

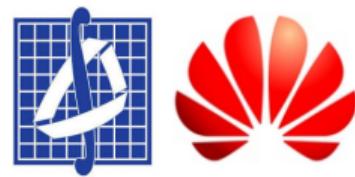
Введение в искусственный интеллект. Современное компьютерное зрение

Лекция 6. Методы семантической и объектно-чувствительной сегментации

Бабин Д.Н., Иванов И.Е., Петюшко А.А.

кафедра Математической Теории Интеллектуальных Систем

12 ноября 2019

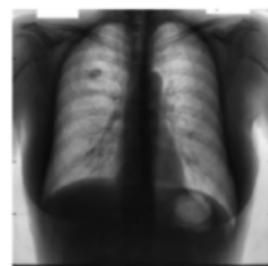


План лекции

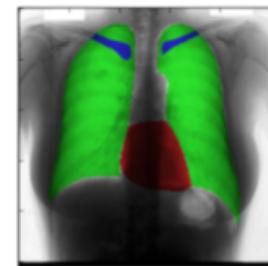
- ① Постановка задачи
- ② Метрики качества
- ③ Методы семантической сегментации
 - ① FCN
 - ② SegNet
 - ③ DeconvNet
 - ④ U-net
 - ⑤ DeepLab
- ④ Методы объектно-чувствительной сегментации
 - ① Mask R-CNN
 - ② Pose2Seg

Сегментация

- На предыдущих лекциях уже было рассказано о задачах классификации и обнаружения объектов
- Локализация объекта в виде содержащего его прямоугольника бывает не всегда достаточна для решения практических задач
- Например, при обнаружении образований на медицинских снимках очень важен их точный размер
- Поэтому возникает необходимость делать попиксельную маску объектов, то есть сегментацию
- Задача поиска попиксельной маски и есть задача сегментации



Input Image



Segmented Image

1

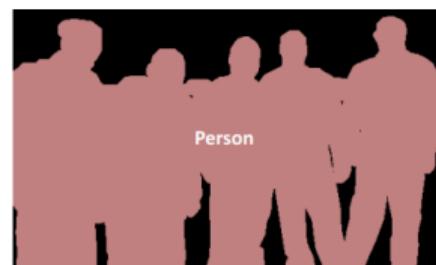
¹image from <https://arxiv.org/abs/1701.08816>

Виды сегментации в компьютерном зрении

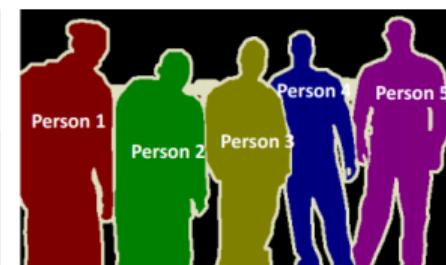
- Семантическая сегментация — классификация каждого пикселя изображения
- При семантической сегментации разные объекты одного класса попадают в одну маску
- Если есть необходимость различать разные объекты одного класса, то это задача называется объектно-чувствительная сегментация



Object Detection



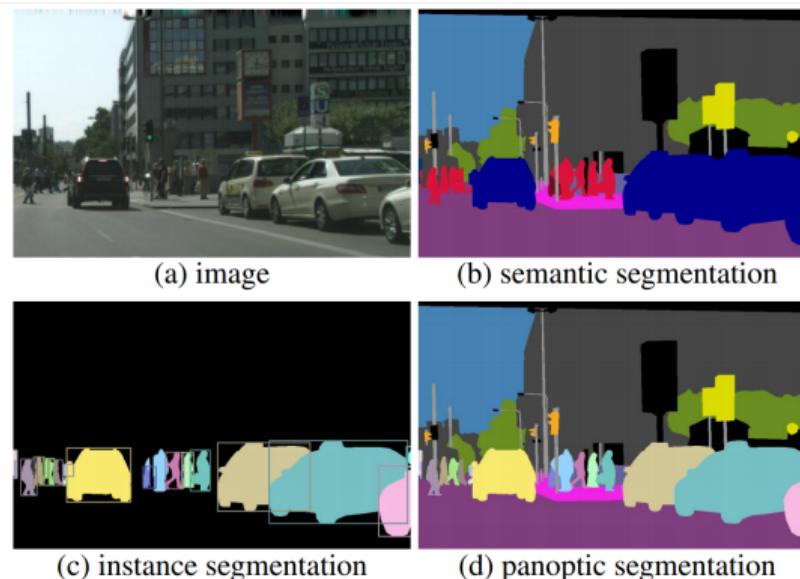
Semantic Segmentation



Instance Segmentation

Паноптик сегментация²

- В 2018 году была представлена новая задача, объединяющая семантическую и объектно-чувствительную сегментацию



²<https://arxiv.org/abs/1801.00868>

Зачем нужна сегментация: приложения

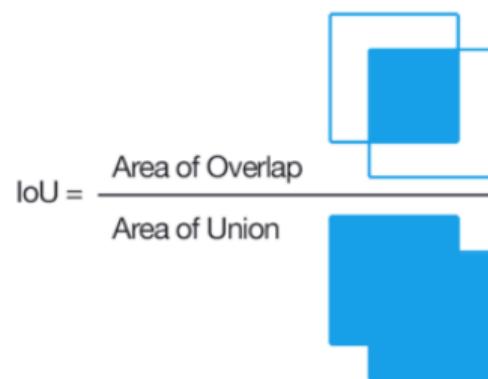
- Медицина: анализ различных снимков
- Анализ снимков из космоса: сегментация домов, кораблей и др. объектов
- Автомобили без водителя: необходимо точно оценивать есть ли поблизости другие участники движения
- Производство: анализ качества продукции, поиск дефектов
- Индустрия развлечений: различные фильтры для социальных сетей

Метрики качества семантической сегментации: Accuracy, mIoU

- Точность (Accuracy) — процент правильно классифицированных пикселей

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

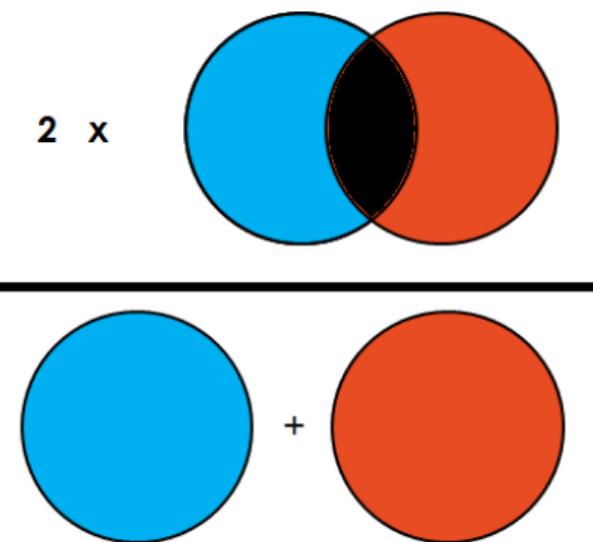
- Для масок произвольной формы мы можем аналогичным образом определить Intersection over Union (IoU, Jaccard Index)
- Для оценки качества масок можно использовать mIoU (mean Intersection over Union) — среднее значение IoU по всем маскам



