

ДИСТАНЦИОННОЕ ЗОНДИРОВАНИЕ АТМОСФЕРЫ, ГИДРОСФЕРЫ И ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ

УДК 551.345

Применение сверточной нейронной сети U-Net и ее модификаций для сегментации тундровых озер на спутниковых оптических изображениях

И.А. Абрамова¹, Д.М. Демчев², Е.В. Харюткина^{2,3},
Е.Н. Савенкова^{2,4}, И.А. Судаков^{5*}

¹Арктический и антарктический научно-исследовательский институт
199397, г. Санкт-Петербург, ул. Беринга, 38

²ООО «ЦНИР» — Центр исследований и разработок
173008, г. Великий Новгород, ул. Рабочая, 18

³Институт мониторинга климатических и экологических систем СО РАН
634055, г. Томск, пр. Академический, 10/3

⁴Российский государственный гидрометеорологический университет
192007, г. Санкт-Петербург, ул. Воронежская, 79

⁵School of Mathematics and Statistics, The Open University
Milton Keynes MK7 6AA, UK

Поступила в редакцию 27.09.2023;
после доработки 24.10.2023;
принята к печати 01.11.2023

Тундровые озера — важный индикатор изменений климата, поэтому анализ динамики их размеров представляет особый интерес. В работе представлены результаты применения сверточной нейронной сети U-Net для сегментации тундровых озер по спутниковым оптическим изображениям на примере данных Landsat. Выполнена сравнительная оценка точности сегментации с помощью оригинальной архитектуры U-Net и ее модификаций: U-Net++, Attention U-Net и R2 U-Net, в том числе с использованием весов, полученных на основе предварительно обученной сети VGG16. Точность сегментации оценивается путем сравнения с результатами ручного картирования тундровых озер в северных районах Сибири. Обучение сетей производилось на основе 500 изображений Landsat; 250 изображений использовались в качестве тестовых. Показано, что более современные модификации U-Net не дают практически значимого выигрыша в точности сегментации, но увеличивают вычислительные затраты. Наилучший результат показывает конфигурация на основе классической реализации U-Net (средний коэффициент Жаккара IoU = 0,88). Предложенный метод и полученные оценки могут использоваться при изучении тенденций современного климата Земли.

Ключевые слова: тундровые озера, U-Net, Арктика, дистанционное зондирование, вечная мерзлота; tundra lakes, U-Net, Arctic, remote sensing, permafrost.

Введение

Тундровые озера являются не только источниками выбросов метана в атмосферу и накопителями двуокси углерода, но и важным индикатором изменений климата. Поэтому изучение динамики их геометрических размеров представляет особый интерес. Дистанционное зондирование, в том числе со спутников, — эффективный инструмент мониторинга состояния тундровых озер, так как натурные наблюдения в труднодоступных районах Арктики

в силу затратности развертывания и поддержания оборудования не могут выполняться систематически на больших площадях. К настоящему времени накоплены ряды спутниковых наблюдений за 30 и более лет, что позволяет анализировать динамику особенностей типов подстилающей поверхности в области многолетней мерзлоты, в том числе на климатических масштабах.

В работе [1] показана возможность автоматической сегментации тундровых озер по спутниковым оптическим изображениям на основе применения пороговых значений яркости и метода growing region.

Перспективным источником данных о состоянии тундровых озер являются и спутниковые радиолокационные изображения, по которым также возможно выделять области, относящиеся к тундровым озерам, причем вне зависимости от наличия

* Ирина Александровна Абрамова (calluna.mirabilis@gmail.com); Денис Михайлович Демчев (demchev@gmail.com); Елена Валерьевна Харюткина (kh_ev@mail2000.ru); Елена Николаевна Савенкова (savenkova.en@mail.ru); Иван Алексеевич Судаков (sudakow.ivan@gmail.com).

облачного покрова и освещенности земной поверхности. Это является существенным преимуществом радиолокационных данных перед оптическими.

