

# Архитектуры нейронных сетей для детекции объектов



---

**Шпилевский Яромир**

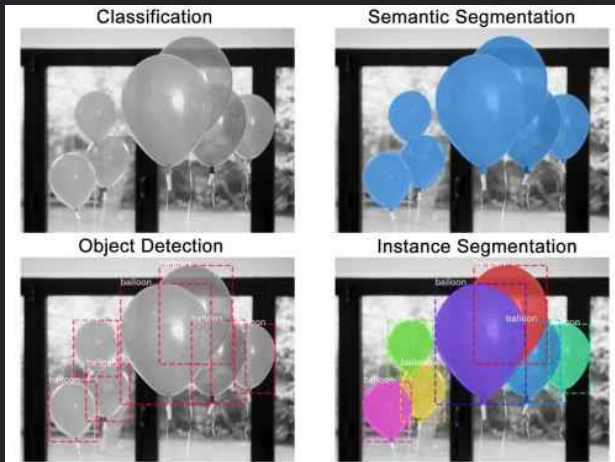
Ведущий разработчик First Line Software

## Agenda

- Задачи компьютерного зрения.
- Взаимосвязь задач.
- Эволюция Mask R-CNN.
  - R-CNN
  - Fast R-CNN
  - Faster R-CNN
  - Mask R-CNN
- YOLO.

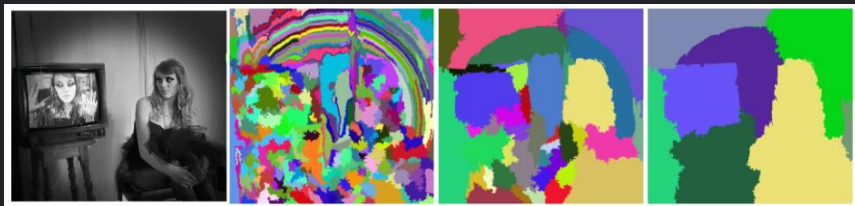
## Задачи компьютерного зрения

- Классификация – определяется принадлежность всего изображения к некоторому классу.
  - «Это фотография с шариками» – классификатор не знает ничего о границах объектов и о том, сколько объектов на изображении.
- Семантическая сегментация – отделить пиксели объектов заданного класса от пикселей других объектов и пикселей фона.
- Детекция объектов – определить рамки объектов.
- Сегментация экземпляра / объекта – внутри рамки отделить пиксели объекта от пикселей фона.

