

# Глубокие методы в компьютерном зрении

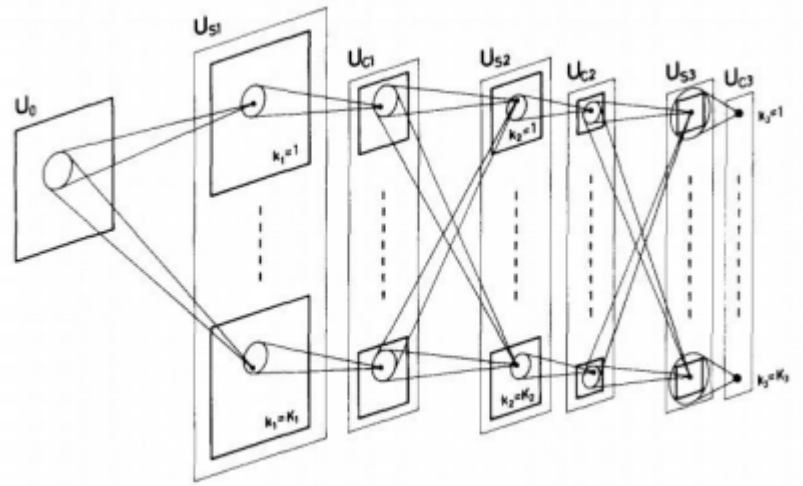
Александр Петюшко, к.ф.-м.н.

Илья Иванов, к.ф.-м.н.

# Нейронные сети: история вопроса

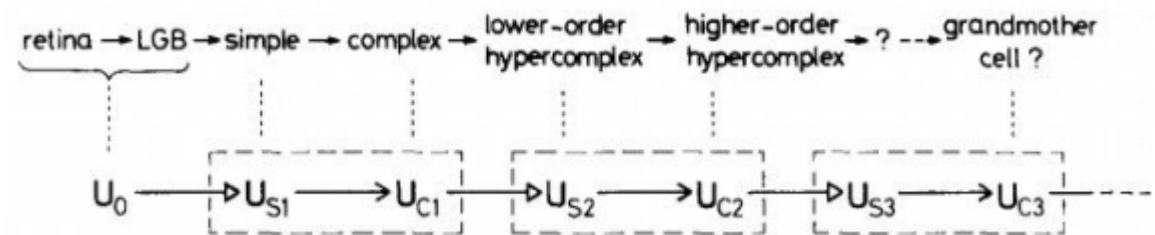
- 1943 - Маккалок и Питтс формализуют понятие нейрона
- 1949 – Хебб предлагает первый алгоритм обучения
- 1958 – Розенблатт изобретает однослойный персептрон и демонстрирует его способность решать задачу классификации
- 1974 – Галушкин и Вербос предлагают метод обратного распространения ошибки (впоследствии улучшен Хинтоном в 1986)
- 1975-1980 – Фукусима предлагает мощную модель когнитрона и неокогнитрона (прототип сверточных сетей)
- 1989 – Лекун представляет первую работу по объединению свёрточных нейросетей и метода обратного распространения ошибки
- 2012 – Крижевский и Хинтон публикуют первую успешную глубокую классификационную нейросеть **AlexNet**





