

## СОВРЕМЕННЫЕ ТЕХНОЛОГИИ АВТОМАТИЧЕСКОГО КАРТОГРАФИРОВАНИЯ ОБЛАЧНОСТИ

*Андрей Владимирович Тарасов*

Пермский государственный национально-исследовательский университет, 614990, Россия, г. Пермь, ул. Букирева, 15, аспирант кафедры картографии и геоинформатики, e-mail: andrew.tarasov.study1993@gmail.com

В статье рассматривается опыт применения методов машинного обучения (градиентного бустинга, глубоких сверточных нейронных сетей) для детектирования и картографирования облачности на примере территории Пермского края по снимкам высокого пространственного разрешения в видимом диапазоне (Sentinel-2). Проведено сравнение с существующими алгоритмами выявления облачности (Fmask, Sen2Cor), а также с базовой маской облачности, которая предоставляется вместе со снимками Sentinel-2. Для автоматического формирования картографических материалов с помощью языка программирования Python в программном комплексе ArcGIS был написан скрипт.

**Ключевые слова:** космические снимки, машинное обучение, картографирование, Sentinel-2.

## CURRENT TECHNOLOGIES OF AUTOMATIC CLOUD MAPPING

*Andrey V. Tarasov*

Perm State University, 15, Bukireva, Perm, 614990, Russia, Ph. D. Student, Department of Cartography and Geoinformatics, e-mail: andrew.tarasov.study1993@gmail.com

The article discusses the experience of using machine learning methods (gradient boosting, deep neural networks) for cloud detection and mapping on the example of the Perm Region using high spatial resolution images in visible range (Sentinel-2). With existing cloud detection algorithms (Fmask, Sen2Cor) Comparisons are made, as well as with the basic cloud mask provided with Sentinel-2 images. For automatic production of cartographic data Python script in the ArcGIS environment was coned fad.

**Key words:** satellite images, machine learning, Sentinel-2.

### *Введение*

На сегодняшний день появляются все новые источники данных дистанционного зондирования Земли, которые обладают высоким пространственным и временным разрешением. Это позволяет увеличить точность и оперативность мониторинга различных природных и антропогенных процессов. Для решения этих задач необходимо использовать автоматические методы анализа данных. Для успешного решения этой задачи необходимо создавать безоблачные покрытия территории, что делает задачу выявления и картографирования облачности актуальной. С появлением новых методов обработки данных дистанционного зондирования Земли, возникает вопрос их применимости для решения этих

задач. На сегодняшний день существует множество алгоритмов машинного обучения, начиная от традиционных (SVM, KNN, RandomForest) и заканчивая современными (сверточные нейронные сети, полносвязные сети, рекуррентные и др.) [1], которые адаптируются и для детектирования облаков. Для автоматического формирования картографического материала на основе полученных данных выделения облачности, можно использовать языки программирования, что позволяет минимизировать участие пользователя в этом процессе и оперативно получать картографический материал.

### ***Территория исследования***

В качестве тестового участка была выбрана часть территории Гайнского района Пермского края. Для анализа был выбран один снимок съемочной системы Sentinel-2 за 14 июня 2019 года (рис.1). Выбор снимка был обусловлен наличием на нем разных типов облаков: кучевых и полупрозрачных перистых. Возможности детектирования дымки в данном исследовании не рассматривались.

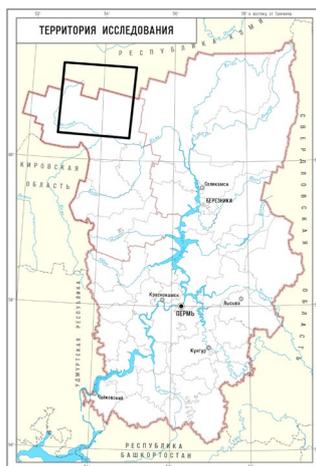


Рис. 1. Территория исследования

### ***Методы***

Для выявления облачности использовались две группы методов. Первая группа – «традиционные», которая включает в себя известный алгоритм Fmask [2] и алгоритм Sen2Cor [3]. Вторая группа методов, основана на методах машинного обучения, которая включает в себя алгоритм, разработанный Европейским космическим агентством (ESA) на основе градиентного бустинга (LightGBM) [4], обученного для детектирования облачности (далее ESABoosting) и алгоритм распознавания облачности на основе глубоких нейронных сетей [5] (далее CCNN).

