

Исследование применимости сверточной нейронной сети U-Net к задаче сегментации изображений авиационной техники

Д.А. Гаврилов^{1,2}

¹ Институт точной механики и вычислительной техники им. С. А. Лебедева РАН,
119991, Россия, Москва, Ленинский проспект, д. 51,

² Московский физико-технический институт (национальный исследовательский университет),
141701, Россия, Московская область, г. Долгопрудный, Институтский переулок, д. 9

Аннотация

В работе проведено исследование применимости сверточной нейронной сети U-Net к задаче сегментации изображений авиационной техники. Метод нейросетевой сегментации изображений основан на реализации Carvana с архитектурой типа U-Net, для распознавания ориентации использована нейросеть, построенная в открытой нейросетевой библиотеке Keras на основе предобученной нейронной сети VGG16. Рассмотренный подход позволяет осуществлять сегментацию изображений. Результаты проведенных экспериментов показали возможность достаточно точного (0,94–0,96) выделения объекта интереса. Полученные бинарные маски позволяют визуально судить о классе самолета, имеющегося на изображении.

Ключевые слова: техническое зрение, обнаружение, локализация, нейронная сеть, распознавание, обработка изображений.

Цитирование: Гаврилов, Д.А. Исследование применимости сверточной нейронной сети U-Net к задаче сегментации изображений авиационной техники / Д.А. Гаврилов // Компьютерная оптика. – 2021. – Т. 45, № 4. – С. 575-579. – DOI: 10.18287/2412-6179-CO-804.

Citation: Gavrilov DA. Investigation of the applicability of the convolutional neural network U-Net to a problem of segmentation of aircraft images. Computer Optics 2021; 45(4): 575-579. DOI: 10.18287/2412-6179-CO-804.

Введение

В настоящее время спутниковые аэрокосмические изображения широко используются для решения множества различных прикладных задач межотраслевого характера [1–5]. Одной из важнейших задач анализа визуальной информации является поиск объектов на спутниковых аэрокосмических снимках. Семантическая сегментация изображений заключается в выделении на изображении локальных областей, соответствующих различным классам объектов. Снимки, полученные с помощью аэрофотосъемки, обеспечивают высокую детальность изображений поверхности, тем не менее, поиск и выделение объектов на таких изображениях требует активного участия человека [6]. Характерными признаками объектов, по которым они могут быть опознаны, являются дешифровочные признаки, к которым относятся форма, размер, характерные детали, цветовые особенности, тени, положение и взаимосвязь объектов. Каждый признак несет информацию о некотором свойстве, присущем объекту.

Для решения задачи классификации объектов на изображениях используются различные методы. Методы морфологического анализа позволяют создать математическое описание изображения, передающее его содержание. Морфология объекта подразумевает описание его внешнего строения, формы, размера, а также расположение образующих его структурных элементов относительно друг друга [7]. Важную роль в классификации объектов играет анализ их формы и

контура. Контурный анализ основан на учете изменения яркости и сравнении ее с пороговой [8]. При этом результат как морфологического, так и контурного анализа зависит от качества построения формы или контура объекта [9].

Несмотря на наличие достаточно большого числа методов детектирования и локализации объектов на изображениях, решение данной проблемы в полной мере пока представляет собой достаточно трудоемкую задачу, при решении которой, как правило, применяется ручной труд операторов-экспертов, что требует больших временных затрат и может сказываться на эффективности. Нейросети являются наиболее эффективным подходом к решению задачи семантической сегментации [10].

Целью настоящей работы является повышение эффективности выделения сложных объектов интереса на фоне для дальнейшего анализа формы этих объектов.

1. Метод анализа изображений

На вход нейросети подается изображение в видимом диапазоне волн в оттенках серого с фоном, на котором присутствуют объекты интереса типа «самолеты». На выходе получается бинарная маска, отделяющая фон от объекта интереса.