# Неокогнитрон

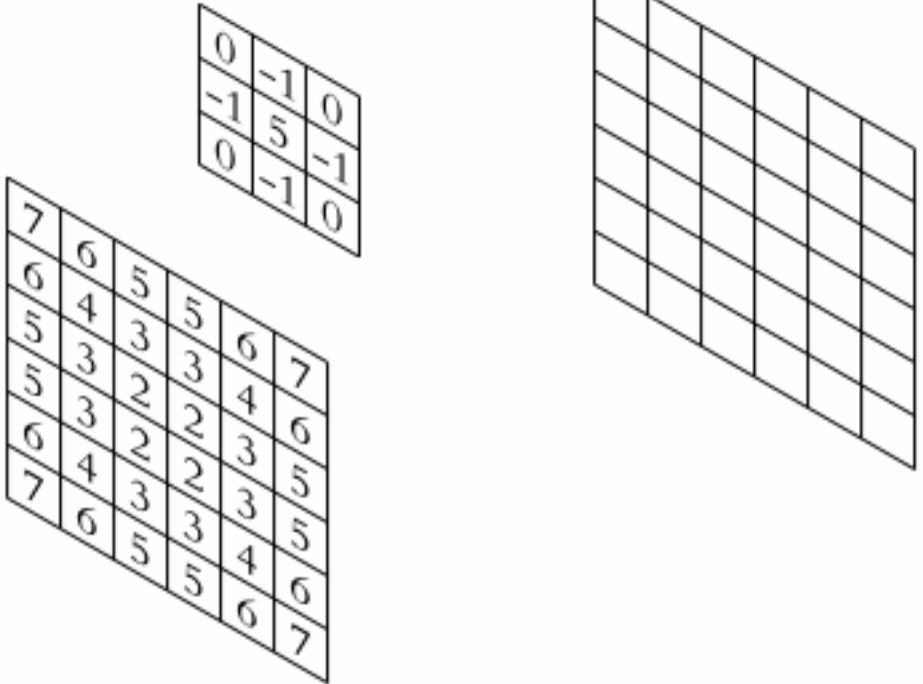
- Фукусима предложил практически современный метод построения архитектуры нейросетей, заимствованный им из модели первичной зрительной коры
- Два вида нейронов:
  - Простые (Simple), отвечающие за локальные признаки
  - Сложные (Complex), отвечающие за компенсацию искажения
  - Организованы в каскадную структуру **SCSCSC...**
  - В сверточной сети **S**=свертка, **C**=субдискретизация
- Главный минус: не было предложено метода обратного распространения ошибки для обучения



## Нейронные сети: предпосылки успеха

- *Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году*
- Основные причины начала новой эры в нейронных сетях
  - Появились большие наборы данных для обучения ( $\sim 10^6$  изображений)
  - Появились мощные видеокарты для обучения нейронных сетей
  - Открытость исследований ([arxiv.org](https://arxiv.org), [github.com](https://github.com))

