

УДК 621.396

МОДИФИКАЦИЯ АРХИТЕКТУРЫ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ U-NET В ЗАДАЧЕ СЕГМЕНТАЦИИ МУЛЬТИКАНАЛЬНЫХ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ

Хрящев В.В., к.т.н., доцент Ярославского государственного университета им. П.Г. Демидова, e-mail: vhr@yandex.ru;

Приоров А.Л., д.т.н., профессор Ярославского государственного университета им. П.Г. Демидова, e-mail: andcat@yandex.ru;

Павлов В.А., аспирант Ярославского государственного университета им. П.Г. Демидова, e-mail: vladimir@1pavlov.com;

Ларионов Р.В., магистрант Ярославского государственного университета им. П.Г. Демидова, e-mail: rv.larionov@yandex.ru.

MODIFICATION OF THE U-NET CONVOLUTIONAL NEURAL ARCHITECTURE IN THE MULTI-CHANNEL SATELLITE IMAGES SEGMENTATION PROBLEM

Khryashchev V.V., Priorov A.L., Pavlov V.A., Larionov R.V.

The results of training a convolutional neural network for segmentation of four-channel satellite images are presented in this article. The algorithm is trained to segment the images of buildings and structures. We used a modification of the U-Net architecture convolutional neural network with two encoders. The values of segmentation quality metrics for 16 regions of the Russian Federation with different building densities are shown in the article. The training based on Spacenet images allowed us to achieve a Sorensen coefficient of 0,783 and a Jacquard index of 0,649.

Key words: Earth remote sensing, segmentation, satellite images, convolutional neural network, deep learning.

Ключевые слова: дистанционное зондирование Земли, сегментация, спутниковые изображения, сверточная нейронная сеть, глубокое обучение.

Введение

Задачи обнаружения и классификации различных объектов на двумерных данных, получаемых путем дистанционного зондирования земной поверхности, являются одновременно и классическими, и современными проблемами в области построения систем обработки визуальной информации [1-3]. Решение заключается в определении наличия или отсутствия конкретных признаков на спутниковых изображениях. После обнаружения этих признаков объект может быть дополнительно классифицирован как принадлежащий к одному из заранее определенных классов.

Большинство современных подходов к решению данной проблемы предполагают разработку алгоритмов глубокого обучения, а именно – использование сверточных нейронных сетей [4]. Например, построенный таким образом алгоритм автоматической сегментации позволяет быстро и достаточно точно выделять области интереса человека, например, сельскохозяйственные угодья, карьеры и другие объекты [5].

В настоящее время важной задачей является выделение зданий и сооружений на спутниковых изображениях. Интерес к указанной проблеме вызван увеличением пространственного разрешения двумерной информации, получаемой путем дистанционного зондирования земной поверхности, что делает возможным обнаружение сравнительно небольших объектов с линейными размерами порядка 5-10 метров.

Представлены результаты обучения сверточной нейронной сети для сегментации четырехканальных спутниковых изображений. Алгоритм обучен выделять на изображениях здания и сооружения. Использовалась модификация архитектуры сверточной нейронной сети U-Net с двумя кодировщиками. Показаны значения метрик качества сегментации для 16 регионов Российской Федерации с различной плотностью застройки. Проведение предварительного обучения на базе изображений Spacenet позволило достичь значения коэффициента Серенсена 0,783 и индекса Жаккарда 0,649.

ружение сравнительно небольших объектов с линейными размерами порядка 5-10 метров.

Анализ современной научно-технической литературы применительно к решению указанной задачи позволяет выявить особенности, предъявляемые к алгоритмам сегментации мультисканальных спутниковых изображений зданий и сооружений [6-10]:

– Объекты зданий и сооружения могут сильно различаться по своим линейным размерам: от небольших частных домов, размером в несколько пикселей, до огромных жилых домов в спальных районах. Разрабатываемый алгоритм должен одинаково хорошо выделять объекты разных линейных размеров. Данной цели можно достигнуть при использовании нескольких кодировщиков в составе сверточной нейронной сети.