В работе [2] продемонстрирована возможность применения «метода k -средних» автоматической кластеризации в сочетании с предварительной низкочастотной фильтрацией для сегментации областей, покрытых льдом или водой, в северных районах Сибири. В недавнем исследовании [3] показана перспективность использования спутниковых радиолокационных данных Sentinel-1 для сегментации тундровых озер и последующего анализа их геометрических и фрактальных размеров. Алгоритм сегментации в [3] основан на применении сверточной нейронной сети U-Net [4]. Эта сеть относится к методам глубокого обучения и изначально была разработана для обработки медицинских изображений, но затем стала активно использоваться и в других областях, где требуется сегментация изображений, в частности, для выделения на снимках, полученных по данным дистанционного зондирования, областей с ледяным покровом или водными образованиями в приполюсных районах [5, 6].

Цель настоящей работы — анализ применения метода на основе U-Net [4] для автоматического распознавания спутниковых изображений тундровых озер, полученных в оптическом диапазоне. Такой тип задач относится к классу сегментации изображений.

Процесс сегментации изображения представляет собой выделение из изображения одной или нескольких связных областей, удовлетворяющих критерию однородности признаков, которые вычисляются из значений нескольких цветовых компонентов. Для изображений оптического диапазона такими признаками могут выступать спектральное альбедо и яркость.

Основные требования при решении задач сегментации следующие: области должны быть однородны относительно значений цветовых компонент; внутренние части областей не должны содержать большого количества пустот; границы каждого сегмента должны быть пространственно точными. В случае спутниковых данных выполнение этих требований может быть затруднено из-за наличия облачного покрова, нечетких границ исследуемых объектов, недостаточной освещенности земной или водной поверхности, наличия инструментальных шумов и пр. Современные методы глубокого обучения являются устойчивыми к подобным шумовым проявлениям даже при использовании небольших выборок тренировочных данных, что позволяет получать надежные описания особенностей на изображениях.

Эффективность применения U-Net для сегментации тундровых озер оценивается в настоящей работе путем сравнения с ее более современными модификациями на примере оптических данных спутниковой системы Landsat. Предложенный метод сегментации и полученные оценки могут использоваться в исследованиях тенденций современного климата Земли. Даются практические рекомендации

по выбору оптимальной архитектуры U-Net с учетом вычислительных затрат и точности результата сегментации.

1. Объекты и методы исследования

1.1. Спутниковые данные и их обработка

В работе используются спутниковые изображения Landsat, полученные над северными районами Сибири летом 2016 г., отобранные с помощью системы Google Earth Engine по критерию наличия фракции облачности до 5%. Изображения представляют собой RGB-композиции каналов B4(0,64–0,67 мкм) + B3(0,53–0,59 мкм) + B2(0,45–0,51 мкм) с пространственным разрешением 30 м в 1 пикселе. На изображениях интерактивным методом размечались области, покрытые тундровыми озерами (одна область — одно озеро), в бинарном формате, где значение пикселя 255 соответствуют областям тундровых озер, 0 — остальным элементам земной поверхности или облачности. Далее исходные спутниковые изображения и изображения-маски тундровых озер разбивались на серии изображений размером 256×256 точек для нейронной сети. Всего для обучения и валидации было подготовлено 750 изображений, из которых 250 использовались в качестве тестовых для оценки точности сегментации.

1.2. Сегментация тундровых озер на основе U-Net

В настоящей работе применяется метод сегментации спутниковых оптических изображений тундровых озер на основе U-Net [4] с помощью алгоритма, предложенного в [3]. В этой статье была продемонстрирована перспективность данного подхода на примере спутниковых данных радиолокационного диапазона; в качестве признаков использовались интенсивность обратного рассеяния при различных поляризационных режимах радиолокатора с синтезированной апертурой (РСА) и угол зондирования. Полученные результаты позволили выполнить оценки геометрических и фрактальных размеров озер.

