УДК 004.93'1

СРАВНЕНИЕ НЕЙРОННЫХ СЕТЕЙ, ПРИМЕНЯЕМЫХ ДЛЯ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

А. А. Воронов[1], С. В. Минчук[2]
[1]Объединенный институт проблем информатики НАН Беларуси, Минск;
[2]Белорусский государственный университет, Минск

Приведены результаты решения задачи сегментации спутниковых снимков Земли. В качестве базовых нейронных сетей для сегментации используются нейросетевые архитектуры U-Net и FPN, DeepLabv3+, которые позволяют выделить пиксели мультиспектрального изображения Земли, принадлежащие классу «водоем». Описаны исследуемые нейросетевые модели и оценки точности для решения задачи семантической сегментации изображений дистанционного зондирования Земли.

1. Методы сегментации изображений и использование нейронных сетей

С развитием технологии дистанционного зондирования Земли (ДЗЗ) качество и разрешение спутниковых изображений значительно улучшились. Однако анализ таких изображений является трудоемким процессом.

Сегментация изображений – один из этапов автоматизации анализа изображений ДЗЗ. Выделяют три основные группы методов сегментации: основанные на алгоритмах наращивания областей, выделения границ, статистическом анализе распределения цвета точек изображения. Методы, основанные на алгоритмах наращивания областей, используют разбиение входного изображения на элементарные области, из которых затем составляются искомые области (сегменты). Гранично-ориентированные методы основаны на поиске изменений интенсивности и цвета точек на границах сегментов. При статистическом методе точки из одного кластера в некотором цветовом пространстве изображения принадлежат сегментам с одинаковой цветовой характеристикой.

Существуют методы сегментации полутоновых изображений, комбинирующие названные выше подходы. Критерием разделения на области выступает яркость (интенсивность), а выделение контуров осуществляется на основе дискретного ортогонального преобразования Уолша, что позволяет получить более тонкую контурную линию. Данные методы хорошо разделяют изображения, имеющие две градации яркости: объект и фон, которые принадлежат разным градациям [1–4].

Нейронные сети (НС) также используются для различных типов сегментации: семантической, паноптической, Instance-сегментации. Семантическая сегментация изображений заключается в выделении на изображении локальных областей (сегментов), соответствующих различным классам объектов без выделения отдельных экземпляров объектов. При сегментации объектов, или Instance-сегментации, добавляется выделение экземпляров объектов каждого класса. Паноптическая сегментация объединяет две перечисленные и обеспечивает более полное понимание изображения. Анализ изображений ДЗЗ заключается в вычислении геометрических, яркостных и текстурных характеристик областей изображения, выбор которых основывается на предыдущем опыте анализа изображений того же типа. Основой данного анализа является, как правило, использование семантической сегментации. Сегментация снимков ДЗЗ находит применение в геоинформационных системах и позволяет выделять водоемы, леса или городскую застройку, что полезно при создании карт местности и мониторинге изменений в окружающей среде.

Основными архитектурами НС для сегментации изображений являются: FPN, SegNet, DeepLabv1-3+, PSPNet, U-Net, HRNet [1–3], в которых применяются некоторые из следующих приемов:

– операция, обратная свертке, – развертка (deconvolution);

– расширенная свертка (dilated convolutions);

– многомасштабное агрегирование контекста (multi-scale context aggregator);

– формирование сети на основе пирамидального объединения (pyramid pooling network);

– марковские случайные поля (Conditional Random Field) и постпроцессинг (postpro-cessing);

– пропуск соединения (skip-connection).

В первой архитектуре сети, выбранной для исследования FPN, основой является построение пирамиды признаков, сохраняющей семантическую репрезентативность на всех уровнях пространственного разрешения. Структурные решения FPN активно применяются при проектировании декодеров для задач покадрового анализа изображений. Структура сети FCN представлена на рис. 1, в качестве экстрактора признаков для изображения применяется ResNet [2].

