



# MACHINE LEARNING AND AI

Лекция n+1

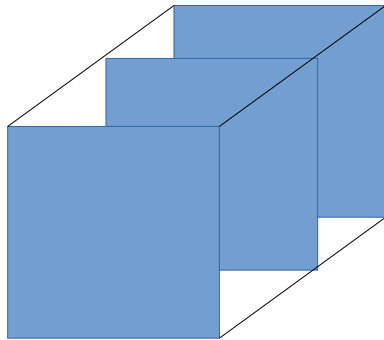


# ГЛУБОКОЕ ОБУЧЕНИЕ — ЧТО НУЖНО ЗНАТЬ?

- Много чего нужно знать
- И это тоже нужно знать
- А это будет на экзамене
- Всё нужно знать

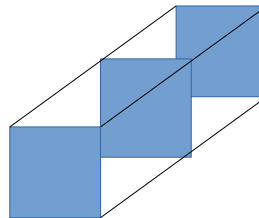
# 1x1 CONVOLUTION

Input



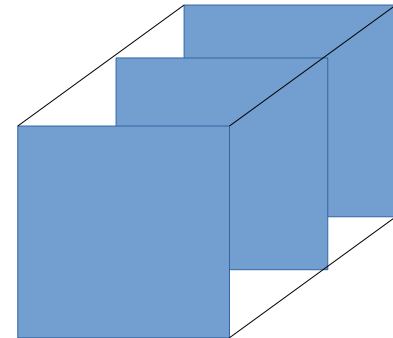
28x28x64

Filters



5x5x64  
128 шт.

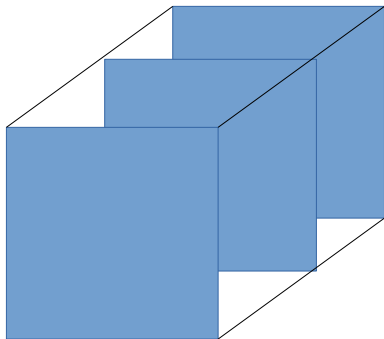
Output



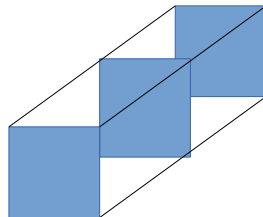
28x28x128

\*

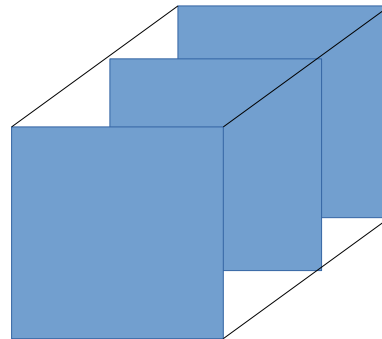
=



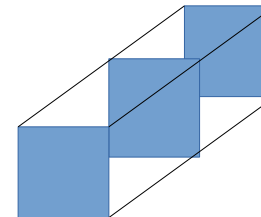
28x28x64



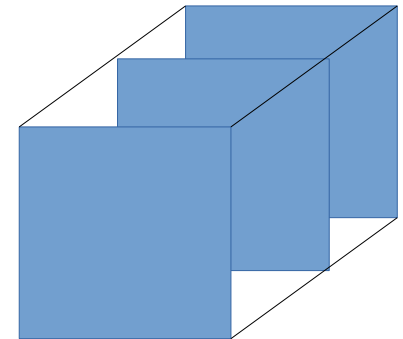
1x1x64  
10 шт.