Метод, предложенный в [4], базируется на классической реализации полносвязной сверточной сети, которая получила название U-Net из-за схожести своей архитектуры с буквой U (рис. 1, а). Данная сеть по сути является кодировщиком-декодером с глубоким обучением, в котором подсети кодера и декодера соединены серией вложенных слоев. U-Net демонстрирует хорошие результаты при решении задач сегментации на ограниченных наборах обучающих образов; она не требовательна к вычислительным ресурсам.

В настоящей работе классическая архитектура U-Net является базовым методом сегментации. Ее эффективность оценивается в сравнении с более поздними модификациями оригинальной архитектуры U-Net: U-Net++ [7], Attention U-Net [8]

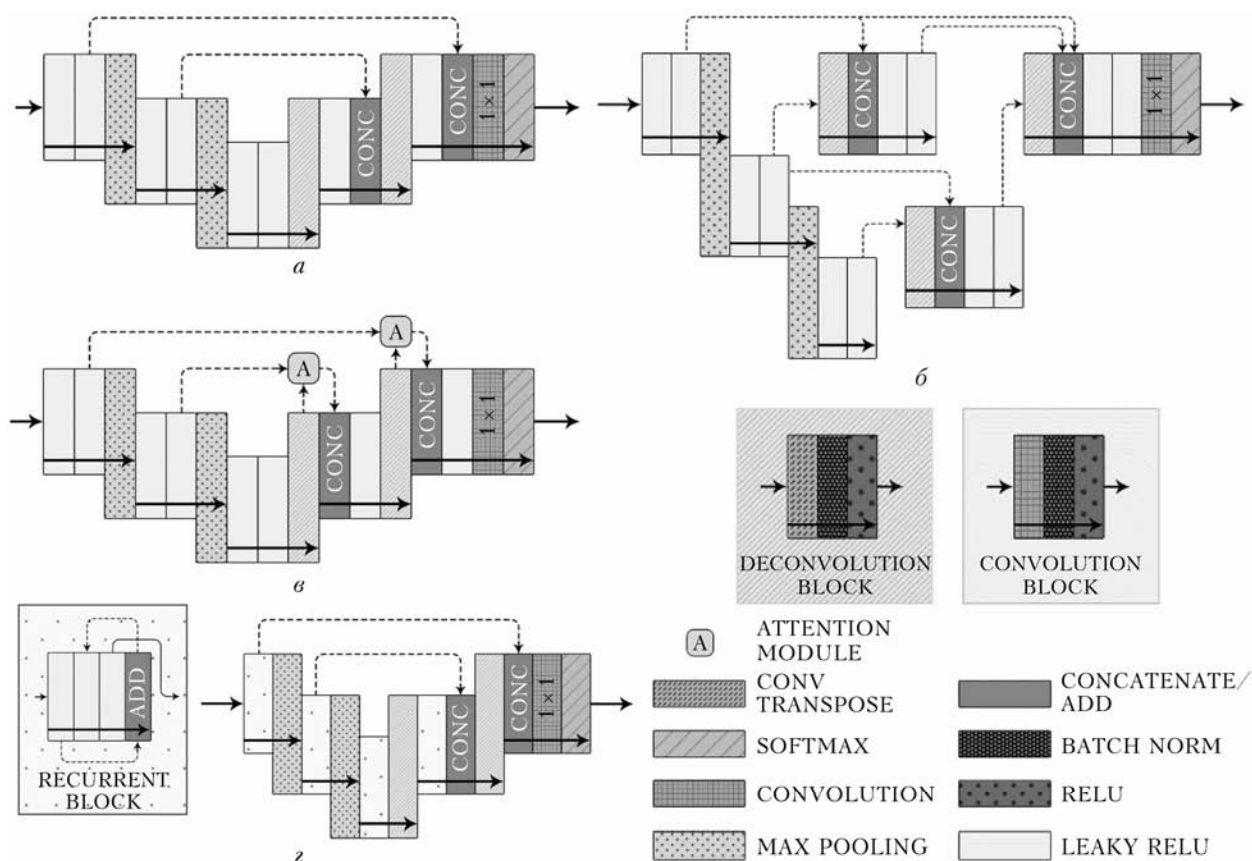


Рис. 1. Архитектуры модификаций U-Net, рассмотренные в данном исследовании: *a* – классическая архитектура U-Net; *б* – U-Net++; *в* – Att. U-Net; *г* – R2 U-Net. Стрелками обозначены потоки данных, а прямоугольниками – соответствующие сверточные слои (ATTENTION MODULE – формирование матрицы весов важности; CONV TRANSPOSE – свертка с транспонированием; SOFTMAX – функция преобразования вектора чисел в вектор вероятностей; CONVOLUTION – операция свертки; MAX POOLING – поиск локальных максимумов; CONCATENATE/ADD – конкатенция/сложение; BATCH NORM – нормализация слоев; RELU – линейная функция активации, определенная для положительной области значений; LEAKY RELU – функция активации на основе RELU, характеризующаяся ненулевыми значениями в отрицательной области; DECONVOLUTION BLOCK – операция транспонированной свертки с обучением; CONVOLUTION BLOCK – операции свертки)

и R2 U-Net [9], которые потенциально могут улучшить качество выделения тундровых озер на спутниковых изображениях. Эти модификации широко применяются в практике сегментации изображений, но характеризуются более сложной архитектурой и большими вычислительными затратами. Перечисленные модификации U-Net схематически изображены на рис. 1, б–г. Кратко рассмотрим особенности каждой из них.

U-Net++ (рис. 1, б) использует набор плотных пропускных соединений, что позволяет улучшить агрегирование функций на разных семантических масштабах.