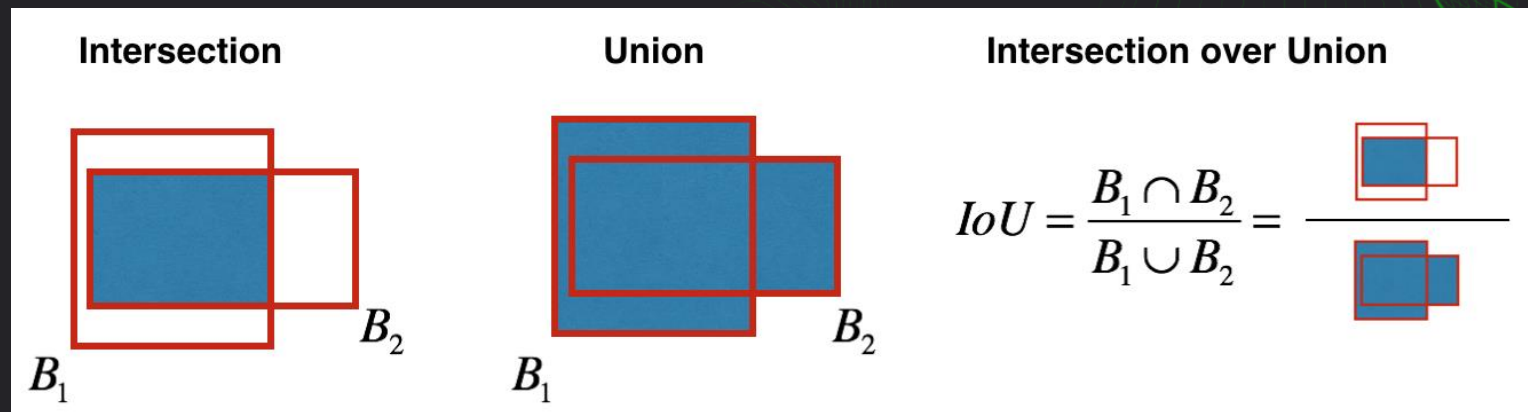


## Взаимосвязь задач

- Часто одна задача является вспомогательным этапом для другой.
- Между задачами нет строгой иерархии сложности.
- Например, сегментацию изображения можно сделать, не зная принадлежности сегментов к классу.



