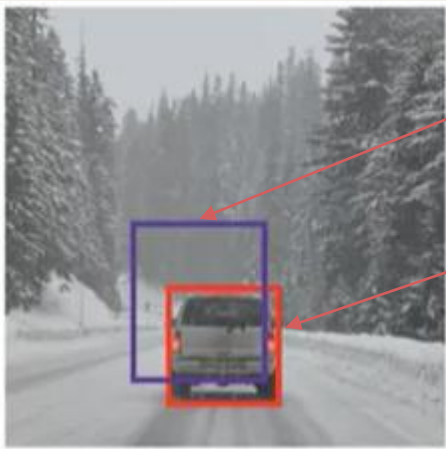


Излаз из мреже



Детекција – метрика (Intersection over Union)

- Мера преклапања два објекта (нпр. Bounding-boxes)



Предикција модела

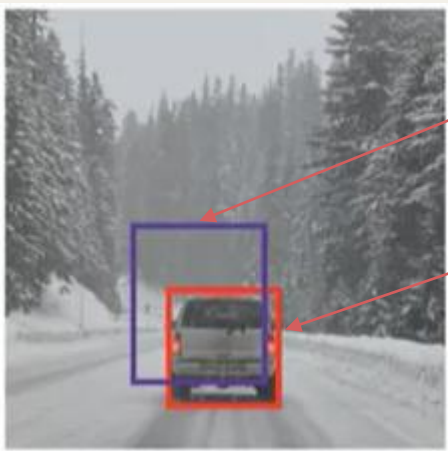
Анотација (шта желимо да модел предвиди) енг. groundtruth

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


<https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>

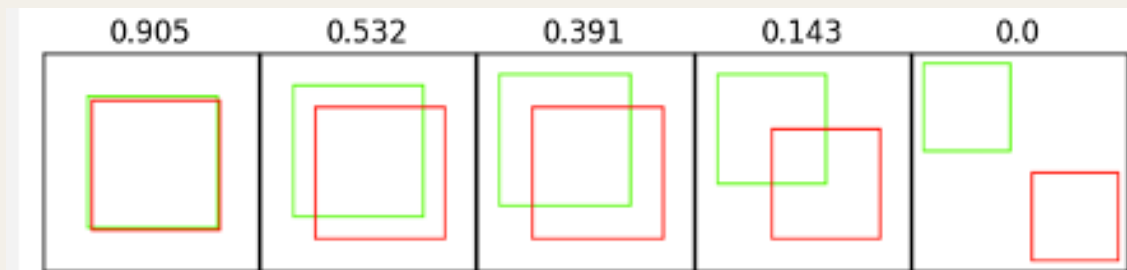
Детекција – метрика (Intersection over Union)

- Мера преклапања два објекта (нпр. Bounding-boxes)