Attention U-Net (далее Att. U-Net) (рис. 1, в) использует модуль Attention в каждом пропускном соединении. Он позволяет сделать акцент на областях, относящихся к исследуемому классу объектов, которые являются более релевантными для обучения сети, чем фоновые области.

Рекуррентно-остаточная модификация (recurrent-residual) R2 U-Net (рис. 1, г) сочетает в себе преимущества остаточного обучения и рекуррентных соединений. Последнее позволяет улучшить

накопление функций и использование параметров. В R2 U-Net остаточное обучение выполняется на каждом из слоев, добавляя короткие соединения, которые направлены на оптимизацию градиентного потока и работают в соответствии с теорией, согласно которой обучение на остаточном образе более эффективно, чем на исходном. Таким образом, R2 U-Net делает возможным накопление описательных признаков и более эффективное использование параметров в сочетании с преимуществами остаточного обучения.

Для некоторых конфигураций на базе U-Net применялись тренировочные коэффициенты (веса), полученные с помощью предобученной сети VGG16 на основе массива размеченных данных Imagenet [10].

Обучение сетей производилось путем минимизации перекрестной энтропии для 100 эпох с использованием оптимизации Адама [11] (10^{-3}) и размером набора (batch size) 8. Коэффициент Серенсена (Dice) применялся для оценки схожести двух выборок данных – в нашем случае предсказанной и полученной экспертным картированием. Он вычисляется как

$$\text{Dice} = \frac{2\text{TR}}{2\text{TR} + \text{FP} + \text{FN}}, \quad (1)$$

где TR — истинно классифицированные пиксели; FP — ложно классифицированные как принадлежащие определенному классу пиксели; FN — ложно классифицированные как не принадлежащие к определенному классу пиксели.

Оценку точности сегментации для каждой из архитектур проводили с помощью коэффициента Жаккара

$$\text{IoU} = \frac{\text{TR}}{\text{TR} + \text{FP} + \text{FN}}. \quad (2)$$

Все эксперименты осуществлялись на ОС Linux Ubuntu 18.04.6 с использованием библиотек Tensorflow 2.10.0 и keras_unet_collection, реализованных на Python 3.10.9. Вычисления выполнялись на 40 процессорах Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz.