– Сложным моментом является разделение объектов зданий и сооружений при их высокой плотности. Это достигается путем тщательного выбора функции потерь при обучении нейронной сети.

– Инвариантность к поворотам и цветовым искажениям, вызванные влиянием атмосферы при получении данных дистанционного зондирования земной поверхно-

сти. Эта проблема может быть решена путем аугментации (размножения) данных.

– Спутниковые изображения местности могут иметь различное пространственное разрешение. На практике существенным моментом является создание алгоритма, обладающего наилучшей обобщающей способностью.

– Спутниковые изображения могут быть очень объемными. Нередко размер файла превышает 150 МБ, а разрешение – больше 16000×16000 пикселей. Для упрощения реализации нейронной сети и увеличения скорости работы алгоритма изображения нередко разрезают на фреймы размером 512×512 пикселей и меньше. Итоговый алгоритм должен решать задачу обратной сборки полученных результатов.

– Алгоритм должен быть робастным к различного рода шумам и помехам. Причинами таких искажений могут быть: небольшие облака, блики от крыш зданий, изменение погодных условий, осадки и пр.

– На практике обычно получают мультиспектральные изображения в разных спектральных диапазонах: от ультрафиолетового до теплового инфракрасного. Использование одних и тех же методов для различных спектральных каналов не всегда приводит к наилучшему результату.

Автоматическая сегментация спутниковых изображений является важным этапом их обработки. В [11] показан алгоритм сегментации таких изображений с применением методов последовательного квадратичного программирования и анализа форм спектральных компонент. Сверточные нейронные сети позволяют получать признаки искомым объектам автоматически и классифицировать пиксели с учетом окружающей их области, то есть учитываются непосредственно свойства самого объекта [12]. Таким образом, повышается робастность алгоритма сегментации к шумам и помехам.

Современные нейронные сети позволяют детектировать в видеопотоке такие непростые объекты, как автомобили и пешеходы, практически в режиме реального времени (время детектирования составляет менее 100 мс) [13]. Структура модели сверточной нейронной сети параллельна, что делает идеальным применение

графических процессоров (GPU) для их обучения [14].

Использование сверточных нейронных сетей позволяет детектировать на мультиспектральных изображениях такие объекты как водные ресурсы, леса, облака над водой и над землей [15-17]. В работе [18] представлена архитектура нейронной сети U-Net, показавшая свою эффективность в решении задачи сегментации медицинских изображений. Сеть U-Net показала свою эффективность и в автоматической сегментации спутниковых изображений при малом количестве обучающих данных [19].

Целью данной работы является представление результатов обучения сверточной нейронной сети U-Net для сегментации спутниковых мультиспектральных изображений с целью выделения зданий и сооружений. Предложена модификация архитектуры данной нейронной сети. Работа является продолжением исследований авторов [20-21].

Подготовка данных для обучения нейронной сети

Сверточная нейронная сеть предварительно обучена на базе изображений Spacenet [22], содержащей изображения 6 городских агломераций размером 650×650 пикселей: Рио-де-Жанейро (Бразилия), Лас-Вегас (США), Париж (Франция), Шанхай (Китай), Хартум (Судан) и Атланта (США). Восьмиканальные изображения получены спутниками WorldView-2 и WorldView-3 с различным пространственным разрешением. Spacenet содержит 2 подмножества общей площадью покрытия 3011 км² и 5555 км² для детектирования зданий и сооружений. Примеры изображений из базы Spacenet приведены на рис. 1.

Набор данных, на которых проходило дообучение модели, состоит из изображений 16 регионов Российской Федерации различной степени застройки. Каждое изображение имеет свою двоичную маску, размеченную вручную экспертами. Набор данных охватывает примерно 30 квадратных километров. Изображения являются четырехканальными: синий, зеленый, красный и ближний инфракрасный (NIR) каналы с пространственным разрешением 3 м/пиксель. Пример изображений в видимом диапазоне приведен на рис. 2.