Предикција модела

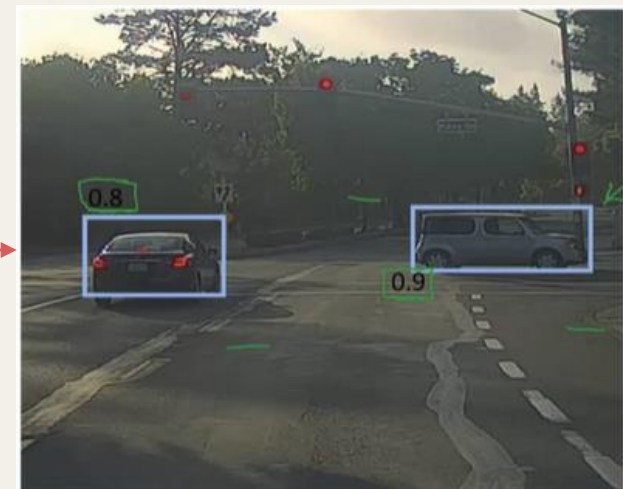
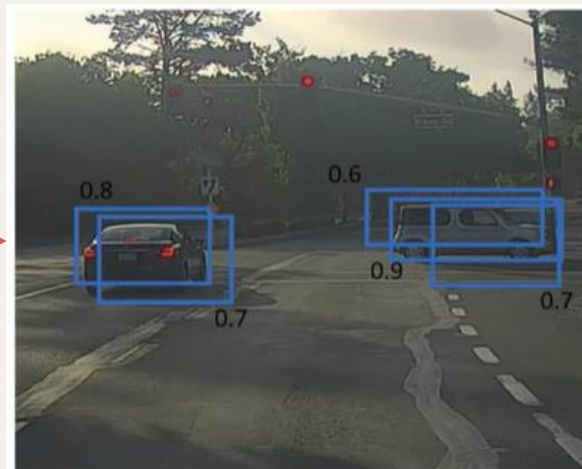
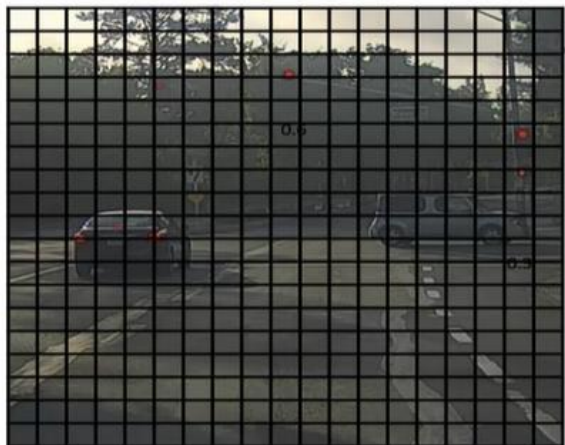
Анотација (шта желимо да модел предвиди) енг. `goundtruth`



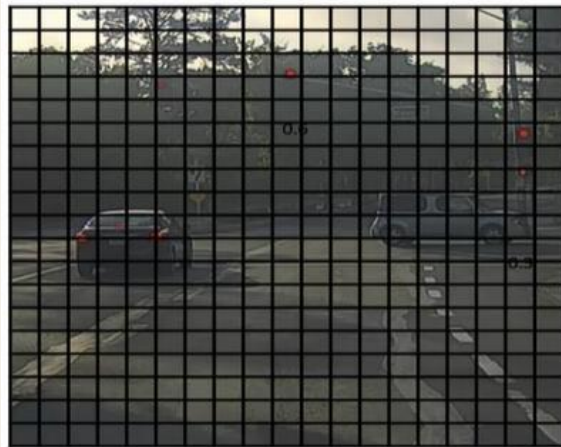
http://ronny.rest/tutorials/module/localization_001/iou/

Детекција – *Non-Maximum Suppression*

- Један исти објект може бити детектован више пута на истој слици нпр. модел га успешно класификује кроз више прозора
- Non-Maximum Suppression је начин одабира само једне детекције по објекту



Детекција – Non-Maximum Suppression



Each output prediction is:

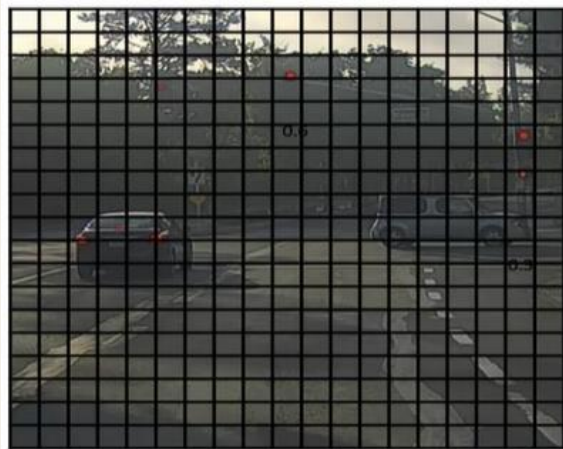
$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

Discard all boxes with $p_c \leq 0.6$

While there are any remaining boxes:

- Pick the box with the largest p_c
Output that as a prediction.
- Discard any remaining box with $\text{IoU} \geq 0.5$ with the box output in the previous step

