

ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ ДЛЯ АВТОМАТИЗИРОВАННОГО ДЕШИФРИРОВАНИЯ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

Смирнов А.А.¹

Научный руководитель: Друки А.А.²

¹НИ ТПУ, ИШИТР, студент, гр. 8ВМ32, e-mail: aas161@tpu.ru

²НИ ТПУ, ИШИТР, к.т.н., доцент ОИТ, e-mail: druki@tpu.ru

Аннотация

В работе представлены результаты разработки и сравнения архитектур сверточных нейронных сетей (СНС) для семантической сегментации объектов земной поверхности на изображениях, полученных с беспилотных летательных аппаратов (БПЛА). Исследование включает подготовку датасета, реализацию моделей U-Net, U-Net++, SegNet и ENet, их обучение и оценку качества сегментации. Наилучшие результаты показала модель SegNet (F1-мера – 0,93, mIoU – 0,88).

Ключевые слова: Сверточная нейронная сеть, семантическая сегментация, компьютерное зрение, БПЛА, дистанционное зондирование Земли, U-Net.

Введение

Актуальность исследования обусловлена возрастающим объемом высокодетальных данных дистанционного зондирования Земли (ДЗЗ), получаемых с беспилотных летательных аппаратов. Эти данные используются в таких областях, как сельское хозяйство, экологический мониторинг, градостроительство и управление природными ресурсами. Однако ручная обработка изображений требует значительных временных и трудовых затрат, подвержена субъективным ошибкам [1]. Автоматизация процесса дешифрирования объектов земной поверхности с помощью нейросетевых алгоритмов позволяет повысить точность и скорость анализа данных, что делает результаты работы востребованными для практических применений.

Современные методы семантической сегментации, основанные на нейросетевых моделях, демонстрируют высокую эффективность в задачах обработки изображений [2]. Однако большинство существующих исследований сосредоточено на данных с высоким балансом классов, тогда как реальные изображения ДЗЗ часто характеризуются значительным дисбалансом, что усложняет задачу сегментации. Это подчеркивает необходимость разработки адаптированных алгоритмов, способных работать с неравномерно распределенными классами объектов.

Целью данной работы является разработка и сравнение архитектур сверточных нейронных сетей для автоматизированного дешифрирования объектов земной поверхности на изображениях ДЗЗ, учитывающих дисбаланс классов. Новизна исследования заключается в применении и адаптации моделей U-Net, U-Net++, SegNet и ENet для обработки данных с высоким дисбалансом, а также в использовании методов аугментации и взвешенной функции потерь для улучшения качества сегментации.

Набор данных для исследования

В качестве объекта исследования использовались изображения с БПЛА и соответствующие им карты сегментации. Съемка проводилась в высоком пространственном разрешении на севере Красноярского края в видимом диапазоне электромагнитного спектра.

Набор данных состоит из 80 изображений с разрешением 37 000×37 000 пикселей. Пространственным разрешение изображений – 6 см/пиксель. Разметка проводилась специалистами инженерно-изыскательских работ в ручном режиме. Для каждого изображения в наборе есть соответствующая ему маска, где каждый класс объектов окрашен в определенный цвет (рис. 1).