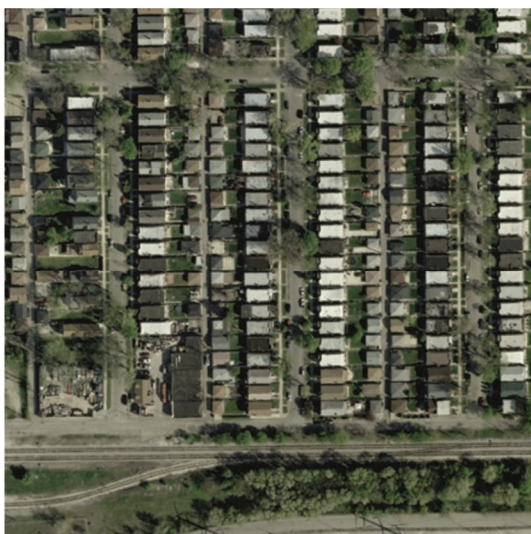


Рис. 1. Пример изображений базы Spacenet



Рис. 2. Пример изображений, использовавшихся для дообучения

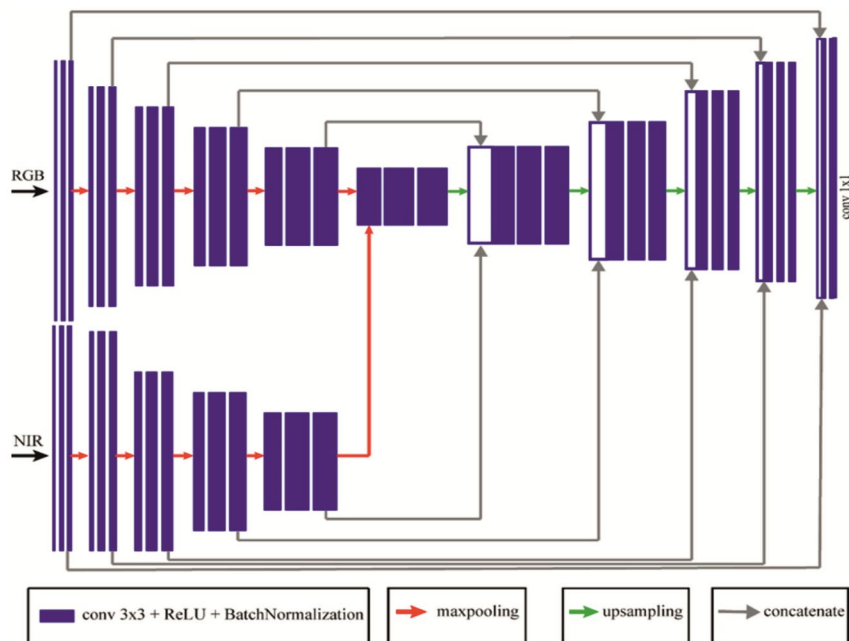


Рис. 3. Архитектура нейронной сети U-Net с двумя кодировщиками

Разработанная модель в качестве входных данных принимает фреймы размером 256×256 пикселей. Для создания обучающей и тестовой выборки каждое изображение из базы данных разделялось на две непересекающиеся полосы (одна шириной 256 пикселей, вторая – все оставшиеся), затем каждая полоса разделялась на фрагменты 256×256 пикселей с шагом 128, чтобы соседние фрагменты пересекались наполовину. Полоса шириной 256 пикселей предназначалась для тестов.

Для увеличения размера обучающей базы изображений проводилась аугментация данных. Она подразумевала выполнение следующих операций:

1. Повороты на 90, 180, 270 градусов и зеркальные отражения. Данные преобразования увеличили набор изображений в 8 раз.

2. Хроматические искажения. Изображения переводились из цветового пространства RGB в цветовое пространство HSV, и к новым цветовым координатам HSV добавлялись случайные значения. Для ближнего инфракрасного канала вместо хроматических искажений к нормализованному в интервале $[0, 1]$ значениям добавлялись случайные величины из интервала $[-0,06, +0,06]$.

