

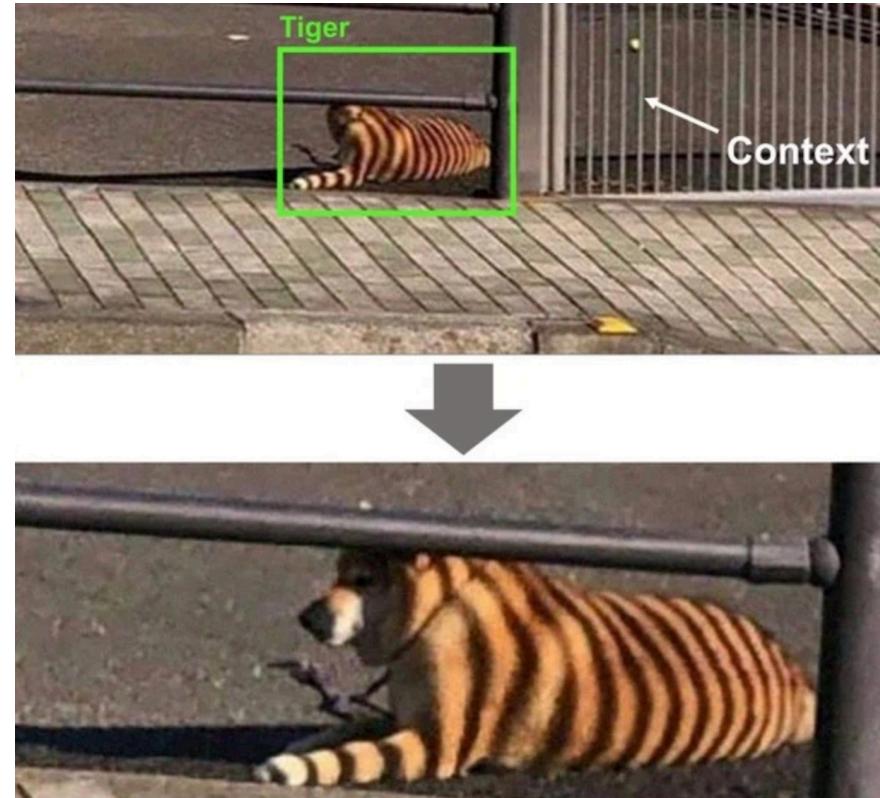
Фаза 2 • Неделя 2 • Вторник

Детекция объектов • Object detection

- Метрика mAP – mean average precision
- Концепция одноэтапного подхода к детекции
- Общая идея архитектуры YOLO
- Использование YOLOv5

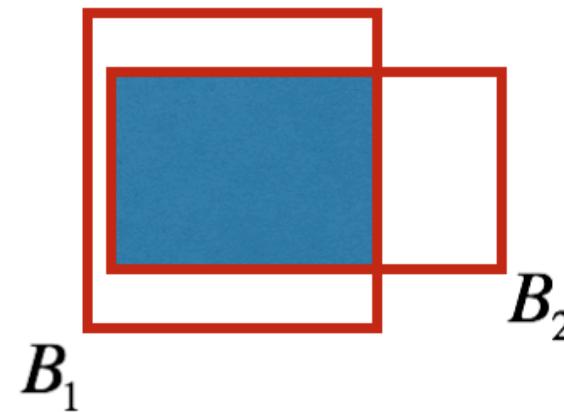
Оценка качества работы детекторов

Как измерить, что детектор хорошо справляется с задачей?

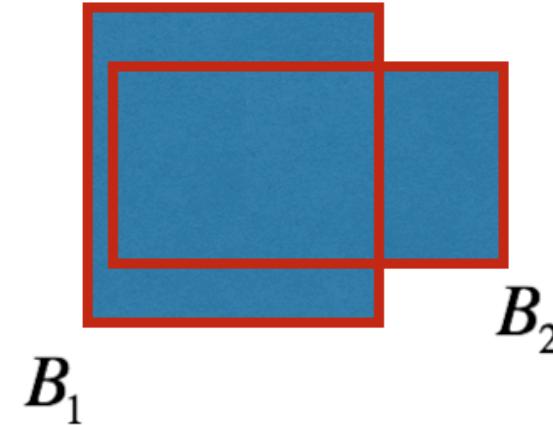


IoU • Intersection over union • Jaccard similarity

Intersection

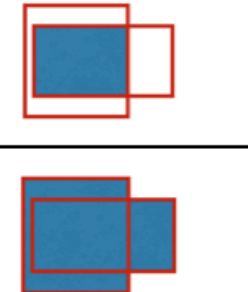


Union



Intersection over Union

$$IoU = \frac{B_1 \cap B_2}{B_1 \cup B_2} = \frac{\text{Intersection Area}}{\text{Union Area}}$$



Ошибки предсказаний

- Объекта нет, а бокс нарисовали – это 

Ошибки предсказаний

- Объекта нет, а бокс нарисовали – это **False Positive**
- **Объект есть, но бокс не нарисовали** – это **False Negative**

Ошибки предсказаний

- Объекта нет, а бокс нарисовали – это **False Positive**
- Объект есть, но бокс не нарисовали – это **False Negative**
- **Объект есть, бокс правильный** – это

Ошибки предсказаний

- Объекта нет, а бокс нарисовали – это **False Positive**
- Объект есть, но бокс не нарисовали – это **False Negative**
- Объект есть, бокс правильный – это **True Positive**

Метрики

Precision = ?