Метрики качества семантической сегментации: Dice coefficient

- Индекс Дайса (Dice coefficient)

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN}$$

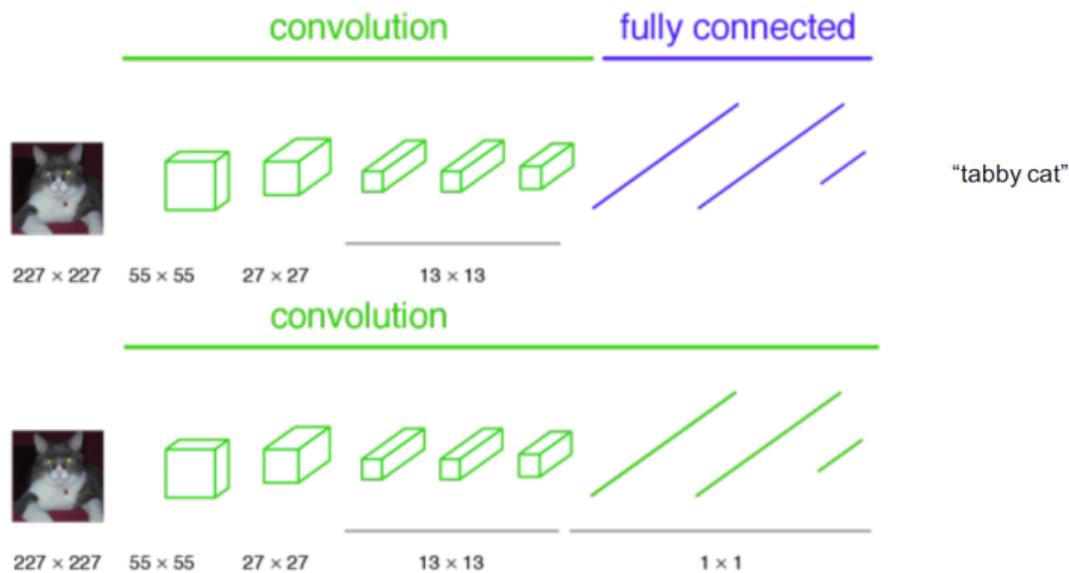


Метрики качества объектно-чувствительной сегментации: Average Precision

- Так как объектно-чувствительная сегментация является прямым обобщением задачи обнаружения, то разумно адаптировать метрики обнаружения для этой задачи
- Для оценки качества работы сегментационный модели можно использовать Average Precision (AP) при фиксированном пороге для IoU
- Обычно считают AP при различных порогах (например от 0.5 до 0.95 с шагом 0.05), а потом усредняют

Идея

Адаптировать классификационную сеть для задачи сегментации

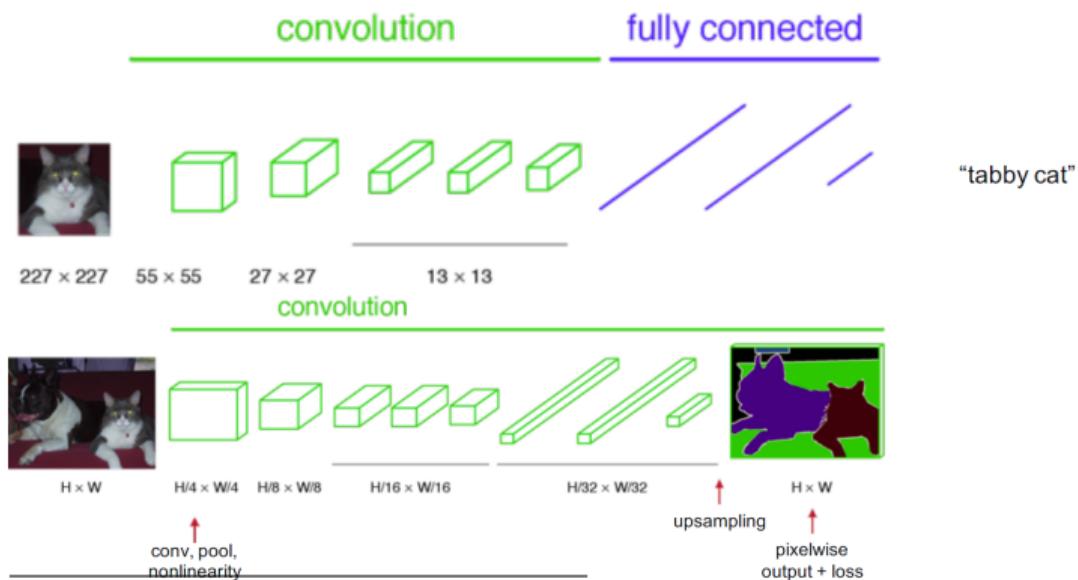


³Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038



Идея

Адаптировать классификационную сеть для задачи сегментации



³Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038



Upsampling

В классификационных сетях происходит постепенное уменьшение пространственной размерности. Поэтому чтобы на выходе получить маски такого же размера как и вход, необходима процедура увеличения пространственной размерности признаков.