Для решения задачи использована предметная модификация сверточной нейронной сети на основе реализации Carvana с архитектурой типа U-Net [11–12], предназначенная для сегментации изображений. Архитектура сети представляет собой последовательность слоёв свёртки и пулинга. В слоях сети сначала

происходит уменьшение пространственного разрешения исходного изображения, затем изображения объединяются и пропускаются через другие слои свертки, после чего выполняется увеличение изображения. Нейронная сеть играет роль фильтра. Обоснованием выбора архитектуры U-Net является возможность использования ограниченного количества данных для обучения. Рассматриваемая сеть способна успешно обучаться в средах с низким и средним количеством данных обучения.

Архитектура сети U-Net состоит из следующих блоков: *ConvBnRelu* – блок, содержащий операции свертки (*Conv2D*), нормализации (*BatchNorm*) и функция активации (*Relu*); *StackEncoder* – совокупность операций сжатия, включая операции *ConvBnRelu* и уменьшения размерности (*MaxPool*); *StackDecoder* – блок, содержащий операции декодирования [11]. В декодере обеспечивается не только восстановление данных после сжатия, но и взаимодействие между данными до передачи их в узкое место, что обеспечивает возможность расширения контекста для следующих слоев декодирования. В данном случае нейросеть обучается обобщению не только сжатых представлений, но также имеет возможность восстанавливать скрытые обобщения в пространственном представлении с правильной семантической ориентацией каждого пикселя. При обучении нейросети использована «энергетическая функция» в сочетании с функцией потерь перекрестной энтропии (*cross entropy*) [11]:

$$p_k(x) = \frac{\exp(a_k(x))}{\left(\sum_k \exp(a_k(x))\right)},$$

где $a_k(x)$ – активация функции в канале k в позиции пикселя $x \in \Omega$ и $\Omega \subset Z^2$, K – число классов, $p_k(x)$ – «энергетическая функция» – аппроксимированная максимальная функция.

Функция потерь имеет вид:

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)),$$

где $l: \Omega \rightarrow \{1, \dots, K\}$ – истинная метка каждого пикселя, а $w: \Omega \rightarrow R$ – карта весов.

Для распознавания ориентации самолета использовалась нейросеть, построенная на открытой нейросетевой библиотеке Keras [13, 14] на основе предобученной нейронной сети *Visual Geometry Group (VGG16)* [12]. *VGG16* – сверточная нейронная сеть, предназначенная для распознавания объектов на изображениях, содержащая 16 слоев. Модель нейросети получена путем прямого преобразования модели *Caffe* [15] и состоит из двух частей. Первая часть сети выделяет характерные признаки и состоит из чередующихся каскадов свертки и подвыборки. Во всех слоях размер узла свертки составляет 3×3 . При

подвыборке производится выбор максимального значения из квадрата размером 2×2 . Вторая часть обеспечивает классификацию и содержит 3 полносвязных слоя.

Основными задачами определения ориентации объекта являются: первичное определение ориентации объектов с помощью нейронной сети по результатам работы алгоритма сегментации объектов, разворот найденных объектов в нужном направлении для удобства визуализации и дальнейшей обработки, выявление и исправление ошибок в определении ориентации объектов, выполняемые в ходе постобработки и коррекции результатов работы нейросети.

2. Подготовка обучающих выборок и обучение нейронной сети

Была выполнена предварительная подготовка набора обучающих выборок в виде изображений техники, а также бинарных масок объектов интереса. В качестве объектов интереса использовались объекты вида «авиационная техника». Подготовка обучающей выборки осложнялась наличием ограниченного количества реально полученных фотоснимков, получение большего числа которых не представлялось возможным [16].