3. Случайные сдвиги, перемасштабирование и повороты на малые углы от -15 до $+15$ градусов.

В результате всех проведенных операций аугментации получена база из 9784 изображений, разрешением 256×256 пикселей.

Архитектура нейронной сети

В работе использовалась модификация известной архитектуры сверточной нейронной сети U-Net [18]. Оригинальная U-Net состоит из двух частей: кодировщика и декодировщика.

Кодировщик представляет собой сверточную нейронную сеть, состоящую из пяти блоков. Каждый такой блок состоит из двух сверточных слоев с фильтрами 3×3 , с функцией активации ReLU и batch normalization, применяемой к каждому из них, а также слоя понижения дискретизации с окном 2×2 . Декодировщик имеет такое же количество блоков, что и кодировщик. Каждый блок декодировщика состоит из слоя повышения дискретизации с окном 2×2 , слияния с соответствующим набором признаков из кодировщика, двух сверточных слоев с фильтрами 3×3 и функцией активации ReLU, примененной к

каждому из них. Последний слой использует сигмоидную функцию активации признаков для классификации на уровне пикселей [23].

Оригинальная архитектура сверточной нейронной сети U-Net была модифицирована: использовалось два кодировщика для RGB и NIR каналов с объединением их на центральном блоке (рис. 3). Итоговая нейронная сеть имела 47 сверточных слоев, 47 функций активации ReLU, 47 операций нормализации батчей, 1 функцию активации сигмоида, 10 операций повышения дискретизации, 5 операций понижения дискретизации, 11 операций слияния признаков.

Численные результаты

Обучение нейронной сети проводилось на суперкомпьютере NVIDIA DGX-1, имеющемся в центре искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова. Длительность процедуры обучения составила около 3 часов. В качестве оптимизатора для обучения модели использовался алгоритм Adam [24], в качестве функции потерь выбрана бинарная кросс-энтропия с оригинальными значениями весов и коэффициентом Серенсена.

Для оценки качества работы алгоритма сегментации использовались специальные метрики в задаче сегментации изображений: коэффициент подобия Серенсена (*dice*) и индекс Жаккарда (*IoU*).

Коэффициент подобия Серенсена принимает значения из отрезка $[0, 1]$ и фактически является двоичной мерой сходства двух множеств. Данный показатель рассчитывается по следующей формуле:

$$dice = \frac{2I}{S},$$

где $I = |X \cap Y|$ – мощность пересечения, а $S = |X| + |Y|$ – сумма мощностей экспертной разметки X и предсказаний Y . В нашей задаче числитель I и знаменатель S можно вычислить по формулам:

$$I = \sum_{\substack{x \in X \\ y \in Y}} xy, \quad S = \sum_{\substack{x \in X \\ y \in Y}} (x + y),$$

где $x, y \in [0, 1]$ – значения вероятностей принадлежности пикселей к тому или иному классу экспертной разметки X и предсказаний Y соответственно.

Индекс Жаккарда определяет степень сходства сравниваемых объектов и вычисляется по формуле:

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|},$$

где A и B – классы экспертной разметки и предсказаний нейронной сети, соответственно.

В задаче сегментации значение коэффициента Серенсена 0,7 является своего рода порогом. Превышение данного значения означает, что алгоритм хорошо справляется со своей задачей [2]. Исходя из этого, ставилась задача превысить данное значение метрики.

Вначале исследовалась целесообразность использования всех четырех каналов спутникового изображения. Для этого нейронная сеть независимо обучалась на RGB-изображениях, на NIR-изображениях и на 4-канальных изображениях, на которых изображена одна и та же территория, а именно 16 регионов РФ с различной степенью застройки. Использовалась классическая архитектура сети U-Net. Результаты представлены в табл. 1. Для удобства, использованы следующие сокращения типов местности: городская застройка – ГЗ, сельская местность – СМ, малый город – МГ.