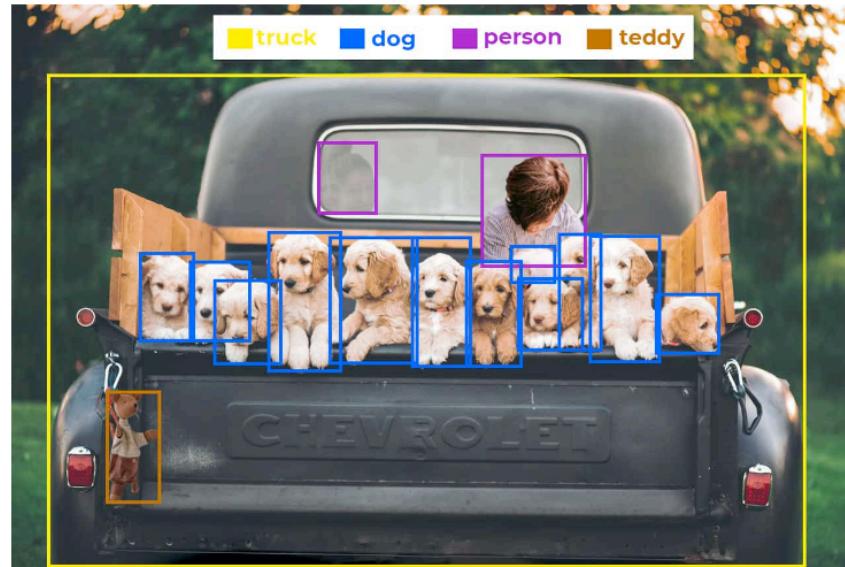
Recall = ?

Метрики

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Mean average precision • mAP

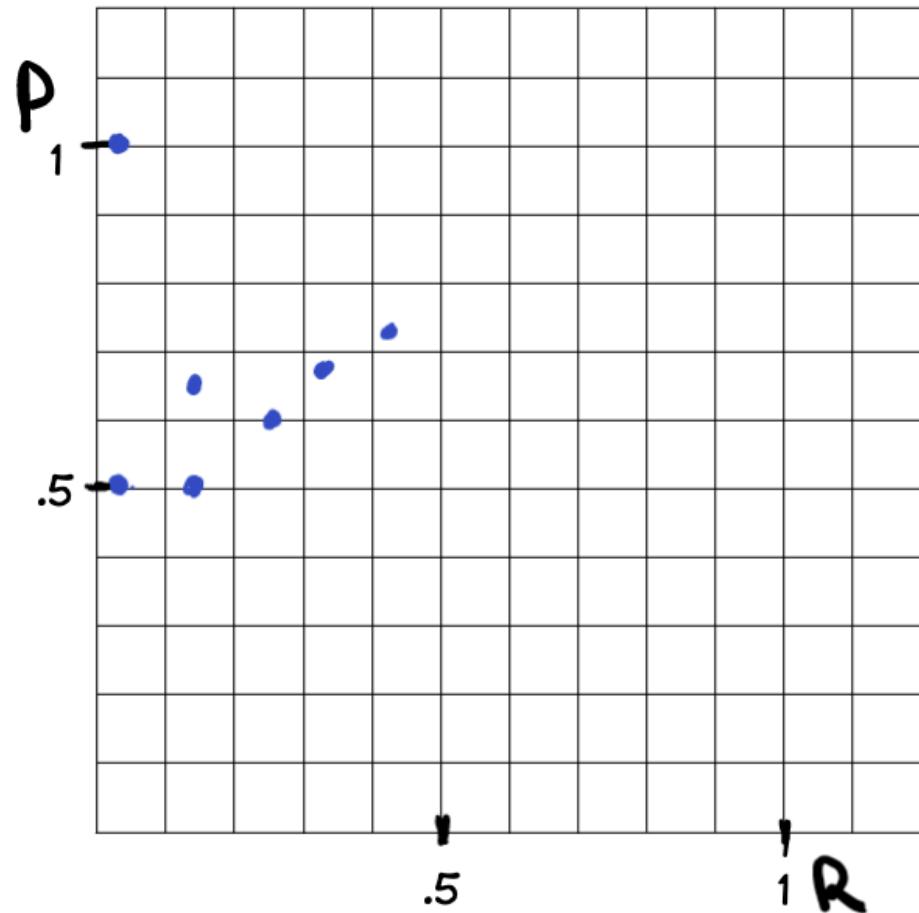


Ground Truth =
16
(Total objects)

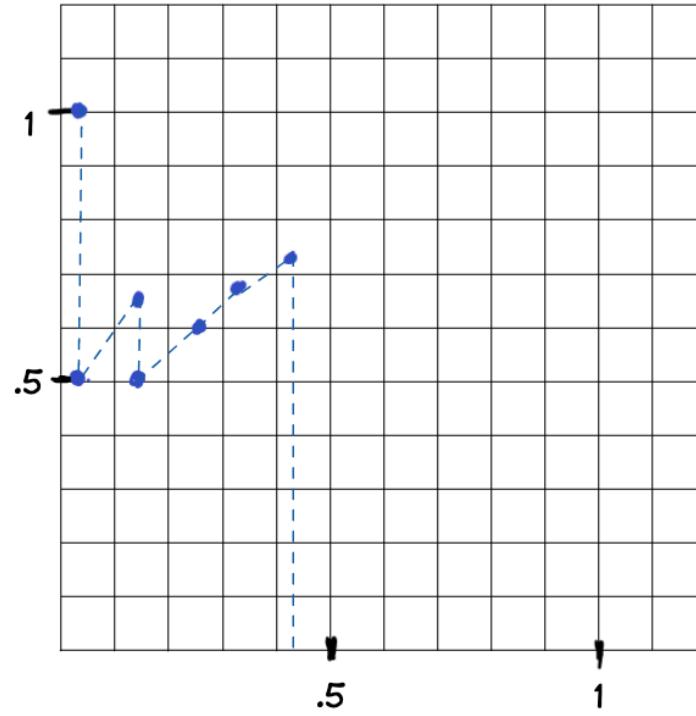
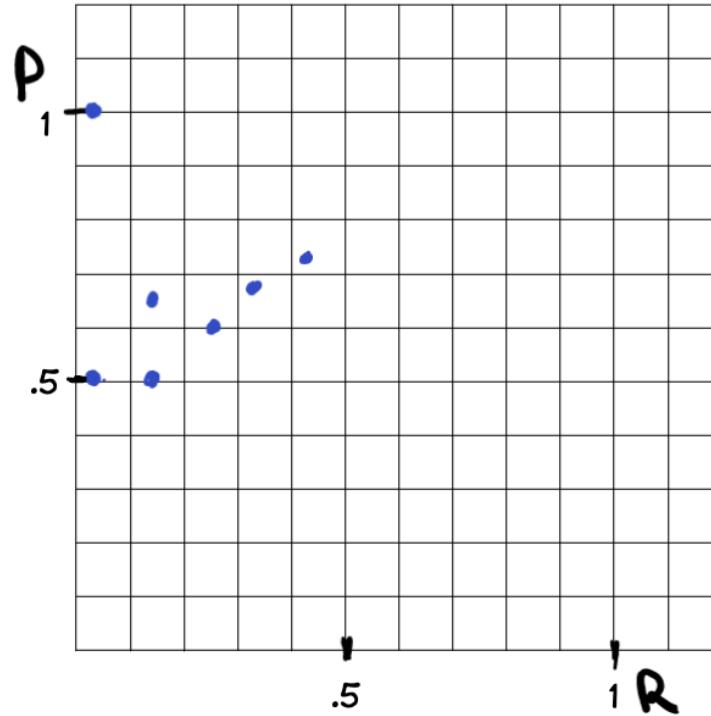
Preds.	Conf.	Matches	Cumulative TP	Cumulative FP	Precision	Recall
	0.92	TP	1	0	1/(1+0) = 1	1/16 = 0.06
	0.91	FP	1	1	1/(1+1) = 0.5	1/16 = 0.06
	0.88	TP	2	1	2/(2+3) = 0.66	2/16 = 0.12
	0.86	FP	2	2	0.5	0.12
	0.77	TP	3	2	0.6	0.25
	0.63	TP	4	2	0.66	0.33
	0.58	TP	5	2	0.71	0.41

