

УДК 004.855

ИССЛЕДОВАНИЕ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ ДЛЯ СЕГМЕНТАЦИИ ВОДНЫХ УЧАСТКОВ НА СПУТНИКОВЫХ SAR ИЗОБРАЖЕНИЯХ

Сенников А.В., аспирант кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: alexeysennikov76@yandex.ru

Ларионов Р.В., аспирант кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: rv.larionov@yandex.ru

Хрящев В.В., к.т.н., доцент кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: v.khryashchev@uniyar.ac.ru

Приоров А.Л., д.т.н., профессор кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: pri@uniyar.ac.ru

INVESTIGATION OF NEURAL NETWORK ALGORITHMS FOR SEGMENTATION OF WATER AREAS ON SATELLITE SAR IMAGES

Sennikov A.V., Larionov P.V., Xryashev V.V., Priorov A.L.

The article considers an algorithm for water area segmentation using satellite SAR images. The study used a set of 27 satellite images with a spatial size of approximately 200 by 300 kilometers with a resolution of 10 meters per pixel. Three models with the ResNet 34+U-Net, SegFormer_b5 and SegNeXt_I architectures are used as neural network models. The corresponding algorithm accepts patches with 2 channels – VV and VH – as input and produces a binary segmentation mask at the output. To evaluate the performance of the models, such metrics as Dice, F measure, accuracy and recall were used. The highest Dice value was 0.9. However, all models have difficulties in accurate segmentation of images at the boundaries of water surfaces, which leads to a large number of false positives. Also, within the framework of this study, an assessment of the impact of speckle noise on the quality of the neural network model was carried out, which showed that even with a noticeable increase in noise, measured by the PSNR metric, dropping to values of 9.65-9.86 dB, the model does not lose accuracy. Both for the original set and for the noisy one, the value of the Dice metric remains within 0.96-0.97, the F1 metric – within 0.81-0.82 and the Recall metric – within 0.97-0.98.

Key words: satellite image segmentation, SAR images, deep machine learning, neural networks, speckle noise.

Ключевые слова: сегментация спутниковых изображений, SAR изображения, глубокое машинное обучение, нейронные сети, спекл-шум.

Введение

В настоящее время использование данных дистанционного зондирования земли (ДЗЗ) нашло широкий спектр применения в различных отраслях экономики, например, для оценки лесной биомассы, управления сельскохозяйственными и природными ресурсами, в задачах геологии, для предупреждения стихийных бедствий [1-3]. Для решения подобных задач в настоящее время широко используются алгоритмы автоматической обработки изображений со спутников, полученных в разных диапазонах электромагнитного спектра, реализованные с помощью современных нейросетевых моделей [4-10].

В данной работе рассматривается задача определения водных поверхностей на спутниковых изображениях, полученных в радиодиапазоне. Она относится к задаче мониторинга водоемов, входящей в группу задач мониторинга земной поверхности, к которой относится также мониторинг сельскохозяйственных полей и лесного покрова. Одной из целей подобных исследований является мониторинг наводнений в режиме онлайн, а также оценка их масштаба.

Рассматривается алгоритм сегментации водных участков с использованием спутниковых SAR изображений. Для проведения исследования применялся набор из 27 спутниковых изображений, обладающих пространственным размером приблизительно 200 на 300 километров с разрешением 10 метров на пиксель. В качестве нейросетевых моделей используются три модели с архитектурами ResNet-34+U-Net, SegFormer_b5 и SegNeXt_I. Соответствующий алгоритм принимает на вход патчи с 2 каналами – VV и VH, а на выходе выдает бинарную маску сегментации. Для оценки работы моделей использовались такие метрики, как Dice, F-мера, точность и полнота. Наибольшее значение Dice составило 0,9. Однако все модели имеют трудности в точной сегментации изображений на границах водных поверхностей, что приводит к большому числу ложных пропусков. Также в рамках данного исследования проведена оценка влияния спекл-шума на качество работы нейросетевой модели, которая показала, что даже при заметном увеличении зашумленности, измеренной метрикой PSNR, опускающейся до значений 9,65-9,86 dB, модель не теряет точность работы. Как по исходному набору, так и по зашумленному значение метрики Dice остается в пределах 0,96-0,97, метрики F1 – в пределах 0,81-0,82 и метрики Recall – в пределах 0,97-0,98.