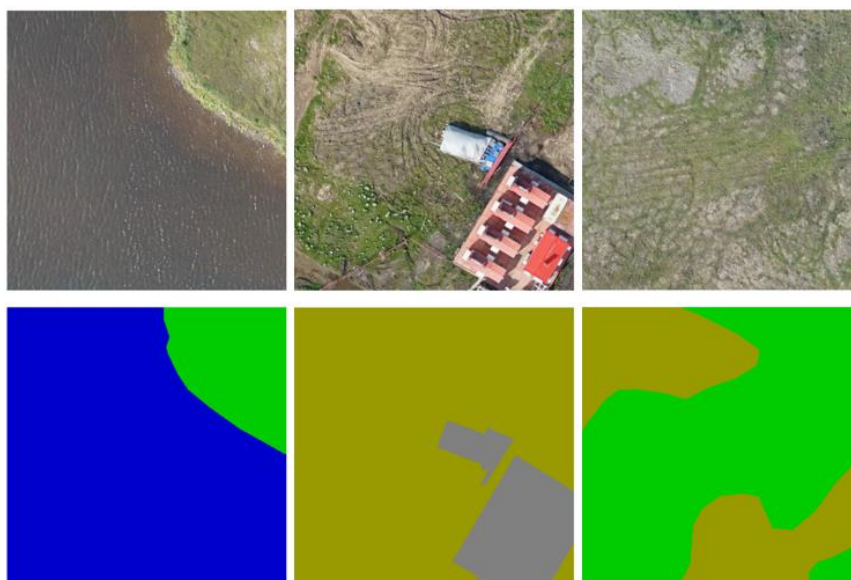


Рис. 1. Примеры изображений и масок из выборки

Набор содержит 5 классов объектов: растительность (зеленый цвет), гидрография (синий цвет), грунт (коричневый цвет), дороги (черный цвет) и антропогенные объекты (серый цвет).

Для формирования датасета исходные изображения были нарезаны на фрагменты размером 512×512 пикселей с перекрытием соседних фрагментов 50 пикселей. В результате было получено 16 200 обучающих и 1 800 тестовых примеров.

Набор данных обладает существенным дисбалансом в пользу растительности – на них приходится 91 % всех пикселей, гидрография – 5 %, дороги – 3 %, грунт – 1 %, антропогенные объекты – около 0,07 %.

Предобработка включала растеризацию векторных масок, создание растровых тайлов, нормализацию и аугментацию изображений. В работе использовалась аугментация данных «на лету» (случайные изменения масштаба, яркости, контраста, отражение по вертикальной и горизонтальной осям, эластические преобразования), расширяющая обучающую выборку синтетическими данными [3]. Цель состоит в том, чтобы на этапе обучения модель никогда не увидела одно и то же изображение дважды. Это поможет модели выявить больше особенностей данных и достичь лучшей степени обобщения.

Архитектуры нейронных сетей

U-Net. Считается одной из стандартных архитектур СНС для задачи сегментации изображений (рис. 2).

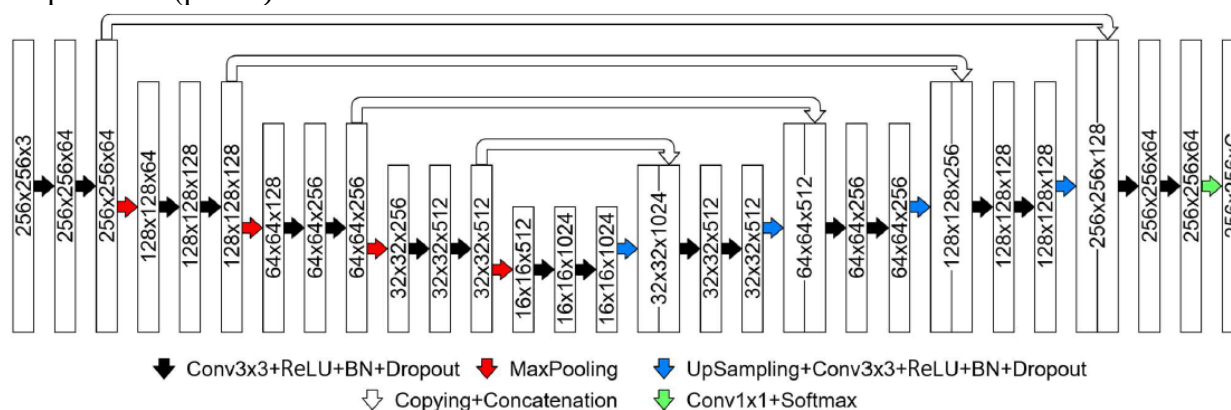


Рис. 2. Архитектура U-Net

Особенностью U-Net является использование сквозных соединений (skip-connections) для объединения деталей из энкодера и контекста из декодера, что особенно полезно для сегментации мелких объектов.