Для решения проблемы получения обучающей выборки достаточного объема для обучения нейросети, обладающей высокой точностью и устойчивостью к условиям съемки, реализовано искусственное расширение обучающей выборки с помощью синтезированных изображений объектов интереса. Включение в обучающие выборки синтезированных изображений является распространенным приемом, применяемым при отсутствии достаточного количества данных или для редко встречающихся классов объектов. Получение синтетических данных осуществлялось с помощью симулятора виртуальной среды Unity3D, предоставляющего множество функциональных возможностей. Для синтеза изображений использовался снимок фона, на который случайным образом накладывалась модель объекта интереса (рис. 1).

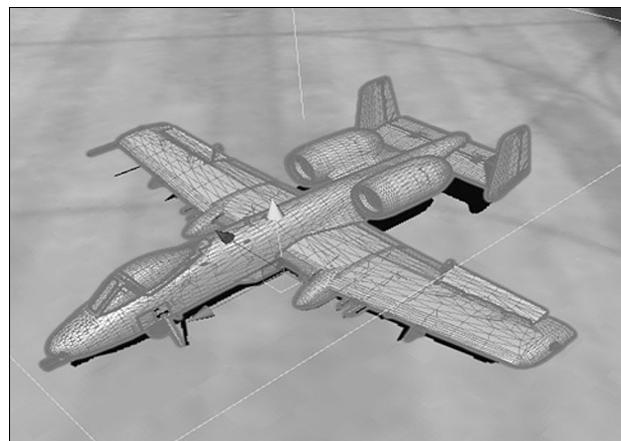


Рис. 1. Формирование синтетического изображения

Комбинированные снимки представлены на рис. 2.



Рис. 2. Комбинирование реальных и синтезированных изображений

Для обучения нейросети вместо «энергетической функции» в качестве функции потерь рассчитывался коэффициент Дайса (*Dice coefficient*), также называемый коэффициентом Сёрнсена (*Sorensen–Dice coefficient*), Жаккара (*Jaccard similarity coefficient*) или метрикой *Intersection Over Union*, который показывает меру сходства. Данный коэффициент представляет меру площади правильно отмеченных сегментов (отношение площади пересечения к площади объединения):

$$Dice = (X, Y) = \frac{2|X \cap Y|}{|X| + |Y|},$$

где X – предсказание; Y – правильно размеченная маска на текущем объекте; $|X|$ – количество элементов в данном множестве X ; $|Y|$ – количество правильно найденных объектов; \cap – оператор нахождения пересечения между X и Y .

3. Экспериментальные исследования

Эксперимент №1

Первый эксперимент был реализован для двух классов авиационной техники: А-4 – штурмовик; А-10 – штурмовик.

Скрипт обучения подготовлен с использованием продукта с открытым кодом Jupyter IDE [16]. Для обучения в рамках эксперимента №1 использованы обучающие выборки, содержащие в общей сложности 200 139 изображений, в том числе по 100 000 синтетических изображений, содержащих маски самолетов А-4 и А-10, 23 реальных изображения А-4 и 116 реальных изображений А-10. На вход нейросети подавались изображения объектов в видимом диапазоне волн в оттенках серого из обучающей выборки массивами данных определенного размера. Реальные и синтетические данные случайным образом смешивались в соотношении 30 к 70.

Аугментация реальных изображений осуществлялась с помощью искажений в виде: случайных поворотов в пределах 360°, сжатия и растяжения по осям абсцисс и ординат с коэффициентом 0,1, а также масштабирования с коэффициентом 0,1. Аугментация реальных изображений с помощью преобразователя данных и дальнейшее их смешивание с искусственно синтезированными данными выполнялись с помощью специально разработанного генератора, подающего изображения на вход нейросети. Массив реальных данных полностью хранился в оперативной памяти в процессе обучения нейросети, синтетические изображения загружались по мере необходимости.

В ходе эксперимента было проведено 20 эпох обучения сети. Каждая эпоха представляет собой единичную итерацию в процессе обучения и включает предъявление всех примеров из обучающего множества. В результате обучения были получены следующие значения точности $Dice = 0,96$.