28x28x10

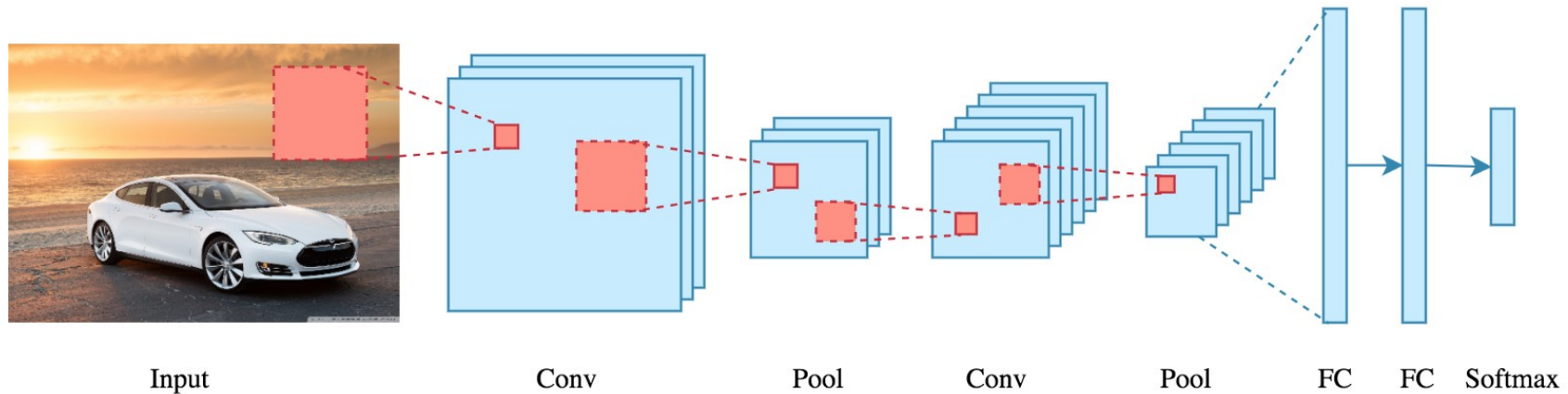


5x5x10  
128 шт.



28x28x128

# BATCH NORMALIZATION



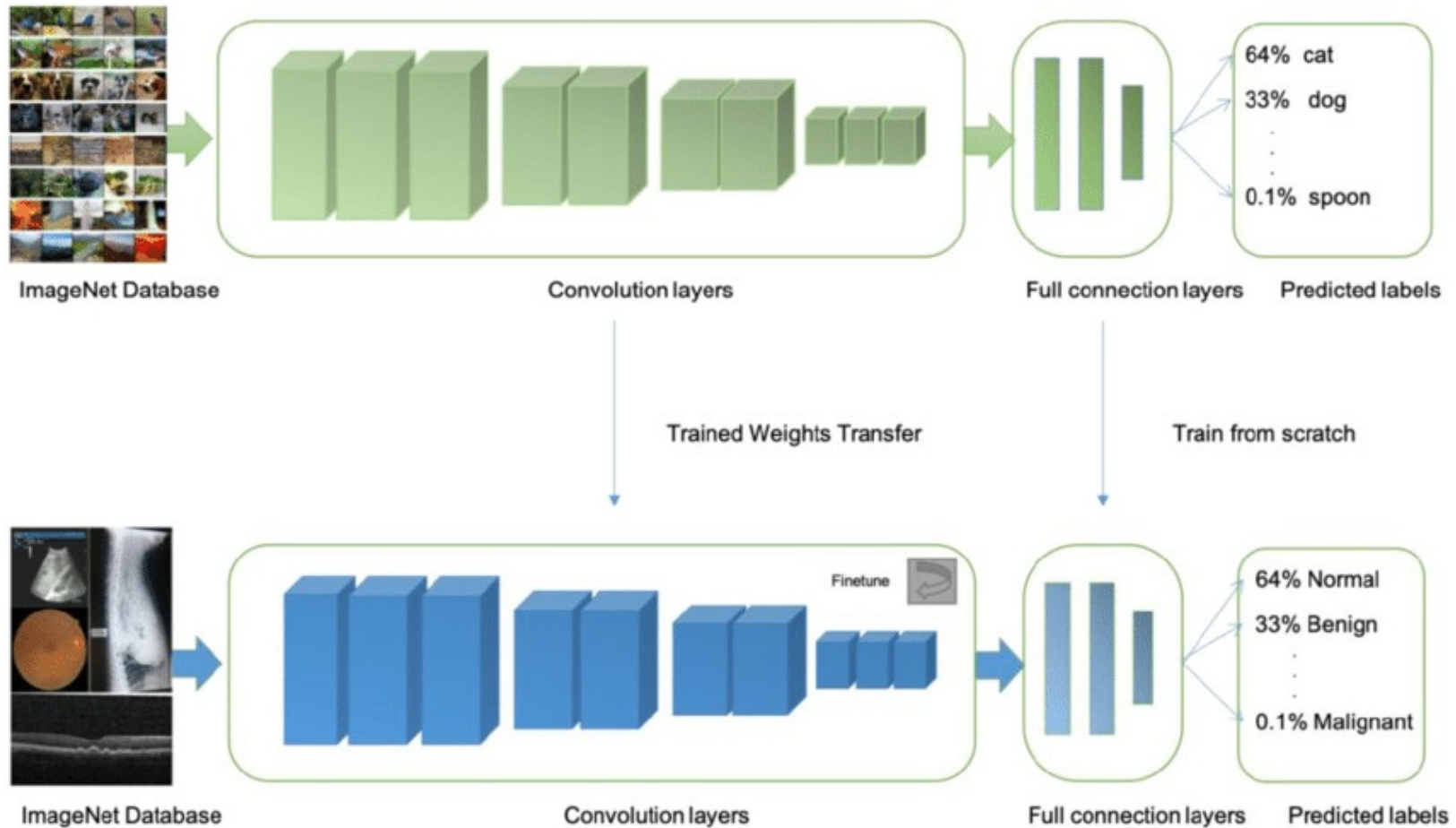
$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad \sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2$$