Для тренировки использовались шесть конфигураций на основе различных архитектур U-Net:

- 1) U-Net (с весами);
- 2) U-Net++ (с весами);
- 3) Att. U-Net (с весами);
- 4) U-Net (без весов);
- 5) R2 U-Net (без весов);
- 6) Att. U-Net (без весов).

2. Результаты и обсуждение

Результаты сегментации тундровых озер оценивались для всех шести конфигураций сети на основе обучения по 500 предварительно размеченным изображениям Landsat размером 256×256 точек.

Графики коэффициента Dice обучающих и тестовых данных для 100 эпох приведены на рис. 2. Для всех конфигураций была получена хорошая сходимость результатов, кроме конфигурации 2 (U-Net++). При этом минимальное время обучения ~ 10 ч было получено для конфигураций 1 и 4, так как они базируются на оригинальной версии U-Net, в которой нет дополнительных этапов обработки, присутствующих в более поздних модификациях U-Net.

Оценка точности сегментации выполнялась на тестовой выборке из 250 изображений, не использовавшихся при обучении. Результаты расчета средних значений IoU приведены в таблице.

Точность сегментации на основе данных Landsat

Номер конфигурации	Тип сети	IoU	Dice
1	U-Net (с весами)	0,873	0,844
2	U-Net++ (с весами)	0,710	0,595
3	Att. U-Net (с весами)	0,871	0,837
4	U-Net	0,879	0,854
5	R2 U-Net	0,860	0,837
6	Att. U-Net	0,829	0,783

Наибольшая точность сегментации была получена с помощью конфигураций оригинальной версии U-Net (1 и 4) и конфигурации 3; наименьшая — с использованием конфигурации 2.

Для иллюстрации эффективности конфигураций на основе U-Net рассмотрим результаты сегментации одного из тестовых изображений, приведенного на рис. 3, а. Результат ручной разметки области тундровых озер изображен на рис. 3, б, а сегментации — на рис. 3, в–з.