# Свертка: основа компьютерного зрения



## Субдискретизация

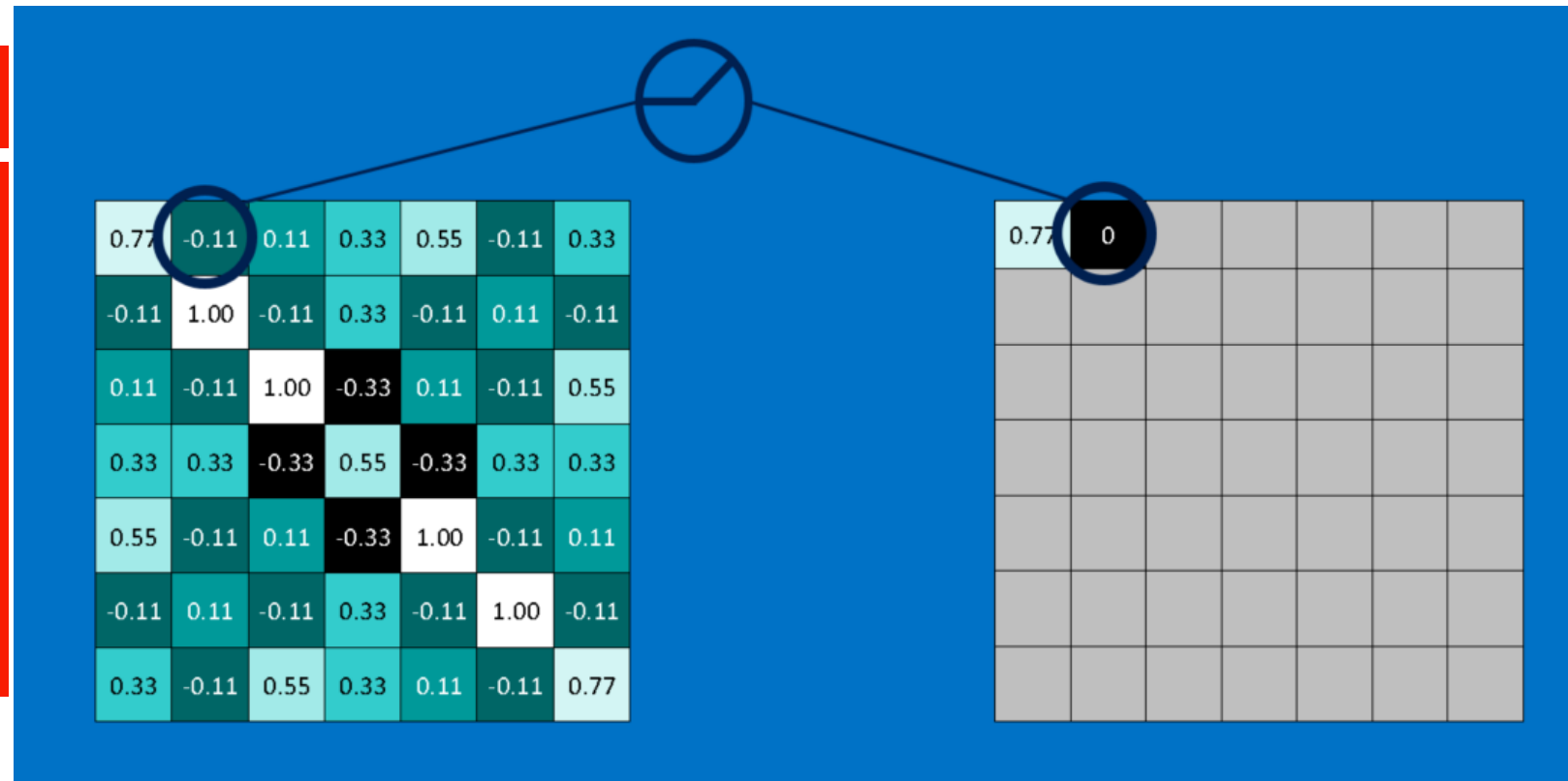
- Слой субдискретизации решает две проблемы:
  - Снижает пространственную размерность
  - Помогает не переобучаться