Специфика задачи мониторинга наводнений заключается в том, что они обычно сопровождаются сопутствующими негативными погодными условиями – дождями, сильным ветром, облачностью. В связи с этим

данные со спутника, полученные в оптическом диапазоне, где атмосферные помехи особенно значимы, являются не самым подходящим источником информации для мониторинга. Поэтому для такого рода задач часто используют спутниковые изображения в радиодиапазоне, полученные с помощью радаров с синтезированной апертурой (PCA, SAR). Благодаря им можно проводить исследования, игнорируя атмосферные условия [11-15]. Однако при этом на изображениях, получаемых с помощью систем PCA, присутствует высокий уровень шума, обусловленного природой обратного рассеяния радара (радиолокационной станции, РЛС). Самым большим источником шума на изображении PCA традиционно является спекл-шум [16], поэтому в данной работе также исследуется его влияние на качество работы нейросетевой модели.

Целью работы является разработка и исследование алгоритма сегментации водных участков на спутниковых изображениях в радиодиапазоне с использованием современных нейронных сетей и методов глубокого машинного обучения.

Описание набора спутниковых изображений

Для проведения исследования создан набор 16-битных радиолокационных изображений с использованием открытой электронной базы проекта Copernicus [17]. Каждое изображение охватывало пространственную область размером приблизительно 200 на 300 километров с разрешением 10 метров на пиксель. В состав каждого снимка входили два канала – VV и VH, предоставляющих информацию о вертикальной и горизонтальной поляризации соответственно. С целью улучшения качества изображений и повышения их информативности удалены тепловые шумы и шумы по краям, а также выполнена радиометрическая калибровка и геометрическая коррекция. Каждому изображению присвоена экспертная бинарная маска, определяющая области с водными поверхностями. Фрагменты из набора данных с соответствующими масками представлены на рис. 1. Для обучения из исходных больших изображений с ша-

ром 256 пикселей сформированы патчи размером 512x512 пикселей. Изображения переведены из 16-битного в 8-битный формат. Полученный набор данных поделен на обучающую и тестовую выборки в соотношении 100 к 8 (или 25 к 2). Более подробная информация об используемом наборе изображений представлена в табл. 1.

Описание архитектур используемых нейросетевых моделей

Для решения задачи сегментации водных участков использовано 3 различных нейросетевых модели. В качестве первой использована широко применяемая в различных задачах сегментации данных ДЗЗ сверточная нейронная сеть U-Net, на входе которой находится базовая сеть с архитектурой ResNet-34 [18]. В качестве двух других алгоритмов выбраны нейронные сети на базе трансформеров, в основе которых лежит наличие механизма внимания: SegFormer [19] и SegNeXt [20].

Первая относится к типу нейронных сетей, полностью работающих с использованием трансформеров. Их главной особенностью является наличие механизма внимания, позволяющего ускорить обучение и связать все вводимые данные между собой. Эта особенность дает возможность выполнять параллельную обработку данных и учитывать контекст. Сеть SegFormer состоит из кодировщика и декодировщика. Кодировщик выполняет роль извлечения грубых и точных признаков из входного изображения и состоит из блока разбиения на патчи и 4 блоков-трансформеров. Декодировщик в SegFormer объединяет многоуровневые признаки, выработанные кодировщиком, для прогнозирования маски сегментации.

