Оценка точности алгоритмов выделения маски облачности по данным Sentinel-2 и PlanetScope

А. В. Тарасов

Пермский государственный национальный исследовательский университет
Пермь, 614990, Россия
E-mail: gis@psu.ru

В последнее десятилетие были предложены новые методы обработки данных дистанционного зондирования Земли, основанные на машинном обучении, которые успешно применяются в том числе и при маскировании облачности. В настоящей работе на примере территории Пермского края проведена оценка точности маскирования облачности по снимкам Sentinel-2 разных сезонов с применением традиционных алгоритмов Fmask и Sen2Cor, а также алгоритма S2cloudless, основанного на машинном обучении. Для данных Sentinel-2 наилучший результат показал алгоритм S2cloudless (средняя точность 83 %), а худший — Fmask (70 %), при этом сезонное изменение точности маскирования облачности не превышает 6 %. Оценена применимость перечисленных алгоритмов для данных PlanetScope, имеющих только видимые и ближний инфракрасный спектральные каналы. Проведены эксперименты по выделению облачности по снимкам PlanetScope. В результате установлено, что алгоритм S2cloudless может применяться для зимних снимков PlanetScope (качество выделения облачности оказалось выше, чем на основе стандартной маски).

Ключевые слова: маскирование облачности, машинное обучение, PlanetScope, Sentinel-2

Одобрена к печати: 05.11.2020
DOI: 10.21046/2070-7401-2020-17-7-26-38

Введение

Облачность — основной фактор, ограничивающий применимость данных космической съёмки Земли в оптическом диапазоне спектра. Наличие облаков и сопутствующих им теней влияет на информационную ёмкость снимков. Качественное маскирование облачности на снимках имеет всё большее значение по мере роста объёмов данных с различных спутников, увеличения площади мониторинга за природными и техногенными объектами, развития информационных систем дистанционного мониторинга (СДМ), в частности для решения задач лесного и сельского хозяйства.

Традиционные методы маскирования облачности основаны на последовательном применении нескольких фильтров, в том числе с использованием средних и дальних инфракрасных (ИК) каналов. Наиболее известны алгоритмы Fmask (Zhe, Curtis, 2012), разработанный для данных Landsat и впоследствии адаптированный к данным Sentinel-2 (Qiu et al., 2019), и Sen2cor, разработанный для данных Sentinel-2 (Main-Knorn et al., 2017).

Однако в последние годы возникло существенное ограничение: новые группировки спутников, обеспечивающие оперативный мониторинг с высокой повторяемостью и детальностью, либо ведут съёмку в среднем ИК-канале с более низким пространственным разрешением (Sentinel-2), либо не имеют такого канала вовсе (PlanetScope). Таким образом, традиционные алгоритмы маскирования облачности имеют ограничения для данных Sentinel-2 и вообще не применимы для данных PlanetScope. Следовательно, требуется разработка новых алгоритмов, использующих только видимый и ближний инфракрасный диапазоны спектра, а также текстурные и геометрические признаки, значимость которых растёт с повышением пространственного разрешения данных. Лучше всего для этого подходят методы машинного обучения.

В настоящее время методы машинного обучения всё более успешно применяются для обработки данных дистанционного зондирования Земли (ДДЗЗ) (Андреев и др., 2019; Aas et al., 2018; Hethcoat et al., 2019; Li et al., 2019; Syrris et al., 2019). Они позволили по-новому подойти

к решению классических задач компьютерного зрения: классификации и сегментации изображений. Одно из преимуществ использования методов машинного обучения — это совместный учёт текстурных, геометрических и спектральных характеристик.

Цель настоящей работы заключается в получении сравнительной оценки точности маскирования облачности с применением традиционных алгоритмов (Fmask, Sen2Cor) и методов машинного обучения (S2cloudless). Также рассмотрена изменчивость точности маскирования облачности по сезонам года. Проведена оценка применимости перечисленных алгоритмов к данным Sentinel-2.