Mean average precision • mAP

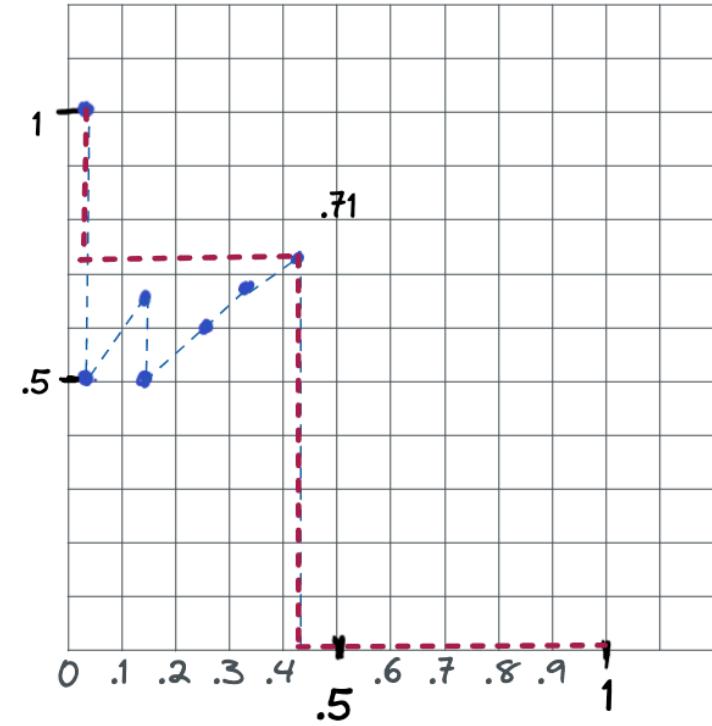
Preds.	Conf.	Matches	Cumulative TP	Cumulative FP	Precision	Recall
	0.92	TP	1	0	$1/(1+0) = 1$	$1/16 = 0.06$
	0.91	FP	1	1	$1/(1+1) = 0.5$	$1/16 = 0.06$
	0.88	TP	2	1	$2/(2+1) = 0.66$	$2/16 = 0.12$
	0.86	FP	2	2	0.5	0.12
	0.77	TP	3	2	0.6	0.25
	0.63	TP	4	2	0.66	0.33
	0.58	TP	5	2	0.71	0.41