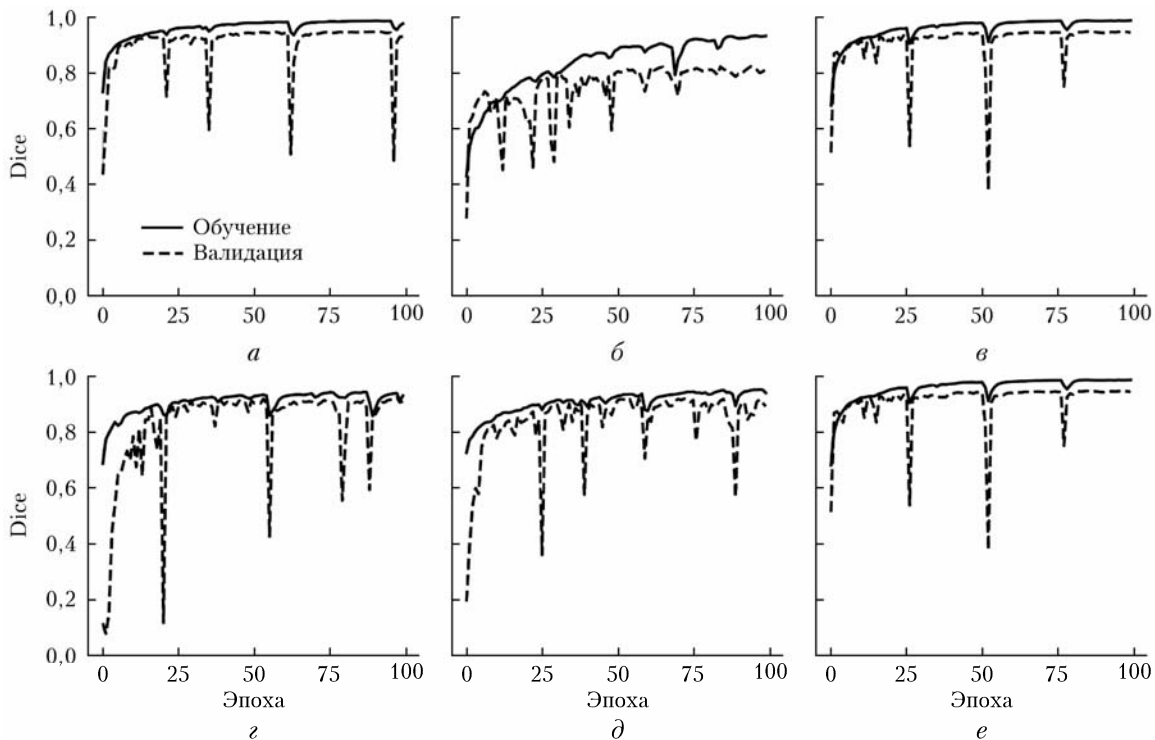


Рис. 2. Ход обучения для шести конфигураций на основе U-Net: а — U-Net (с весами); б — U-Net++ (с весами); в — Att. U-Net (с весами); г — U-Net; д — R2 U-Net; е — Att. U-Net

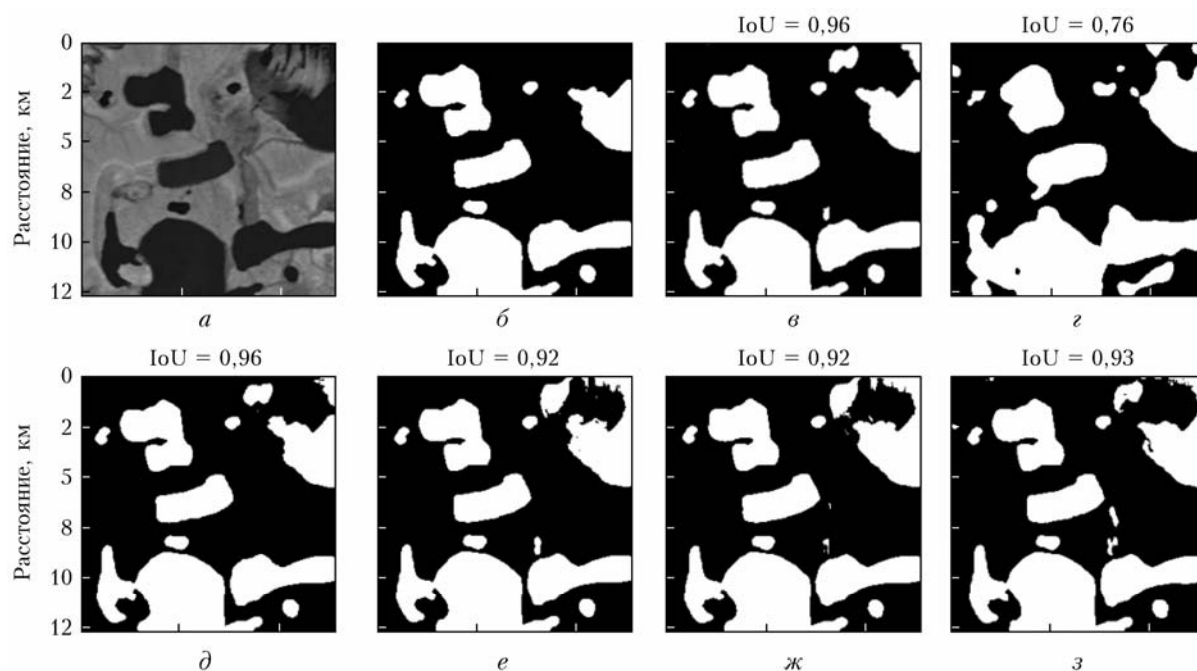


Рис. 3. RGB-изображение Landsat (*а*); результаты ручного картирования тундровых озер на его основе (*б*) и автоматической сегментации с использованием конфигурации 1 (*в*), 2 (*г*), 3 (*д*), 4 (*е*), 5 (*ж*) и 6 (*з*)