1	3	2	9
7	4	1	5
8	5	2	3
4	2	1	4

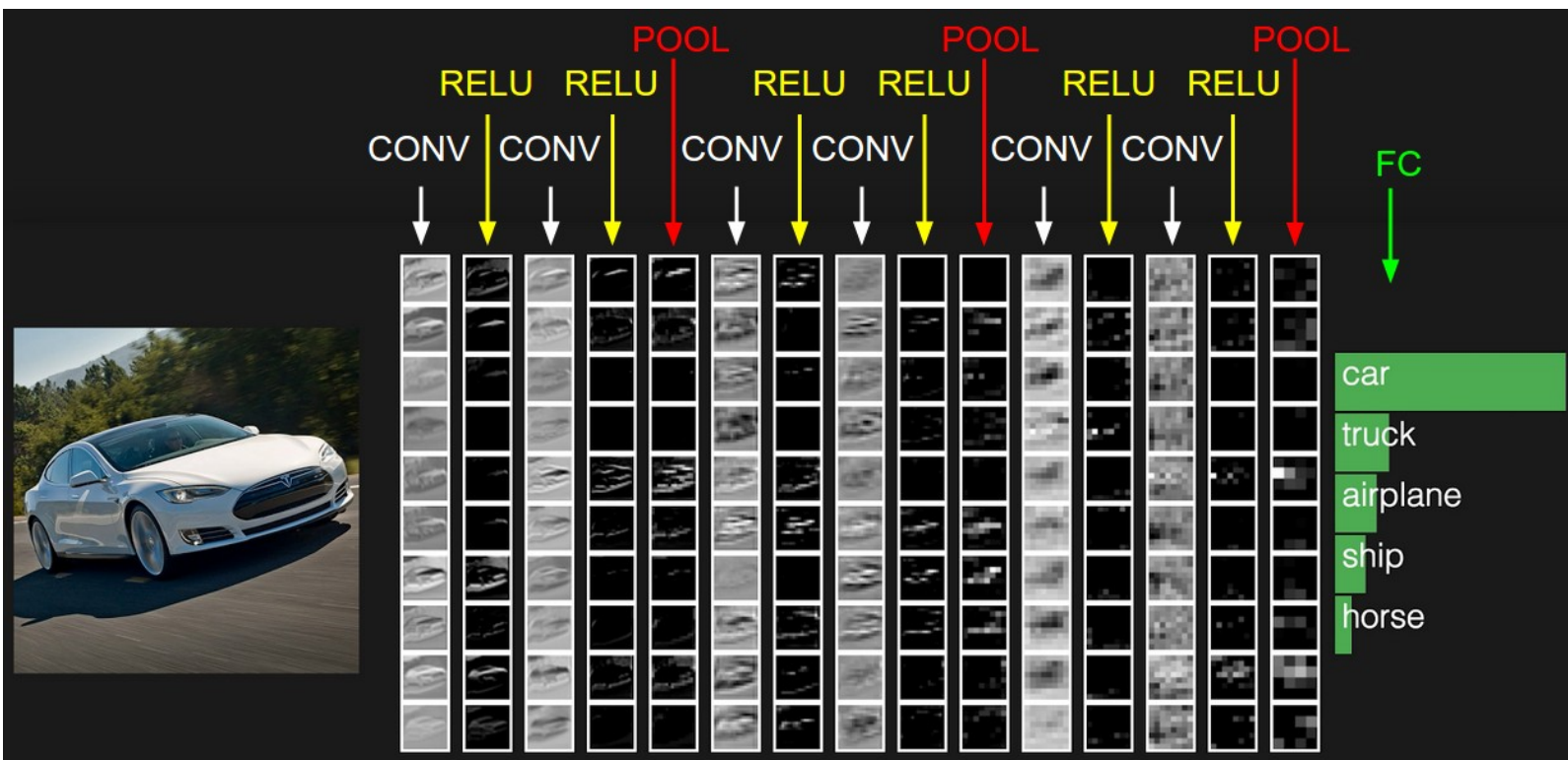
7	9
8	

$$\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x})$$

Активация:  
оставляем  
наиболее  
значимую  
информацию



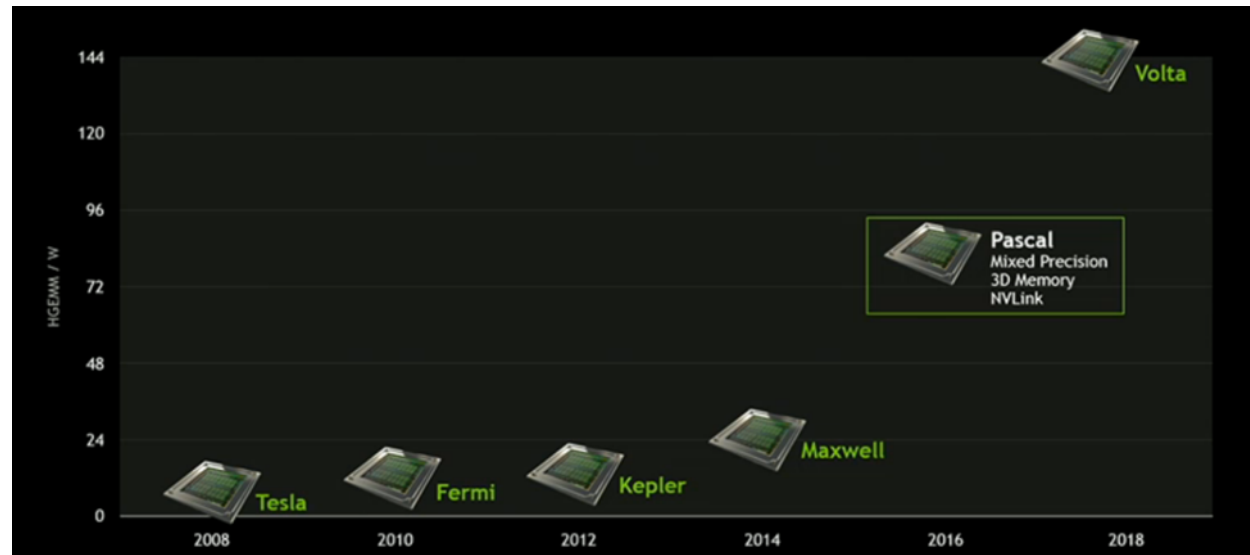
# Визуализация работы сверточной сети



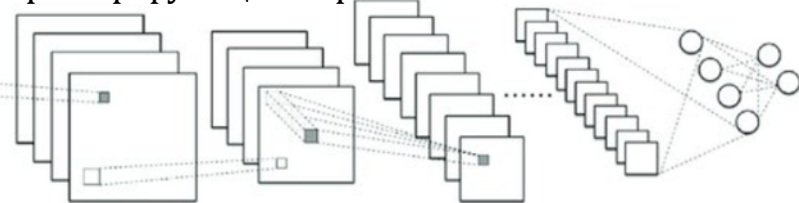


## Необходимое оборудование: NVIDIA- видеокарты

- На данный момент обучение нейросетей немыслимо без использования видеокарт (GPU, graphical processing unit) от NVIDIA
- Для обучения промышленных нейросетей (например, как Mask R-CNN) необходим объем памяти GPU от 12-16 ГБ
  - В силу этого подходящие карты на данный момент: Titan X, GTX 1080 Ti, P100, V100