В качестве третьей модели использовалась нейросетевая модель с архитектурой SegNeXt. Она также относится к типу трансформеров и является улучшением архитектуры SegFormer. В кодировщике SegNeXt вместо блока внимания из SegFormer используется многомасштабный модуль сверточного внимания – MSCA (multi-scale convolutional attention). Совместно с 2 слоями нормализации батча и сетью прямого распространения, мо-

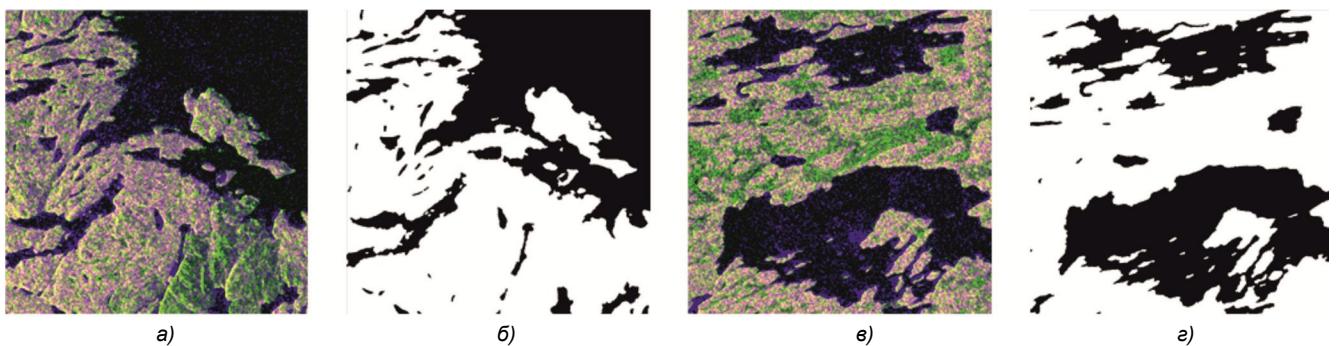


Рис. 1. Фрагменты изображений из набора (а, в) и соответствующие им маски (б, г)

Таблица 1. Статистика сформированного набора изображений

	Обучающая выборка	Тестовая выборка
Количество изображений	25	2
Количество патчей	178817	13714
Количество патчей с водными участками	97679	7202
Количество патчей без водных участков	72226	6512

дуль MSCA образует блок MSCAN, являющийся основным структурным элементом кодировщика в SegNeXt, аналогичный трансформерному блоку в SegFormer. Блок MSCAN также повторен 4 раза. Декодировщик в SegNeXt отличается тем, что объединяет признаки, полученные только с 3 последних этапов, а не со всех 4, как в SegFormer, что повышает производительность и отбрасывает низкоуровневую информацию с первого этапа. Кроме того, он отличается наличием операции разложения матриц извлеченных признаков.

Результаты компьютерного моделирования алгоритмов сегментации

Рассмотрим результаты обучения и тестирования представленных моделей. Все они перед итоговым обучением имели веса, полученные при их обучении на доступных наборах изображений ImageNet и DeepGlobe Land Cover [21]. Обучение на наборе ImageNet дало способность решать задачу поиска низкоуровневых признаков, а на наборе DeepGlobe – решать задачу сегментации на спутниковых снимках. Обучение происходило с размером батча 4. Каждая из моделей обучалась тремя разными способами. Первые два – с балансированной батчом (*Balance50x50*), а также без его балансировки (*noBalance*). Третий способ (*AI*) заключался в подходе повышения инвариантности модели к различным аугментациям. Для него брался тот способ балансировки, который показал наиболее высокие результаты.

Специфика спутниковых изображений заключается в том, что они не привязаны к определенным углам поворота, а также могут подвергаться достаточно серьезным фотометрическим искажениям из-за угла наклона камеры. Поэтому стандартные методы обучения нейронных сетей могут давать худшие результаты в задачах сегментации спутниковых изображений. Подход повышения инвариантности к таким преобразованиям заключается в том, что предсказания по исходному изображению должны совпадать с предсказаниями по аугментированному после их обратного преобразования [22]. Для обучения на основе такого подхода использовалась общая функция потерь, которую можно выразить как:

$$L_{tot} = L_{seg}(x, y) + L_{seg}(A(x), A(y)) + \lambda L_{AI}(x, A(x)),$$

$$L_{AI}(x, A(x)) = \frac{1}{I} \sum_{i \in I} [f_i(x) - A^{-1}(f_i(A(x)))]^2,$$

где L_{tot} – общая ошибка потерь, L_{seg} – ошибка сегментации, L_{AI} – разница между предсказаниями по изоб-

ражению с аугментациями и без них, выраженная как среднеквадратичная ошибка, A – набор аугментаций, x – исходное изображение, y – экспертная разметка, f_i – предсказание модели, λ – модулирующий параметр. В качестве аугментаций использовалось: изменение яркости, случайные повороты, искажение перспективы, отражение по вертикали и горизонтали.