Исходные данные

Космические аппараты Sentinel-2A/B оснащены оптико-электронным мультиспектральным сенсором MSI (*англ*. Multi-Spectral Imager) для съёмок с разрешением от 10 до 60 м в 13 спектральных каналах в видимой, ближней ИК (*англ*. visible and near-infrared — VNIR) и коротковолновой ИК (*англ*. short wave infrared — SWIR) зонах спектра. Уникальность миссии Sentinel-2 связана с сочетанием большого территориального охвата, частых повторных съёмок и, как следствие, систематическим получением полного покрытия всей Земли мультиспектральной съёмкой высокого разрешения. Частота получения снимков составляет один раз в пять дней на экваторе и один раз в два-три дня в средних широтах, а ширина полосы съёмки — 290 км (Sentinel-2..., 2017).

Данные со спутника Sentinel-2 доступны с разным уровнем обработки: Level 1C и Level 2A (Sentinel-2..., 2017). Снимки распространяются в виде тайлов размером ~100×100 км. Полное описание спецификаций продуктов Sentinel-2 представлено в документации (Sentinel-2..., 2017).

Крупнейшая в мире группировка PlanetScope компании Planet Labs позволяет получать снимки любой части территории земного шара с высоким пространственным разрешением (3 м). Большое количество спутников (больше 120, каждый весом 4,7 кг) равномерно распределены по орбите, а их простая конструкция позволяет существенно снизить стоимость снимков. Съёмка ведётся в четырёх спектральных каналах: красном, зелёном, голубом, ближнем инфракрасном. Повторяемость съёмки — ежедневная (возможно получение снимков несколько раз в день). Ширина полосы съёмки — от 24 до 3 км. Архив доступен с 2016 г. (в зависимости от территории). Снимки распространяются на коммерческой основе (Planet..., 2020).

Данные со спутников PlanetScope поставляются в формате нескольких продуктов: PlanetScope Basic Scene Product, PlanetScope Ortho Scene Product, PlanetScope Ortho Tile Product. В настоящем исследовании использовался PlanetScope Ortho Scene Product. Этот продукт доступен в виде сцен размером от 24×8 до 32×16 км в зависимости от конкретного спутника. В состав продукта входят снимки, для которых проведены ортотрансформирование, радиометрическая калибровка, геометрическая коррекция, атмосферная коррекция, проведено слияние каналов, а также конвертация в формат GeoTiff. В поставку включены два растровых файла: исходный (MS) и с выполненной атмосферной коррекцией (SR). Полное описание алгоритмов обработки и спецификации поставляемых продуктов Planet приведено в документации (Planet..., 2020).

Существующие методы выделения маски облачности

В настоящей работе рассматриваются два традиционных алгоритма маскирования облачности, основанные на пороговых значениях (Fmask, Sen2Cor), а также алгоритмы с применением машинного обучения (S2Cloudless, CCNN (*англ*. Convexified Convolutional Neural Networks)).

Fmask — алгоритм детектирования облачности, который изначально был разработан для съёмочной системы Landsat, а в последней версии адаптирован для Sentinel-2 (Qiu et al., 2019). В основе метода — дерево принятия решений, которое учитывает совместно яркость в спектральных каналах (2, 3, 4, 7, 8, 8A, 10, 11, 12) (*табл. 1*) и их производные значения

(например, индексы NDVI (*англ*. Normalized Difference Vegetation Index — нормализованный разностный вегетационный индекс), NDSI (*англ*. Normalized Difference Snow Index — нормализованный разностный снежный индекс), NDBI (*англ*. Normalized Difference Built-Up Index), Whiteness Index), каждый из них приводится к пространственному разрешению 20 м. В результате последовательного применения правил дерева решения к входному снимку выделяются несколько классов: облачность, тени, вода, снег и безоблачная территория. Заявленная точность алгоритма — 90–95 %.