Детекција – Non-Maximum Suppression



Each output prediction is:

$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

Discard all boxes with $p_c \leq 0.6$

While there are any remaining boxes:

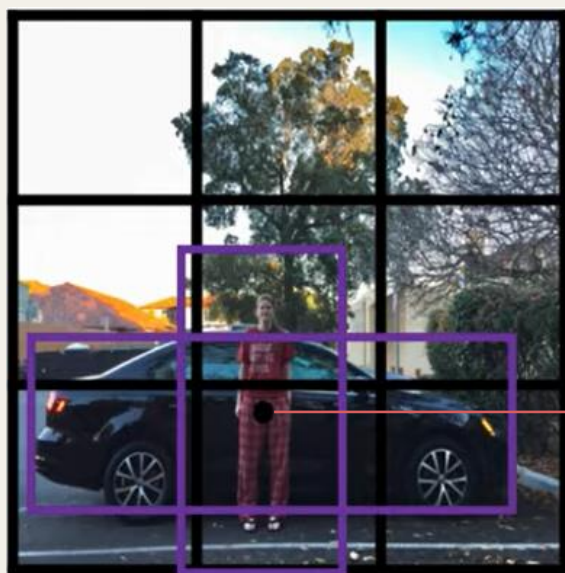
- Pick the box with the largest p_c
Output that as a prediction.
- Discard any remaining box with $\text{IoU} \geq 0.5$ with the box output in the previous step

Вектор је облика али је овде део везан за предикцију класа изостављен јер се алгоритам примењује за сваку класу посебно

$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ w \\ h \\ \cancel{c_1} \\ \cancel{c_2} \\ \cancel{c_3} \end{bmatrix}$$

Детекција – Anchor box

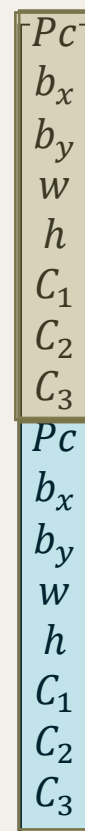
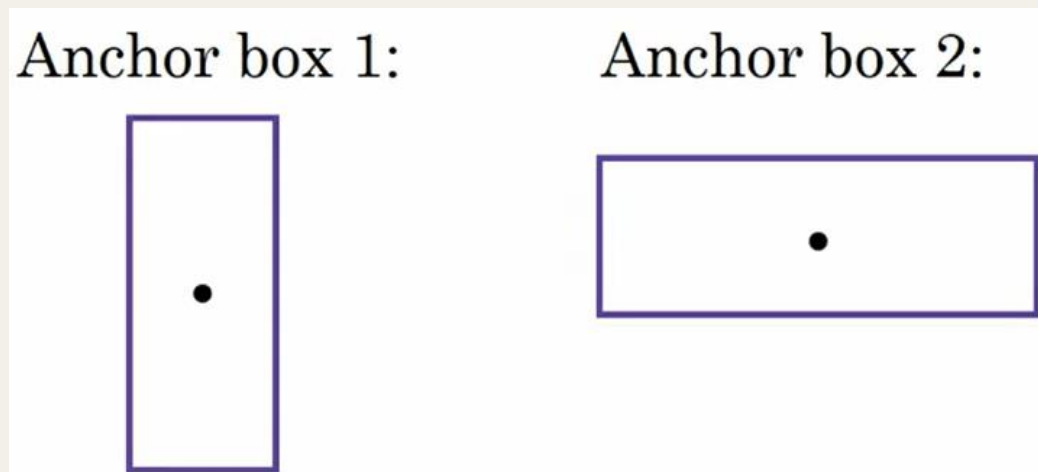
- Сваки прозор (од прозора на које је подељена улазна слика) може да детектује само један објекат
- Шта се дешава уколико се у истом прозору налазе два или више објеката?


$$\begin{bmatrix} P_c \\ b_x \\ b_y \\ w \\ h \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

На овакав начин је могуће доделити само један објекат овом прозору

Детекција – Anchor box

- Предефинисати облик Anchor box-а (два или више)