$$\hat{x}_i^{(k)} = \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{\sigma_B^{(k)2} + \epsilon}}$$

$$k \in [1, d] \quad i \in [1, m]$$

$$y_i^{(k)} = \gamma^{(k)} \hat{x}_i^{(k)} + \beta^{(k)}$$

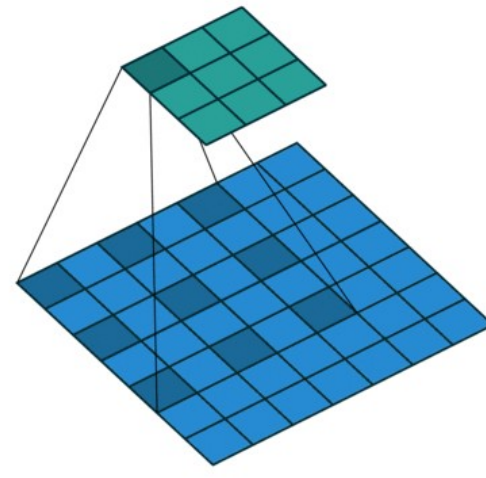
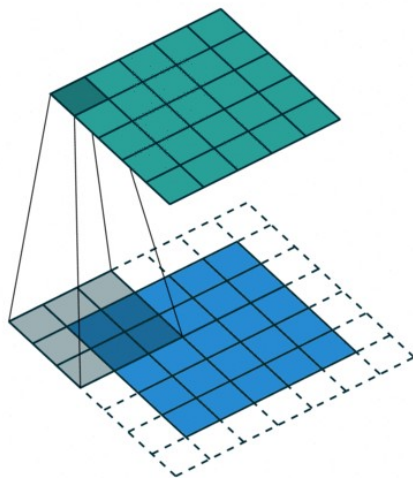
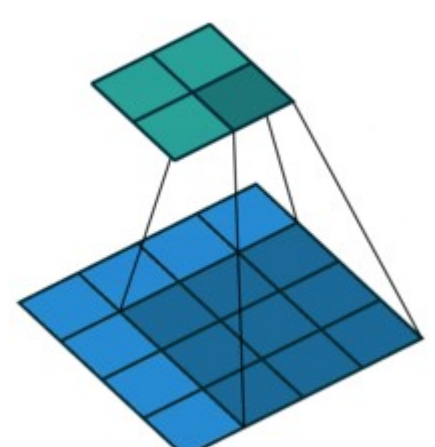
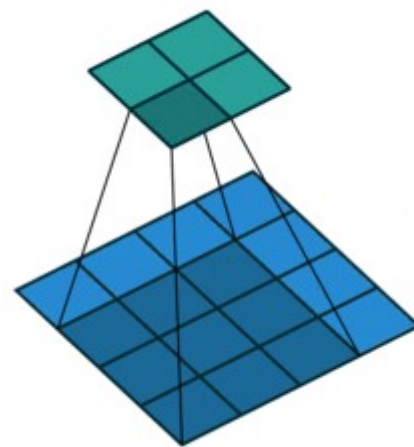
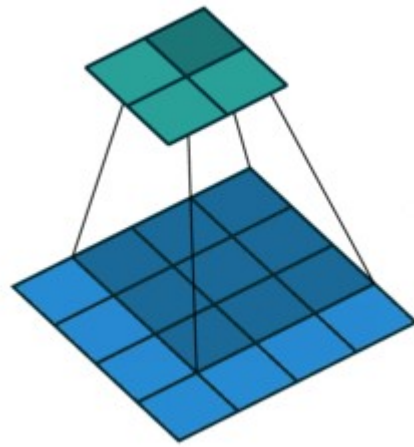
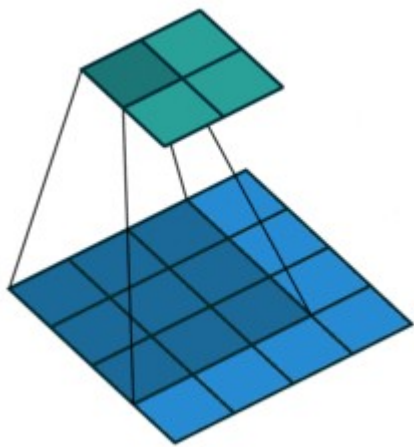
# TRANSFER LEARNING