Таблица 1. Значения коэффициента Серенсена и индекса Жаккарда в 16 регионах РФ для классической U-Net с RGB-, NIR- и 4-канальными изображениями на входе

Регион	Тип застройки	Коэффициент Серенсена			Индекс Жаккарда		
		RGB+NIR	RGB	NIR	RGB+NIR	RGB	NIR
1	ГЗ	0,639	0,586	0,580	0,486	0,436	0,419
2	ГЗ	0,860	0,713	0,773	0,761	0,591	0,641
3	СМ	0,704	0,617	0,547	0,549	0,457	0,392
4	ГЗ	0,754	0,670	0,707	0,626	0,535	0,567
5	МГ	0,601	0,572	0,429	0,436	0,407	0,275
6	МГ	0,627	0,592	0,551	0,458	0,421	0,381
7	СМ	0,582	0,617	0,550	0,443	0,509	0,418
8	СМ	0,669	0,715	0,720	0,633	0,672	0,676
9	СМ	0,613	0,520	0,612	0,466	0,38	0,469
10	СМ	0,533	0,479	0,493	0,367	0,321	0,320
11	МГ	0,697	0,651	0,674	0,567	0,509	0,540
12	ГЗ	0,865	0,847	0,821	0,763	0,735	0,697
13	ГЗ	0,744	0,689	0,708	0,597	0,530	0,553
14	ГЗ	0,659	0,643	0,519	0,551	0,531	0,395
15	ГЗ	0,536	0,527	0,618	0,400	0,395	0,472
16	ГЗ	0,588	0,622	0,619	0,456	0,489	0,478
Средние	–	0,668	0,633	0,627	0,544	0,506	0,478

Анализ результатов показывает, что использование 4-канальных изображений дает лучшие результаты в большинстве случаев. Более сложными наборами оказывается сельская местность (область маленьких зданий и сооружений), где использование всех 4-х каналов не дает преимущества. В то же время среднее значение коэффициента Серенсена остается ниже порога в 0,7. Для улучшения качества работы алгоритмов использовалось два варианта модификации классической сети U-Net. Первый из них заключается в использовании двух сетей U-Net для RGB- и NIR-каналов по отдельности с их объединением на последнем классифицирующем слое. Второй состоит в модификации U-Net с двумя кодировщиками, как показано на рис. 3. Сравнение работы данных алгоритмов по значениям коэффициента Серенсена изображено на рис. 4. Из приведенных данных следует, что схема с двумя кодировщиками является предпочтительной в большинстве случаев.

Дополнительно проведено предварительное обучение сети U-Net с двумя кодировщиками на базе изображения Spacenet. Во время предобучения значение коэффициента Серенсена достигло значения 0,84, индекса Жаккарда – 0,77.

Сравнение значений коэффициента Серенсена и индекса Жаккарда для классической архитектуры U-Net с 4-канальными изображениями на входе и модификации с двумя кодировщиками для RGB- и NIR-каналов без предобучения и с предобучением на базе изображений Spacenet приведены в табл. 2. Использованы следующие сокращения: сеть с двумя кодировщиками – U-Net2, сеть с двумя кодировщиками с проведенным предварительным обучением на базе изображений Spacenet – U-Net2P. Как следует из приведенных данных – сеть U-Net2P показывает лучший результат в 13 случаях из 16, повышая при этом среднее значение коэффициента Серенсена до 0,783 и индекса Жаккарда до 0,649.

На рис. 5 и рис. 6 приведены примеры работы алгоритма в условиях городской застройки и сельской местности соответственно. Коэффициент Серенсена для полученного изображения составил 0,782 для городской застройки и 0,713 для изображения сельской местности. Меньшее значение коэффициента для регионов с сельской местностью связано с небольшим размером выделяемых объектов, которые зачастую занимают несколько пикселей при пространственном разрешении спутникового изображения – 3 метра на пиксель.