*Fmask* – алгоритм детектирования облачности, который изначально был разработан для съемочной системы Landsat, а в последней версии адаптирован для

Sentinel-2 [6]. В основе метода – дерево принятия решений, которое учитывает совместно яркости в спектральных каналах (2, 3, 4, 7, 8, 8А, 10, 11, 12) и их производных значений (например, индекс NDVI), а все каналы приводятся к пространственному разрешению 20 м. Последовательно применяя правила дерева решения к входному снимку, выделяются несколько классов – облачность, тень, вода, снег и безоблачная территория.

*Sen2Cor* – набор инструментов для предварительной обработки снимков Sentinel-2 (атмосферной коррекции и маскирования облачности), разработанный совместно несколькими университетами Германии при участии Европейского космического агентства. В составе набора есть классификатор, который выделяет на снимке 11 классов, в том числе несколько классов облачности (облачность с высокой вероятностью (*cloud\_high\_probability*), средней вероятностью (*cloud\_medium\_probability*), перистые облака (*thin cirrus*), тени от облаков (*cloud shadows*)). Такая классификация может быть получена с разным пространственным разрешением (20 м/пиксель, 60м/пиксель). Алгоритм основан на последовательном применении определенных пороговых значений к значениям спектральных каналов Sentinel-2. Кроме того, пороговые значения применяются к расчетным показателям (индексам), таким как Normalized Difference Vegetation Index (NDVI) и Normalized Difference Snow Index (NDSI).

*ESABoosting* разработан Европейским космическим агентством (ESA). В основе – алгоритм градиентного бустинга (LightGBM [4]), который относится к ансамблевым методам машинного обучения. Эта группа методов позволяет обучать несколько моделей для итогового предсказания. В градиентном бустинге каждая новая модель учится на ошибках предыдущих, тем самым делая общее предсказание точнее с каждой итерацией. Алгоритм LightXBM также отличается скоростью работы (в два раза быстрее своего аналога XGBoost).

*CCNN* отличается от вышеописанных. Во-первых, он не получает точную маску облаков, а классифицирует тайлы размером  $\approx 400 \times 400$  метров, что обосновано невозможностью абсолютно четкого выделения границ облачности разного типа. Алгоритм использует ансамбль из глубоких сверточных нейронных сетей (DenseNet [6], ResNet50 [7], VGG10 [8]) и объединяет их предсказания, используя гребневую регрессию (*ridge regression*).

Алгоритмы были запущены со стандартными параметрами. Результаты работы алгоритмов были переведены в векторный вид, генерализованы и загружены в программу ArcMap для дальнейшей оценки точности.

### ***Сравнение результатов методов и формирование картографического изображения***

В качестве контрольного результаты экспертным путем была создана маска облачности на исследуемой территории. Были выделены объекты двух типов – сплошная облачность и полупрозрачная облачность. Критерием разделения была видимость поверхности Земли и наличие тени от объекта. Дополнительно был проведен анализ стандартной маски облачности, которая доступна при загрузке

снимка Sentinel-2 (далее – Sentinel-2 standard mask). Она рассчитывается на основе соотношения яркости в голубом и средних инфракрасных каналах (рис. 2). Объекты в маске облачности также делятся на два класса – непрозрачная облачность (opaque) и перистые (cirrus) [10].