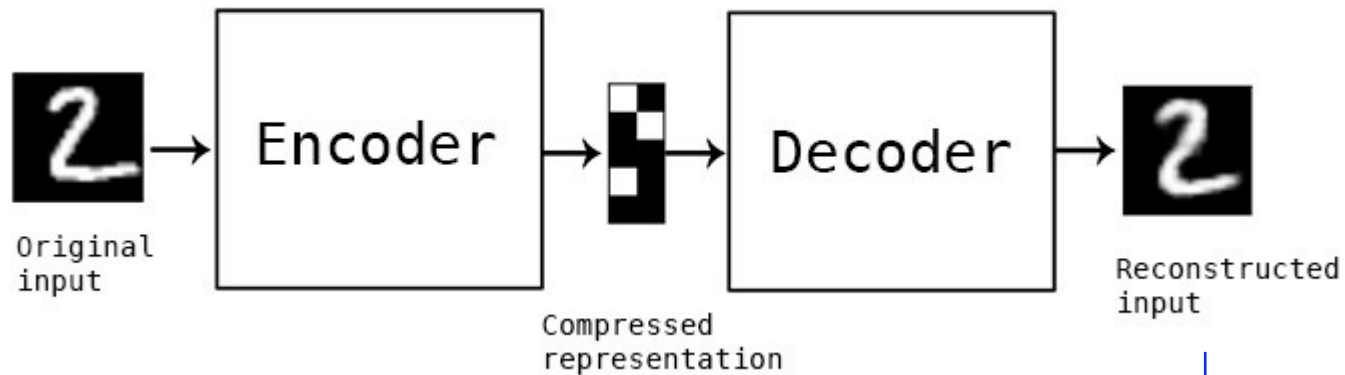
# END-TO-END LEARNING

- Пример обучения end-to-end — создание транскрипта по записанным аудиоклипам
- Ещё один пример — беспилотные автомобили. Они анализируют визуальную информацию посредством CNN. В этом случае системе сперва предоставляется учитель-человек
- Машинный перевод