Наибольшая точность достигнута с помощью U-Net (с весами) и Att. U-Net (с весами). Она на 4% лучше, чем для U-Net без предобученных весов ($\text{IoU} = 0,92$). Использование весов в случае Att. U-Net также позволило увеличить точность сегментации на 3% по сравнению с версией без предобучения ($\text{IoU} = 0,96$ и $0,93$, соответственно). Наименьшая эффективность сегментации у U-Net++ ($\text{IoU} = 0,76$). С помощью R2 U-Net без предобученных весов была достигнута точность, сопоставимая с конфигурацией 5 ($\text{IoU} = 0,92$).

Эти оценки позволяют сделать вывод, что использование стандартной версии U-Net достаточно для успешной сегментации изображений тундровых озер в оптическом диапазоне ($\text{IoU} = 0,88$). Более современные модификации U-Net не дают практически значимого увеличения точности сегментации, особенно с учетом большей сложности и меньшей скорости вычислений, что может иметь существенное значение при обработке больших массивов данных. Для всех рассмотренных конфигураций отмечена сопоставимая точность. Так, $\text{IoU} = 0,71$ для U-Net++, $0,86$ для R2 U-Net и $0,87$ для Att. U-Net. При этом использование более сложных конфигураций закономерно потребовало больших вычислительных затрат.

Можно также отметить, что применение весовых коэффициентов предварительно натренированной сети VGG16 позволило улучшить результат для Att. U-Net с $0,83$ до $0,87$. Для оригинальной версии U-Net такого увеличения точности не наблюдалось — разница полученных результатов оказалась в пределах $0,006$.

Коэффициенты IoU или Dice при оценке качества сегментирования изображений важны для объ-

ективной оценки результата, однако при решении исследовательских или практических задач, а также при разработке программных реализаций алгоритмов сегментации следует учитывать и другие факторы, например скорость обучения, вычислительные затраты.

Применение оригинальной конфигурации U-Net для сегментации тундровых озер является самым вычислительно эффективным решением с точки зрения обработки обучающего набора данных, так как она обладает наименьшим количеством параметров, а следовательно, самой низкой сложностью, при этом ее производительность сопоставима и даже превосходит рассмотренные более современные модификации. Следует заметить, что потенциальные небольшие улучшения в производительности могут оказаться неэффективными, если они достигаются за счет значительного увеличения сложности, что неизбежно связано со снижением скорости вычислений. Так, точность сегментации с помощью Att. U-Net сопоставима со стандартной U-Net, но использование Att. U-Net связано со значительно более медленным обучением из-за дополнительных вычислений на каждом этапе повторного применения слоев сети.

Заключение

В настоящей работе рассмотрены шесть модификаций сети U-Net с использованием предобученных весов сети VGG16 для решения задачи сегментации тундровых озер по спутниковым оптическим снимкам на примере данных Landsat. Точность сегментации оценивалась метриками IoU и Dice. Все рассмотренные архитектуры U-Net показали хоро-

шие (Att. U-Net, R2 U-Net) или удовлетворительные (U-Net++) результаты. Наилучший результат был достигнут с помощью стандартной реализации U-Net (IoU = 0,88).

Предобученные веса позволили повысить точность для Att. U-Net, но не для U-Net. Наименьшая точность сегментации получена с помощью сети U-Net++ (IoU = 0,71). Мы рекомендуем стандартную версию U-Net как эффективный метод решения рассматриваемой задачи. Применение других, более сложных, архитектурных модификаций U-Net не приводит к повышению точности сегментации и связано с большими вычислительными затратами.