  - Для любого проекта необходимо обеспечить доступ к данному оборудованию (например, удаленно через ssh)



Пример функционирования СНС



# Компьютерное зрение: классификация

Основные представители:

2012 – AlexNet

2014 – VGGNet

2014 – GoogleNet/Inception

2015 – ResNet

2017 – MobileNet

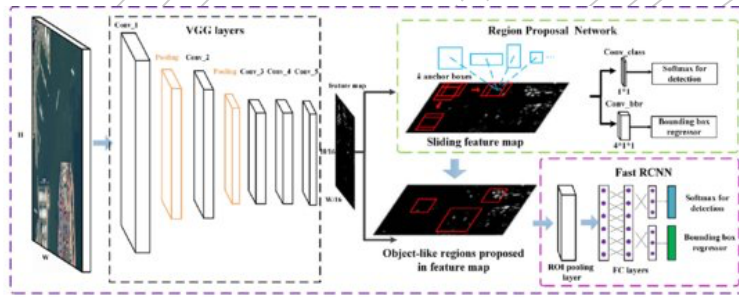
2017 – NASNet/PNASNet

- Задача: отнести входное изображение к одному из классов
  - По фотографии человека определить пол, возраст и т.д
  - По фотографии автомобиля определить марку и модель
- Самая простая задача компьютерного зрения
  - Решается с помощью сверточных нейросетей (СНС)
- Проблема: двусмысленность изображений
  - Не любое изображение можно описать одним классом (см. ниже)

Сено, лошадь или женщина?

