

Обзор литературы

За последние два года глубокие сверточные сети превзошли современный уровень во многих задачах визуального распознавания, например, [1,2]. Хотя сверточные сети уже существуют в течение длительного времени [3], их успех был ограничен из-за размера доступных обучающих наборов и размера рассматриваемых сетей. Прорыв Крижевского и соавт. [1] был связан с контролируемой тренировкой большой сети с 8 слоями и миллионами параметров в наборе данных ImageNet с 1 миллионом обучающих изображений. С тех пор были обучены еще более крупные и глубокие сети [4].

Сверточная нейронная сеть по-прежнему использует те же принципы, что и MLP, однако эта нейронная сеть реализует сверточные уровни. Важно отметить, что сверточные нейронные сети обычно используются для изображений и видео. Изображения - это просто сетка чисел, и каждое число говорит вам, насколько интенсивен определенный пиксель. Зная, что это сетка чисел, мы можем манипулировать этими числами, чтобы найти закономерности и характеристики изображения. Сверточные слои делают это с помощью фильтров.

Фильтр - это определенная сетка чисел $N \times M$ (N & M представляет размер сетки), которая умножается на исходное изображение многократно. Фильтр перемещается по сетке и создает новые значения. Эти новые значения могут представлять края или линии на изображении.

U-Net Сверточные нейронные сети для биомедицинской сегментации изображения

В этой статье [5] представляется сетевая и обучающая стратегия, основанная на активном использовании дополнения данных для более эффективного использования доступных аннотированных выборок. Архитектура состоит из сокращающегося пути для захвата контекста и симметричного расширяющегося пути, который обеспечивает точную локализацию. Показывается, что такая сеть может быть обучена сквозным образом из очень небольшого числа изображений и превосходит предыдущий лучший метод (сверточная сеть со скользящим окном) в задаче ISBI для сегментации нейронных структур в электронно-микроскопических стопках.

Сеть работает быстро. Сегментация изображения 512x512 занимает менее секунды на недавнем графическом процессоре.

FCN для семантической сегментации

В данной статье [6] показывается, что сами по себе сверточные сети, прошедшие сквозную подготовку, пиксели в пиксели, превосходят современные технологии семантической сегментации. Ключевая идея-создание «полностью сверточных» сетей, которые принимают ввод произвольного размера и

производят вывод соответствующего размера с эффективным выводом и обучением. Определяется и детализирует пространство полностью сверточных сетей, объясняется их применение к пространственно плотным задачам прогнозирования и устанавливаются связи с предыдущими моделями. Кроме того, адаптируются современные классификационные сети (AlexNet, VGG net и GoogLeNet) в полностью сверточные сети и переносятся их изученные представления путем точной настройки в задачу сегментации. Затем определяется новая архитектура, которая объединяет семантическую информацию из глубокого, грубого слоя с информацией о внешнем виде из мелкого, тонкого слоя для получения точных и подробных сегментаций. Данная сверточная сеть достигает современной сегментации POSCAL VOC (20% относительного улучшения до 62,2% среднего IU в 2012 году), NYUDv2 и SIFT Flow, тогда как вывод занимает менее одной пятой секунды для типичного изображения.

PAN для семантической сегментации

В данной работе [7] предлагается сеть PAN для использования влияния глобальной контекстной информации в семантической сегментации. В отличие от большинства существующих работ, в данной работе объединяется механизм внимания и пространственная пирамида для извлечения точных плотных объектов для маркировки пикселей вместо сложной дилатации свертки и искусственно созданных сетей декодеров. В частности, вводится модуль Feature Pyramid Attention для выполнения структуры внимания пространственной пирамиды на высокоуровневых выходных данных и комбинирования глобального пула для лучшего представления функций, а также модуль Global Attention Upsample на каждом уровне декодера для обеспечения глобального контекста в качестве ориентира низкого уровня. Предложенный подход позволяет достигать самых современных результатов в тестах PASCAL VOC 2012 и Cityscapes с новым рекордом точности MIOU 84,0% для PASCAL VOC 2012 при обучении без набора данных COCO.

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: *Imagenet classification with deep convolutional neural networks*. In: NIPS. pp. 1106–1114 (2012)
2. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: *Discriminative unsupervised feature learning with convolutional neural networks*. In: NIPS (2014)
3. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: *Backpropagation applied to handwritten zip code recognition*. *Neural Computation* 1(4), 541–551 (1989)
4. Simonyan, K., Zisserman, A.: *Very deep convolutional networks for large-scale image recognition* (2014), *arXiv:1409.1556 [cs.CV]*
5. Olaf Ronneberger, Philipp Fischer, and Thomas Brox : *U-Net: Convolutional Networks for Biomedical Image Segmentation* (2015)
6. Jonathan Long, Evan Shelhamer, Trevor Darrell : *Fully Convolutional Networks for Semantic Segmentation* (2015)
7. Hanchao Li, Pengfei Xiong, Jie An, Lingxue Wang : *Pyramid Attention Network for Semantic Segmentation* (2018)