Результаты работы могут быть использованы для оптимизации климатических расчетов с помощью сети U-Net на основе спутниковых данных, полученных в оптическом диапазоне.

Финансирование. Исследование выполнено при финансовой поддержке гранта РНФ (проект № 21-71-10052).

Список литературы

1. *Sudakov I., Essa A., Mander L., Gong M., Kariyawasam T.* The geometry of large tundra lakes observed in historical maps and satellite images // *Remote Sens.* 2017. V. 9, N 10. P. 1072.
2. *Sobiech J., Dierking W.* Observing lake- and river-ice decay with SAR: advantages and limitations of the unsupervised *k*-means classification approach // *Ann. Glaciol.* 2013. V. 54, N 62. P. 65–72.
3. *Demchev D., Sudakov I., Khodos A., Abramova I., Lyakhov D., Michels D.* Recognizing the Shape and Size of Tundra Lakes in Synthetic Aperture Radar (SAR) Images Using Deep Learning Segmentation // *Remote Sens.* 2023. V. 15, N 5. P. 1298.
4. *Ronneberger O., Fischer P., Brox T.* U-net: Convolutional networks for biomedical image segmentation // *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015.* P. 234–241.
5. *Sudakov I., Asari V.K., Liu R., Demchev D.* MeltPondNet: A Swin Transformer U-Net for Detection of Melt Ponds on Arctic Sea Ice // *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. IEEE*, 2022. V. 15. P. 8776–8784.
6. *Wang Y.-R., Li X.-M.* Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning // *Earth Syst. Sci. Data.* 2021. V. 13, N 6. P. 2723–2742.
7. *Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J.* U-Net++: A nested U-Net architecture for medical image segmentation // *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018.* V. 11045. P. 3–11.
8. *Oktay O., Schlemper J., Folgoc L. Le, Lee M., Heinrich M., Misawa K., Mori K., McDonagh S., Hammerla N.Y., Kainz B., Clodcker B., Rueckert D.* Attention U-Net: Learning where to look for the pancreas // *Arxiv: Computer Vision and Pattern Recognition.* DOI: 10.48550/arXiv1804.03999.
9. *Alom M.Z., Hasan M., Yakopcic C., Taha T.M., Asari V.K.* Recurrent residual convolutional neural network based on U-Net (R2 U-Net) for medical image segmentation // *Arxiv: Computer Vision and Pattern Recognition.* DOI: 10.48550/arXiv1802.06955.
10. *Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.* Imagenet: A large-scale hierarchical image database // *IEEE Conf. Comp. Vision and Patt. Recog.* 2009. P. 248–255.
11. *Kingma D.P., Ba J.* Adam: A method for stochastic optimization // *Arxiv Machine Learning.* DOI: 10.48550/arXiv1412.6980.

I.A. Abramova, D.M. Demchev, E.V. Kharyutkina, E.N. Savenkova, I.A. Sudakov. **Use of the U-Net convolutional neural network and its modifications for segmentation of tundra lakes in satellite optical images.**

Tundra lakes are an important indicator of climate change; therefore, the analysis of the dynamics of their size is of particular interest. This paper presents the results of using the U-Net convolutional neural network for tundra lakes segmentation in satellite optical images using Landsat data as an example. The comparative assessment of segmentation accuracy is performed for the original U-Net design and its modifications: U-Net++, Attention U-Net, and R2 U-Net, including with weights derived from a pre-trained VGG16 network. The segmentation accuracy is assessed based on the results of manual mapping of tundra lakes in northern Siberia. It is shown that more recent U-Net modifications do not provide a practically significant gain in segmentation accuracy, but increase the computational costs. A configuration based on the classic U-Net gives the best result in most cases (the average Soerens coefficient IoU = 0.88). The technique suggested and the resulting estimates can be used in analysis of modern climate trends.