## Метрика качества Intersection over Union







**SKILLFACTORY**

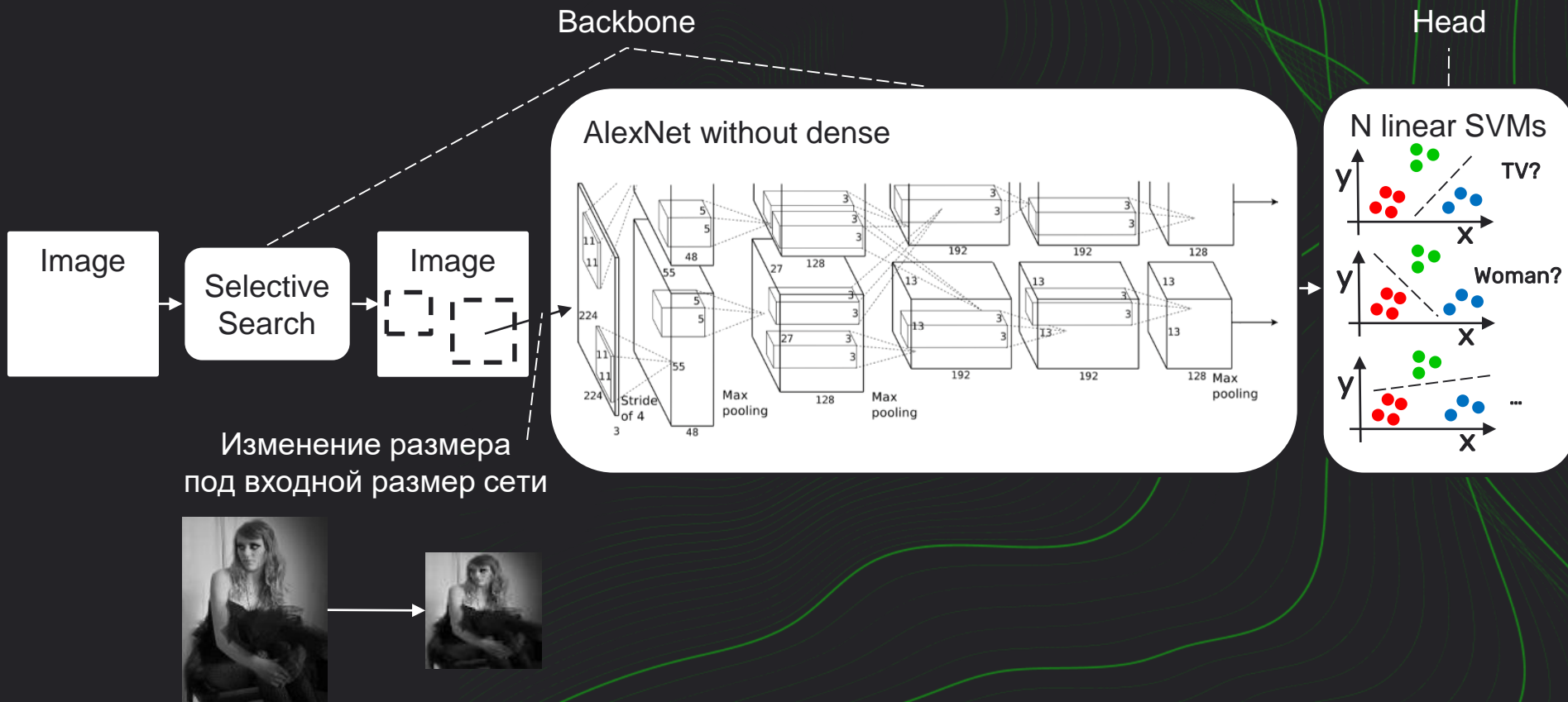
# Эволюция Mask-RCNN



## R-CNN. Общая идея

- Convolutional Neural Network (CNN) – свёрточная нейронная сеть, выделяет признаки.
- Regional CNN (R-CNN).
  - Выделяем регионы изображения.
  - Вычисляем свёртки (признаки).
  - По признакам пытаемся классифицировать объект в регионе.
- Как подбирать регионы?
- Можно использовать разные подходы, авторы выбрали Selective Search. [1]
- AlexNet для вычисления свёрток (верхнеуровневых признаков).
  - Можно использовать другие известные архитектуры.
- N линейных SVM классификаторов для бинарной классификации.
- N – количество классов, объекты которых должна уметь детектировать вся сеть.
- Каждый SVM классификатор отвечает за свой класс.
  - Это / не это.

## R-CNN. Архитектура





## R-CNN. Selective Search

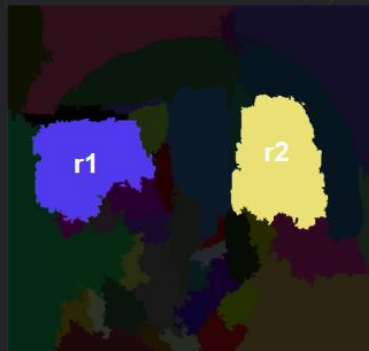
- Сгенерировать субсегментацию.
- Объединить сегменты на основе похожести.
- Вычислить границы.



- Метрика похожести?

## R-CNN. Selective Search. Метрика

- Подробно описана в [1].
- На основе схожести по цвету, текстуре, размеру и совместимости форм.



## R-CNN. Недостатки.

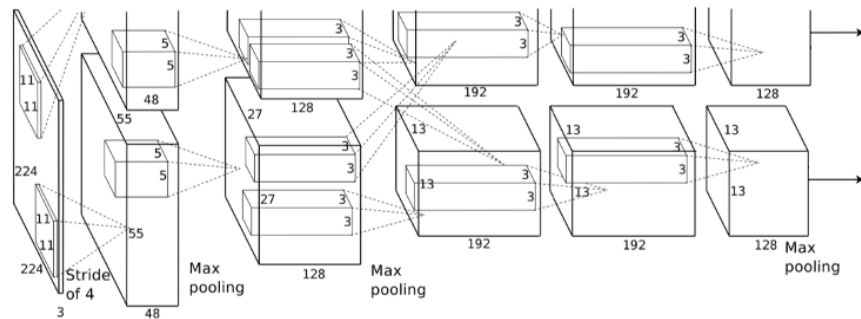
- Большая вычислительная сложность обучения.
  - Сеть должна независимо обучиться на каждый сегмент каждого изображения из обучающей выборки.
- Относительно большой объём хранения на диске.
  - N SVM классификаторов занимают довольно много места.
  - SVM применены как раз для уменьшения вычислительной сложности обучения, чтобы не обучать ещё и коэффициенты полносвязного слоя.