U-Net содержит 23 сверточных слоя, плюс в каждом из них применяются пакетная нормализация и функция активации ReLU, 4 слоя Max pooling, 4 слоя Upsampling, в конце функция активации Softmax. В сумме получается 31 слой [1].

U-Net++. Архитектура U-Net++ является развитием идей U-Net (рис. 3).

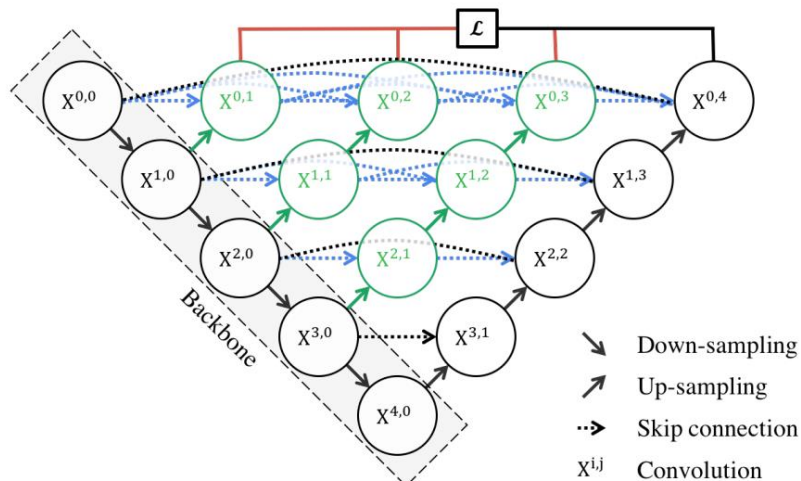


Рис. 3. Архитектура U-Net++

В U-Net++ используется 30 сверточных слоев и несколько уровней пространственных разрешений для более эффективного извлечения признаков. Это позволяет сети лучше адаптироваться к различным размерам объектов на изображениях. В целом, U-Net++ предлагает более сложную архитектуру с дополнительными механизмами для более эффективного извлечения признаков и улучшения качества сегментации изображений.

SegNet. Архитектура SegNet представлена на рис. 4.

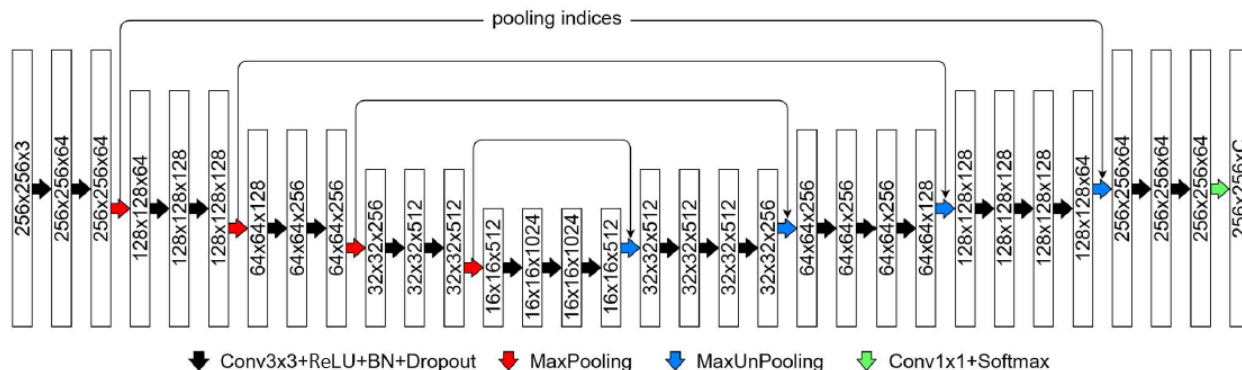


Рис. 4. Архитектура SegNet

Принципиальным отличием архитектуры SegNet от архитектуры U-Net является способ восстановления исходного разрешения карты сегментации в декодере: вместо операций увеличения масштаба и полного копирования карт признаков, SegNet использует операции MaxUnPooling, выполняемые с использованием сохраненных позиций пикселей, выбранных операциями MaxPooling в энкодере [5].