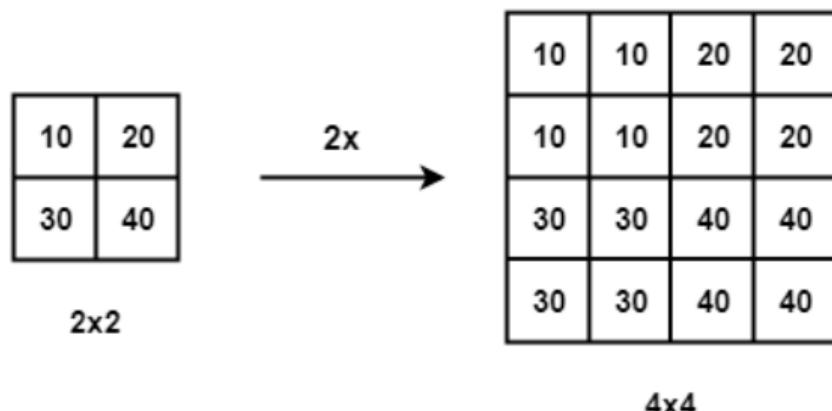
Стандартные подходы следующие:

- Билинейная (или любая другая) интерполяция
- Транспонированная свёртка
- Субпиксельный слой



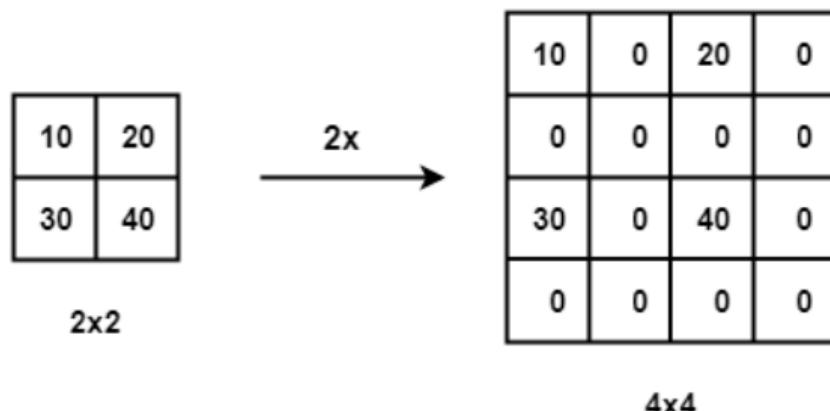
Методы восстановления изображения (upsampling): Интерполяция

- Метод ближайшего соседа (nearest neighbor)
- Unpooling "Bed of nails"
- Max Unpooling
- Билинейная интерполяция



Методы восстановления изображения (upsampling): Интерполяция

- Метод ближайшего соседа (nearest neighbor)
- Unpooling "Bed of nails"
- Max Unpooling
- Билинейная интерполяция



Методы восстановления изображения (upsampling): Интерполяция

- Метод ближайшего соседа (nearest neighbor)
- Unpooling "Bed of nails"
- **Max Unpooling**
- Билинейная интерполяция

Max Pooling

Remember which element was max!

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8



Max Unpooling

Use positions from pooling layer

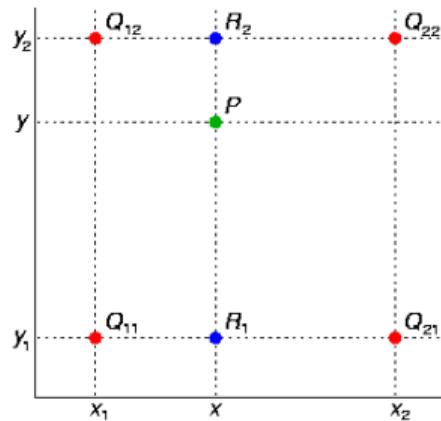
1	2
3	4

0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4



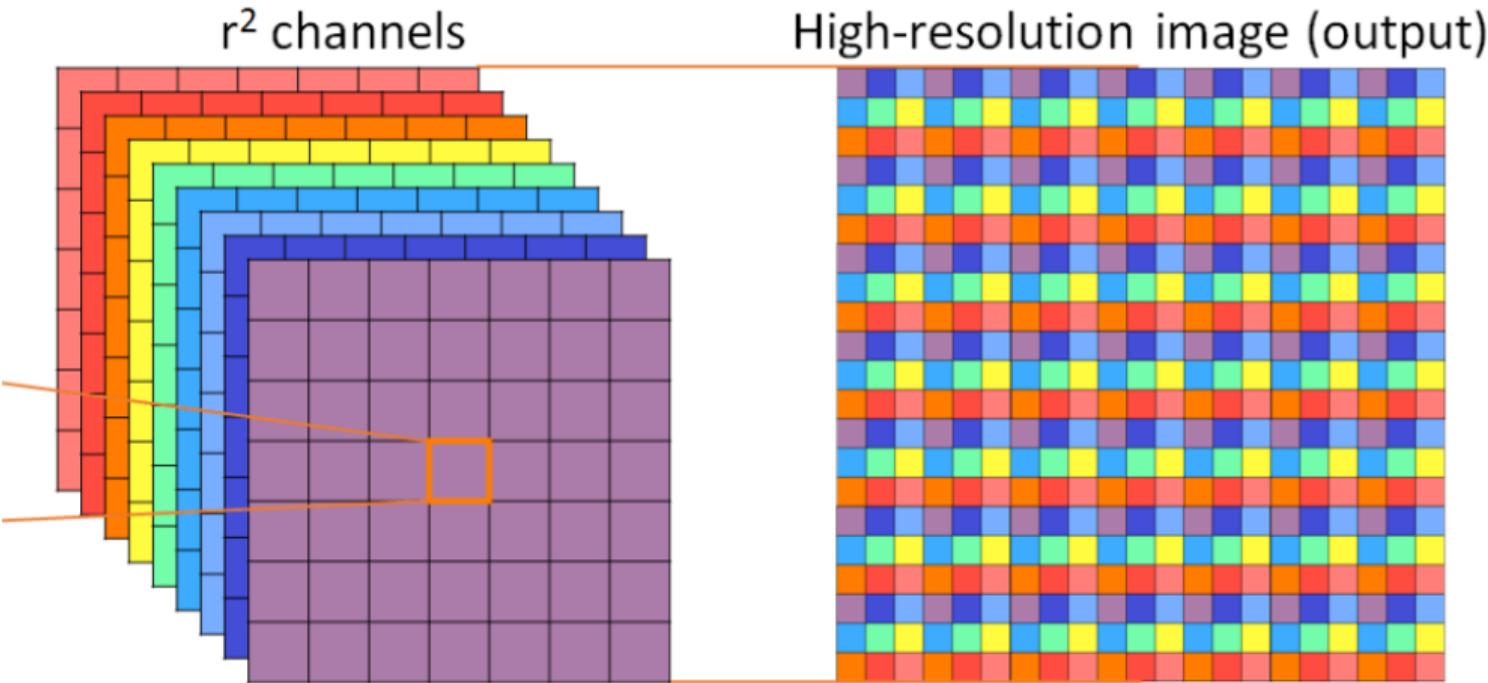
Методы восстановления изображения (upsampling): Интерполяция

- Метод ближайшего соседа (nearest neighbor)
- Unpooling "Bed of nails"
- Max Unpooling
- Билинейная интерполяция ⁴