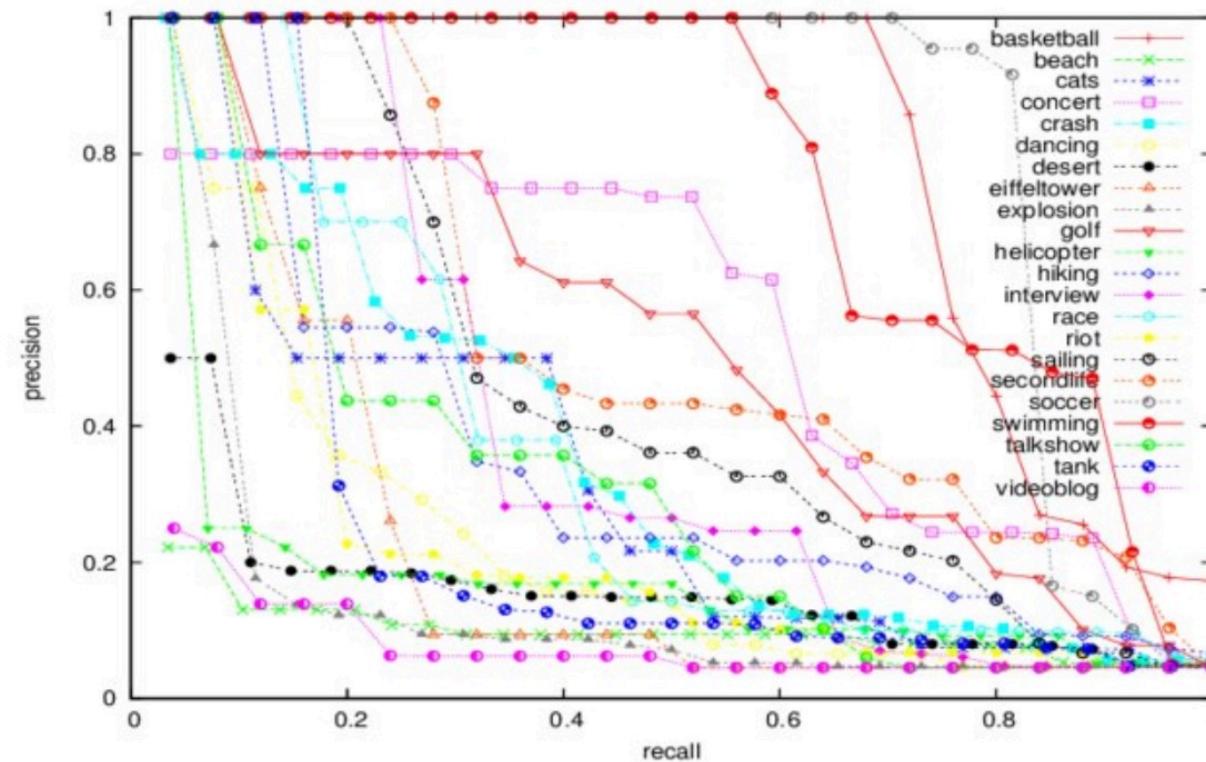
Mean average precision • mAP



Interpolation over 11 points



Mean average precision • mAP



Семейство YOLO • You only look once



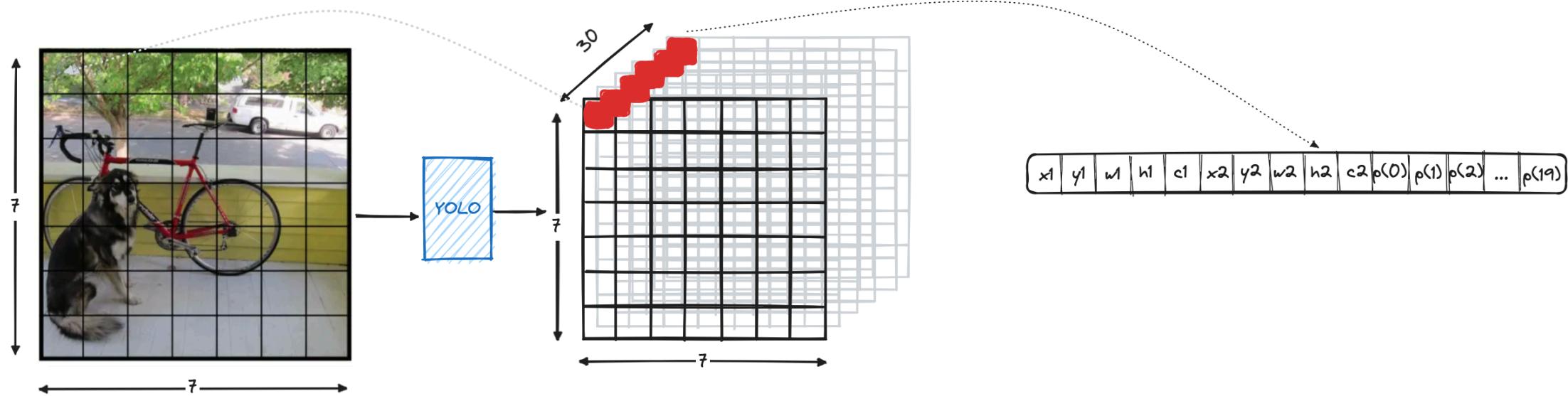
- Текущие модели детекции работали на 17 fps (frame per second)
- Хотелось быстрее – в 2016 году предложили модель YOLO
- Сейчас только номерных версий 9, различных модификаций еще больше

Архитектура YOLO

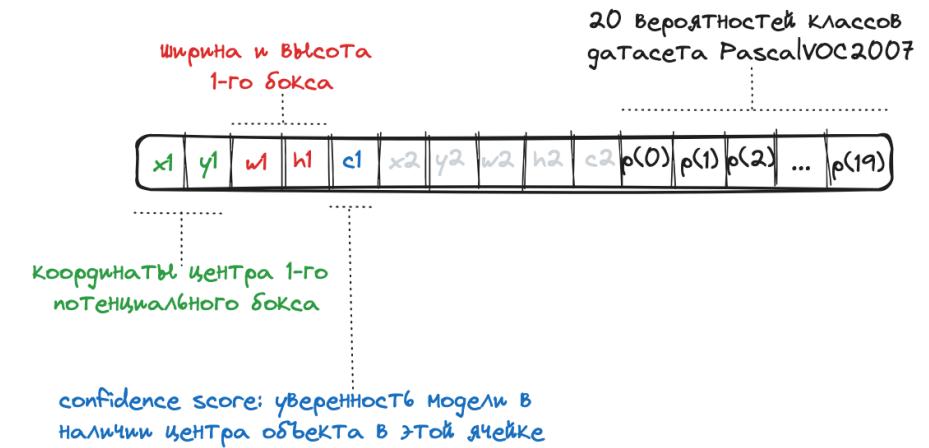
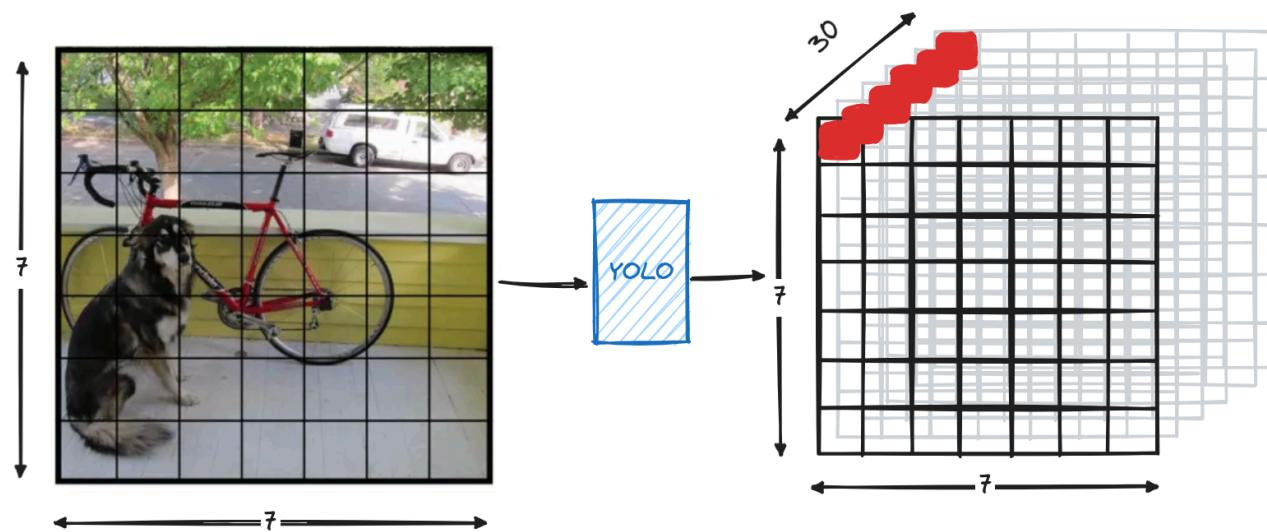
Состоит из двух частей:

- Классификационная часть (немного модифицированный GoogLeNet, а в общем случае любая предобученная на ImageNet модель)
- Детекционная часть (состоит из нескольких полно связных), выход которой тензор $7 \times 7 \times 30$ (30 матриц размера 7×7)

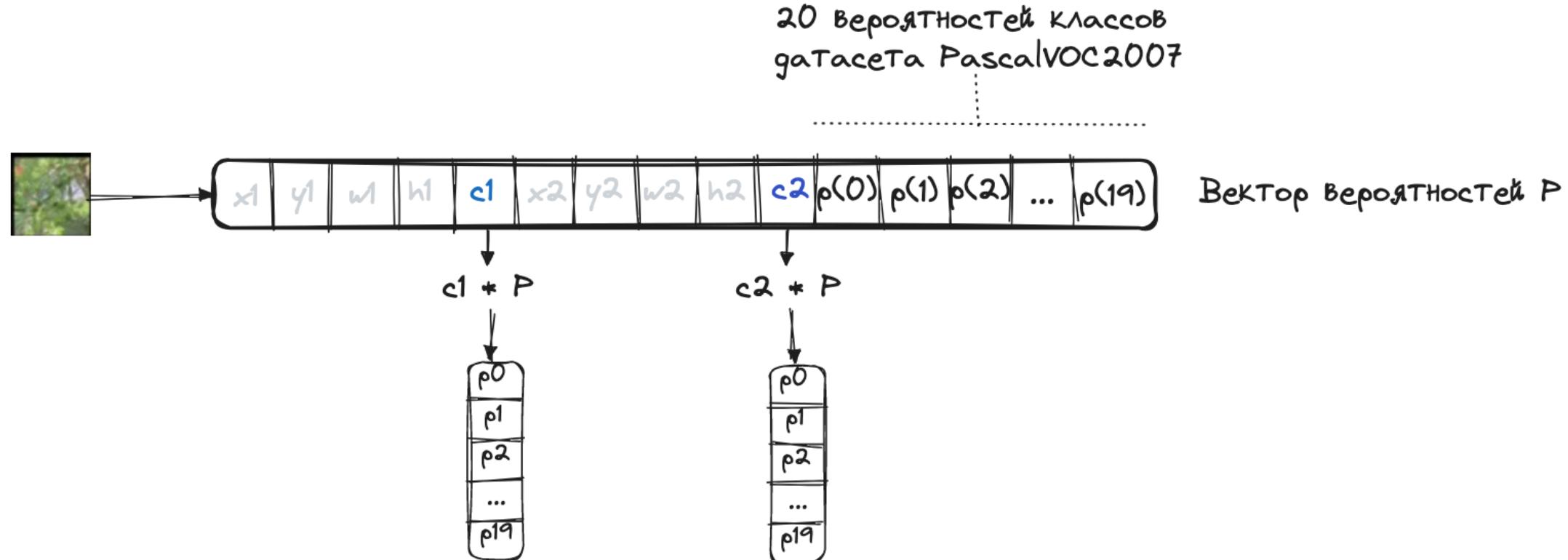
Архитектура YOLO



Архитектура YOLO



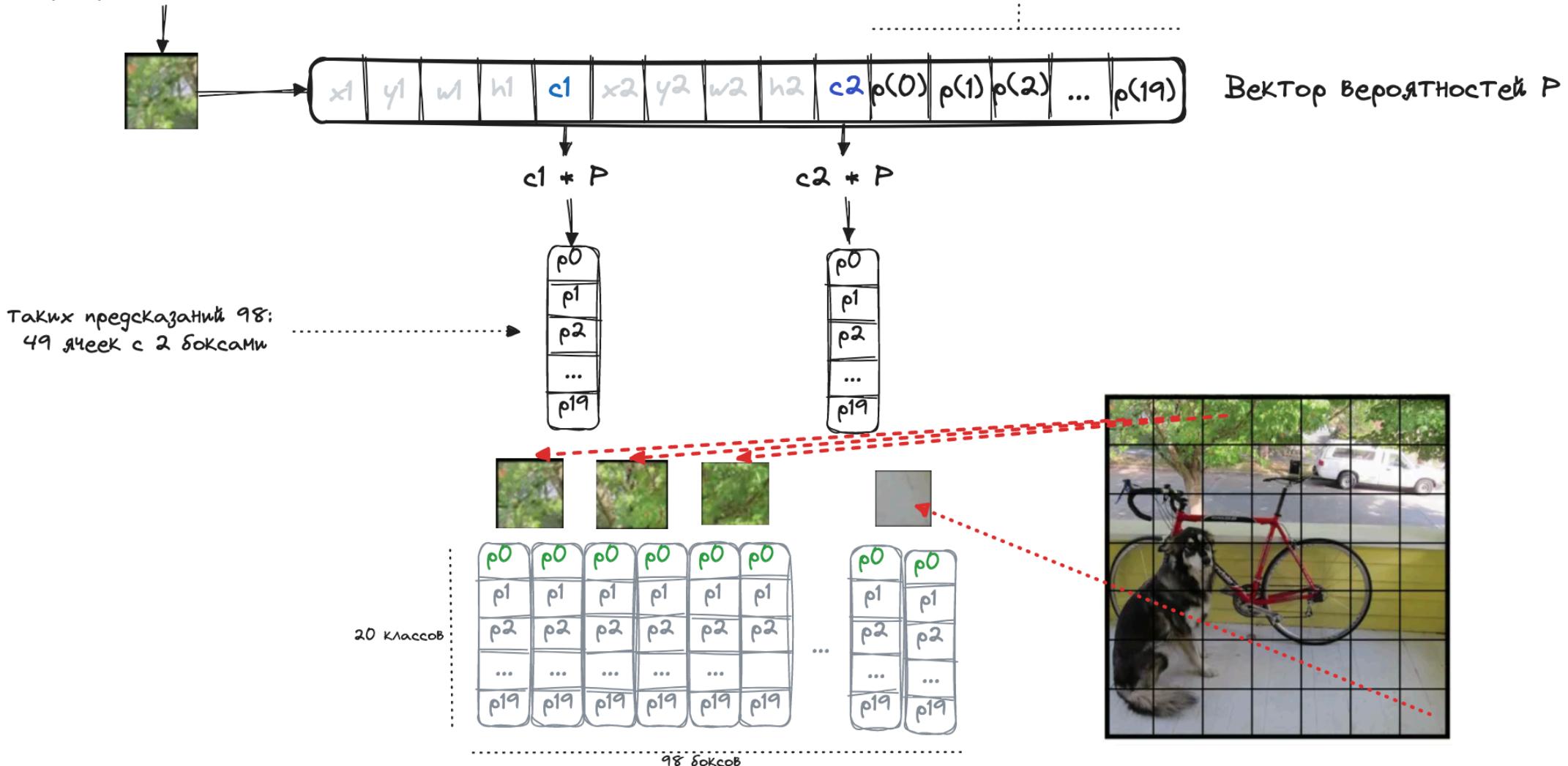
Архитектура YOLO



Архитектура YOLO

Таких ячеек $7 \times 7 = 49$
для каждого из 98 потенциальных бокса

20 вероятностей классов
датасета PascalVOC2007

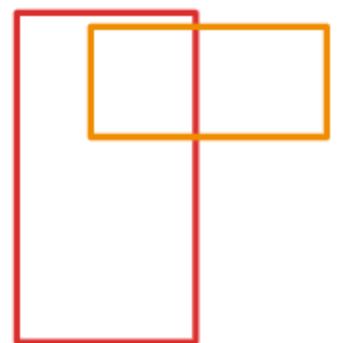
