

## ПРИМЕНЕНИЕ МОДЕЛИ «КОДЕР-ДЕКОДЕР» ДЛЯ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

*Нишчхал<sup>1</sup>, Помыткина Диана Павловна<sup>2</sup>*  
*Студент*

*<sup>1,2</sup>Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева  
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31*

## APPLICATION OF ENCODER-DECODER MODEL FOR IMAGE SEGMENTATION

*Nishchhal<sup>1</sup>, Pomytkina Diana Pavlovna<sup>2</sup>*  
*Student*

*<sup>1,2</sup>Reshetnev Siberian State University of Science and Technology  
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation*

**Аннотация:** Рассматривается проблема сегментации реальных изображений, включая спутниковые изображения, которые могут быть зашумленными или включать перекрытие визуальных объектов. В докладе предлагается решение на основе модели «кодер-декодер».

**Abstract:** The problem of image segmentation in real-life including satellite images, which can suffer from noise or object obstruction, is considered. The paper proposes a solution based on an encoder-decoder model.

**Ключевые слова:** глубокие нейронные сети, кодер-декодер.

**Keywords:** deep neural network, encoder-decoder.

### Введение

Сегментация является важным этапом системы распознавания, обнаружения и классификации изображений, поскольку она извлекает интересующие нас объекты для дальнейшей обработки. Сегментация изображения используется на практике для классификации пикселей изображения [1][4][5]. Методы сегментации используются для выделения желаемого объекта из изображения с целью выполнения анализа объекта [2]. С этой точки зрения интерес представляют методы машинного обучения без учителя.

Хотя многочисленные работы за последние десятилетия внесли существенный вклад в улучшение методов сегментации изображений, объекты с похожими визуальными/спектральными характеристиками и однородные объекты с различными визуальными/спектральными характеристиками представляют собой проблему для алгоритмов сегментации. Существующие методы сегментации изображений имеют ограничения, связанные с излишней детализацией и высокой чувствительности к шуму.

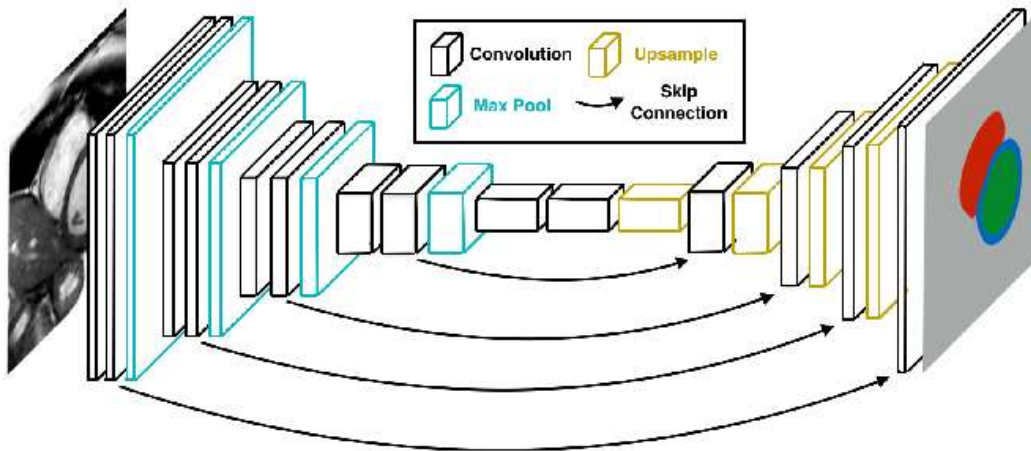
Кроме того, процесс сегментации требует больше времени для обработки, а некоторые признаки являются избыточными и перекрываются [3]. Определение признаков всего изображения также является сложной задачей. Из-за наличия избыточных и перекрывающихся признаков невозможно получить точный результат.

Непосредственное использование сверточных глубоких сетей для попиксельной сегментации дает плохие результаты, главным образом потому, что слои максимального объединения и слои подвыборки уменьшают разрешение карты объектов и, следовательно, выходное разрешение уменьшается. Даже при экстраполяции до исходного разрешения создается изображение с потерями.

### Модель «кодер-декодер»

В качестве решения этой проблемы предлагается использовать глубокую нейронную сеть на основе модели «кодер-декодера», которая использует не только традиционные слои, но улучшенную архитектуру и типы слоев, такие как свертка с разделением по глубине. Кроме того, в данной модели используются остаточные связи, которые могут обеспечить улучшение по сравнению с классическими последовательными сетями и связи с пропуском соединений (skip connection).

В качестве базовой сети была выбрана сеть SegNet, потому что ее архитектура использует структуру «кодер-декодер». Слои кодера и декодера симметричны друг другу. Операция повышающей дискретизации слоев декодера использует индексы максимального объединения соответствующих слоев кодера. Модель включает блоки свертки, как показано на рисунке 1, с обоснованием увеличения рецептивного поля. За каждой операцией свертки следует пакетная нормализация и нелинейная активация с использованием функции ReLU. После завершающей свертки следует активация softmax. Количество параметров в части кодировщика и части декодера также огромно. Таким образом, все обычные свертки заменяются глубинно отделимой сверткой. В то же время количество каналов в кодере и декодере также уменьшается.



*Рис 1. Кодер-декодер с пропуском соединения*

Особенность реализуется на стадии подвыборки (subsampling), когда слой Max-pooling используется для достижения трансляционной инвариантности при небольших пространственных сдвигах изображения. В сочетании с подвыборкой это приводит к тому, что каждый пиксел управляет большим контекстом входного изображения (пространственным окном). Эти методы обеспечивают более высокую точность классификации, но уменьшают размер карты признаков, что приводит к представлению изображения с потерями и размытыми границами. Решение этой проблемы заключается в билинейной восходящей выборке и низкоуровневой карте признаков одного слоя модуля кодирования, когда формируется модуль декодирования сети, повышающая выборка возвращается к исходному размеру карты, а соответствующая карта сегментации получается через слой классификации softmax. Это необходимо для разделения пространственной информации и информации о глубине. Обнаружено, что эффект набора деталей 1/2 размера карты объектов и функция декодера сливаются, и, наконец, достигаются хорошие результаты. Перед подвыборкой необходимо захватить и сохранить информацию о границах в картах объектов кодировщика. Для повышения эффективности сеть SegNet хранит только индексы максимального пула, т.е. местоположения максимального значения функции в каждом окне пула запоминаются для каждой карты кодировщика. Для каждого окна 2×2 требуется всего 2 бита с небольшой потерей точности.

#### Заключение

Предлагаемые сети позволяют значительно сократить количество параметров, что эффективно как с точки зрения уменьшения объемов памяти и вычислительного времени. Это может не только улучшить производительность сегментации, но и значительно скомпенсировать дисбаланс обучающих данных. Как способ улучшения способности к обучению этого метода.

Те архитектуры, которые хранят карты сетевых объектов кодировщика в полном объеме, работают лучше всего, но потребляют больше памяти во время логического вывода. С другой стороны, сеть SegNet хранит только индексы максимального объединения карт функций и использует их для достижения хорошей производительности.

В дальнейшем планируется увеличение глубины сети, а также модификация модуля деформируемых сверток для повышения производительности на небольших наборах данных.

#### Список литературы

1. D. C. Cirean et al., "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in Neural Information Processing Systems*, vol. 25, pp. 2852–2860, 2012.
2. R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271618301229>.
3. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, pp. 3431–3440, 2015.
4. P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 75–91.

5. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018.

© Нишчхал , Помыткина Д.П., 2022