## Fast R-CNN. Основная идея.

- Обучать CNN сразу на всём изображении.
- CNN вычисляет карту признаков.
- Selective search предлагает регионы – Regions of Interest (RoI).
- Регионы применяются к карте признаков.
- Регионы разных размеров.
  - Фрагменты карты признаков тоже разных размеров.
  - Для приведения их к одинаковому размеру используется процедура RoI pooling [2]

## Fast R-CNN. Архитектура

AlexNet without dense



Image

Feature map



RoI  
pooling

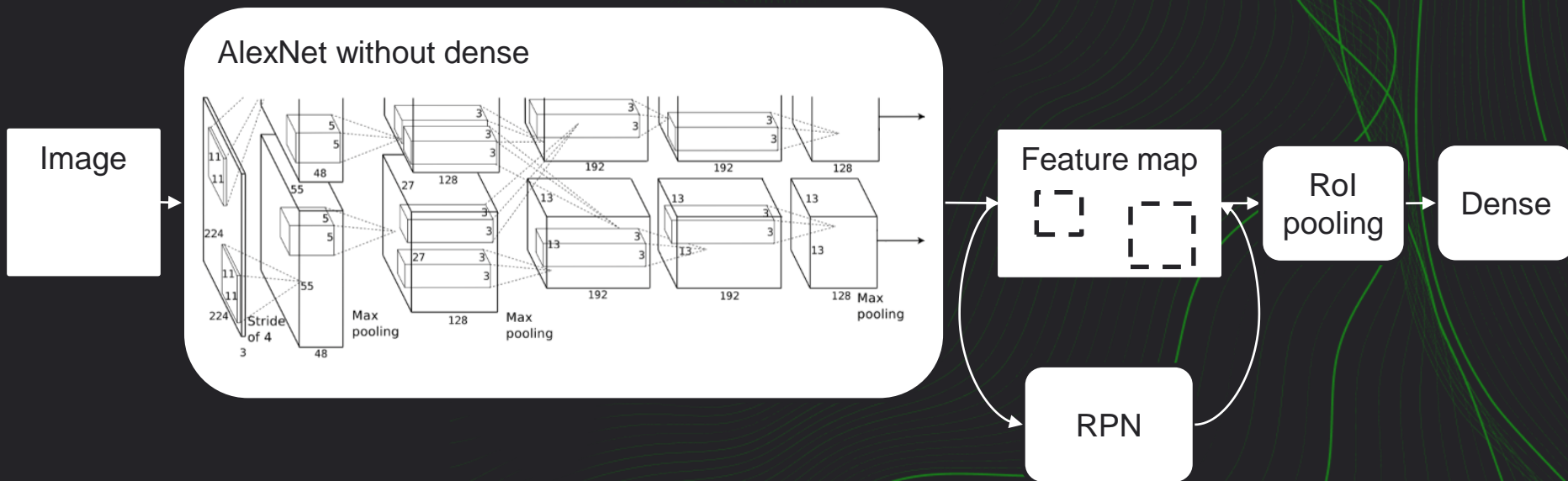
Dense

Selective  
Search



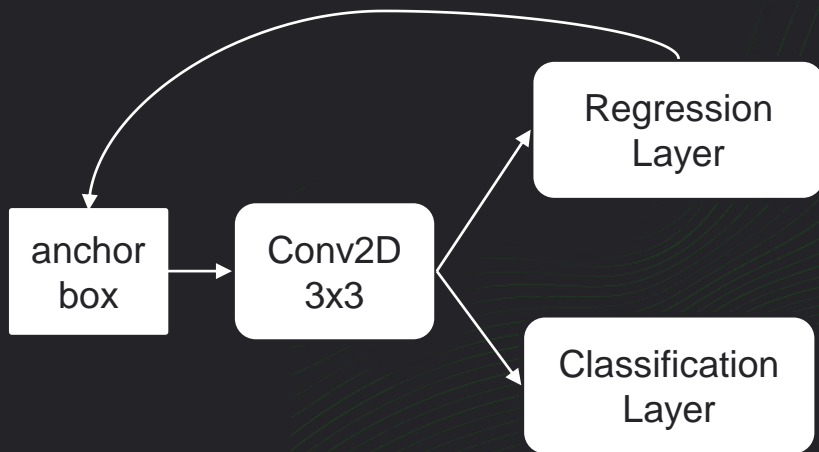
## **Faster R-CNN. Основная идея.**

- «Узкое место» Fast R-CNN – генерация регионов.
- Исследователи из Microsoft Research предложили решение.
- Аналогичное Fast R-CNN.
- Регионы вычислять так же по карте признаков.
- Как сделать?
  - Вспомогательная нейронная сеть – Region Proposal Network (RPN).
- Selective Search работает с пикселями.
- Region Proposal Network – с признаками.