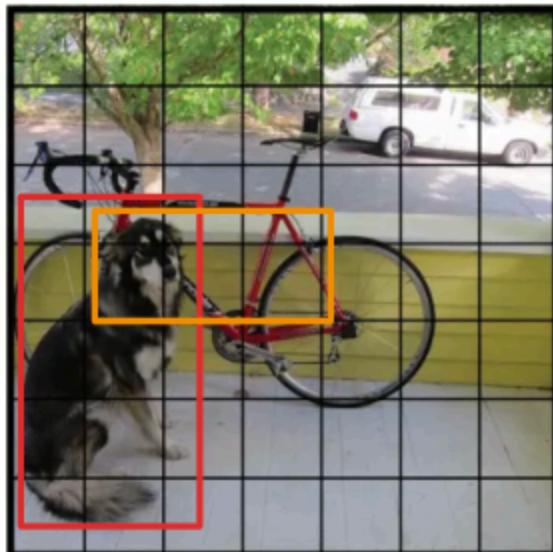


NMS • Non Maximum Suppression

.4	.3	.28	.19	.1	0	.1	0	.1	0	.12	.12	0	.18	026	.14	.13	.8
----	----	-----	-----	----	---	----	---	----	---	-----	-----	---	-----	---	-----	-----	-----	-----	----

затираем всё, что <0.3 (может перебираться) и сортируем

.8	.4	.3	0	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
----	----	----	---	---	---	---	---	---	---	---	---	---	---	---	-----	---	---	---	---

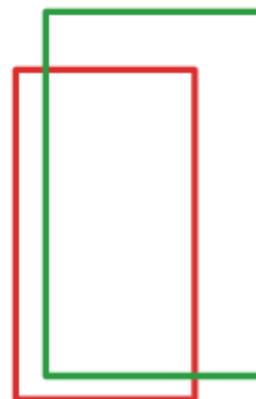
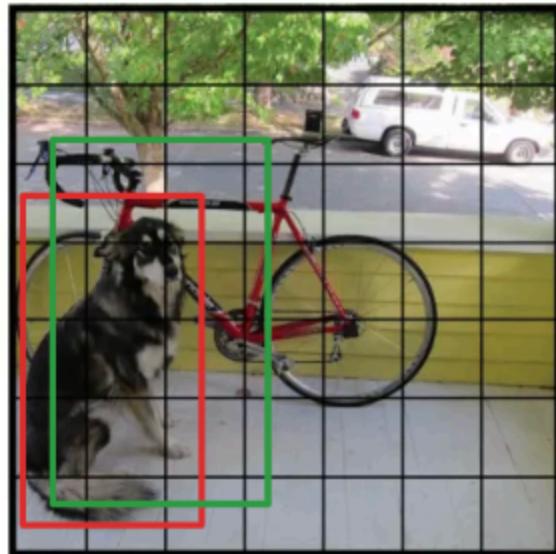
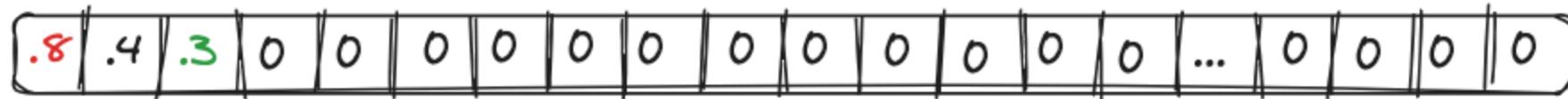