Все обучаемые модели принимают на вход 2-канальный патч размером 512x512 пикселей, а на выходе выдают одноканальную бинарную маску с тем же размером, что и патч на входе.

Для объективной оценки качества работы нейросетевых моделей использовались стандартные метрики, такие как точность (precision, P), полнота (recall, R), F-мера (F1) и коэффициент Сёренсена (Dice). В табл. 2 приведены результаты тестирования трех моделей. Здесь TP (true positive) – количество истинно положительных предсказаний, FP (false positive) – количество ложноположительных предсказаний, FN (false negative) – количество ложных пропусков.

Видно, что все модели достаточно хорошо определяют границы найденных объектов, на что указывает высокое значение коэффициента Сёренсена. Наибольшим значением, равным 0,9, обладают модели SegFormer_b5 и SegNeXt_I, обученные с балансированной батчом 50x50. Лучшее значение полноты получилось у модели ResNet-34+U-Net. Она обладает наибольшим количеством истинных предсказаний. Подход повышения инвариантности, примененный при обучении SegFormer_b5 и SegNeXt_I на данном наборе изображений, привел к увеличению количества ложных срабатываний на 20-25% для обеих моделей. При этом количество истинных ответов и ложных пропусков изменилось незначительно. В связи с этим, у моделей, обученных с данным подходом, метрики точности и полноты хуже, чем у моделей, обученных без его использования. Кроме того, у всех моделей наблюдается большое количество ложных срабатываний. Это можно объяснить тем, что на изображениях много малых водных участков.

На рис. 2 на примере модели SegFormer_b5 представлен результат обработки двух изображений и приведено наглядное сравнение эталонной и предсказанных масок, которое подтверждает значения полученных выше метрик. На верхнем изображении большой участок сегментировался достаточно точно. Ошибки проявились на границе с земной поверхностью. Данный пример показывает ложные пропуски. На нижнем предсказании

Таблица 2. Результаты тестирования нейросетевых моделей

Нейросетевая архитектура	Метрика						
	Dice	F1	Precision	Recall	TP	FP	FN
ResNet-34+U-Net, noBalance	0,73	0,45	0,35	0,61	25157	45732	16270
ResNet-34+U-Net, Balance50x50	0,74	0,44	0,32	0,71	29261	60768	12166
SegFormer_b5, noBalance	0,89	0,39	0,30	0,54	22282	50646	19145
SegFormer_b5, Balance50x50	0,90	0,39	0,31	0,53	21963	49585	19464
SegFormer_b5, AI + Balance50x50	0,87	0,39	0,25	0,50	20682	63119	20745
SegNeXt_I, noBalance	0,89	0,40	0,31	0,58	24447	51684	17980
SegNeXt_I, Balance50x50	0,90	0,42	0,33	0,58	24231	49072	17196
SegNeXt_I, AI + Balance50x50	0,88	0,40	0,30	0,61	25089	59950	16338