# DECONVOLUTION



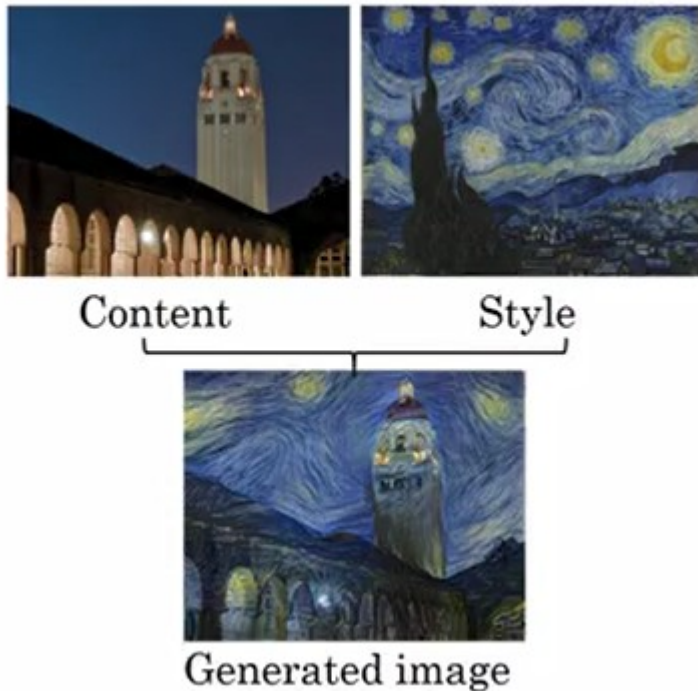
# AUTOENCODER CNN



↓  
Motion



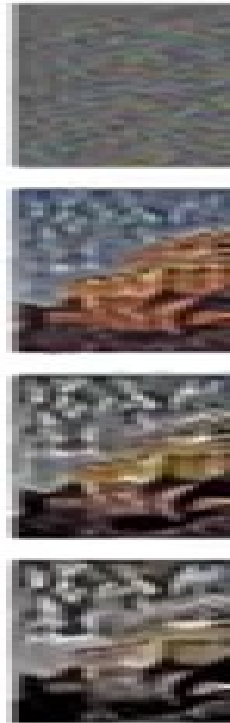
# NEURAL STYLE TRANSFER



$$J(G) = \alpha J_{\text{content}}(C, G) + \beta J_{\text{style}}(S, G)$$

1. Генерируем  $G$  случайно  
 $G: 100 \times 100 \times 3$
2. Используем градиентный спуск

$$G := G - \frac{\partial}{\partial G} J(G)$$



# NEURAL STYLE TRANSFER

$$J(G) = \alpha J_{\text{content}}(C, G) + \beta J_{\text{style}}(S, G)$$

Используем скрытый слой  $l$  для расчёта  $J_{\text{content}}(C, G)$

Используем обученную CNN (например VGG)

Пусть  $a^{[l](C)}$  и  $a^{[l](G)}$  — активация слоя  $l$  от изображений  $C$  и  $G$

Если  $a^{[l](C)}$  и  $a^{[l](G)}$  почти равны, то обе картинки имеют похожий контент

$$J_{\text{content}}(C, G) = \frac{1}{2} \|a^{[l](C)} - a^{[l](G)}\|^2$$



# FACE RECOGNITION/VERIFICATION

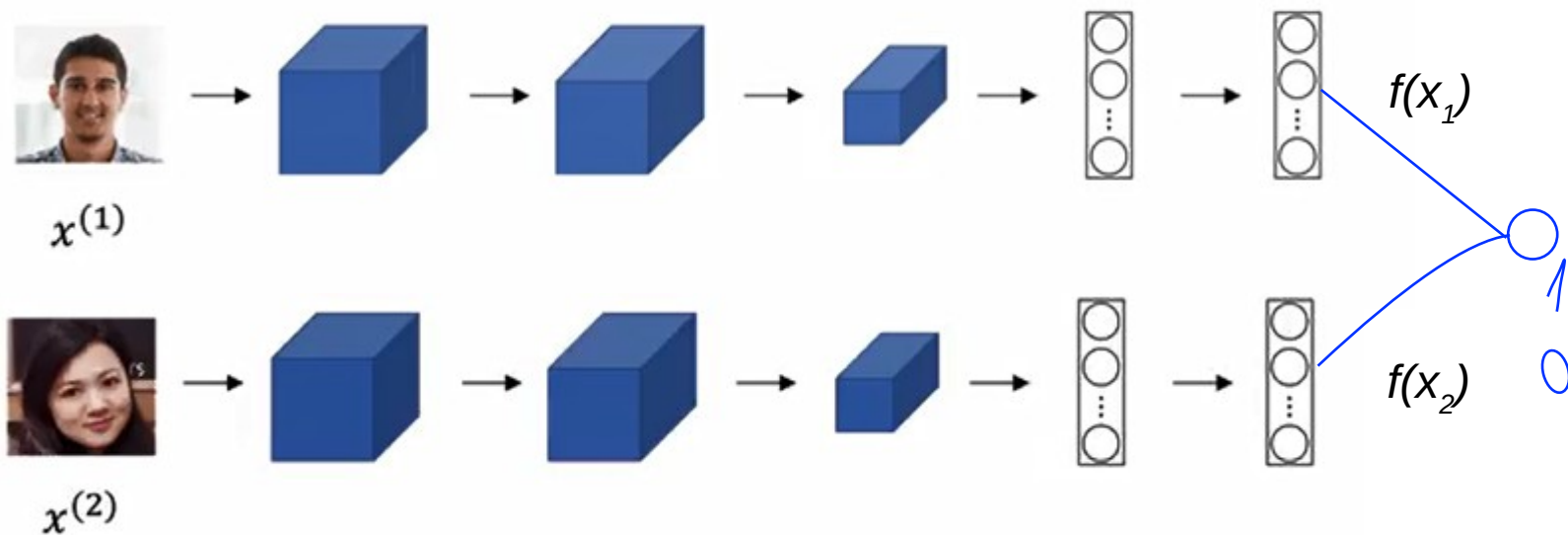
## Верификация

- Input — фотография, имя/ID
- Output — определить, принадлежит ли фото заявленному лицу