IoU > .5? — нет → оставляем бокс

ga

затираем бокс

NMS



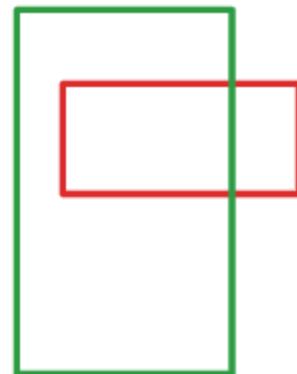
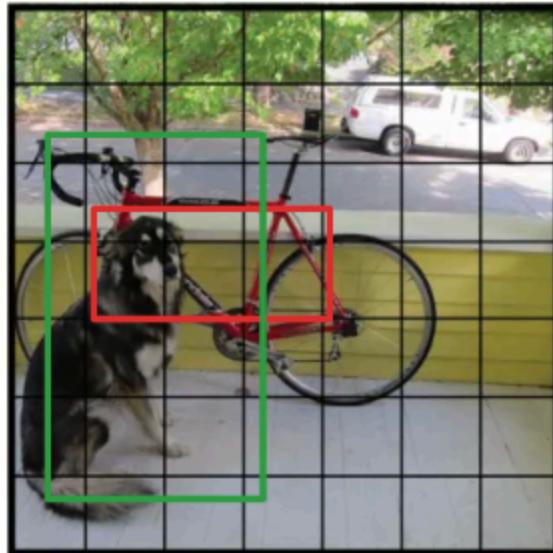
$\text{IoU} > .5?$ — нет → оставляем бокс

9

затягиваем бокс

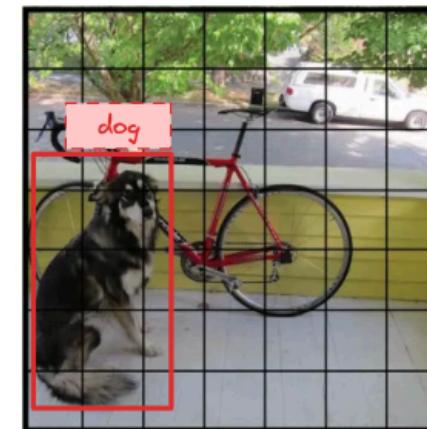
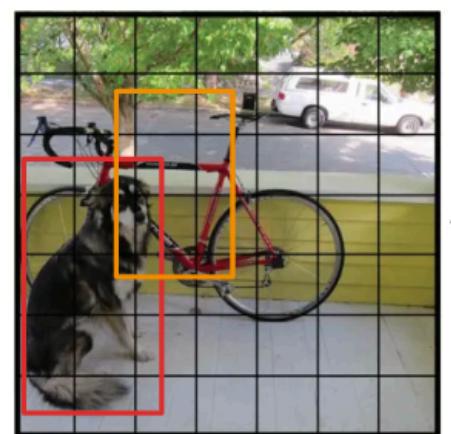
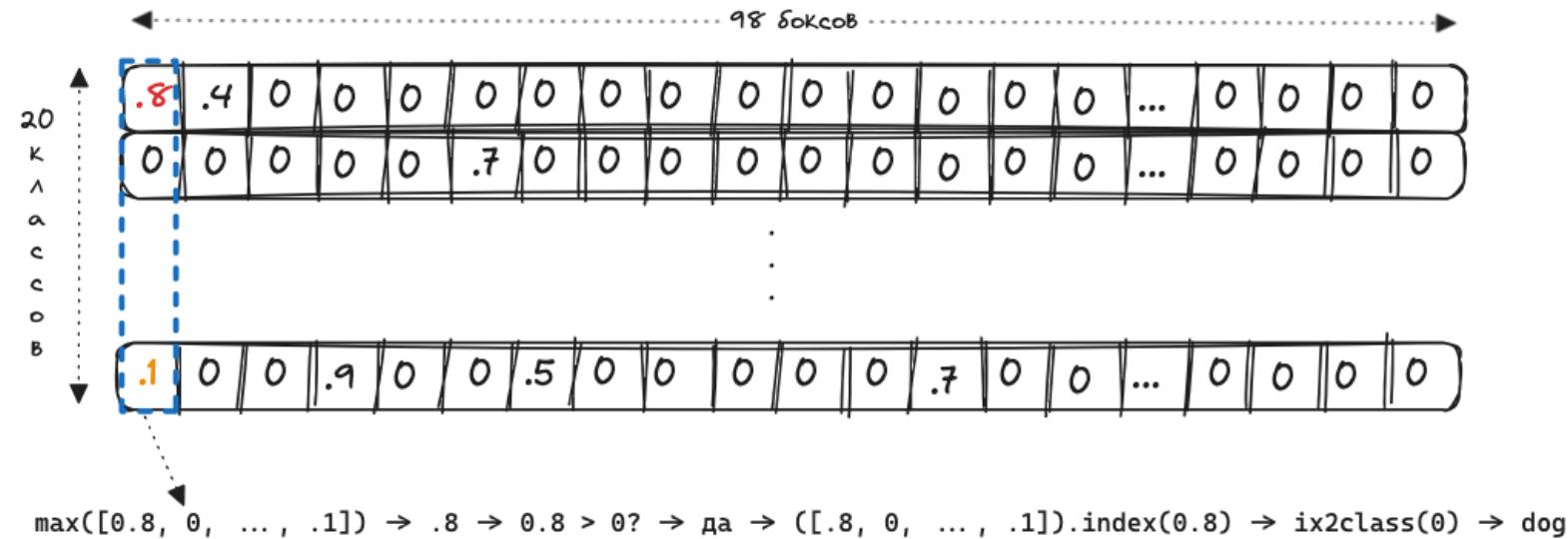
NMS