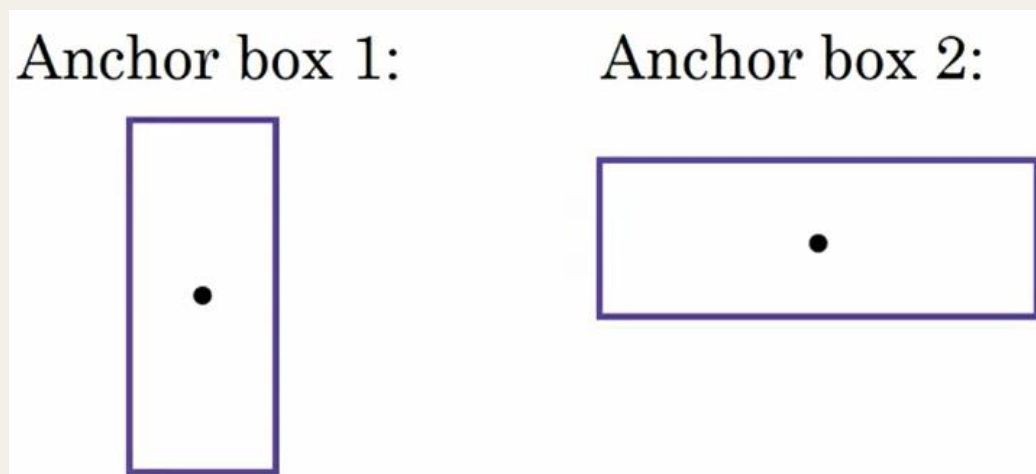


Излази мреже
који одговарају
објекту који
стаје у облик
Anchor box 1

Излази мреже
који одговарају
објекту који
стаје у облик
Anchor box 2

Детекција – *Anchor box*

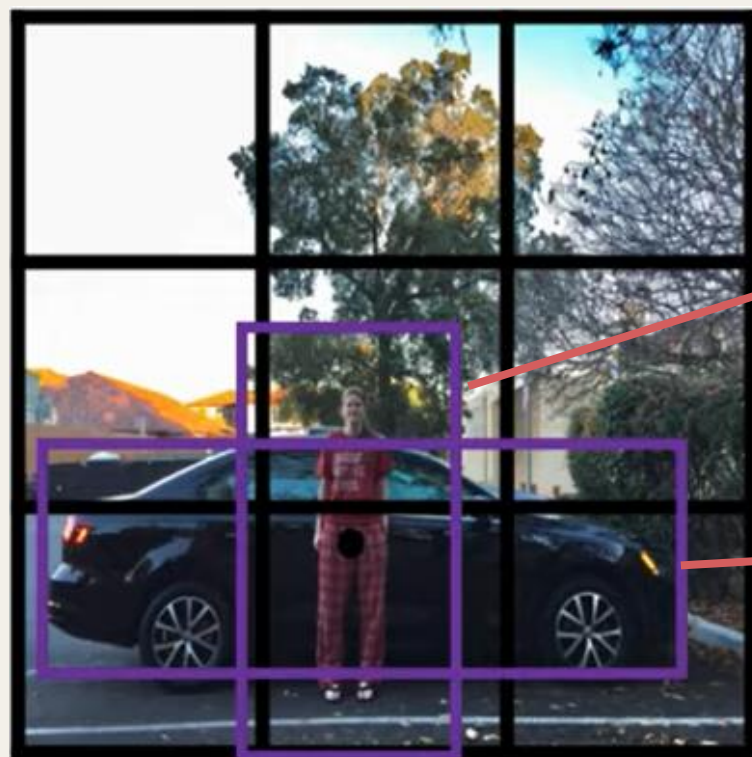
- Предефинисати облик Anchor box-а (два или више)



- Могу да се дефинишу ручно на основу облика већине објеката који постоје у неком скупу података
- Примена неког алгоритма кластеровања којим би се одредили облици који су погодни за коришћење

Детекција – Anchor box

- Предефинисати облик Anchor box-а (два или више)