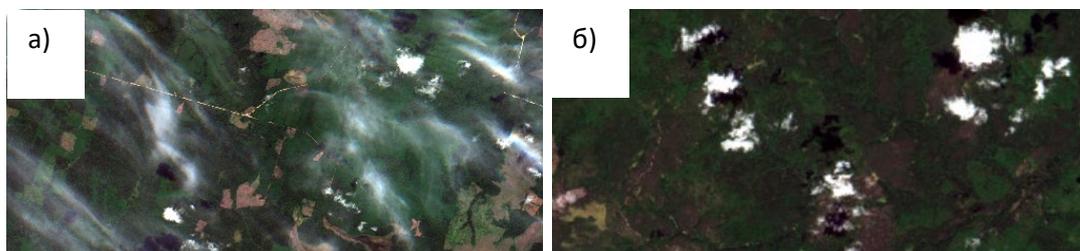


Рис. 2. Типы облачности (а) – частичная, (б) – сплошная

Сначала проведено сравнение выделенной облачности без разделения типов (табл. 1). В качестве показателей использовались общая точность, точность производителя (producer accuracy) и точность пользователя (user accuracy) [11].

Таблица 1

Точность классификации облачности

Критерий	Fmask	Sen2Cor	EsaBoosting	CCNN	Sentinel-2 standart mask
Общая точность, %	<b>44,5</b>	40,2	30,1	8,2	33,4
Точность пользователя, %	<b>46,7</b>	25,2	32,9	8,7	33,4
Точность производителя, %	91,0	99,7	78,0	63,4	<b>99,7</b>

По всем критериям лучший результата показал алгоритм Fmask. Высокая точность Sen2Cor обусловлена низким числом ложно выявленных объектов, при этом их контуры правильно выделены и совпадают с эталонами. Далее рассчитывались значения точности для отдельных классов облачности (табл. 2).

Таблица 2

Точность классификации полупрозрачной облачности

Критерий	Fmask	Sen2Cor	EsaBoosting	CCNN	Sentinel-2 standart mask
Общая точность, %	<b>43,8</b>	28,2	32,4	05,1	37,2
Точность пользователя, %	<b>45,2</b>	28,2	33,2	05,1	37,3
Точность производителя, %	93,2	<b>99,2</b>	92,6	91,1	96,5

Для второго типа лучший результат показал алгоритм Fmask по двум критериям. Однако, точность производителя оказалось выше у алгоритма Sen2Cor, который чуть точнее выделяет площади.

Для второго типа облачности были подсчитаны дополнительные критерии. Для каждого экспертно выделенного полигона подсчитывалось количество детектированных алгоритмами объектов, которые пересекают этот полигон. Если хотя бы один объект попадает на полигон, считалось что алгоритм правильно выделил объект. Количество правильно выделенных объектов нормировалось по общему числу экспертно выделенных полигонов (табл. 3).

Таблица 3

Точность классификации сплошной облачности

Критерий	Fmask	S2Cor	EsaBoosting	CCNN	Sentinel-2 standart mask
Доля правильно выделенных объектов, %	<b>96,8</b>	88,2	50,0	76,7	2,0
Доля ложно выделенных объектов, %	3,0	10,5	<b>0,0</b>	5,8	<b>0,0</b>
Точность пользователя, %	<b>95,1</b>	41,2	63,9	77,6	5,0
Точность производителя, %	45,8	<b>96,2</b>	23,0	89,1	95,0

Стоит отметить почти двукратное преимущество алгоритма Fmask над остальными методами. Этот алгоритм выделяет и небольшие по площади участки сплошной облачности.

Для формирования картографического материала использовались модули `arcsw.mapping` и `arcsw.da`. Первый позволяет настраивать порядок слоев, отображение и формировать готовую карту, а второй считывать данные и записывать их в компоновку. Результаты работы инструмента представлен на рис. 3.