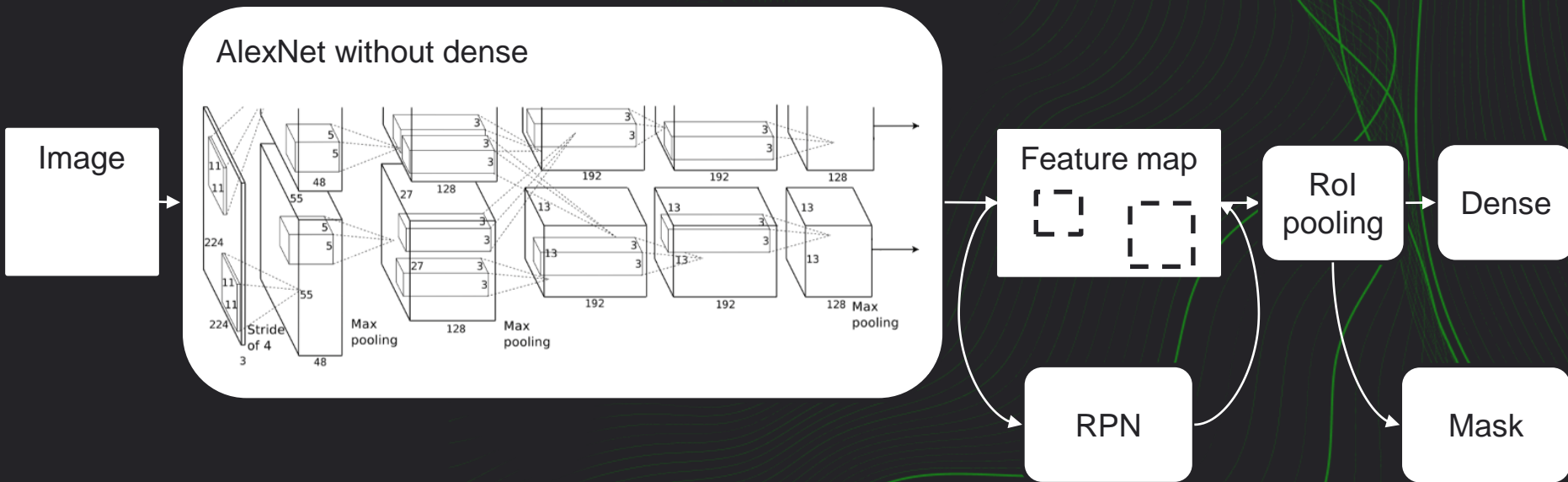
## Faster R-CNN. Region Proposal Network.

- k anchor boxes.
- CNN слой с окном 3x3.
- Регрессионный слой – настраивает anchor boxes.
- Слой классификации – определяет, есть ли в предложенном anchor boxes какой-нибудь объект или нет.



## Mask R-CNN

- Осталось только добавить параллельно вспомогательную нейронную сеть для наложения маски, чтобы получить сегментацию объектов.