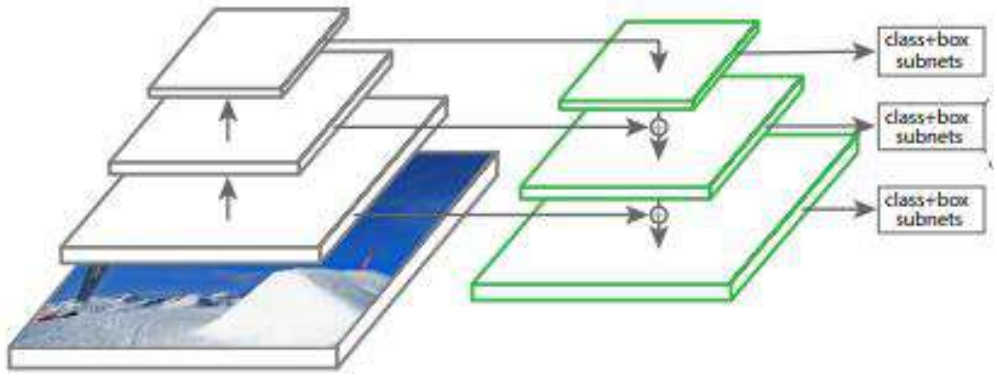

Рис. 1. Общий вид сети FPN

Архитектура же СНС для сегментации U-Net отличается тем, что использует первый и последний из перечисленных приемов: развертка (deconvolution) и пропуск соединения (skip-connection) для связи элементов кодера и декодера. Типичная структура сети U-Net изображена на рис. 2.

Кодер использует уровень объединения для постепенного уменьшения пространственного измерения входных данных, в то время как декодер постепенно восстанавливает детали и соответствующее пространственное измерение на уровне сети, таком как уровень деконволюции. От кодировщика к декодеру обычно существует связь, которая помогает декодеру лучше восстанавливать целевые детали. U-Net хорошо справляется с сегментацией, однако для очень мелких объектов или объектов, сильно перекрывающихся, сеть может сталкиваться с трудностями.

Архитектура DeepLabV3+ реализует подход к сегментации, основанный на извлечении разномасштабного контекста с сохранением разрешения. В отличие U-Net семейство моделей DeepLab решает проблему масштабирования признаков с помощью модулей пространственного пирамидального пулинга [3] и атрозных сверток [5].

Пространственный пирамидальный пулинг для извлечения признаков для различных масштабов изображения востребован при анализе спутниковых снимков с высокой вариативностью размеров объектов одного класса и совмещен со структурой кодировщика-декодировщика. В версии DeepLabV3+ применяется упрощенный декодер с од-

ним уровнем конкатенации признаков, что позволяет уточнять границы мелких объектов, помогает решать различные задачи сегментации лучше, чем U-Net [3], однако повышает вычислительную сложность.

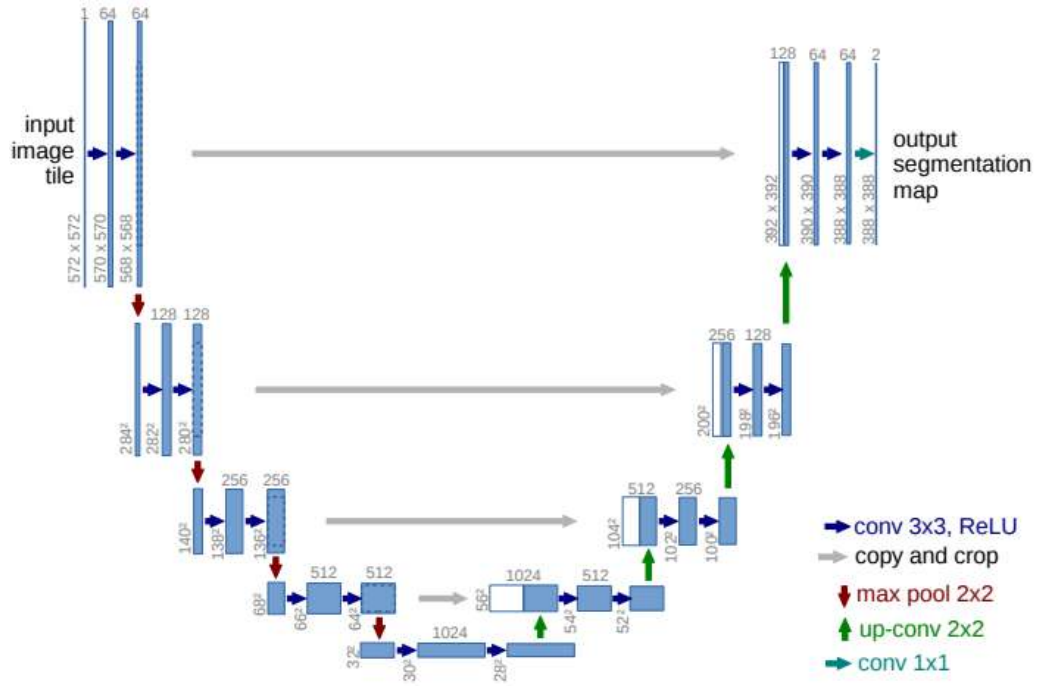

Рис. 2. Общий вид нейронной сети U-Net

По результатам анализа следует отметить перспективность применения для семантической сегментации изображений ДЗЗ СНС U-Net, FCN и DeepLabV3+. U-Net имеет простую архитектуру и, как следствие, малое потребление ресурсов при использовании в рабочем режиме и при обучении – даже при небольшой обучающей выборке достигаются приемлемые по качеству результаты.