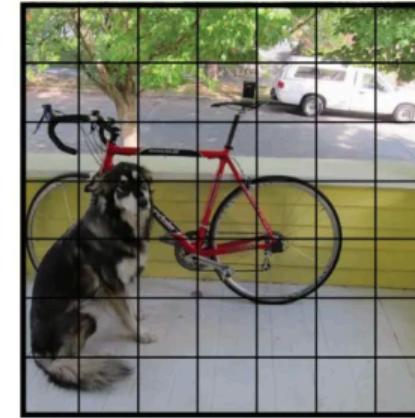
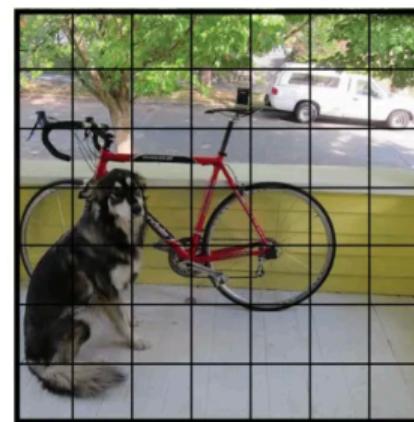
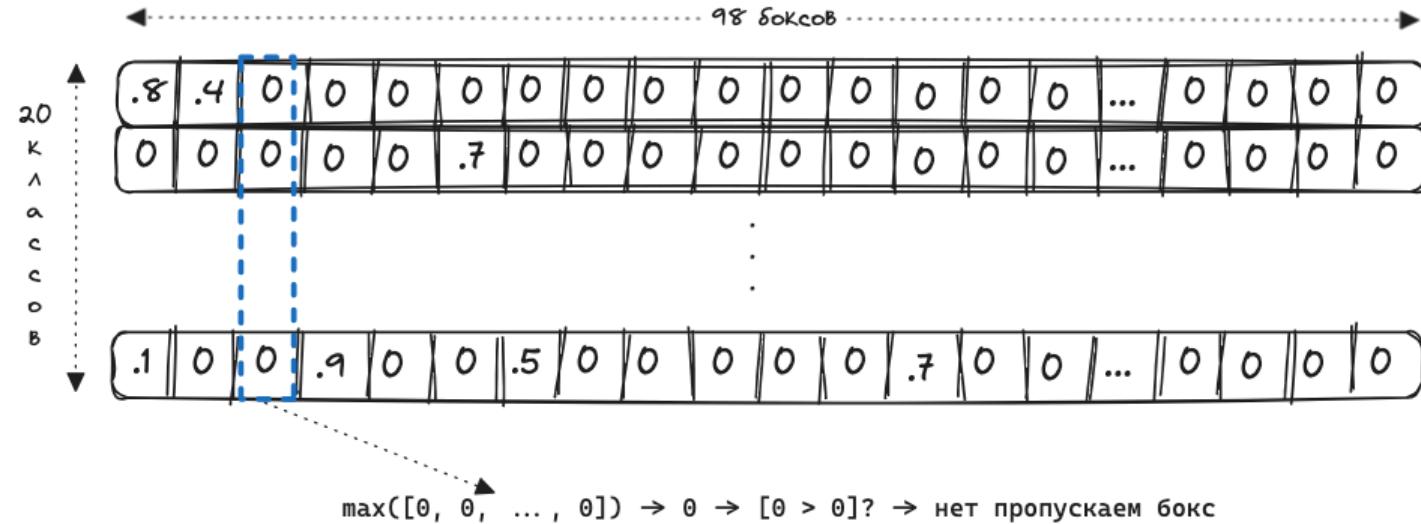
.8 .4 .3 0 0 0 0 0 0 0 0 0 0 0 0 ... 0 0 0 0



$\text{IoU} > .5?$ — нет \rightarrow оставляем бокс

затыляем бокс





Функция потерь

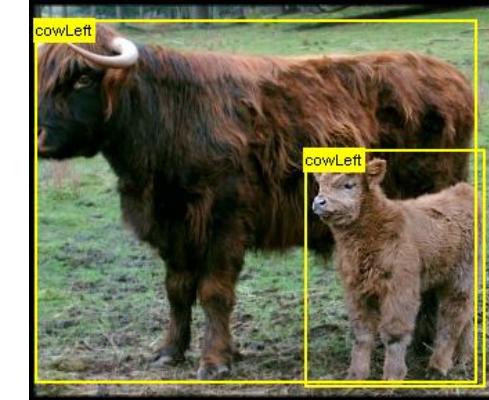
У YOLO весёлый лосс, но разбирать мы его не будем

loss function:

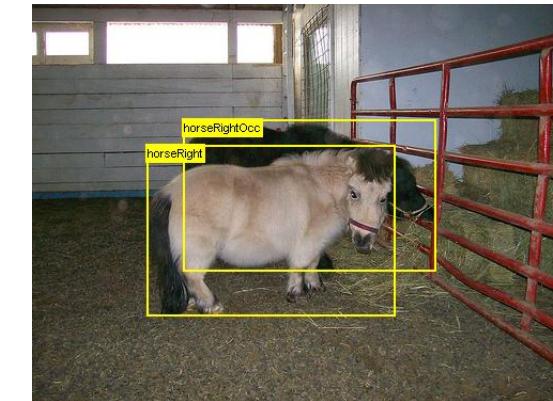
$$\begin{aligned}
 & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
 & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
 & + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (3)
 \end{aligned}$$

ИТОГИ

- mAP $\approx 63\%$ на PASCAL VOC 2007
- скорость: 45 fps, то есть в реальном времени
- но проиграла другой модели (о ней завтра) на датасете PASCAL VOC 2012
- история YOLO по частям: 1 → 2 → 3 → 4 → 5



PASCAL VOC 2007



PASCAL VOC 2012