Результаты эксперимента №1 представлены на рис. 3. Как видно из рис. 5, форма масок, полученных сетью на тестовой выборке, позволяет визуально судить о классе самолета, имеющегося на изображении.

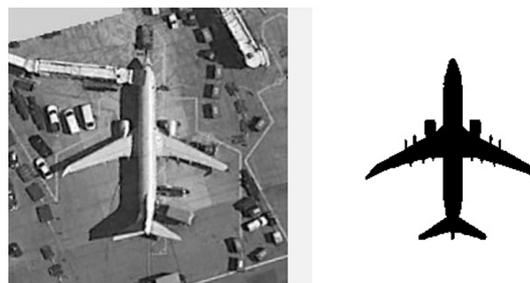


Рис. 3. Результаты эксперимента №1

Эксперимент №2

Для реализации второго эксперимента использовалось большее число классов авиационной техники. В общей сложности было представлено 12 классов техники: А-4 – штурмовик; А-10 – штурмовик; АV-8В – штурмовик; F-4 – тактический истребитель; F-14 – тактический истребитель; F-15 – тактический истребитель; F-16 – тактический истребитель; F-18 – тактический истребитель; F-22 – тактический истребитель; F/A-18 – тактический истребитель; Е-3 – самолет системы дальнего радиолокационного обнаружения и управления (ДРЛОУ); С-17 – стратегический военно-транспортный самолет.

Для обучения использовалось в общей сложности 6313 изображений, в том числе 5000 синтетических изображений каждого класса объектов и 1313 реальных изображений указанных объектов. Тестовая выборка состояла из 230 изображений. На вход нейронной сети, так же как и в первом эксперименте, подавались случайным образом смешанные реальные и синтетические изображения в соотношении 30 к 70. В результате обучения были получены следующие значения точности $Dice = 0,94$.

В результате обучения сети были получены сегментированные маски объектов интереса (рис. 4). Как видно из рис. 4, контуры масок сглажены по сравнению с обучающими масками и не содержат неровностей в виде двигателей, хвостового оперения. Полученные маски позволяют осуществлять классификацию объектов.

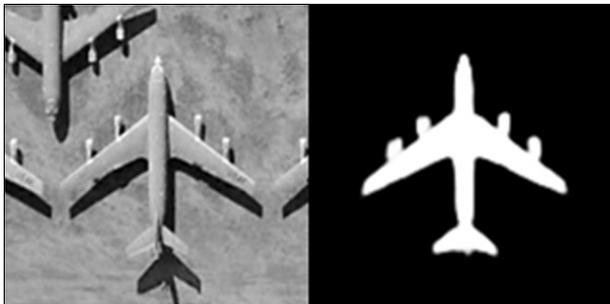


Рис. 4. Результаты эксперимента №2

Эксперимент №3

Реализация третьего эксперимента осуществлялась с помощью нейронной сети, обученной на распознавание угла поворота самолета. Предварительно обученная модель была переобучена. Аугментация данных для подготовки обучающей выборки проводилась изменением угла поворота самолета на величину погрешности сети, равную 10 градусам. При таких искажениях качество маски достаточно сильно зависит от угла поворота.

Экспериментальное тестирование выполнялось в два этапа.

На первом этапе с помощью нейронной сети VGG16 осуществлялся разворот самолета в нулевое положение. Сеть была предварительно обучена на черно-белых изображениях размером 64 × 64. На вход нейросети подавалось изображение, содержащее один объект, расположенный по центру. На выходе обеспечивалось определение значения синуса и косинуса угла поворота самолета относительно вертикальной оси на изображении. Результат работы алгоритма определения ориентации представлен на рис. 5.