⁴https://en.wikipedia.org/wiki/Bilinear_interpolation

Методы восстановления изображения (upsampling): Субпиксельный слой⁵

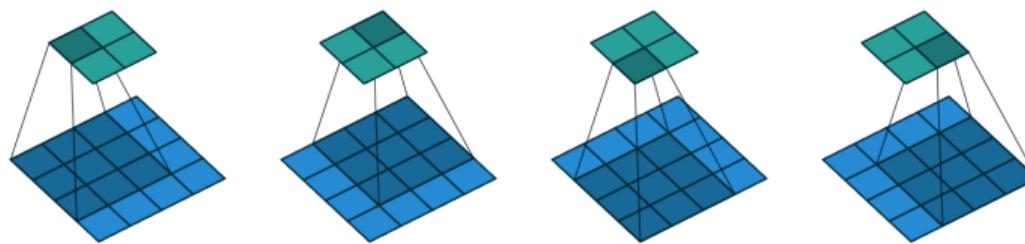


⁵<https://arxiv.org/abs/1609.05158>

Методы восстановления изображения (upsampling): Транспонированная свёртка⁶

Матричное представление свёртки

Свёртка является частным случаем полно связного слоя и может быть представлена в виде умножения на матрицу.



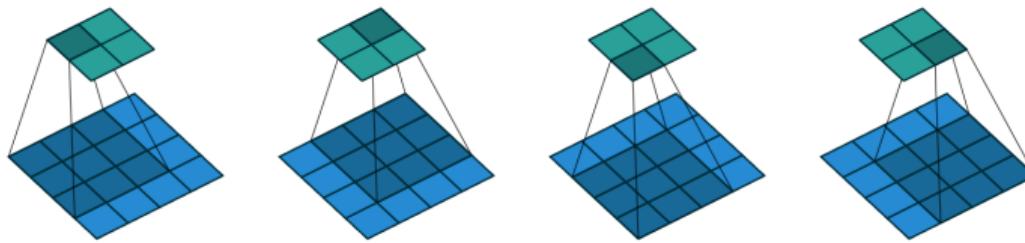
$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

⁶Отличная статья по свёрточной арифметике <https://arxiv.org/pdf/1603.07285.pdf>

Методы восстановления изображения (upsampling): Транспонированная свёртка

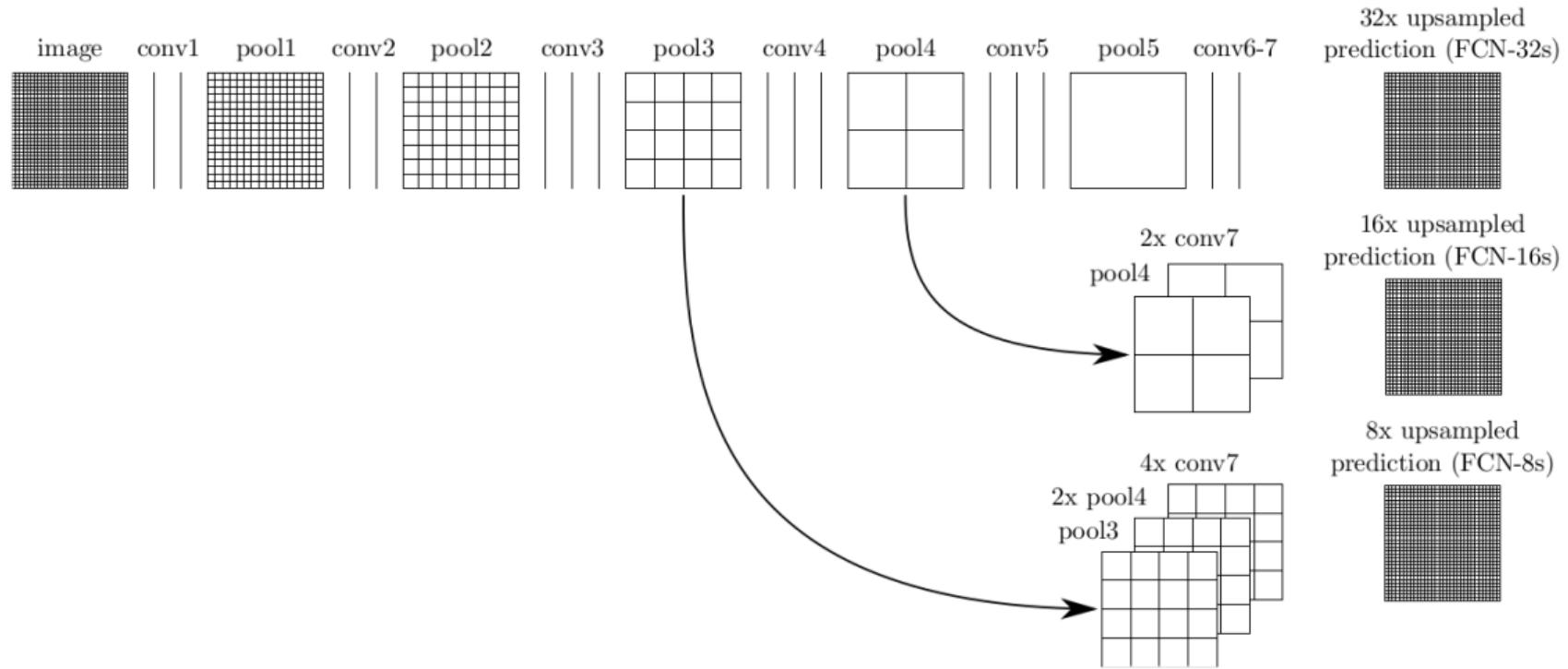
Идея транспонированной свёртки

Один и тот же набор параметров задаёт матрицу весов W и W^T . Применение матрицы W^T ведет к увеличению пространственной размерности и называется транспонированной свёрткой