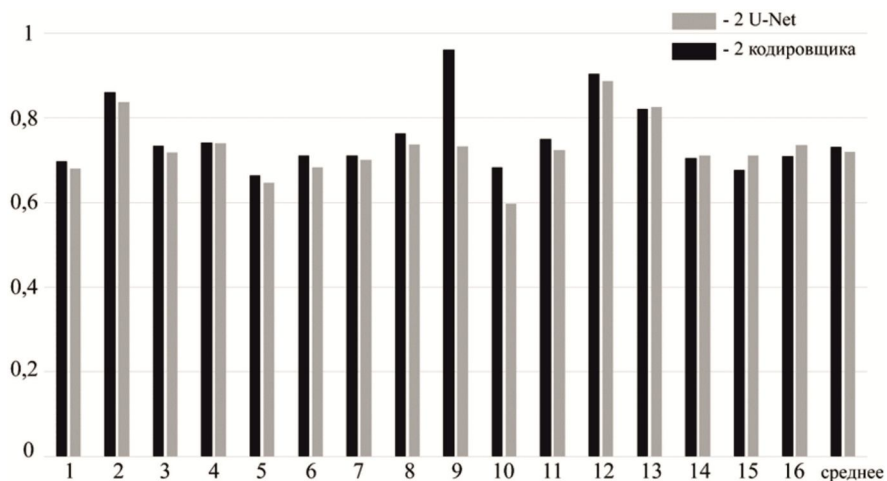


Рис. 4. Значения коэффициента Серенсена для сети с двумя кодировщиками и двумя U-Net при работе с 4-канальными изображениями



Рис. 5. Пример работы алгоритма в городской застройке. Коэффициент Серенсена для изображения равен 0,782

Таблица 2. Значения коэффициента Серенсена и индекса Жаккарда в 16 регионах РФ для классической U-Net, U-Net с двумя кодировщиками с предобучением и без предобучения

Регион	Тип застройки	Коэффициент Серенсена			Индекс Жаккарда		
		U-Net	U-Net2	U-Net2P	U-Net	U-Net2	U-Net2P
1	ГЗ	0,639	0,698	0,768	0,486	0,549	0,638
2	ГЗ	0,860	0,860	0,887	0,761	0,760	0,803
3	СМ	0,704	0,733	0,625	0,549	0,583	0,471
4	ГЗ	0,754	0,742	0,842	0,626	0,616	0,738
5	МГ	0,601	0,664	0,667	0,436	0,500	0,504
6	МГ	0,627	0,711	0,737	0,458	0,552	0,584
7	СМ	0,623	0,654	0,747	0,448	0,485	0,597
8	СМ	0,669	0,763	0,797	0,633	0,697	0,745
9	СМ	0,613	0,753	0,644	0,466	0,637	0,480
10	СМ	0,533	0,684	0,714	0,367	0,416	0,582
11	МГ	0,697	0,749	0,880	0,567	0,624	0,788
12	ГЗ	0,865	0,904	0,916	0,763	0,825	0,846
13	ГЗ	0,744	0,821	0,810	0,597	0,700	0,687
14	ГЗ	0,659	0,704	0,770	0,551	0,602	0,672
15	ГЗ	0,536	0,677	0,716	0,400	0,543	0,587
16	ГЗ	0,588	0,710	0,783	0,456	0,590	0,671
Средние	–	0,668	0,732	0,783	0,544	0,612	0,649



Рис. 6. Пример работы алгоритма в сельской местности. Коэффициент Серенсена для изображения равен 0,713

Заключение

Описан процесс обучения модифицированной сверточной нейронной сети U-Net, предназначенной для сегментации мультиканальных спутниковых изображений. Обучение проводилось на базе из 9784 изображений разрешением 256×256 пикселей. Одна итерация обучения на суперкомпьютере NVIDIA DGX-1 составляла около 3 часов.

Показано, что использование ближнего инфракрасного канала повышает точность сегментации в задаче выделения зданий и сооружений. Предложено использование двух кодировщиков в сверточной нейронной сети при работе с 4-канальными изображениями.