ENet. Разработана для работы в реальном времени, что достигается за счет уменьшения количества параметров (рис. 5). ENet позволяет осуществлять сегментацию изображений в реальном времени, что полезно при использовании на мобильных устройствах или при реализации систем компьютерного зрения на программируемых логических интегральных схемах [6].

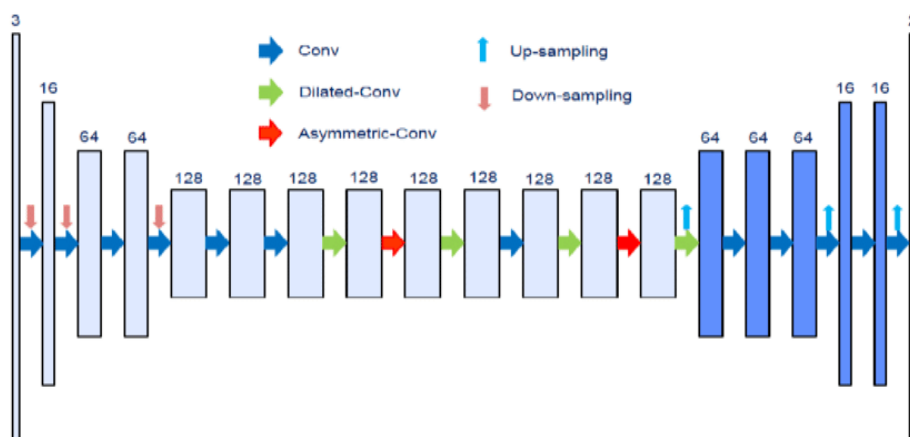


Рис. 5. Архитектура ENet

Гиперпараметры и метрики

В качестве функции потерь была выбрана взвешенная категориальная кросс-энтропия. Функция активации скрытых слоев – ReLU, выходного слоя – Softmax. Размер пакета – 32. Метод оптимизации – Adam со скоростью обучения 0,0005. Обучение проводилось на 100 эпохах с уменьшением скорости обучения в 2 раза после 5 эпох без улучшений потерь.

Для оценки качества сегментации использовались метрики: индекс Жаккара (intersection over union, IoU), точность (precision), полнота (recall) и F1-мера.

Результаты

Были обучены 4 модели CHC: U-Net, U-Net++, SegNet и ENet.

На рис. 6 представлен график зависимости потерь и точности от эпохи для тренировочного и тестового наборов модели ENet. Обе кривые потерь (тренировочная и тестовая) снижаются с увеличением количества эпох, что свидетельствует о корректном обучении модели. Кривые потерь и точности стабилизируются после 20 эпох, что может указывать на достижение предела обучения. Разница между тренировочной и тестовой точностью незначительна после 8 эпох, что говорит об отсутствии явного переобучения. После 17 эпох тестовая точность незначительно отличается от тренировочной (на 0,01–0,04), что может указывать на достижение предела обучения и хорошую обобщающую способность модели. После 26 эпох обучения тестовая точность составила 0,87, что является достаточно высоким показателем для данной задачи сегментации.