## Распознавание

- Имеется БД на K персон
- Подаётся на вход фото
- Выводим ID если фото принадлежит кому-либо из K персон, в противном случае — «не распознано»







# FACE RECOGNITION/VERIFICATION



Если  $x(i)$ ,  $x(j)$  принадлежат одному человеку, то  $\frac{1}{2}\|f(x^{(1)}) - f(x^{(2)})\|$  мало

Если  $x(i)$ ,  $x(j)$  принадлежат разным людям, то  $\frac{1}{2}\|f(x^{(1)}) - f(x^{(2)})\|$  велико

# FACE RECOGNITION/VERIFICATION

$x$		$y$
		1
		0
		0

# CLASSIFICATION AND DETECTION

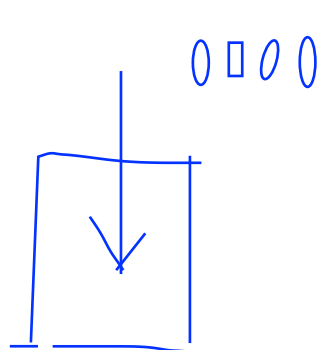
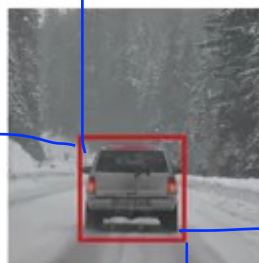


Image classification



car

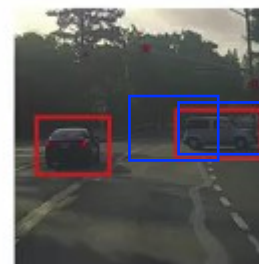
Classification with localization



car

0,5

Detection



Несколько объектов



## Классификация с локализацией



Softmax (4)

$b_x, b_y, b_h, b_w$   
(рамка)

1. машина
2. пешеход
3. мотоцикл
4. фон

# CLASSIFICATION AND DETECTION

$$y = [p_c, b_x, b_y, b_h, b_w, c_1, c_2, c_3]$$



Есть ли объект?



$[1, b_x, b_y, b_h, b_w, 0, 1, 0]$

$[0, ?, ?, ?, ?, ?, ?, ?]$

# LANDMARK DETECTION

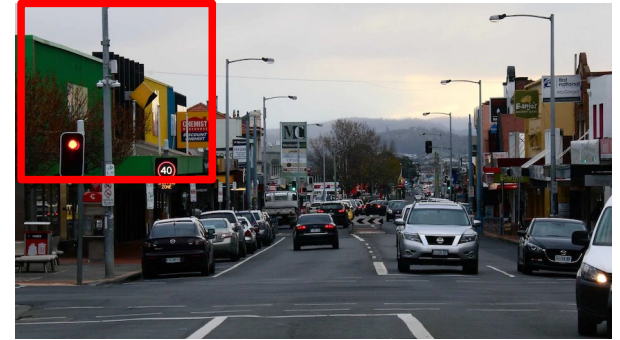
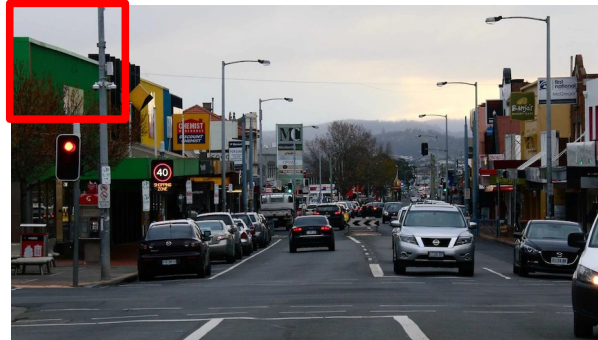


$$\begin{matrix} l_{1x}, l_{1y} \\ l_{2x}, l_{2y} \\ \dots \\ l_{64x}, l_{64y} \end{matrix}$$

$$y = [\text{face?}, l_{1x}, l_{1y}, l_{1x}, l_{1y}, \dots, l_{64x}, l_{64y}]$$



# OBJECT DETECTION



Sliding window