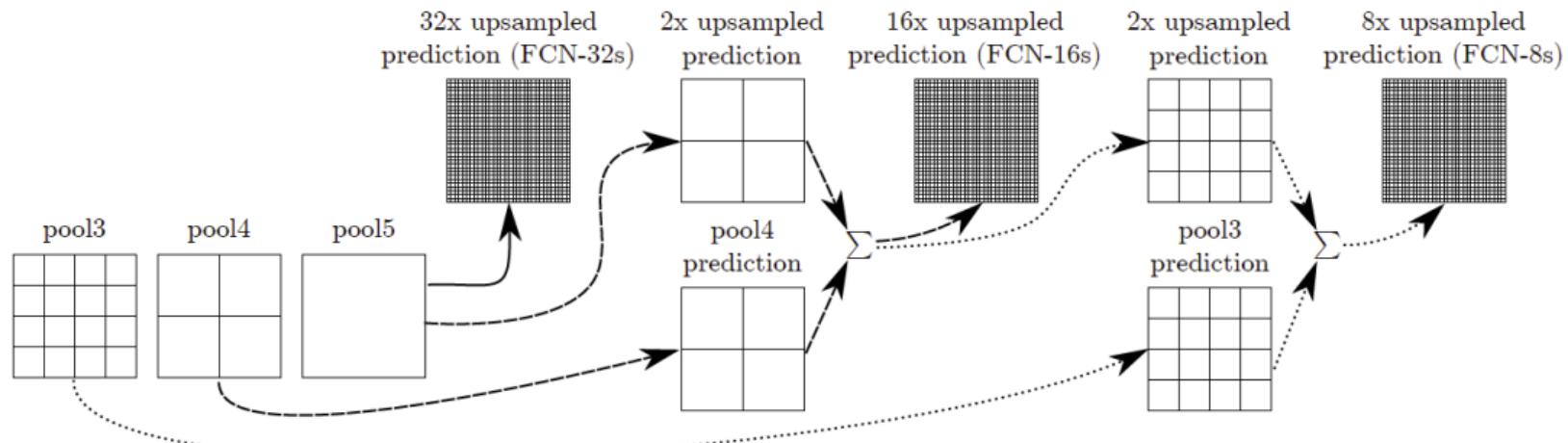
$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

FCN upsampling: Транспонированная свёртка⁷

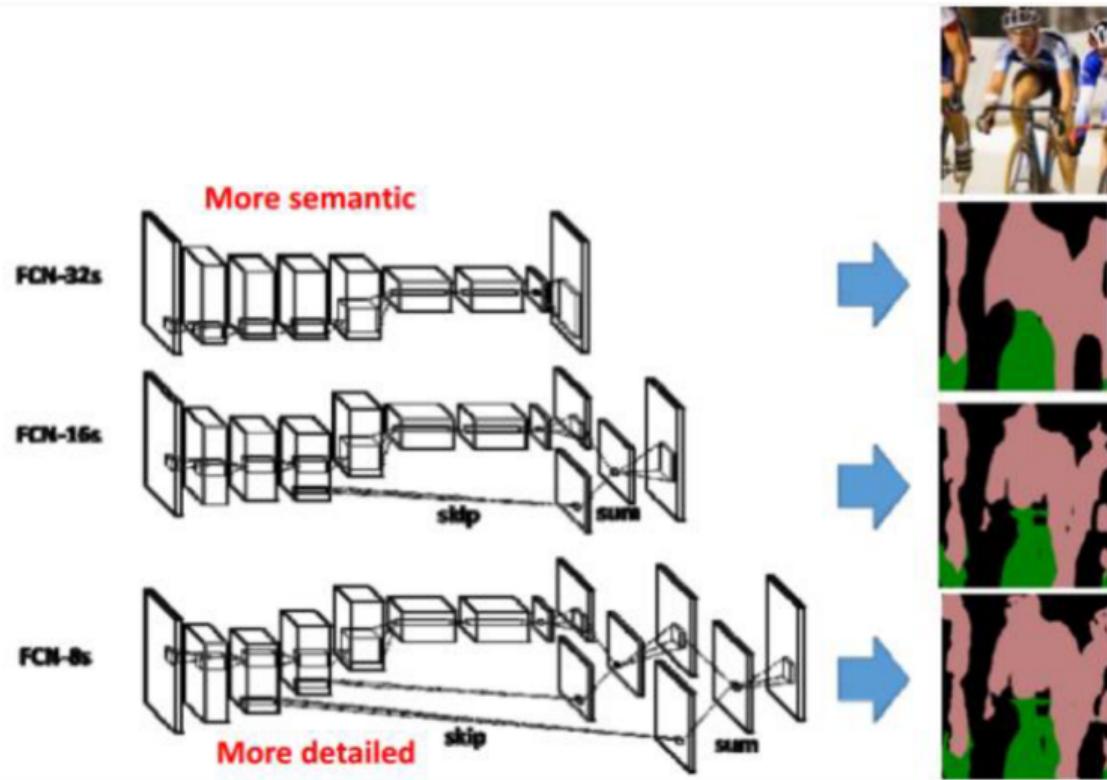


⁷http://deeplearning.net/tutorial/fcn_2D_segm.html

FCN upsampling: Транспонированная свёртка



Эффект прокидывания связей



Предобучение

Так как основная идея получить сегментационную сетку из классификационной, то кажется разумным использовать предобученную на ImageNet модель.

Функция потерь

Для каждого пикселя решается классификационная задача, поэтому обычно используется средняя кросс-энтропия для всего изображения.



Результаты работы FCN

На момент выхода статьи модель достигала SOTA результатов на нескольких датасетах