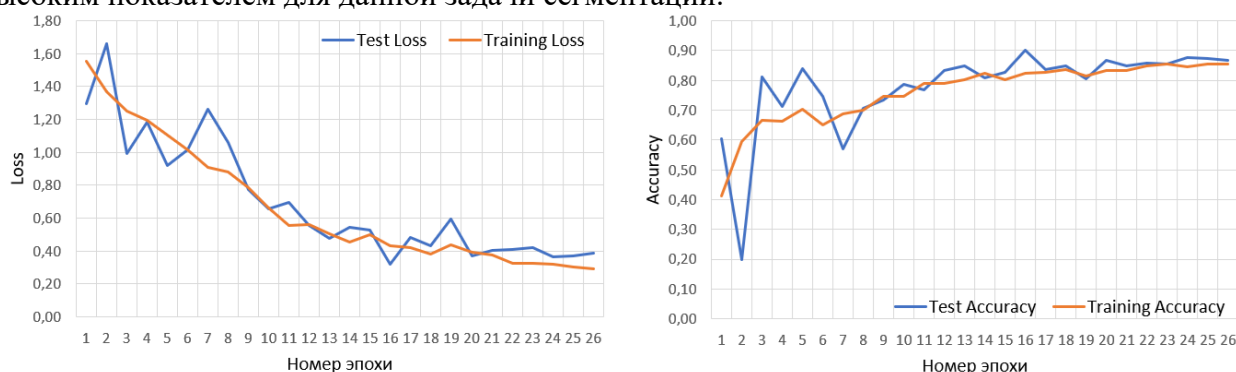


Рис. 6. График изменения потерь (слева) и точности (справа) модели ENet

В таблице 1 представлены результаты тестирования полученных моделей на тестовом наборе. Полужирным выделены лучшие модели по каждой метрике, усредненной по всем классам. Как видно, лучшие значения метрик F1 и mIoU показала модель SegNet.

Таблица 1. Метрики качества сегментации изображений

Модель	F1-мера	Precision	Recall	mIoU
U-Net	0,91	0,96	0,88	0,86
U-Net++	0,90	0,96	0,87	0,83
SegNet	0,93	0,96	0,90	0,88
ENet	0,92	0,96	0,90	0,87

В таблице 2 представлено распределение метрик по пяти классам сегментации. Полу жирным выделены лучшие значения метрики F1 для каждого класса.

Все модели показали высокие результаты для растительности ($F1 = 0,92\text{--}0,95$), что ожидаемо для доминирующего класса. Для редких классов (грунт, дороги, антропогенные объекты) метрики значительно ниже ($F1 = 0,20\text{--}0,58$), особенно для антропогенных объектов (наивысший $F1 = 0,58$ у SegNet). SegNet лидирует по F1 для 3 из 5 классов (растительность, гидрография, антропогенные объекты). ENet лучше всего справилась с грунтом ($F1 = 0,43$), но показала худший результат на антропогенных объектах ($F1 = 0,20$). U-Net++ показала лучший результат для дорог ($F1 = 0,32$), но это всё равно крайне низкое значение. Причиной является малый объем обучающих данных для редких классов и ошибки ручной разметки.

Таблица 2. Распределение метрик по классам сегментации

Класс	U-Net			U-Net++			SegNet			ENet		
	prec	recall	f1	prec	recall	f1	prec	recall	f1	prec	recall	f1
Растительность	0,99	0,89	0,94	0,99	0,86	0,92	0,99	0,91	0,95	0,99	0,90	0,95
Грунт	0,23	0,76	0,35	0,20	0,87	0,33	0,28	0,76	0,41	0,31	0,68	0,43
Дороги	0,16	0,80	0,26	0,20	0,75	0,32	0,15	0,78	0,26	0,14	0,74	0,23
Гидрография	0,87	0,96	0,91	0,73	0,99	0,84	0,90	0,96	0,93	0,76	0,97	0,85
Антропогенные	0,28	0,97	0,43	0,26	0,99	0,42	0,43	0,90	0,58	0,11	0,97	0,20

На рис. 7 представлены результаты сегментации моделью U-Net++. Видно, что модель довольно точно выделяет крупные объекты (растительность и водоёмы). При этом на мелких и редких объектах (дороги и грунт) модель часто ошибается.