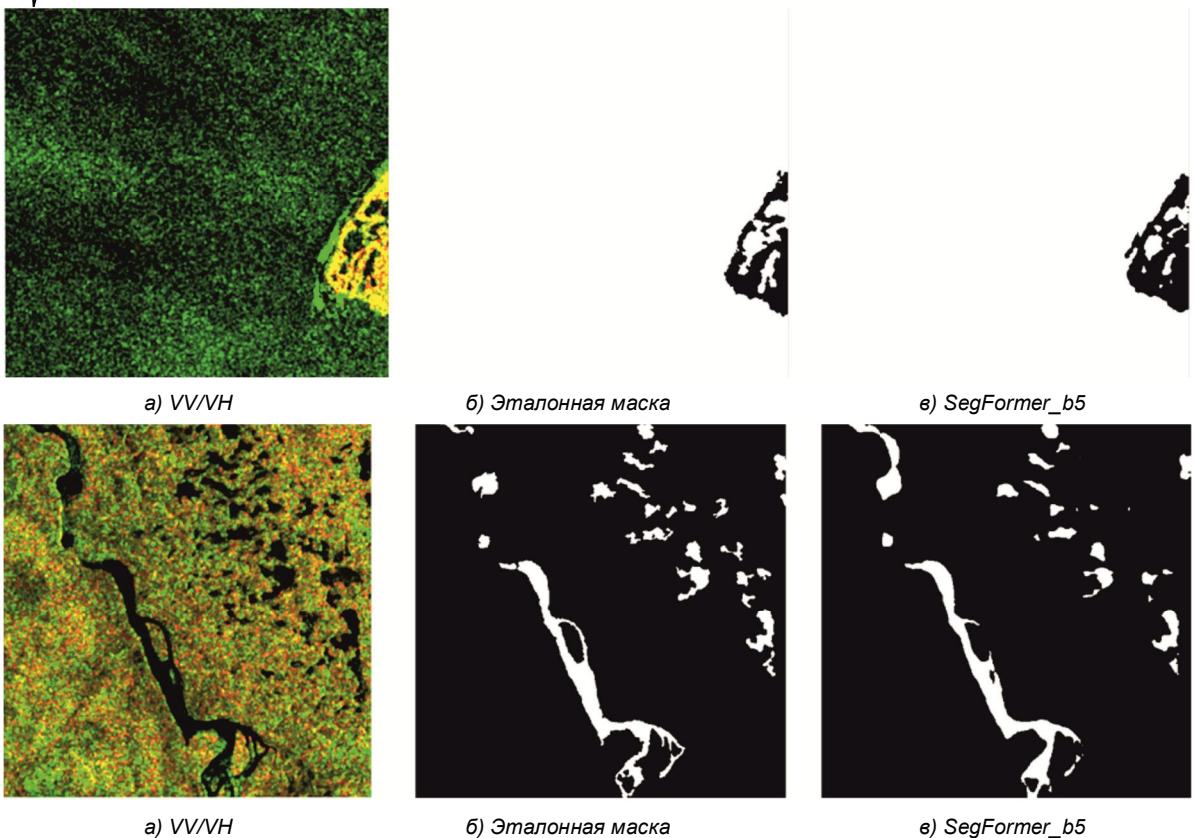


Рис. 2. Пример работы нейросетевого алгоритма для двух изображений

присутствуют ложные срабатывания, выраженные в том, что маленькие участки воды сегментируются более слажено, чем в эталонной маске.

Оценка влияния спекл-шума на работу алгоритмов сегментации

Проведено также исследование влияния спекл-шума на работу нейросетевых алгоритмов. Для исследования использовалась модель спекл-шума, выраженная следующей формулой:

$$\text{noise}_{\text{img}} = \text{img} + \text{img} * \text{gauss}(\text{mean}, \text{var}),$$

где $\text{noise}_{\text{img}}$ – зашумленное изображение, img – исходное изображение, $\text{gauss}(\text{mean}, \text{var})$ – случайный набор пикселей с размерностью изображения и гауссовским распределением с математическим ожиданием mean и дисперсией var . Для моделирования использовалось нулевое математическое ожидание и значения дисперсии – 0,01; 0,05; 0,1; 0,5; 1. Для каждого значения математического ожидания и дисперсии создавался соответствующий набор изображений. Примеры изображения с разной степенью зашумленности представлены на рис. 3. На изображениях var – значение дисперсии, используемой при зашумлении.

Далее на каждом таком наборе тестировалась модель SegFormer_b5. Результаты тестирования приведены в табл. 3.

Тестирование на данных наборах показало, что при увеличении шума заметно падает количество ложных срабатываний. Отсюда можно сделать вывод о том, что это происходит из-за того, что изображения имеют большое количество малых объектов, на которых модель часто ошибается, а зашумление дало эффект того,

что модель стала чаще пропускать эти объекты.