FCN-32s



FCN-16s



FCN-8s

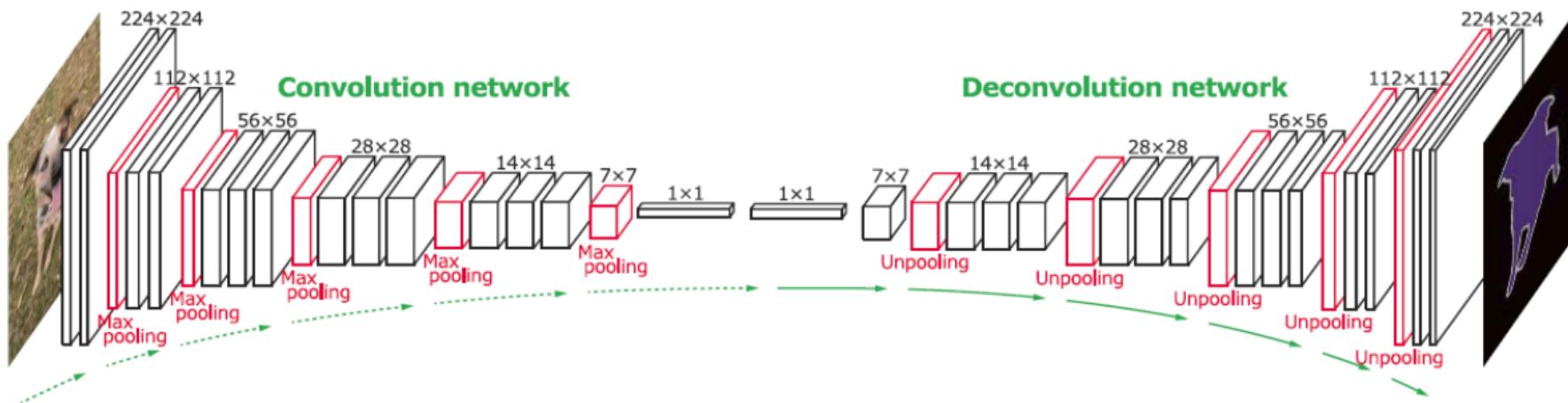


Ground truth



DeconvNet⁸

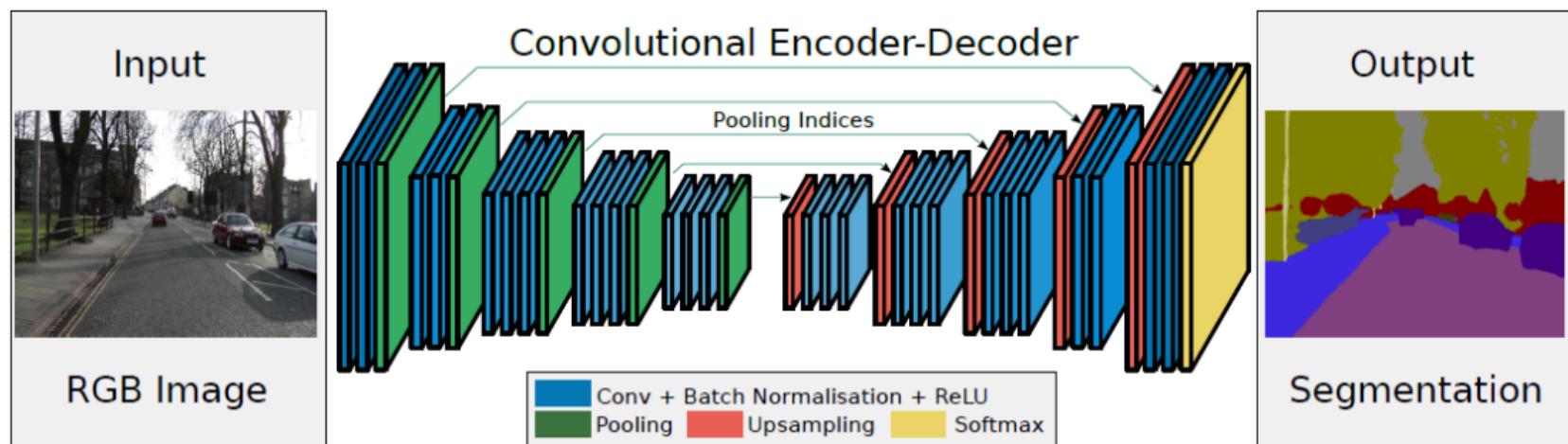
- Для восстановления изображения применяется max unpooling с запоминанием позиции
- Энкодер-декодер архитектура



⁸<https://arxiv.org/pdf/1505.04366.pdf>

SegNet⁹

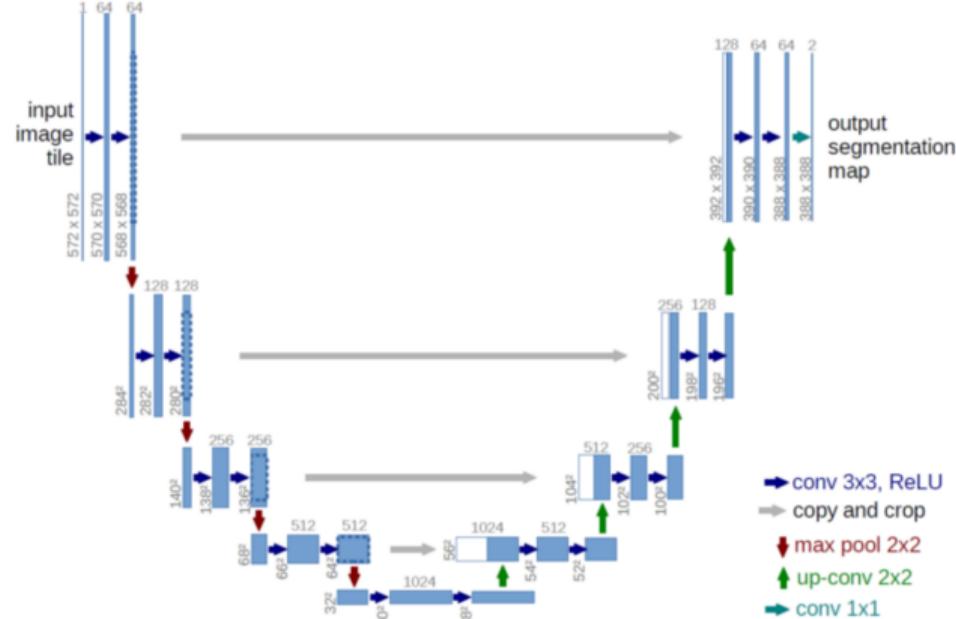
- Энкодер-декодер архитектура
- Энкодер состоит из 13 сверточных из VGG-16
- Для восстановления изображения применяется max unpooling с запоминанием позиции



⁹V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," arXiv:1511.00561, 2015.

U-Net¹⁰

- Одна из самых популярных сегментационных моделей на Kaggle.com
- Изначально была придумана для анализа медицинских изображений
- Модель без полносвязных слоёв (FCN) с энкодер-декодер архитектурой
- Транспонированные конволюции в декодере

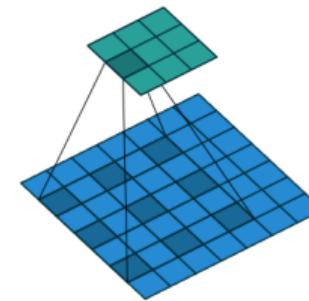
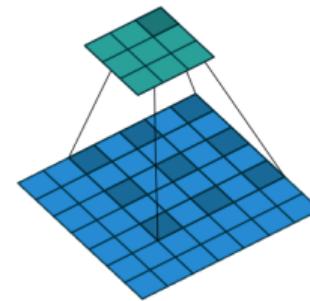
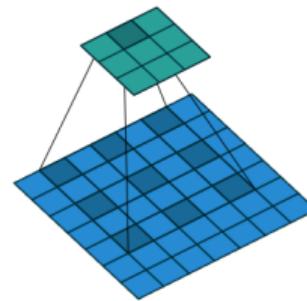
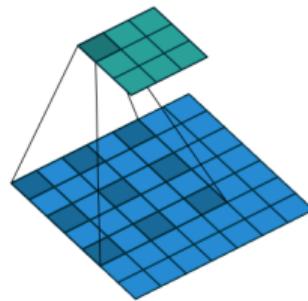