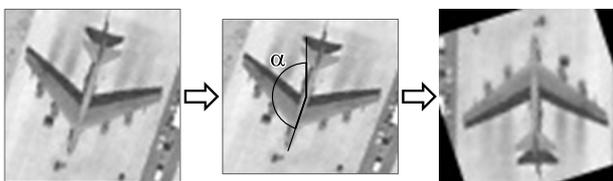


Рис. 5. Результат работы алгоритма определения ориентации

На втором этапе производилась сегментация. Произведенные операции позволили существенно улучшить качество получаемых бинарных масок.

Медианная (ME) и среднеквадратическая ошибки (MSE) рассчитывались в соответствии с выражениями

$$ME = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i);$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2,$$

где N – количество объектов, X_i – фактическое значение угла поворота, \hat{X}_i – найденное значение угла поворота.

В результате тестирования медианная ошибка составила 1°, среднеквадратическая ошибка (MSE) – 6,6°.

Заключение

Исследован метод сегментации изображений при помощи нейросети, основанной на реализации Carvana с архитектурой типа U-Net. Для распознавания ориентации самолета использовалась нейросеть, построенная в открытой нейросетевой библиотеке Keras на основе предобученной нейронной сети VGG16. Представленный алгоритм позволяет осуществлять детектирование объектов интереса на изображениях.

В процессе выполнения работ выделены принципы формирования обучающих выборок для нейронных сетей. Сформирована обучающая выборка и произведено ее расширение за счет аугментации исходных изображений с помощью специально разработанного генератора. На основе обучающей выборки проведено обучение нейросети выполнению задачи сегментации объектов на изображениях.

Проведены экспериментальные исследования эффективности разработанного метода. Результаты проведенных экспериментов показали возможность достаточно точного выделения объекта интереса. Полученные бинарные маски позволяют визуально судить о классе самолета, имеющегося на изображении.

Литература

1. **Белов, А.М.** Алгоритм выявления случайных искажений в составе сцены на серии разновременных изображений ДЗЗ одной и той же территории / А.М. Белов, А.Ю. Денисова // Компьютерная оптика. – 2019. – Т. 43, № 5. – С. 869-885. – DOI: 10.18287/2412-6179-2019-43-5-869-885.
2. **Борзов, С.М.** Исследование эффективности классификации трудноразличимых типов растительности по гиперспектральным изображениям / С.М. Борзов, М.А. Гурьянов, О.И. Потатуркин // Компьютерная оптика. – 2019. – Т. 43, № 3. – С. 464-473. – DOI: 10.18287/2412-6179-2019-43-3-464-473.
3. **Lovtsov, D.A.** Automated special purpose optical electronic system's functional diagnosis / D.A. Lovtsov, D.A. Gavrilov // 2019 International Seminar on Electron Devices Design and Production (SED). – 2019. – 4 p. – DOI: 10.1109/SED.2019.8798409.
4. **Gavrilov, D.A.** Quality assessment of objects detection and localization in a video stream / D.A. Gavrilov // Вестник МГТУ им. Н.Э. Баумана. Серия Приборостроение. – 2019. – Т. 125, № 2. – С. 40-55.
5. **Гаврилов, Д.А.** Поточная аппаратная реализация алгоритма SURF / Д.А. Гаврилов, А.В. Павлов // Известия