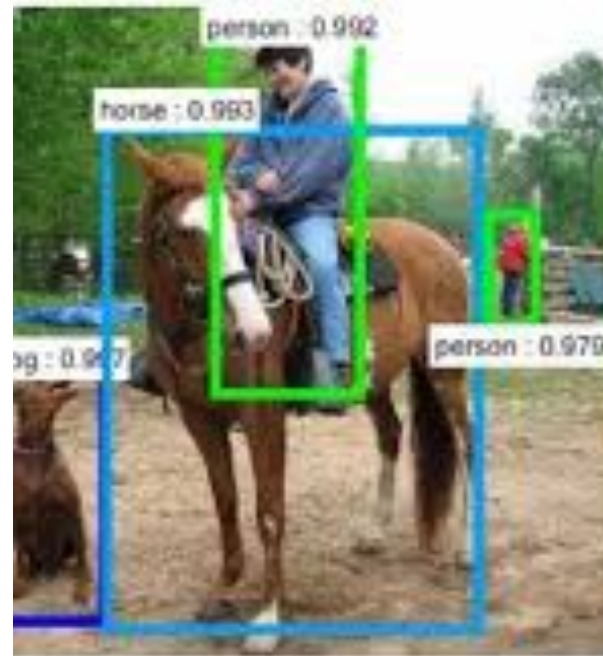


## Faster R-CNN



# Компьютерное зрение: детекция

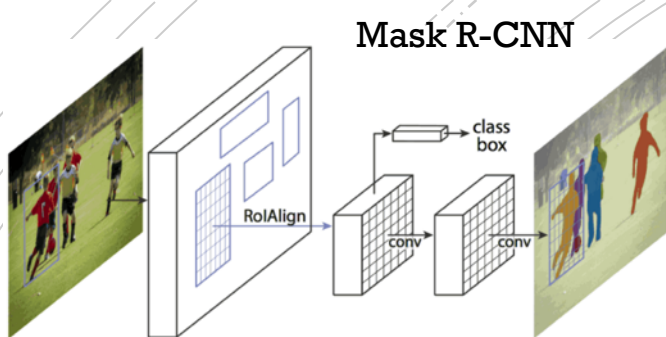
- Задача: по входному изображению требуется найти объекты заранее заданных классов и обвести их прямоугольником
- Сложнее, чем классификация, так как необходимо не только распознать класс, но и локализовать объект
- Существуют алгоритмы, которые решают эту задачу (в т.ч. в реальном времени)
  - Faster R-CNN, YOLO, SSD



Основные представители:

- 2013 - Overfeat
- 2014 - R-CNN
- 2015 - Faster R-CNN
- 2015 - YOLO
- 2015 - SSD
- 2016 - R-FCN





# Компьютерное зрение: сегментация

Основные представители:

2014 – FCN

2015 – U-Net

2015 – SegNet

2016 – RefineNet

2016 – PSPNet

2017 – Mask R-CNN

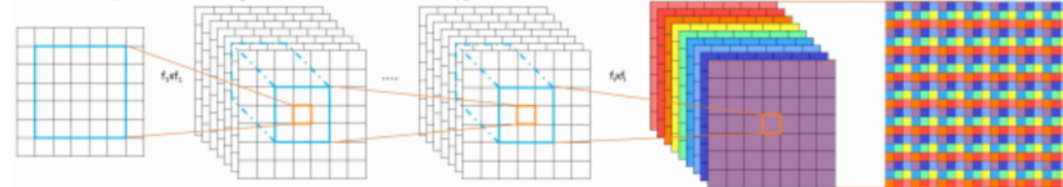
- Задача: по входному изображению найти объекты заранее заданных классов и выдать попиксельную маску объекта
- Сложнее, чем детекция, так как необходимо точно определять границы объекта (а не просто заключить в прямоугольник)
- На данный момент существуют алгоритмы-комбайны (решающие одновременно задачи распознавания, детекции и сегментации)
  - Пока не работают в реальном времени (Mask R-CNN)

Выход решения-комбайна



# Улучшение изображений: НейроСверх-разрешение

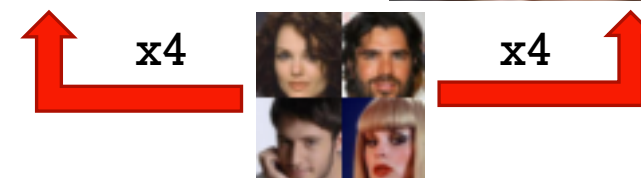
Метод сверх-разрешения	PSNR
Бикубическая интерполяция	24.32 dB
Бикубическая интерполяция + GIMP	24.69 dB
Бикубическая интерполяция + Photoshop	24.72 dB
НейроСверх-разрешение	27.20 dB