¹⁰O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation <https://arxiv.org/abs/1505.04597>

Atrous (dilated) convolution

Свойства

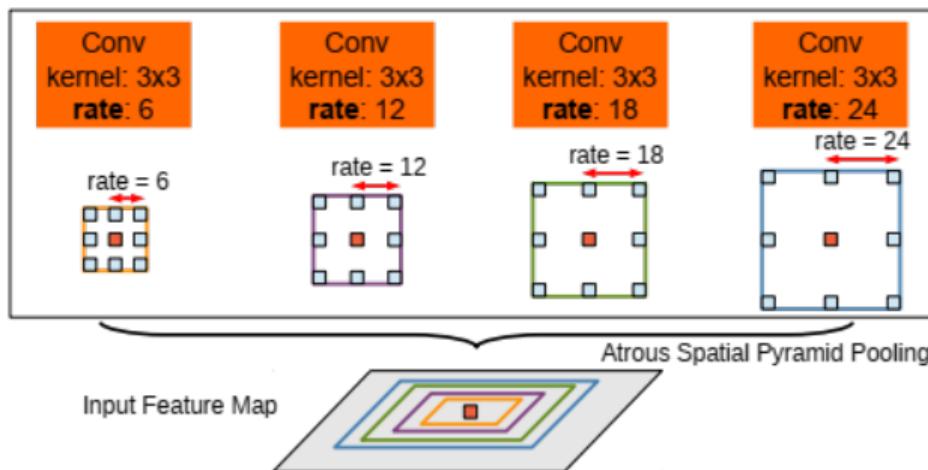
- Для качественной сегментации является важной информация об объекте целиком. Этого можно достичь увеличением receptive field
- Atrous convolution позволяют достичь увеличения receptive field при ограниченном наборе параметров нейронной сети
- При dilation rate = 1 имеем обычную свёртку



Atrous Spatial Pyramid Pooling

Идея

- При помощи delayed convolution построить пирамиду признаков разных масштабов



От семантической сегментации к объектно-чувствительной

Идея

Instance Segmentation = Detection + Semantic Segmentation

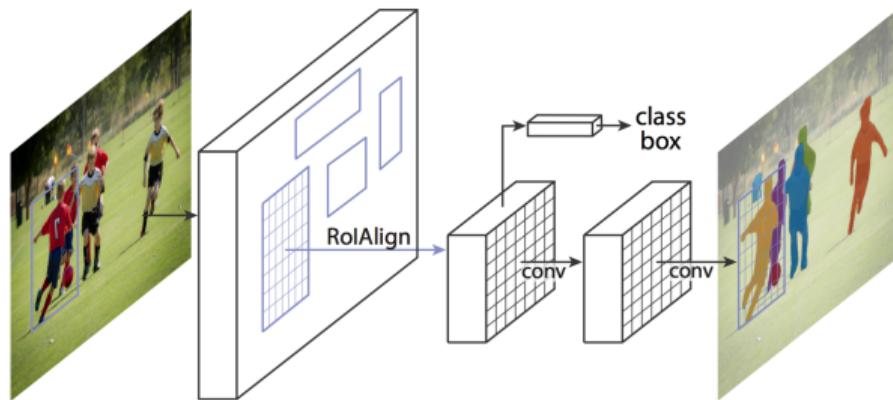
- Сначала грубо разделить объекты, используя алгоритм обнаружения
- Для каждого объекта уточнить его границы при помощи сегментационной модели

Недостатки подхода

- ➊ Если объекты находятся близко друг к другу, то в обрамляющий прямоугольник может содержать несколько объектов и разделить их, выполнив семантическую сегментацию почти невозможно
- ➋ Опять же из-за не точного обнаружения маска объекта может в итоге получиться обрезанной

Идея

- Еще одна статья из цикла "R-CNN" от тех же авторов
- Основная идея — добавить еще одну голову к Fast R-CNN модели с сегментационной маской
- Может также использоваться для поиска особох точек



¹¹<https://arxiv.org/abs/1703.06870>

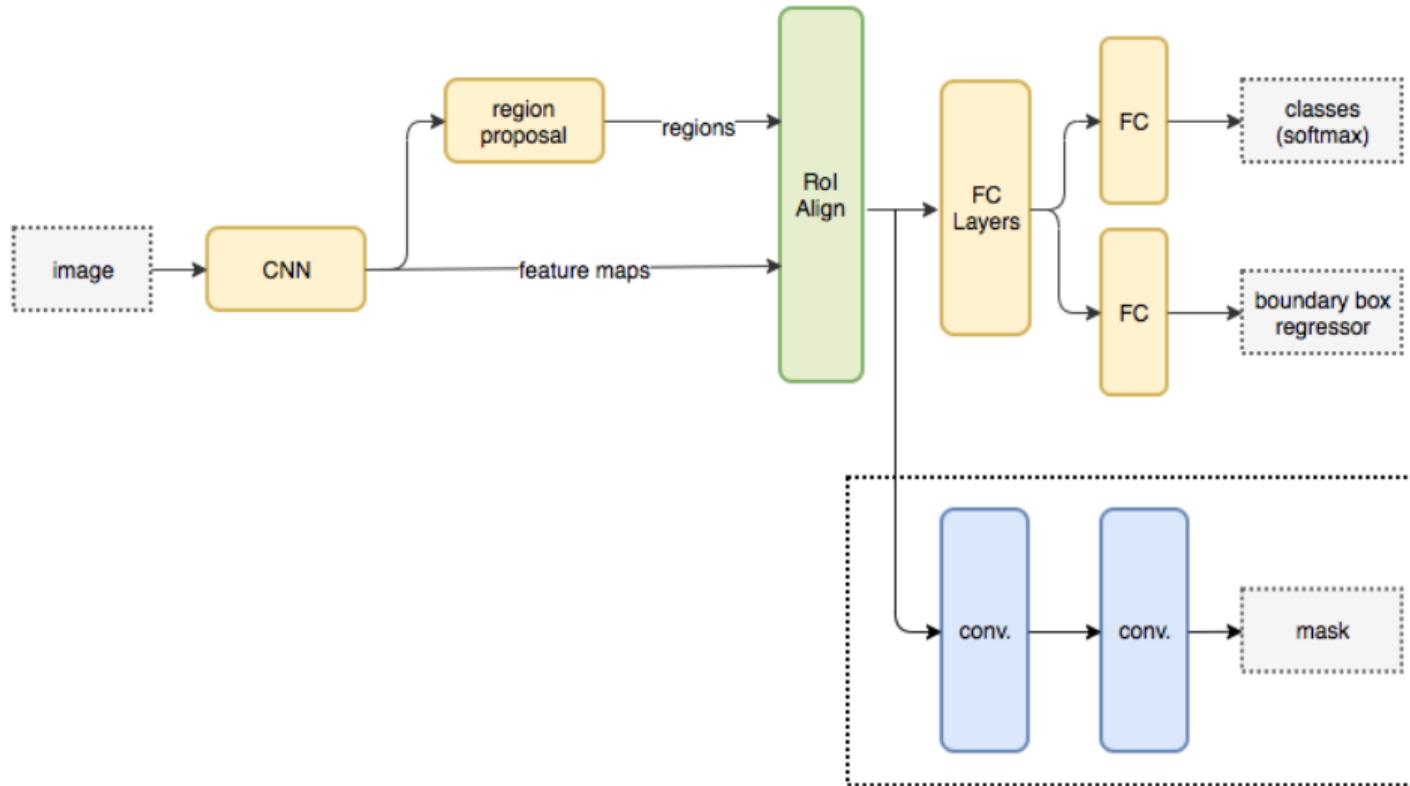
- Для добавления маски вместо слоя RoIPooling используется слой RoIAvg
- Оптимизация остаётся многозадачной и лосс выглядит следующим образом