- высших учебных заведений. Электроника. – 2018. – Т. 23, № 5. – С. 502-511.
6. **Паркалов, А.В.** Применение нейронных и семантических сетей для сегментации растровых изображений земной поверхности / А.В. Паркалов // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2012). – 2012. – С. 527-530.
 7. **Ревякин, А.М.** Подходы к разработке системы распознавания для решения задачи определения контента цифровых изображений [Электронный ресурс] / А.М. Ревякин, А.В. Скурнович // Наукоедение. – 2016. – Т. 8, № 4. – URL: <https://naukovedenie.ru/PDF/30TVN416.pdf> (дата обращения 02.11.2020).
 8. **Дюдин, М.В.** Методы и алгоритмы контурного анализа для задач классификации сложноструктурируемых изображений / М.В. Дюдин, А.Д. Поваляев, Е.С. Подвальный, Р.А. Томакова // Вестник Воронежского государственного технического университета. – 2014. – Т. 3, № 1. – С. 54-59.
 9. **Местецкий, Л.М.** Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры / Л.М. Местецкий. – Москва: Физматлит, 2009.
 10. **Друки, А.А.** Семантическая сегментация данных дистанционного зондирования земли при помощи нейросетевых алгоритмов / А.А. Друки, В.Г. Спицын, Ю.А. Больтова, А.А. Башлыков // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2018. – Т. 329, № 1. – С. 59-68.
 11. **Ronneberger, O.** U-Net: Convolutional networks for biomedical image segmentation / O. Ronneberger, Ph. Fischer, Th. Brox. – In: Medical image computing and computer-assisted intervention – MICCAI 2015 / ed. by N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi. – New York: Springer International Publishing Switzerland, 2015. – Part III. – P. 234-241. – DOI: 10.1007/978-3-319-24574-4_28.
 12. Carvana image masking challenge–1st place winner’s interview [Electronical Resource]. – 2017. – URL: <http://blog.kaggle.com/2017/12/22/carvana-image-masking-first-place-interview/> (request date 22.07.2020).
 13. Keras [Electronical Resource]. – URL: <https://keras.io/> (request date 22.07.2020).
 14. **Simonyan, K.** Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // 3rd International Conference on Learning Representations (ICLR 2015). – 2015. – 14 p.
 15. **Nanda, Y.** What is the VGG neural network? [Electronical Resource] / Y. Nanda. – 2018. – URL: <https://www.quora.com/What-is-the-VGG-neural-network> (request date 22.07.2020).
 16. **Gavrilov, D.A.** A method for aircraft labeling in remote sensing images based on continuous morphological models / D.A. Gavrilov, L.M. Mestetskiy, A.B. Semenov // Programming and Computer Software. – 2019. – Vol. 45, Issue 6. – P. 303-310.

Сведения об авторе

Гаврилов Дмитрий Александрович, 1985 года рождения, в 2009 году окончил Московский физико-технический институт (государственный университет) (МФТИ) по специальности 03.03.01 «Прикладные математика и физика», работает ведущим научным сотрудником в Институте точной механики и вычислительной техники им. С. А. Лебедева РАН, руководителем лаборатории цифровых систем специального назначения МФТИ. Область научных интересов: системы распознавания образов, системы глубокого обучения, компьютерное моделирование, параллельные вычисления, виртуализация, искусственные нейронные сети.
E-mail: gavrilov.da@mipt.ru.

ГРНТИ: 28.23.37

Поступила в редакцию 1 сентября 2020 г. Окончательный вариант – 19 апреля 2021 г.

Investigation of the applicability of the convolutional neural network U-Net to a problem of segmentation of aircraft images

D.A. Gavrilov^{1,2}

¹ Lebedev Institute of Precise Mechanics and Computer Engineering, Russian Academy of Sciences, Russian Federation, Moscow, 51, Leninskiy boulevard, 119991,

² Moscow Institute of Physics and Technology, Russian Federation, 9 Institutskiy per., Dolgoprudny, Moscow Region, 141701

Abstract

The paper investigates the applicability of the convolutional neural network "U-Net" to a problem of segmentation of aircraft images. The neural network image segmentation method is based on the "Carvana" implementation with the "U-Net" architecture. For orientation recognition, a neural network built in the Keras open neural network library based on the pretrained VGG16 neural network is used. The approach considered allows the image segmentation to be conducted. The results of the experiments have shown the possibility of a fairly accurate selection of the object of interest. The resulting binary masks make it possible to visually classify the aircraft in the image.

Keywords: technical vision, detection, localization, neural network, recognition, image processing.

Citation: Gavrilov DA. Investigation of the applicability of the convolutional neural network U-Net to a problem of segmentation of aircraft images. *Computer Optics* 2021; 45(4): 575-579. DOI: 10.18287/2412-6179-CO-804.