Далее использовались только такие тестовые изображения, на которых присутствовали лишь средние и большие объекты, а тестовый набор данных фильтровался по размеру объектов. Подготовлено 2 дополнительных набора данных – набор изображений с минимальным размером объекта в 1000 пикселей и набор с минимальным размером объекта в 3000 пикселей. Для порога в 1000 пикселей количество изображений составило 7562, а для порога в 3000 – 6989 изображений. Результаты тестирования приведены в табл. 4 и табл. 5.

В обоих случаях все метрики, кроме Precision, стали очень высокими. Метрика Precision осталась низкой из-за большого количества ложных срабатываний на малых объектах. Выход модели при данном тестировании не фильтровался по площади найденных объектов. Данные результаты также подтверждают предыдущие выводы о том, что модели плохо работают с малыми объектами. Также можно заметить, что даже при значительном увеличении шума модель является устойчивой к нему при сегментации средних и больших объектов, что подтверждает неизменность значений метрик при разных значениях добавленного шума.

Заключение

Разработан алгоритм сегментации водных участков с использованием SAR-изображений. Для реализации использовались 3 модели глубокого обучения – ResNet-34+U-Net, SegFormer_b5 и SegNeXt_I, каждая из которых обучалась тремя способами: первые два – с балансировкой батча, третий – без балансировки. Использован подход повышения инвариантности к аугментациям. Для