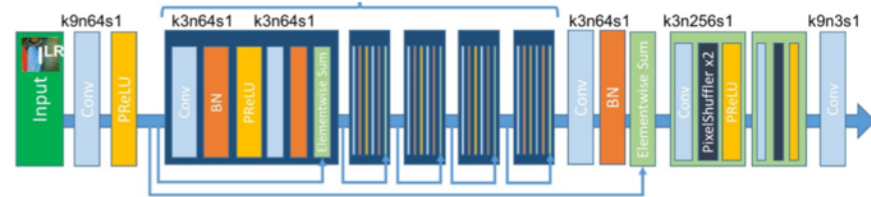
ESPCN, 2016

- Задача: увеличить разрешение фотографии / видео-кадра
  - Применяется нейронная сеть, заранее обученная на множестве картинок / видео (ESPCN)
  - Цель: победить классические алгоритмы

Бикубическая интерполяция: PSNR = 24.32    НейроСверх-разрешение: PSNR = 27.20



Малоразмерный вход



SRGAN, 2016

- Задача: улучшить качество дефектных изображений
  - Устранить смаз и размытие (**ShallowCNN**)
  - Удалить артефакты, являющиеся результатом попытки устранить дефекты защитным в камере алгоритмом (**SRGAN**)

Улучшение  
изображений:  
устранение  
дефектов

Исходное  
размытие



Классический  
**Wiener**



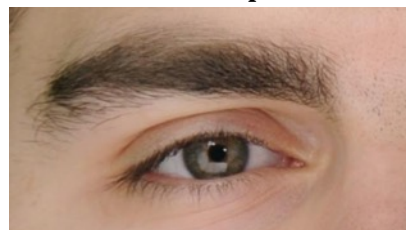
Коммерческое  
решение



Нейросетевое  
решение



«Сырое» изображение  
с камеры



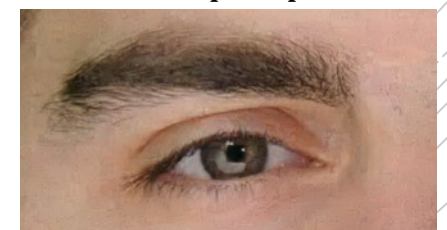
Устранение дефектов  
внутри камеры



Классическая  
нейросеть



Нейросеть с  
генератором





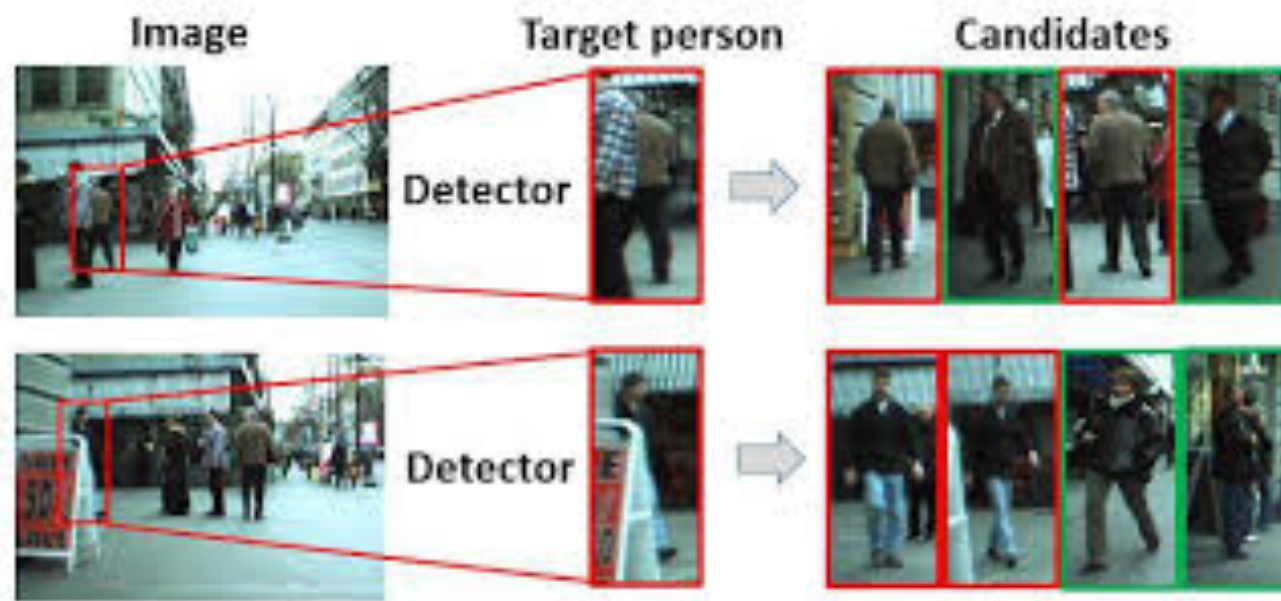
## Пример задачи: распознавание автомобильных номеров