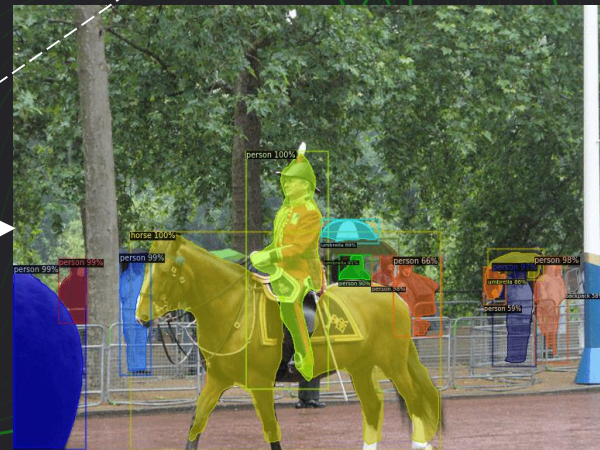
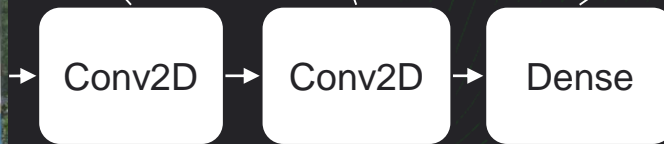
## Mask

Низкоуровневые  
признаки

Высокоуровневые  
признаки

Классификация  
2 класса:

Пиксель фона или пиксель объекта?





## Реализация. Detectron 2

- В Facebook's AI Research (FAIR) Detectron 2 одна из доступных моделей – Mask R-CNN. [3]





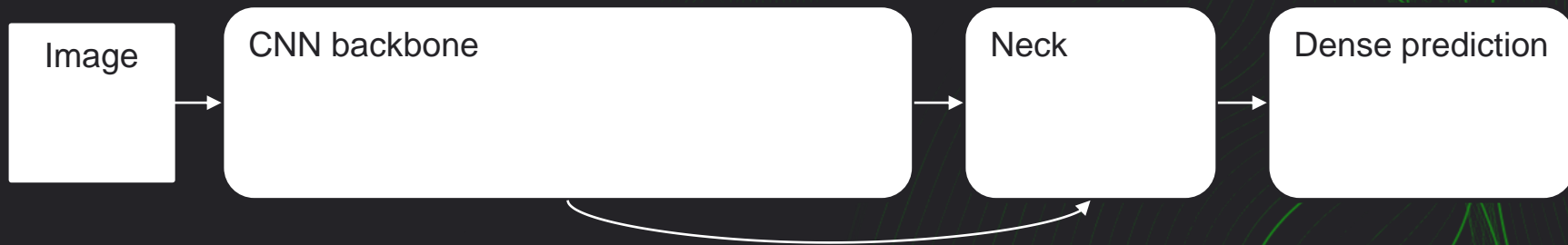
**SKILLFACTORY**

# YOLO

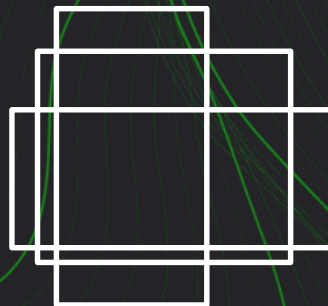


## YOLO

- You Only Look Once
  - Анализирует всё изображение за один проход.