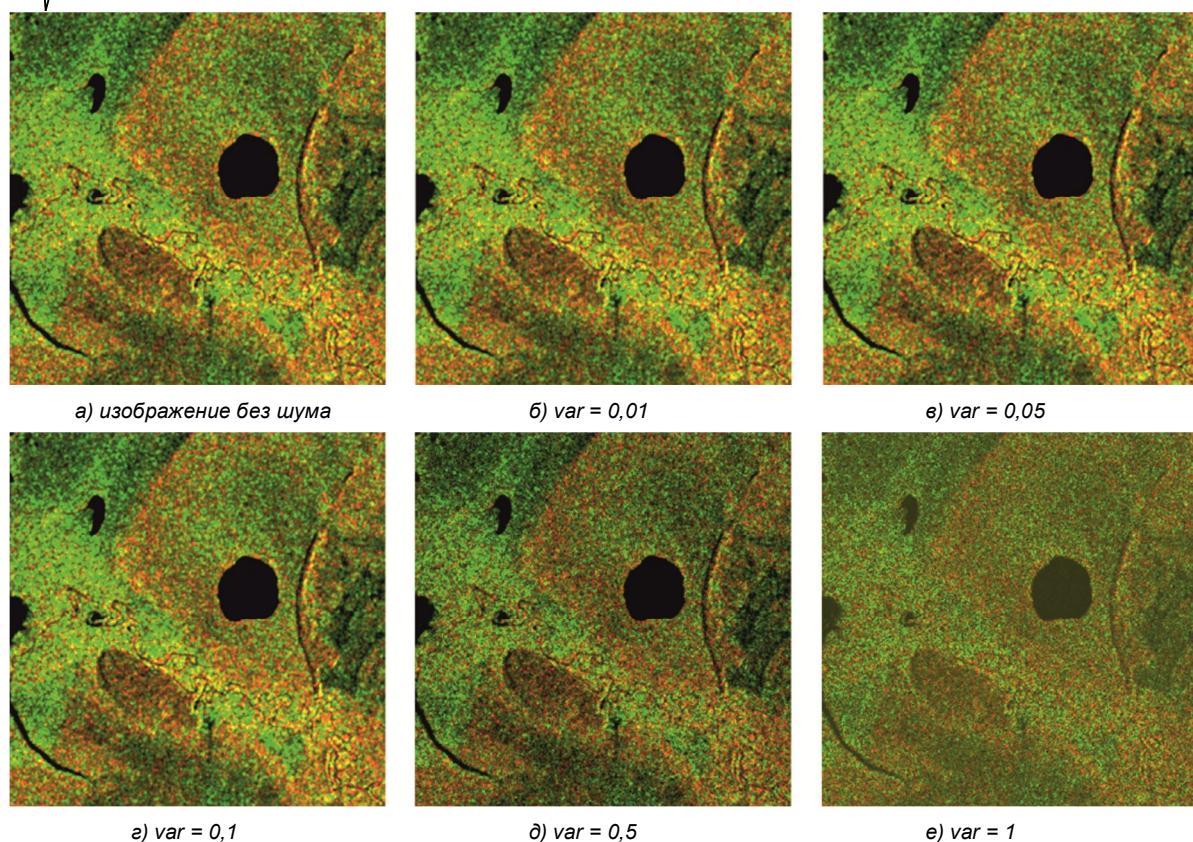


Рис. 3. Пример одного изображения с разной степенью зашумленности

Таблица 3. Тестирование модели SegFormer_b5 на зашумленных изображениях

mean	var	TP	FP	FN	Precision	Recall	F1	Dice
0	0	22282	50646	19145	0,31	0,54	0,39	0,86
0	0,01	21952	49603	19475	0,31	0,53	0,39	0,90
0	0,05	21938	49449	19489	0,31	0,53	0,39	0,90
0	0,1	21921	49482	19506	0,31	0,53	0,39	0,90
0	0,5	20855	44194	20572	0,32	0,50	0,39	0,90
0	1	17443	35738	23984	0,33	0,42	0,37	0,90

Таблица 4. Тестирование модели SegFormer_b5 на изображениях с минимальным размером объекта 1000 пикселей

mean	var	TP	FP	FN	Precision	Recall	F1	Dice
0	0	1487	1836	17	0,45	0,99	0,62	0,97
0	0,01	1487	1840	17	0,45	0,99	0,62	0,97
0	0,05	1487	1808	17	0,45	0,99	0,62	0,97
0	0,1	1487	1814	17	0,45	0,99	0,62	0,97
0	0,5	1484	992	20	0,60	0,99	0,75	0,97
0	1	1479	497	25	0,75	0,98	0,85	0,96

Таблица 5. Тестирование модели SegFormer_b5 на изображениях с минимальным размером объекта 3000 пикселей

mean	var	TP	FP	FN	Precision	Recall	F1	Dice
0	0	567	1294	6	0,30	0,99	0,47	0,98
0	0,01	567	1296	6	0,30	0,99	0,47	0,98
0	0,05	567	1275	6	0,31	0,99	0,47	0,98
0	0,1	567	1287	6	0,31	0,99	0,47	0,98
0	0,5	567	533	6	0,52	0,99	0,68	0,98
0	1	565	207	8	0,73	0,99	0,84	0,97

обучения подготовлен набор из 27 спутниковых изображений в радиодиапазоне, которые поделены на патчи размером 512x512 пикселей. Оценка качества работы обученных моделей происходила с помощью таких метрик как точность, полнота, F-мера и Dice. Исследование показало, что все 3 модели на хорошем уровне опреде-

ляют границы найденных участков, на что указывает высокое значение метрики Dice. Наибольшее ее значение (0,9) у моделей SegFormer_b5 и SegNeXt_I. Использование подхода повышения инвариантности к аугментациям на этом наборе изображений привело к увеличению числа ложных срабатываний на 20-25% для каждой

модели. Также установлено, что все модели в целом обладают большим количеством ложных срабатываний. В основном данные ошибки происходят на границах водных и земных участков, а также на малых водных участках.

Проведенный анализ влияния спектрального шума на качество работы моделей показал, что в данной задаче даже сильное его увеличение, выраженное опускающейся до значений 9,65-9,86 дБ метрикой PSNR, не влияет на значение метрик оценки качества работы модели. Как по исходному набору, так и по зашумленному значение метрики Dice остается в пределах 0,96-0,97, метрики F1 – в пределах 0,81-0,82, а метрики Recall – в пределах 0,97-0,98.

