

Обзор современных методов сегментации изображений.

Руденко Ирина, группа М05-894а.

18 декабря 2018


1 Постановка задачи.

Сегментация — это хорошо известная задача компьютерного зрения, одна из трех важнейших, наряду с классификацией и обнаружением объектов. Задача сегментации изображений заключается в разделении изображения на фрагменты(группы пикселей) по определённому критерию. Существуют различные критерии: выделение общего объекта на изображении(ко-сегментация); выделение указанного пользователем объекта(извлечение объекта); разделение изображения на регионы, сходные по своему содержанию, но различные визуально от соседей(сегментация без учителя). В своей работе я хочу обратить внимание на современные подходы в задаче семантической сегментации.

Семантическая сегментация - это попиксельная классификация изображения, где каждая метка соответствует определённому объекту. Так как цвета и яркости различных групп пикселей могут значительно отличаться, то сегментационной модели необходимо продемонстрировать некоторое «понимание» изображение, объединяя близкие по яркости, цвету, текстуре окрестности кусочки в «семантически» единый сегмент изображения.



Рис. 1: Пример семантической сегментации.

$$IoU(i) = \frac{|X_i \cap Y_i|}{|X_i \cup Y_i|}$$


Методы оценки качества сегментации должны включать оценку соответствия границ сегментов границам объектов и вложения объекта в сегмент. Используются классификационные метрики: accuracy, precision, recall, f-мера. Также существует метрика Intersection over Union (IoU), которая рассчитывается для каждого класса как частное области перекрытия и области объединения.

2 Методы семантической сегментации изображений.

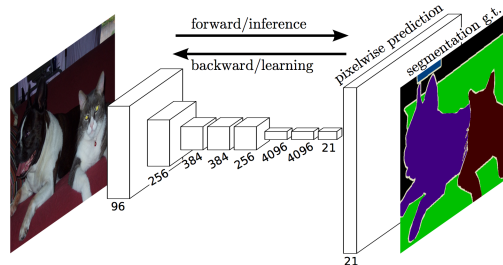
Рассмотрим наиболее популярные подходы.

2.1 Независимая попиксельная классификация.

Самый простой и очевидный подход - это классификационная модель. В таком случае метка класса определяется для каждого пиксела: задача рассматривается как задача «попиксельной» классификации. Например, использование классических классификационных нейросетей и подавать им на вход некоторую окрестность классифицируемого пикселя. Точность сегментации повышается с увеличением размера окна, так как оно охватывает больше контекстной информации, при этом увеличивается перекрытие соседних окон, таким образом возрастает объем избыточных вычислений. Основным недостатком подхода является вычислительная сложность алгоритма, а соответственно низкая скорость работы.

2.2 Полносвёрточные нейронные сети.

Идея заключается в использовании Fully-convolutional нейронных сетей, построенных из AlexNet, VGG или GoogLeNet [2].



Для построения соответствующей сети необходимо обрезать последние слои исходной и добавить слой up-conv для получения карты с вероятностями классов.

Этот подход позволяет эффективно (без избыточных вычислений) получить высокое качество поточечной классификации.

2.3 Графические модели: CRF(Conditional random fields).

Также существует метод уточнения границ, при котором результат какого-либо классификатора улучшают с помощью графических моделей. Предлагаю рассмотреть один из подходов из статьи про CRF[3]. Условное случайное поле (Conditional Random Field, CRF) — это статистический метод классификации, характерным отличием которого является возможность учитывать «контекст» классифицируемого объекта. На множестве классифицируемых объектов необходимо построить граф смежности с заданными унарными и парными потенциалами, где унарные потенциалы - это вероятность класса, а парный потенциал задаёт эвристику оценивающую вероятность того, что два смежных объекта принадлежат одному классу.

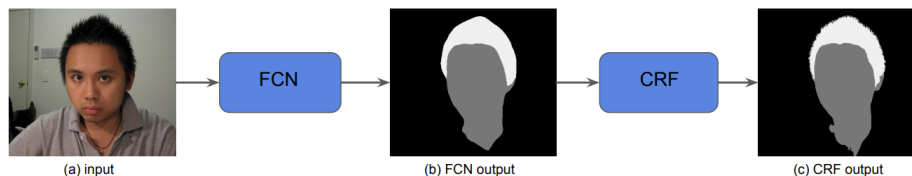


Рис. 2: Использование CRF для уточнения границ.

Далее вычисляется распределение $p(y|x)$ (вероятности метки при условии вероятности класса объекта и контекста) как сумма парных потенциалов по всем смежным вершинам и унарного потенциала в точке, после чего задача предсказания состоит в том, чтобы оптимальным образом восстановить значения y , при условии, что нам даны наблюдаемые x , соответственно найти $\arg\max_y p(y|x)$ (вероятности метки при условии вероятности класса объекта и контекста).

Унарные потенциалы задаются, например, как выходы классификаторов и шкалируются так, чтобы выходы лежали в $[0, 1]$ Парные потенциалы могут состоять из: нормы разности интенсивности, штрафа за разные метки смежных пикселей.

2.4 Полносвёрточные encoder-decoder нейронные сети.

Другим популярным подходом являются encoder-decoder сегментационные сети[1]. Основа работы полносвёрточных сетей – свёртка изображения. Пройдя через требуемое количество слоёв свёртки, изображение затем попадает в слои пулинга и регуляризации. После экспериментально выясненного количества блоков downsampling-а уменьшенное изображение требуется вернуть к изначальному размеру. Слой upsample (upsample layer) выполняет увеличение изображения. На каждый выход имеется два входных изображения: первое – это обработанная картинка с предыдущего слоя (это может быть convolution или pooling), второе – это картинка из соответствующий pooling layer.

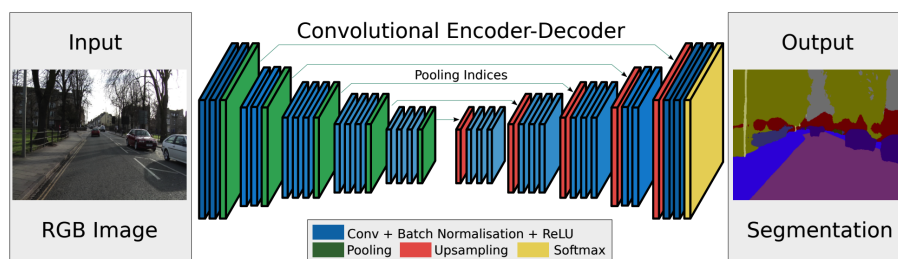


Рис. 3: Архитектура Segnet.

2.5 Заключение.

Большое разнообразие подходов и архитектур для задачи семантической сегментации позволяет решать значимые прикладные задачи, находя оптимальное соотношение качества сегментации против вычислительной сложности алгоритма. Оптимальным выбором для многих приложений сейчас считаются encoder-decoder CNN.

Список литературы

- [1] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation.
- [2] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. pages 3431–3440.
- [3] J. Shotton, J. Winn, C. Rother, and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Computer Vision – ECCV 2006*, Lecture Notes in Computer Science.