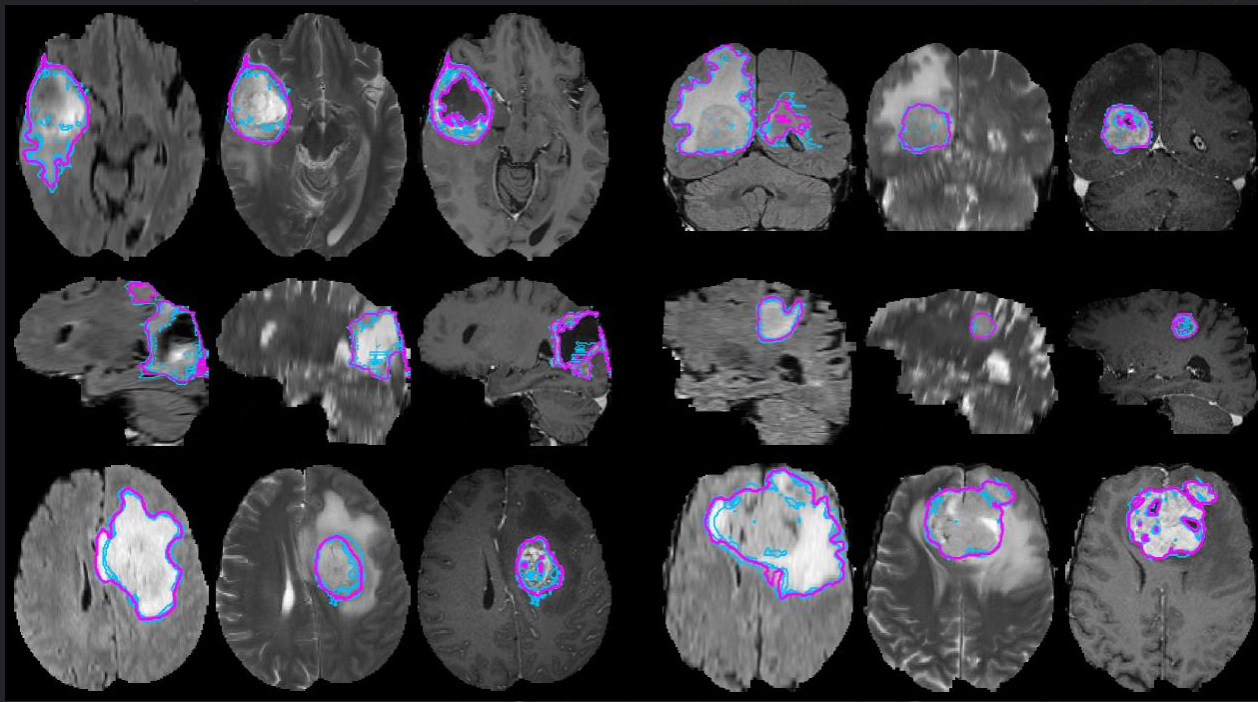


- CNN backbone для извлечения признаков, «хребет» сети.
- Neck («шея») для учёта выходов промежуточных слоёв.
- Dense prediction – для корректировки предсказанных bounding boxes и отбора лучшего кандидата.
  - Кандидаты генерируются с помощью кластеризации по методу К-средних.
  - На основе метрики IoU.



## Пример использования. Детекция опухолей в мозге

- BraTS dataset. [\[4\]](#)
- Пополняется каждый год.
- Можете попробовать свои силы. 😊





## Пример использования. COVID-19

- Можно попробовать сегментировать пневмонию, вызванную вирусом SARS-CoV-2.
  - IEEE 8023 COVID chest X-ray dataset. [\[5\]](#)
- Или хотя бы классифицировать пневмонию, вызванную вирусом SARS-CoV-2, от других пневмоний.
  - COVID-19 X-ray image dataset. [\[6\]](#)



## Ссылки

1. Selective Search for Object Recognition ([http://vision.stanford.edu/teaching/cs231b\\_spring1415/slides/ssearch\\_schuyler.pdf](http://vision.stanford.edu/teaching/cs231b_spring1415/slides/ssearch_schuyler.pdf))
2. Region of Interest Pooling (<https://deepsense.ai/region-of-interest-pooling-explained/>)
3. Facebook's AI Research (FAIR) Detectron 2 (<https://github.com/facebookresearch/detectron2>)
4. Brain Tumor Segmentation (<https://paperswithcode.com/task/brain-tumor-segmentation>)
5. IEEE 8023 COVID chest X-ray dataset (<https://github.com/ieee8023/covid-chestxray-dataset>)
6. COVID-19 X-ray image dataset (<https://www.kaggle.com/c/stat946winter2021/overview>)

## Резюме

- Рассмотрены различные задачи компьютерного зрения и их взаимосвязь.
- Рассмотрена эволюция Mask R-CNN.
  - R-CNN
  - Fast R-CNN
  - Faster R-CNN
  - Mask R-CNN

## Вопросы для самоконтроля

- Какие задачи компьютерного зрения вы можете назвать?
- Как они взаимосвязаны?
- Какие архитектуры нейронных сетей вы можете назвать?
- В чём особенность каждой?

# Спасибо!