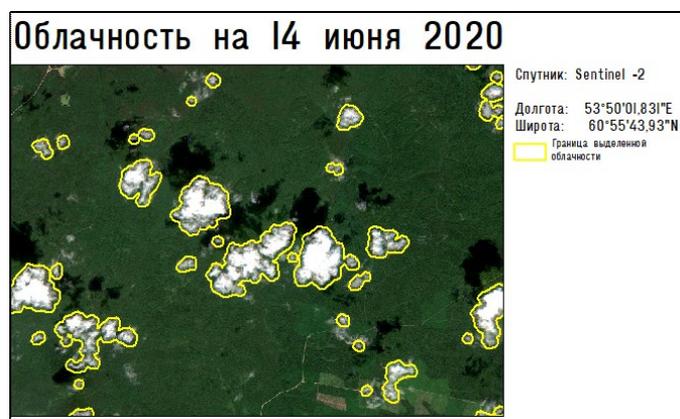


Рис. 3. Пример сформированного картографического изображения

## *Результаты и обсуждение*

В работе проведено сравнение методов выявления облачности на снимках высокого пространственного разрешения с помощью традиционных и современных методов обработки данных дистанционного зондирования Земли. Алгоритм Fmask показал точность выше 90 процентов для отдельных видов облачности. Стандартная маска облачности Sentinel-2 имеет высокую точность производителя по двум типам облачности, но в ней пропущено очень много объектов класса сплошной облачности. Метод ESABoosting лучше выявляет объекты сплошной облачности по сравнению со стандартной маской Sentinel-2. В целом, лучший результат показал алгоритм Fmask. Низкая эффективность методов машинного обучения для решения данной задачи говорит о необходимости усложнения моделей и использования других архитектур для улучшения точности выявления облачности на снимках высокого пространственного разрешения. Более сложная архитектура в данном исследовании не использовалась, в связи с недоступностью вычислительных ресурсов. Языки программирования встроенные в ГИС системы, позволяют автоматизировать процесс создания картографического материала, для оперативного доступа к этой информации.

Таким образом современные методы выявления и детектирования облачности совместно с инструментами автоматизации формирования картографического материала предоставляют возможности выполнения детектирования и картографирования облачности больших территорий в автоматическом и полуавтоматическом режимах.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Camps-Valls, G. Machine Learning in Remote Sensing Data Processing. // Proceedings of the 2009 IEEE Signal Processing Society Workshop. United States, 2009.
2. Zhe Z., Curtis E.W. Object-based cloud and cloud shadow detection in Landsat imagery.// Remote Sensing of Environment. 2012. Vol (118). P. 83–94
3. M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, F. Gascon. Sen2Cor for Sentinel-2. // Proceedings of the Image and Signal Processing for Remote Sensing XXIII, 2017.
4. Guolin K., Qi M., Finley T., Wang T., Chen W., Weidong M., Qiwei Y., Tie-Yan L.. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. // Proceeding of 31st Annual Conference on Neural Information Processing Systems (NIPS 2017). California, United States, 2017. P. 3149-3157.
5. Shendryk Y., Rist Y., Ticehurst C., Thorburn P. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery// ISPRS Journal of Photogrammetry and Remote Sensing, – 2019. – Vol. 157. – P. 124 – 136.
6. Qiu S., Zhu Z. & Binbin H. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4-8 and Sentinel-2 imagery // Remote Sensing of Environment. 2019. Vol (231).
7. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger Densely connected convolutional networks Paper Presented at the CVPR (2017)
8. K. He, X. Zhang, S. Ren, J. Sun Deep residual learning for image recognition. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
9. Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. // Proceeding of ICLR. 2015.

10. Level-1C Cloud masks Sentinel-2 MSI Technical guide. Электронный ресурс. URL <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks>
11. Heenkenda, M., Joyce K., Maier S., Bartolo R.. Mangrove Species Identification: Comparing WorldView-2 with Aerial Photographs. // Remote Sensing. 2012. Vol. (6). P. 6064-6088

© *A. B. Тарасов, 2020*