Јер је облик
објекта
ближи (IoU)
Anchor box
1

Јер је облик
објекта
ближи (IoU)
Anchor box
2

P_c
 b_x
 b_y
 w
 h
 C_1
 C_2
 C_3

Илази мреже
који одговарају
објекту који
стаје у облик
Anchor box 1

Илази мреже
који одговарају
објекту који
стаје у облик
Anchor box 2

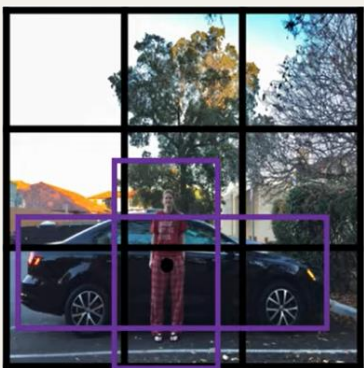
Anchor box 1:



Anchor box 2:



Детекција – Anchor box



Previously:

Each object in training image is assigned to grid cell that contains that object's midpoint.

Излаз је облика $3 \times 3 \times 8$

With two anchor boxes:

Each object in training image is assigned to grid cell that contains object's midpoint and anchor box for the grid cell with highest IoU.

Излаз је облика $3 \times 3 \times 16$ ($3 \times 3 \times 2 \times 8$)

Детекција – Anchor box

Проблеми који остају:

- Ако се у једном прозору налазе три објекта а користе се два anchor box-а
- Ако се у једном прозору налазе два објекта истог облика

... иако се ретко дешавају, поготово ако је улазна слика подељена на више прозора

Детекција – YOLO алгоритам

1. Тренинг модела
2. Примена модела

Детекција – YOLO алгоритам

1. **Тренинг модела**
2. **Примена модела**

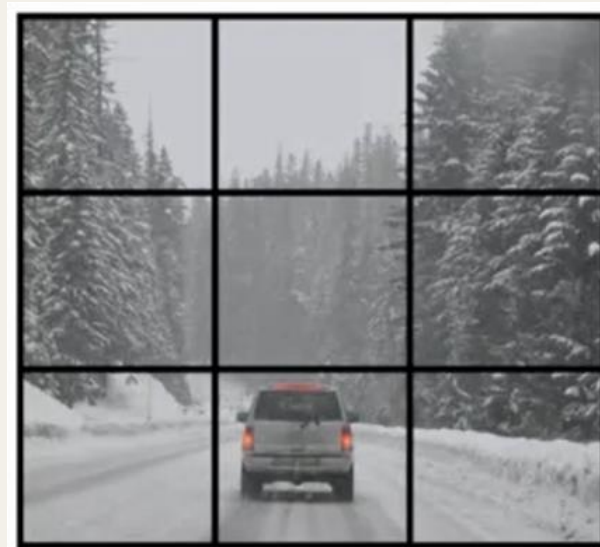
Anchor box 1:



Anchor box 2:



Детекција – YOLO алгоритам

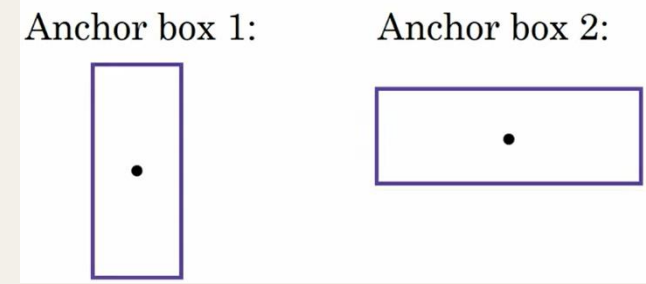


y is $3 \times 3 \times 2 \times 8$

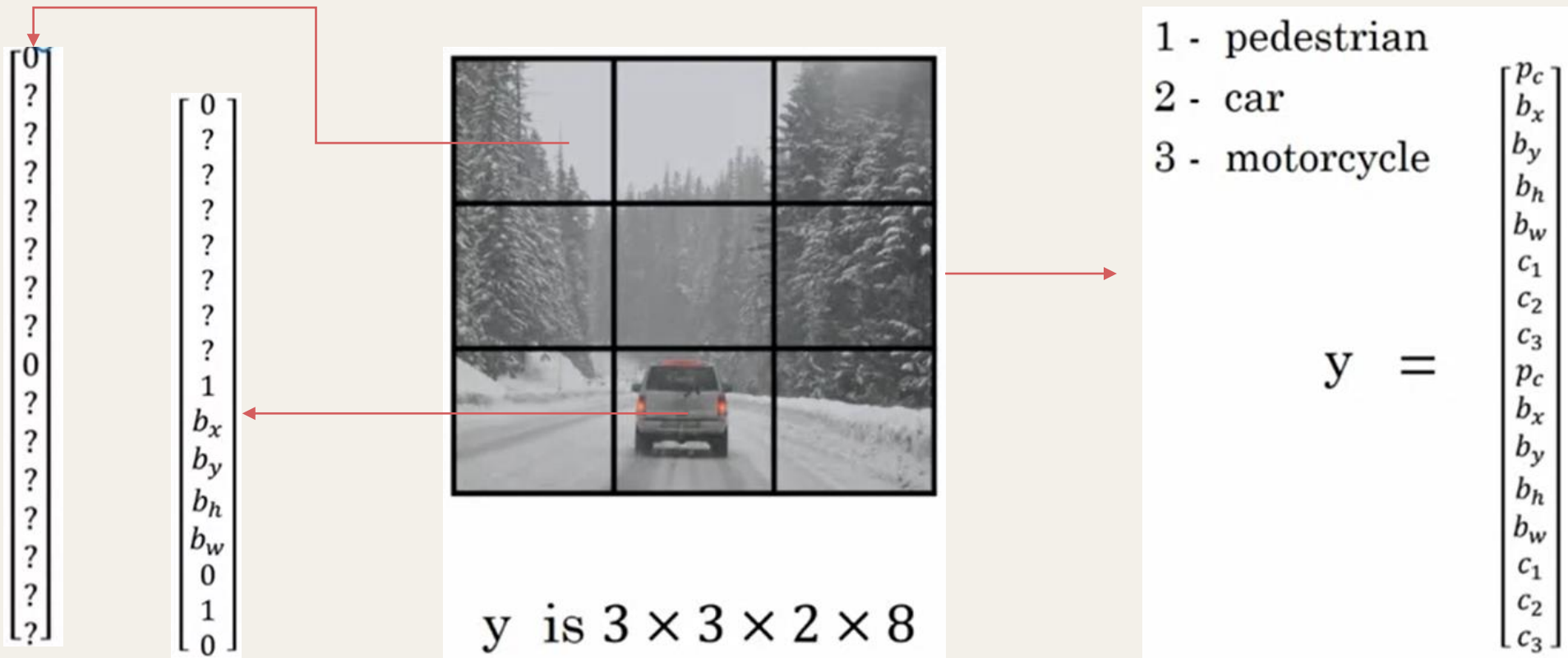
- 1 - pedestrian
- 2 - car
- 3 - motorcycle

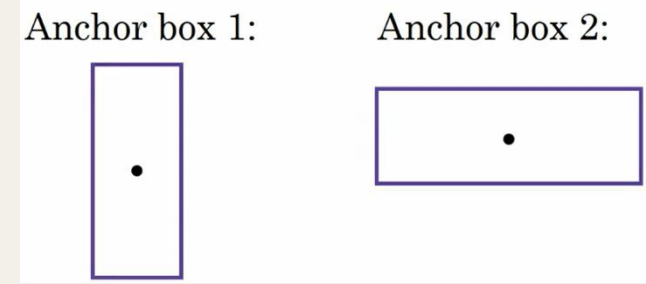
$y =$

$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$



Дет екција – YOLO алгорит ам





Дет екција – YOLO алгорит ам

Јер је
поклапање (IoU)
анотације веће
са anchor box 2

- 0

?

?

?

?

?

?

?