2. Исходные данные и предобработка для сегментации

Источником оптических данных послужили снимки космических аппаратов Sentinel-2 с пространственным разрешением 10 м на пиксель в видимом диапазоне. Процедура подготовки данных и предобработки включала фрагментацию исходных сцен на тайлы фиксированной размерностью 512×512 пикселей. Параллельно осуществлялась фильтрация артефактов: фрагменты с перекрытием облачностью более 10 % исключались из выборки, так как наличие плотных атмосферных образований генерирует неустранимый шум, препятствующий стабильной сходимости градиентного спуска при обучении.

Для каждого валидного фрагмента генерировалась бинарная сегментационная маска. Пиксели водной поверхности кодировались значением 1, элементы суши и фон – 0. Специфика контуров при сегментации определяется естественным ландшафтом, где береговые линии формируют нелинейные градиентные переходы. Завершающим этапом предобработки выступала поканальная нормализация значений интенсивности RGB-изображений к числовому диапазону от 0 до 1.

Входными данными для обучения являются пары, состоящие из оригинальных изображений и размеченных масок. На масках определенным цветом обозначены области, которые соответствуют классу «водоем». Исходные изображения ДЗЗ необходимо научиться сегментировать с помощью предложенной сети. Сформированный набор данных содержит 2 тыс. аннотированных пар «изображение – маска». Массив разделен на обучающую и валидационную выборки в пропорции 80 и 20 % соответственно. Выборка структурно сбалансирована по геометрическим характеристикам целевого класса: она включает как сплошные водные массивы большой площади, так и узкие линейные объекты (реки, каналы). Наличие целевых объектов с разной формой позволяет корректно протестировать модели.

Для сегментации применяются сети U-Net и FPN, DeepLabV3+, реализованные на языке *Python* с использованием библиотек *Scikit-image, NumPy, Matplotlib, PyTorch, Albumentations*.

Также ко всем изображениям была применена аугментация – увеличение обучающей выборки данных посредством модификации этих данных и создания на их основе дополнительных изображений, полученных при помощи поворота исходных на угол, кратный 45º, и аффинных преобразований, случайного изменения контраста и яркости исходных изображений.

В качестве метода оптимизации используется Адам [6] с параметром скорости обучения, равным 0,001. Алгоритм оптимизации является расширением стохастического градиентного спуска, который в последнее время получил широкое распространение для приложений глубокого обучения в области компьютерного зрения. Алгоритм автоматически регулирует скорость обучения для каждого параметра модели, что позволяет ему быстрее сходиться к оптимальному решению.

3. Оценка обучения нейросетевых моделей

В качестве функции потерь для решения задачи сегментации анализировалось поведение Binary CrossEntropy Loss (BCE Loss), или бинарной кросс-энтропии, – одной из самых распространенных функций потерь в области глубокого обучения. Бинарная кросс-энтропия измеряет разницу между истинным классом и предсказанной вероятностью. Она штрафует модель за неправильные предсказания.

В качестве метрики для оценки точности работы алгоритма выбрана Intersection-Over-Union (IoU), которая вычисляется по формуле

$$\mathrm{IoU} = \frac{Si}{Su},$$

где *Si* – площадь пересечения результата сегментации и истинной области, *Su* – площадь объединения результата сегментации и истинной области, что эквивалентно мере Жаккара [6, 7].

Анализировали процесс обучения моделей по значениям функции потерь BCE Loss и метрики качества для нейронной сети U-Net, DeepLabV3+, FPN. В результате обучения максимальную точность продемонстрировала архитектура DeepLabV3+ со значением метрики IoU, равным 0,7846. Модель позволила агрегировать информацию с обширных участков изображения, сохраняя целостность масок крупных водных массивов.

Значение метрики IoU достигло 0,7659 для нейросетевой архитектуры U-Net в аналогичных условиях, но сформировались неразмеченные области внутри сплошных водо-

емов. Модель FPN достигла значения 0,7669 метрики IoU в процессе обучения, успешно сегментируя узкие русла рек.

Визуальное оценить пример сегментации при помощи рассматриваемых моделей можно на рис. 3, где представлено также подтверждение описанных эффектов.