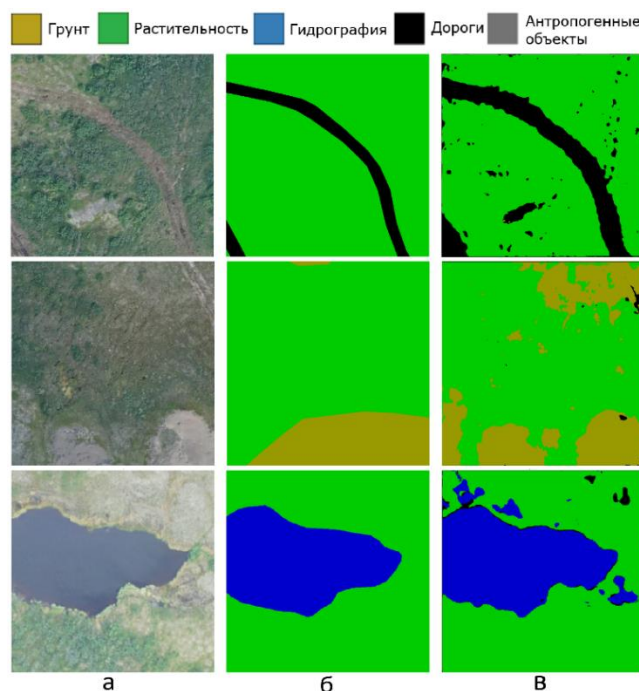


Рис. 7. Результаты сегментации моделью U-Net++: а – исходное изображение, б – эталонная карта сегментации, в – предсказание НС

SegNet продемонстрировала наилучшие результаты по метрикам F1 (0,93) и mIoU (0,88), что свидетельствует о её высокой точности и способности корректно сегментировать объекты даже при дисбалансе классов. U-Net++, несмотря на более сложную архитектуру, оказалась менее эффективной, что может быть связано с переобучением из-за недостаточного количества данных для редких классов [7].

На примере сегментации моделью U-Net++ видно, что сеть корректно выделяет крупные объекты господствующих классов (например, заросли кустарников), но допускает ошибки при разметке мелких объектов, а также на редких классах (дороги, антропогенные объекты) (рис. 7в). Это подтверждает необходимость дополнительной работы с датасетом, что включает увеличение количества обучающих примеров для редких классов и улучшение качества разметки [3]. Полученные результаты согласуются с данными литературы, где SegNet и ENet также отмечаются как эффективные архитектуры для задач сегментации [6]. Однако дисбаланс классов остается серьёзной проблемой, требующей применения методов аугментации, взвешенных функций потерь или синтетических данных. Низкие показатели для редких классов указывают на необходимость дальнейшей оптимизации, например, использования методов oversampling или transfer learning [3].

Заключение

В результате проведенного исследования были реализованы нейросетевые модели автоматизированного дешифрирования объектов земной поверхности на изображениях с БПЛА. Лучшие значения метрик F1 и mIoU получила модель SegNet. Низкие показатели метрик для классов: дороги, грунт и антропогенные объекты можно объяснить малым количеством объектов в обучающем наборе, а также некачественной разметкой. В результате анализа было установлено, что для повышения качества сегментации требуется: увеличение размера датасета для редких классов, использование методов transfer learning и более сложных архитектур, таких как DeepLab или HRNet, а также применение мультиспектральных изображений. Результаты работы могут быть применены в сельском хозяйстве, экологическом мониторинге и градостроительстве для автоматизации анализа изображений ДЗЗ. Это позволит сократить временные затраты и минимизировать субъективные ошибки ручной обработки.

Список использованных источников

1. Друки А.А., Спицын В.Г. Применение нейросетевых алгоритмов для семантической сегментации спутниковых снимков поверхности Земли // Вестник ТГУ. – 2023. – № 63. – С. 62-71.
2. Mo Y., Wu Y., Yang X., Liu F., Liao Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning // Neurocomputing. – 2022. – Vol. 493. – P. 626–646.
3. Yang S., Xiao W., Zhang M., Guo S., Zhao J., Shen F. Image Data Augmentation for Deep Learning: A Survey // arXiv. – [Электронный ресурс]. – URL: arxiv.org/abs/2204.08610 (дата обращения: 25.03.2025).
4. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI-2015). – 2015. – Vol. 93. – P. 234-241.
5. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. – [Электронный ресурс]. – URL: arxiv.org/abs/1511.00561/ (дата обращения: 25.03.2025).
6. Adam Paszke, Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. – [Электронный ресурс]. – URL: arxiv.org/abs/1606.02147 (дата обращения: 25.03.2025).
7. Zhou Z. et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation // IEEE Transactions on Medical Imaging. – 2018. – Vol. 37, № 5. – P. 1234-1244.