Проведение предварительного обучения на базе изображений Spacenet позволило достичь значения

коэффициента Серенсена 0,783 и индекса Жаккарда 0,649.

Разработанный алгоритм сегментации может найти применение при решении задач по оценке уровня урбанизации различных регионов и отслеживанию строительства крупных объектов.

Литература

1. Рассел Д. Спутниковая фотосъемка. М.: Мир, 2013. 157 с.
2. Визильтер Ю.В., Желтов С.Ю. Обработка и анализ изображений в задачах машинного зрения. Курс лекций и практических занятий. М.: Физматкнига, 2010. 672 с.
3. Алпатов Б.А., Ершов М.Д., Фельдман А.Б. Алгоритм обработки изображений для системы комбиниро-

ванного видения летательного аппарата // Цифровая обработка сигналов. 2015. № 3. С. 8-14.

4. Goodfellow Y., Bengio Y., Courville A. Deep Learning. The MIT Press, 2016, 800 p.

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. Zhang Z., Liu Q., Wang, Y. (2018). Road extraction by deep residual U-Net. IEEE Geoscience and Remote Sensing Letters, 15(5), pp. 749-753.

7. Zhang L., Zhang L., Du B. Deep learning for remote sensing data: A technical tutorial on the state of the art // IEEE Geoscience and Remote Sensing Magazine 4.2 (2016): pp. 22-40.

8. Chen Y., Jiang H., Li C., Jia X., Ghamisi P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks, in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

9. Zhu X., Tuia D., Mou L., Xia G-S., Zhang L., Xu F., Fraundorfer F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources, in IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8-36, Dec. 2017.

10. Iglovikov V., Mushinskiy S., & Osin V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. arXiv preprint arXiv:1706.06169.

11. Григорьева О.В. Субпиксельная идентификация объектов местности по гиперспектральным данным на основе последовательного квадратичного программирования и метода анализа формы спектральных компонент // Цифровая обработка сигналов. 2018. № 3. С. 26-31.

12. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2019, 480 с.

13. Ерохин Д.Ю., Ершов М.Д. Современные сверточные нейронные сети для обнаружения и распознавания объектов // Цифровая обработка сигналов. 2018. № 3. С. 64-69.

14. Seferbekov S., Iglovikov V., Buslaev A., Shvets A. Feature Pyramid Network for Multi-Class Land Segmentation. Web: <https://arxiv.org/pdf/1806.03510.pdf>.

15. 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.

16. Khryashchev V., Pavlov V., Priorov A., Ostrovskaya A. Deep Learning for Region Detection in High-Resolution Aerial Images // 2018 IEEE East-West Design & Test Symposium (EWDTS), Kazan, 2018, pp. 1-5.

17. Khryashchev V., Ivanovsky L., Pavlov V., Ostrovskaya A., Rubtsov A. Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation // Proceedings of the FRUCT'23, Bologna, Italy, 13-16 November 2018. pp. 172-179.

18. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, vol. 9351, 2015. pp. 234-241.

19. Соловьев Р., Тельпухов Д., Кустов А. Автоматическая сегментация спутниковых снимков на базе модифицированной сверточной нейронной сети UNET // Инженерный вестник Дона. 2017. Т. 47. № 4.

20. Ivanovsky L., Khryashchev V., Pavlov V., Ostrovskaya A. Building Detection on Aerial Images Using U-NET Neural Networks // 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 2019, pp. 116-122.

21. Хрящев В.В., Приоров А.Л., Павлов В.А., Ивановский Л.И. Сегментация объектов на спутниковых изображениях с использованием сверточных нейронных сетей // Успехи современной радиоэлектроники. 2019. Т. 73. № 6. С. 28-34.

22. SpaceNet Database, Web: <http://explore.digitalglobe.com/spacenet>.

23. Gulli A., Pal S. Deep Learning with Keras, Packt Publishing, 2017, 320 p.

24. Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization, Web: <https://arxiv.org/abs/1412.6980>.