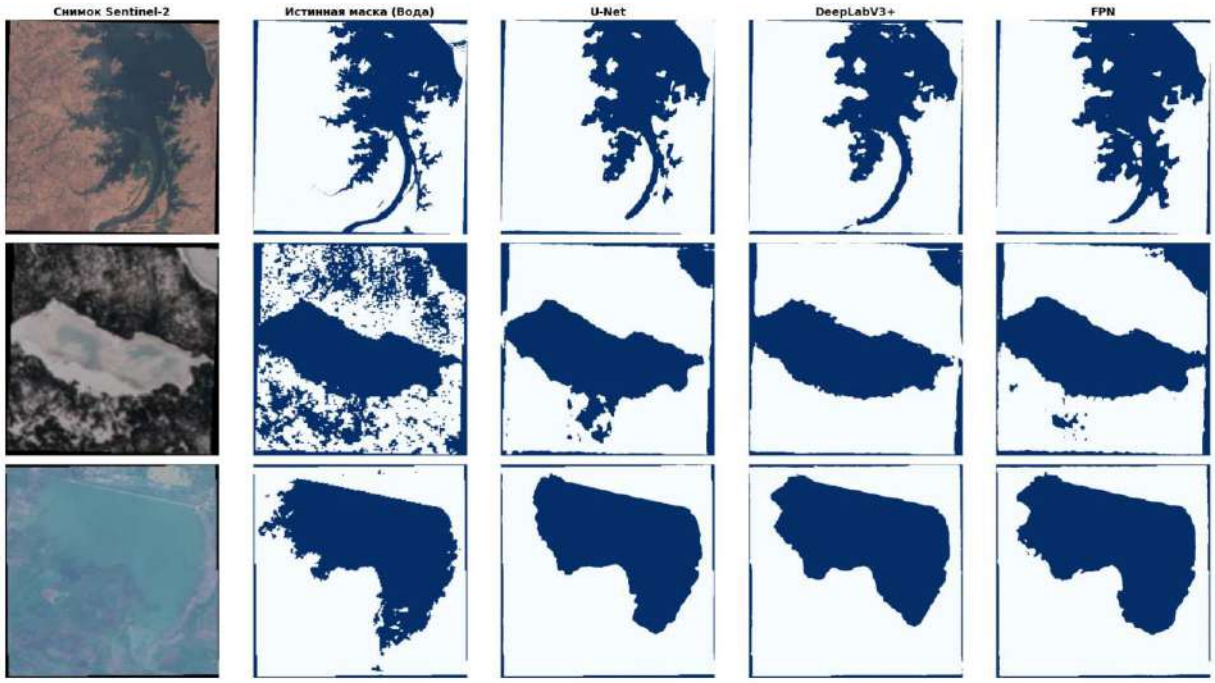

Рис. 3. Три примера исходного изображения, соответствующей маски водоема
и результатов сегментации при помощи нейронных сетей U-Net, DeepLabV3+ и FPN

При обработке спутниковых снимков со стабильным масштабом и ортогональной проекцией наибольшую эффективность демонстрирует архитектура DeepLabV3+, так как нет дробления сплошных водных полигонов и обеспечивается максимальное значение индекса Жаккара. U-Net демонстрирует фрагментацию предсказаний из-за отсутствия механизмов оценки глобального контекста, а иерархическая структура FPN показывает промежуточные метрики качества.

Список литературы

1. Сикорский, О. С. Обзор сверточных нейронных сетей для задачи классификации изображений / О. С. Сикорский // Новые информационные технологии в автоматизированных системах. – 2017. – № 20. – С. 37–42.

2. Feature pyramid networks for object detection / T.-Y. Lin, P. Dollár, R. Girshick [et al.] // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 21–26 July 2017. – Honolulu, 2017. – URL: https://arxiv.org/abs/1612.03144 (date of access: 17.08.2026).

3. Николенко, С. Глубокое обучение / С. Николенко, А. Кадурин, Е. Архангельская. – СПб. : Питер, 2018. – 480 с.

4. Ваткин, М. Е. Методы и алгоритмы обработки и анализа топологических изображений интегральных микросхем на основе нейронных сетей : дис. … канд. техн. наук : 05.13.01 ; 30.05.06 / Ваткин Максим Евгеньевич. – Мн., 2006. – 187 с.

5. Encoder-decoder with Atrous separable convolution for semantic image segmentation / L.-C. Chen, Y. Zhu, G. Papandreou [et al.] // Computer Vision – ECCV 2018 : 15th European Conf., Munich, Germany, 8–14 Sept. 2018. – Cham : Springer, 2018. — P. 833–851. – (Lecture Notes in Computer Science ; vol. 11211). – DOI: 10.1007/978-3-030-01234-2_49.

6. Adam: A Method for Stochastic Optimization. – URL: https://arxiv.org/abs/1412.6980 (date of access: 15.08.2026).

7. Гонсалес, Р. Цифровая обработка изображений : пер. с англ. / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.