1

b_x

b_y

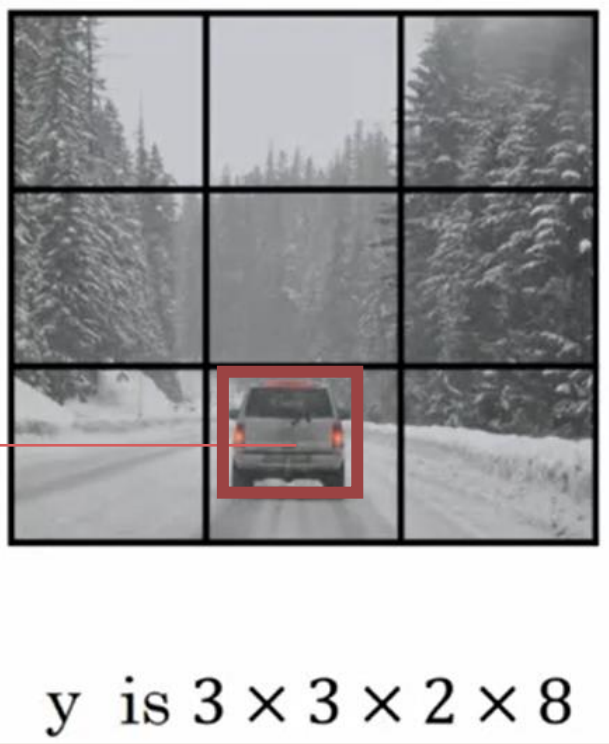
b_h

b_w

0

1

0



- 1 - pedestrian

2 - car

3 - motorcycle

$$y =$$

- p_c

b_x

b_y

b_h

b_w

c_1

c_2

c_3

p_c

b_x

b_y

b_h

b_w

c_1

c_2

c_3

Исто је што $3 \times 3 \times 16$, у пракси се често користи $11 \times 11 \times 16$

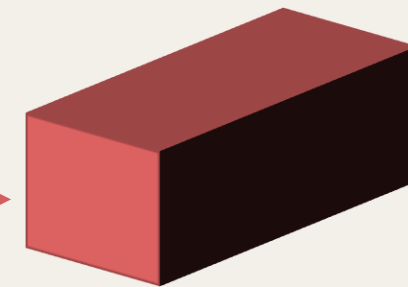
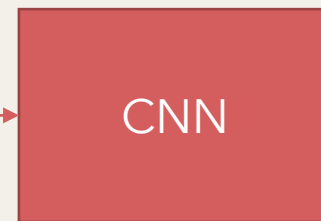
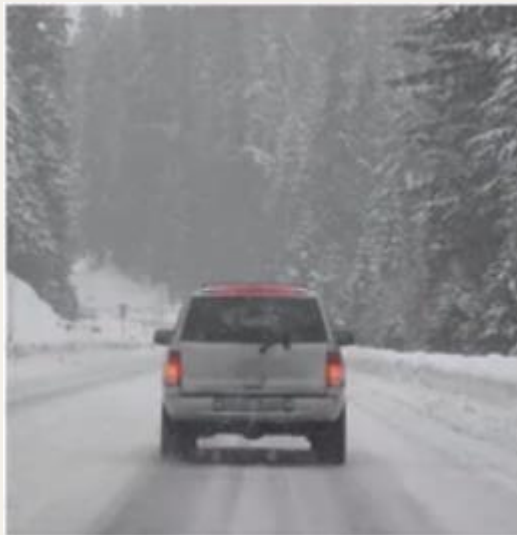
Anchor box 1:



Anchor box 2:



Детекција – YOLO алгоритам



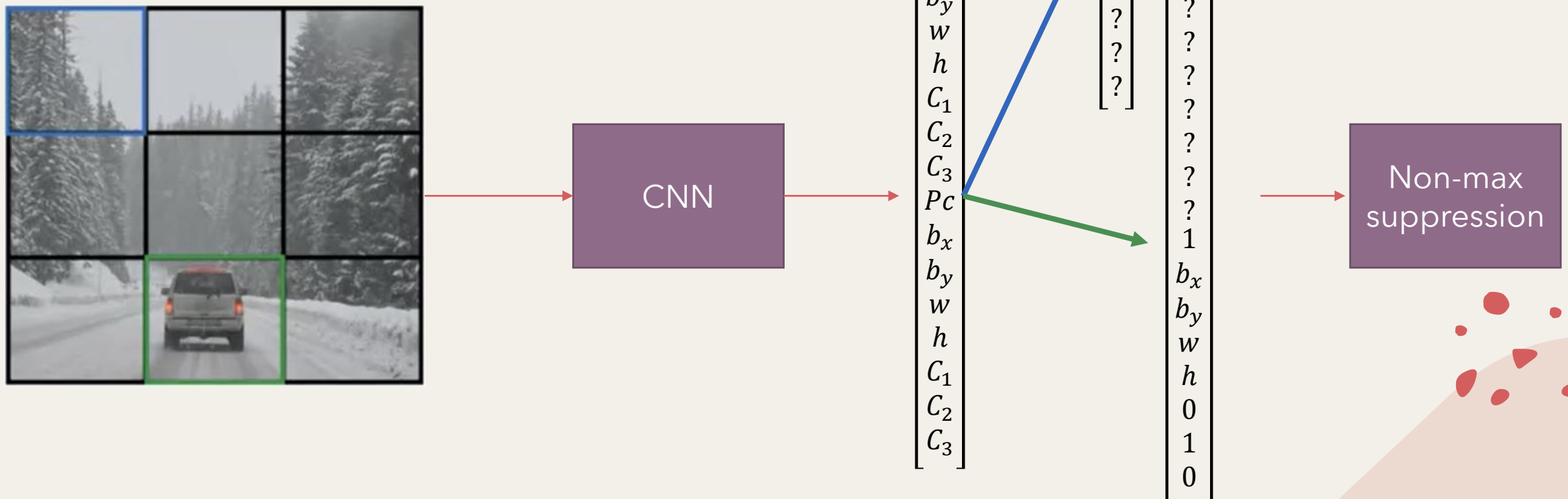
$3 \times 3 \times 16$

Детекција – YOLO алгоритам

1. Тренинг модела

2. Примена модела

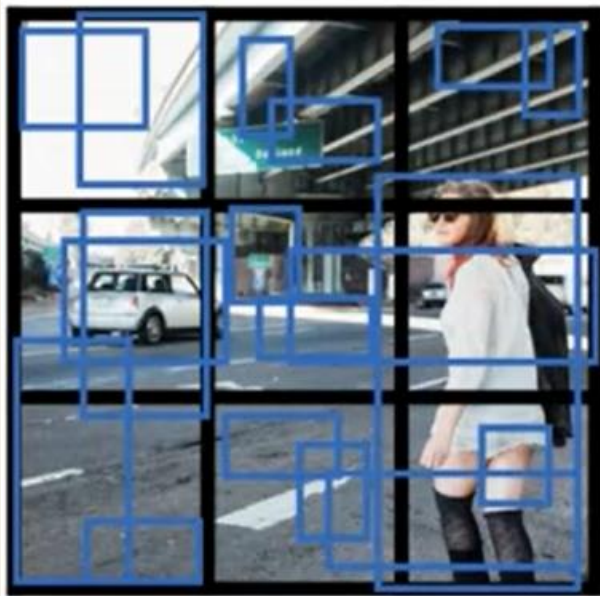
Детекција – YOLO алгоритам



Детекција – YOLO алгоритам



Детекција – YOLO алгоритам



- For each grid cell, get 2 predicted bounding boxes.

Детекција – YOLO алгоритам



- For each grid cell, get 2 predicted bounding boxes.
- Get rid of low probability predictions.

Детекција – YOLO алгоритам



- For each grid cell, get 2 predicted bounding boxes.
- Get rid of low probability predictions.
- For each class (pedestrian, car, motorcycle) use non-max suppression to generate final predictions.

Детекција – YOLO алгоритам

