

Сравнение нескольких нейросетевых архитектур

1 PSPNet

Эта модель была создана для использования в области scene parsing. Главная задача, которую эта модель помогает решить - попиксельная сегментация объектов на изображении при условии наличия большого количества меток (например, датасет ADE20K, о котором идёт речь в статье, содержит изображения с 150 метками).

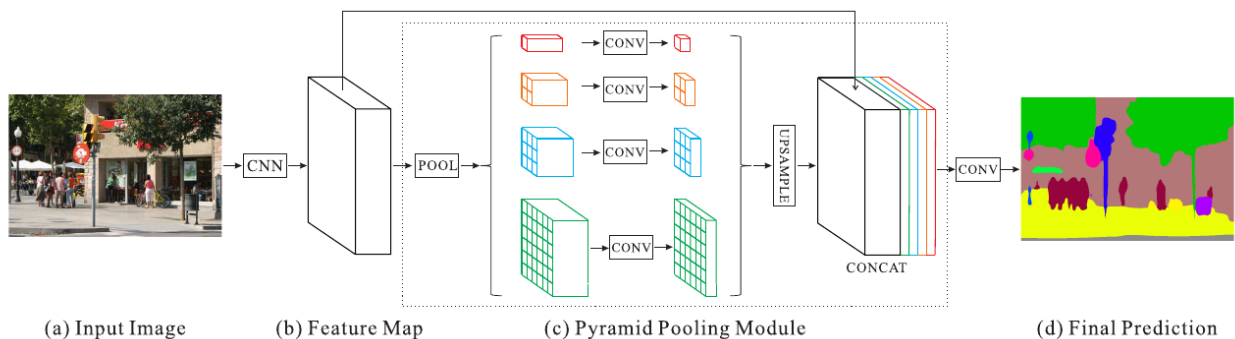
PSPNet помогает решить проблемы:

- взаимосвязи меток (например, объект с меткой «самолёт» скорее всего будет находиться в пространстве с меткой «аэропорт» или «посадочная полоса»)
- совпадающих категорий (наличие в тренировочной выборке объектов вроде «поле» и «земля», «холм» и «гора»)
- небольших объектов (например, «фонарь» или «вывеска», находящиеся на дальнем плане изображения)

Из вышеперечисленных проблем в текущей работе по сегментации скоплений нас может интересовать только последняя - искомое скопление может занимать маленькую площадь относительно всего изображения.

Основная идея PSPNet заключается в использовании Pyramid Pooling Module. Чтобы получить этот модуль, нужно сначала сделать несколько версий изначального изображения в разных масштабах с помощью pooling слоёв разных размеров, первый «грубый» слой собирает данные всего изображения в один многоканальный пиксель, следующий содержит данные нескольких подрегионов, и далее каждый последующий содержит всё меньше глобальной информации и всё больше локальной.

Далее для каждого масштаба проводится свёртка для уменьшения количества каналов и upscaling с помощью билинейной интерполяции, и все масштабы конкатенируются, чтобы с помощью последнего слоя свёртки получить для них итоговую сегментацию.



2 LinkNet

Эта модель решает проблему большого количества параметров и низкой скорости других моделей из похожей области применения. LinkNet хорошо подойдёт для сегментации в реальном времени и сегментации на видео, что тоже имеет очень слабое отношение к нашей области, где количество данных фиксировано и сегментация в реальном времени не требуется, однако скорость нейросетевой модели тем не менее является плюсом.

Общая структура LinkNet очень напоминает UNet, с тем отличием, что в блоках кодировщика добавлены дополнительные skip-connection связи.

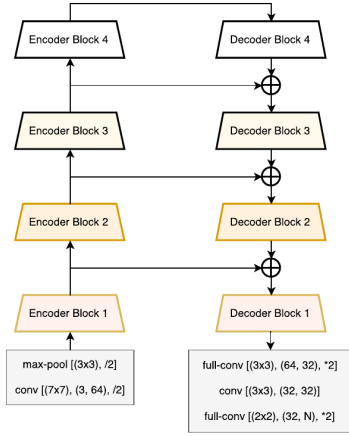


Fig. 1: LinkNet Architecture

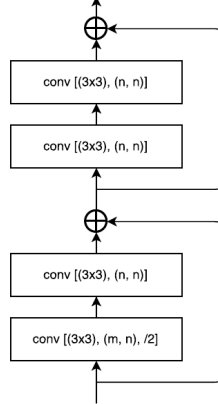


Fig. 2: Convolutional modules in *encoder-block (i)*

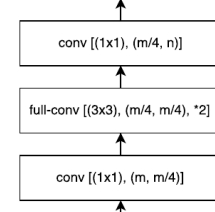


Fig. 3: Convolutional modules in *decoder-block (i)*

3 W-Net

Эта архитектура была призвана улучшить результаты сегментации при условиях:

- без использования сложных архитектур свёрточных нейронных сетей
- при тестировании на данных из разных датасетов

При сегментации космических данных мы обычно имеем общий обзор некоторой области неба в определенном спектре, и эти данные получены с помощью определенного телескопа. Однако возможно появится причина обучать модель на данных одного телескопа, а тестировать на других, что в какой-то степени коррелирует с описанными условиями.

Архитектура W-Net предлагает улучшение модели U-Net: обучение происходит на двух последовательных нейросетевых моделях. Выход первой модели конкатенируется с входом и отправляется на обработку второй моделию.