References

- [1] Belov AM, Denisova AY. Scene distortion detection algorithm using multitemporal remote sensing images. *Computer Optics* 2019; 43(5): 869-885. DOI: 10.18287/2412-6179-2019-43-5-869-885.
- [2] Borzov SM, Guryanov MA, Potaturkin OI. Study of the classification efficiency of difficult-to-distinguish vegetation types using hyperspectral data. *Computer Optics* 2019; 43(3): 464-473. DOI: 10.18287/2412-6179-2019-43-3-464-473.
- [3] Lovtsov DA, Gavrilov DA. Automated special purpose optical electronic system's functional diagnosis. *International Seminar on Electron Devices Design and Production (SED)* 2019; 8798409.
- [4] Gavrilov DA. Quality assessment of objects detection and localization in a video stream [In Russian]. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering* 2019; 125(2): 40-55.
- [5] Gavrilov DA, Pavlov AV. Streaming hardware based implementation of SURF algorithm [In Russian]. *Proceedings of Universities. Electronics* 2018; 23(5): 502-511.
- [6] Parkalov AV. Application of neural and semantic networks in the segmentation of the earth's surface bitmaps [In Russian]. *Open Semantic Technologies for Intelligent Systems (OSTIS-2012)* 2012: 527-530.
- [7] Revyakin AM, Skurnovich AV. Approaches to the development of a recognition system to solve the problem of determining the content of digital images [In Russian]. *Naukovedenie* 2016; 8(4). Source: (<https://naukovedenie.ru/PDF/30TVN416.pdf>).
- [8] Dyudin MV, Povalyaev AD, Podvalny ES, Tomakova RA. Methods and algorithms of contour analysis for classification problems of complexly structured images [In Russian]. *Bulletin of Voronezh State Technical University* 2014; 3(1): 54-59.
- [9] Mestetsky LM. Continuous morphology of binary images: figures, skeletons, circulars [In Russian]. Moscow: "Fizmatlit" Publisher; 2009.
- [10] Druki AA, Spitsyn VG, Boltova YuA, Bashlykov AA. Sematic segmentation of earth remote sensing data using neural network algorithms [In Russian]. *Bulletin of the Tomsk Polytechnic University. Series Engineering of Georesources* 2018; 329(1): 59-68.
- [11] Ronneberger O., Fischer Ph., Brox Th. U-net: Convolutional networks for biomedical image segmentation. In Book: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical image computing and computer-assisted intervention – MICCAI 2015. Pt III*. New York: Springer International Publishing Switzerland; 2015: 234-241. DOI: 10.1007/978-3-319-24574-4_28.
- [12] Carvana image masking challenge – 1st place winner's interview. Source: (<http://blog.kaggle.com/2017/12/22/carvana-image-masking-first-place-interview/>).
- [13] Keras. Source: (<https://keras.io/>).
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ICLR-2015* 2015: 1-14.
- [15] Nanda Y. What is the VGG neural network? Source: (<https://www.quora.com/What-is-the-VGG-neural-network>).
- [16] Gavrilov DA, Mestetskiy LM, Semenov AB. A method for aircraft labeling in remote sensing images based on continuous morphological models. *Program Comput Softw* 2019; 45(6): 303-310.

Author's information

Dmitry Alexandrovich Gavrilov (b. 1985) graduated from Moscow Institute of Physics and Technology in 2009, majoring in Applied Mathematics and Physics. Currently he works as the leading researcher at the Lebedev Institute of Precise Mechanics and Computer Engineering and head of laboratory for special purpose digital systems at the Moscow Institute of Physics and Technology. Research interests are pattern recognition systems, deep learning systems, computer modeling, parallel computing, virtualization, artificial neural networks. E-mail: gavrilov.da@mipt.ru.

Received September 1, 2020. The final version – April 19, 2021.