Литература

1. Гук А.П., Евстратова Л.Г. Дистанционное зондирование и мониторинг территорий. Ч. 1. Дистанционное зондирование. Теоретические основы и технические средства: учебник. М.: КУРС. 2024. 224 с.
2. Шовенгердт Р.А. Дистанционное зондирование. Методы и модели обработки изображений. М.: Техносфера. 2013. 624 с.
3. Khryashchev V., Larionov R. Wildfire Segmentation on Satellite Images using Deep Learning. 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT). pp. 1-5.
4. Аллатов Б.А., Ершов М.Д., Фельдман А.Б. Алгоритм обработки изображений для системы комбинированного видения летательного аппарата. Цифровая обработка сигналов. 2015. № 3. С. 8-14.
5. Барталев С.А., Ховратович Т.С. Анализ возможностей применения методов сегментации спутниковых изображений для выявления изменений в лесах. Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 1. С. 44-62.
6. Ерохин Д.Ю., Ершов М.Д. Современные сверточные нейронные сети для обнаружения и распознавания объектов. Цифровая обработка сигналов. 2018. № 3. С. 64-69.
7. Иванов Е., Тищенко И., Виноградов А. Сегментация мультиспектральных снимков с применением сверточных нейронных сетей. Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 1. С. 25-34.
8. Хрящев В.В., Приоров А.Л., Павлов В.А., Ларинов Р.В. Модификация архитектуры сверточной нейронной сети U-Net в задаче сегментации мультиканальных спутниковых изображений. Цифровая обработка сигналов. 2019. № 3. С. 39-45.
9. Романов А.А., Рубанов К.А. Сравнение методов объектно-ориентированной и нейросетевой классификации данных дистанционного зондирования Земли на основе материалов систем Landsat-5 и Orbview-3. Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9. № 4. С. 29-36.
10. Родионова Н.В. Анализ изображений Sentinel 1 для весеннего паводка в Алтайском крае в апреле 2015 года и Рязанской области в апреле 2016 года. Современные проблемы дистанционного зондирования Земли из космоса. 2017. Т. 14. №. 1. С. 136-146.
11. Colin A. et al. Semantic segmentation of metoceanic processes using SAR observations and deep learning. Remote Sensing. 2022, vol. 14, no. 4, pp. 851.
12. Lv S., Meng L., Edwing D., Xue S., Geng X., Yan X.-H. High-performance segmentation for flood mapping of HISEA-1 SAR remote sensing images. Remote Sens. 2022. pp. 1-19.
13. Paul S., Ganju S. Flood segmentation on sentinel-1 SAR imagery with semi-supervised learning. arXiv preprint arXiv:2107.08369. 2021.
14. Сенинков А.В., Ларинов Р.В., Хрящев В.В. Разработка алгоритма сегментации водных участков на спутниковых изображениях в радиодиапазоне. Цифровая обработка сигналов и ее применение (DSP-2024): докл. 26-й междунар. конф. Москва, 2024. С. 220-225.
15. Khryashchev V., Kotov N., Larionov R., Priorov A. Semantic segmentation of satellite SAR-C images using deep learning based algorithms. 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications "SYNCHROINFO-2022". Arkhangelsk. Russia. Conference Proceedings. 2022. pp. 1-5.
16. Li H., Huang B., Huang X. A level set filter for speckle reduction in SAR images. EURASIP Journal on Advances in Signal Processing. 2010, vol. 2010, pp. 1-14.
17. Copernicus [Электронный ресурс]. URL: <https://www.copernicus.eu/en> (дата обращения 27.10.2024).
18. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. Springer. 2015, vol. 9351, pp. 234-241. doi: 10.1007/978-3-319-24574-4_28.
19. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / Enze X. et al. Neural Information Processing Systems (NeurIPS). 2021, pp. 1-14.
20. SegNeXt: Rethinking convolutional attention design for semantic segmentation / Guo M. et al. Neural Information Processing Systems (NeurIPS). 2022, pp. 1-15.
21. Demir I., Koperski K., Lindenbaum D. (2018) Deepglobe 2018: A challenge to parse the earth through satellite images. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
22. Tavera A., Arnaudo E., Masone C., Caputo B. Augmentation Invariance and Adaptive Sampling in Semantic Segmentation of Agricultural Aerial Images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2022, pp. 1656-1665.