- Задача: по видеопотоку детектировать транспортное средство и распознать госномер
- Как правило, такие системы состоят из следующих моделей:
  - Детектирование автомобиля
  - Распознавание марки и модели (опционально)
  - Трекинг автомобиля
  - Детектирование номера
  - Распознавание номера
  - Объединение результатов с разных кадров / поиск ключевых кадров



## Пример задачи: ре- идентификация пешеходов

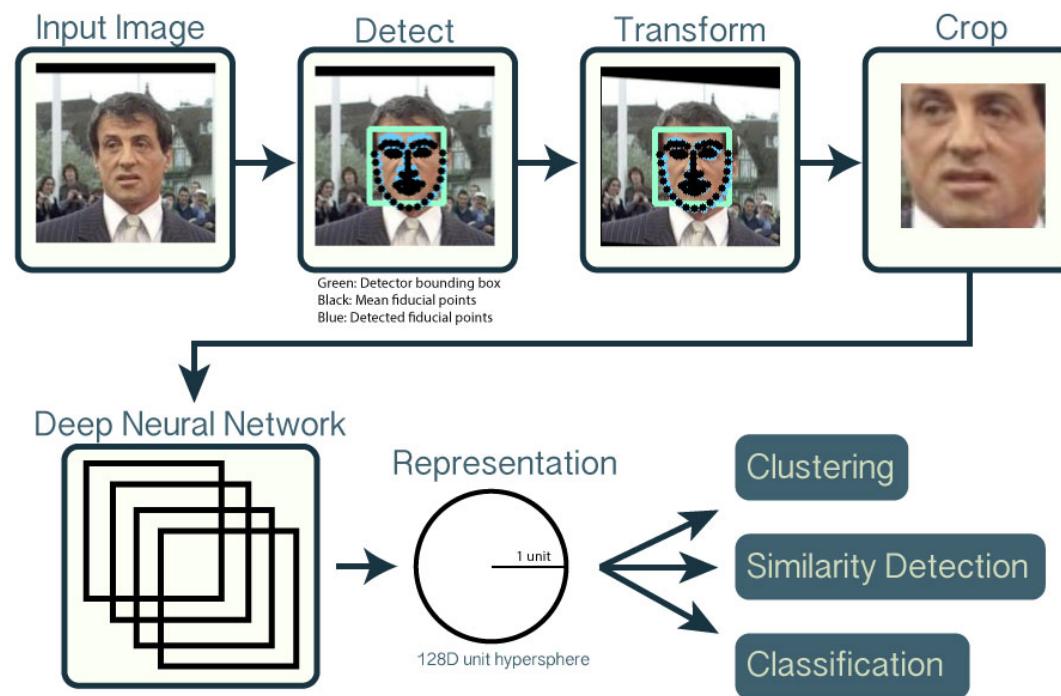
- Задача: идентификация пешеходов на камерах наблюдения
- Задача сводится к решения следующих подзадач
  - Предобработка
    - Детектирование пешеходов
    - Трекинг пешеходов
  - Извлечение признаков
  - Поиск в пространстве признаков





## Пример задачи: распознавание лиц

- Задача: по входному изображению лица распознать человека. Есть два подтипа задач:
  - Верификация – принадлежат ли две входные картинки одному человеку?
  - Идентификация – нахождение в базе данных человека, фотография которого подана на вход
- Решается с помощью СНС, но главная особенность – нахождение N-мерного вектора признаков, соответствующего данному человеку



A red speech bubble graphic with a white outline, containing the text "Спасибо!".

Спасибо!

И до скорых встреч Ж)