$$L = L_{cls} + L_{box} + L_{mask}$$

- Мaska для каждого класса предсказывается независимо (нет конкуренции между классами)
- Новая архитектура backbone-a



Схема Mask R-CNN



Mask

RoiAlign

0.1	0.3	0.2	0.3	0.2	0.6	0.8	0.9
0.4	0.5	0.1	0.4	0.7	0.1	0.4	0.3
0.2	0.1	0.3	0.8	0.6	0.2	0.1	0.1
0.4	0.6	0.2	0.1	0.3	0.6	0.1	0.2
0.1	0.8	0.3	0.3	0.5	0.3	0.3	0.3
0.2	0.9	0.4	0.5	0.1	0.1	0.1	0.2
0.3	0.1	0.8	0.6	0.3	0.3	0.6	0.5
0.5	0.5	0.2	0.1	0.1	0.2	0.1	0.2

0.1	0.3	0.2	0.3	0.2	0.6	0.8	0.9
0.4	0.5	0.1	0.4	0.7	0.1	0.4	0.3
0.2	0.1	0.3	0.8	0.6	0.2	0.1	0.1
0.4	0.6	0.2	0.1	0.3	0.6	0.1	0.2
0.1	0.8	0.3	0.3	0.5	0.3	0.3	0.3
0.2	0.9	0.4	0.5	0.1	0.1	0.1	0.2
0.3	0.1	0.8	0.6	0.3	0.3	0.6	0.5
0.5	0.5	0.2	0.1	0.1	0.2	0.1	0.2

0.8	0.6
0.9	0.6

0.88	0.6
0.9	0.6

Достоинства и недостатки Mask R-CNN

Достоинства

- ① На время выхода алгоритм считался одним из лучших. Сейчас Mask R-CNN можно считать крепким бейзлайном, с которым все сравниваются
- ② Так как в виде масок можно кодировать ключевые точки, то Mask R-CNN легко адаптируется для задачи их поиска

Недостатки

- ① Если объекты находятся близко друг к другу, то обрамляющий прямоугольник может содержать несколько объектов и разделить их, выполнив семантическую сегментацию, почти невозможно
- ② Опять же из-за не точного обнаружения маска объекта может получиться обрезанной
- ③ При стандартных параметрах Mask R-CNN возвращает довольно грубую маску объектов

Pose2Seg — объектно-чувствительная сегментация людей¹²

Идея

Информация о ключевых точках (позе) человека может существенно улучшить качество сегментации

Instance Segmentation = Key Point Detection + Pose2Seg

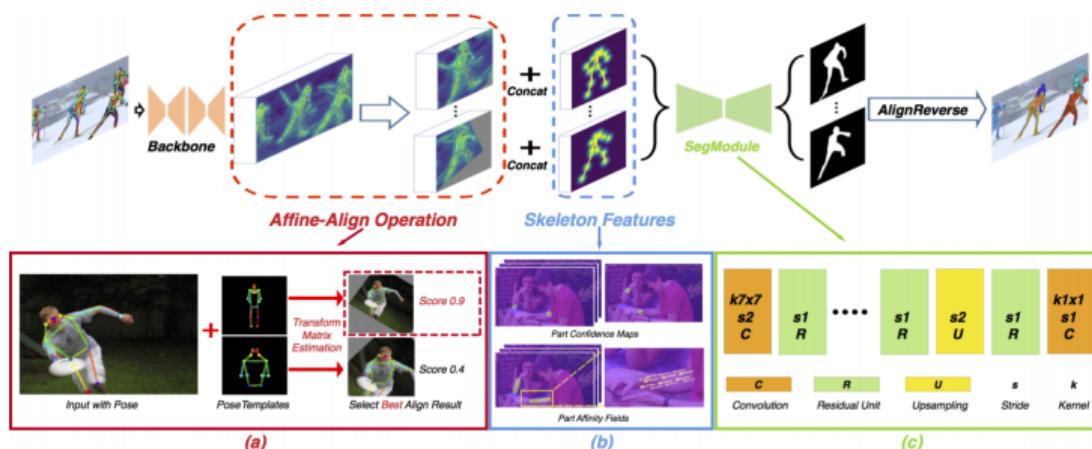


Figure 4: Overview of our network structure (Sec. 4.1). (a) Affine-Align operation (Sec. 4.2). (b) Skeleton features (Sec. 4.3). (c) Structure of SegModule (Sec. 4.4), in which residual unit refers to [15].

¹²<https://arxiv.org/abs/1803.10683>

- Сегментация — одна из ключевых задач компьютерного зрения, которая применима ко многим приложениям
- Сегментация бывает семантической, объектно-чувствительной и паноптической
- Современные сегментационные архитектуры как правило представляют собой энкодер-декодер архитектуру
- Одно из ключевых отличий различных моделей — метода восстановления изображений (upsampling)

Спасибо за внимание!