Таблица 1. Спектральные каналы спутника Sentinel-2

Имя канала	Центр спектрального диапазона, нм	Ширина диапазона, нм	Пространственное разрешение, м/пиксель
1	442,7	21	60
2	492,4	66	10
3	559,8	36	10
4	664,6	31	10
5	704,1	15	20
6	740,5	15	20
7	782,8	20	20
8	832,8	106	10
8A	864,7	21	20
9	945,1	20	60
10	1373,5	31	60
11	1613,7	91	20
12	2202,4	175	20

Sen2Cor — набор инструментов для предварительной обработки снимков Sentinel-2 (атмосферной коррекции и маскирования облачности), разработанный совместно несколькими университетами Германии при участии Европейского космического агентства (*англ*. European Space Agency — ESA) (Main-Knorn et al., 2017). В составе набора есть классификатор, который выделяет на снимке 11 классов, в том числе несколько классов облачности: облачность с высокой вероятностью (cloud high probability), облачность со средней вероятностью (cloud medium probability), перистые облака (thin cirrus), тени от облаков (cloud shadows). Алгоритм основан на последовательном применении определённых пороговых значений к значениям спектральных каналов Sentinel-2, а также к расчётным показателям (индексам), таким как NDVI и NDSI. На выходе алгоритм формирует карту вероятности принадлежности пикселя к определённому классу (S2 MPS..., 2017).

Множество работ также посвящено решению проблемы маскирования облачности с помощью методов машинного обучения (см., например, работы (Drönner et al., 2018; Feng et al., 2017; Liu et al., 2018; Liu et al., 2019)).

Алгоритм S2cloudless разработан Европейским космическим агентством. В его основе лежит метод градиентного бустинга LightGBM (*англ*. Gradient Boosting Machines) (Guolin et al., 2017), который относится к ансамблевым методам машинного обучения. Метод LightGBM также отличается скоростью работы (в два раза быстрее своего аналога XGBoost (*сокр*. *англ*. EXtreme Gradient Boosting)) (Gursky, 2020). Алгоритм S2cloudless попиксельно классифицирует изображения и рассчитывает вероятность отнесения пикселя к облачности на основе нормированных значений яркости в различных каналах (либо во всех 13 каналах, либо в 10 каналах Sentinel-2: 1, 2, 4, 5, 8, 8A, 9, 10, 11, 12). По данным авторов алгоритма, модель была обучена на масках облачности (количество образцов — 14140), выделенных алгоритмом MAJA (Lonjou et al., 2016), который рассчитывает маску на основе последовательных снимков одной

и той же территории (time-series analysis). Для обучения были отобраны только сцены, классифицированные алгоритмом MAJA с точностью выше 90 %, а процент облачности на исходном снимке составлял от 5 до 90. Дополнительно использовались около 1000 образцов, созданных авторами алгоритма с помощью онлайн-классификатора (https://apps.sentinel-hub.com/classificationApp-trial). Заявленная точность алгоритма — 91–99 %.

Метод, описанный в работе (Shendryk et al., 2019) (далее — CCNN), отличается от вышеприведённых. Во-первых, он не получает точную маску облаков, а определяет наличие или отсутствие облачности в тайлах размером 400×400 м, что может быть обосновано невозможностью абсолютно чёткого выделения границ облачности разного типа. Во-вторых, в качестве входных признаков для данного метода используются только четыре спектральных канала: ближний инфракрасный, красный, зелёный, голубой, а также вегетационные индексы NDVI и GNDVI (аналогичный NDVI, но с использованием зелёного канала вместо красного). Таким образом, метод CCNN, в отличие от всех вышеперечисленных, может быть применён для снимков PlanetScope.

Авторы алгоритма CCNN используют ансамбль из глубоких сверхточных нейронных сетей DenseNet (Huang et al., 2017), ResNet50 (He et al., 2016), VGG10 (Simonyan, Zisserman, 2014), а после объединяют их предсказания, применяя гребневую регрессию (ridge regression) (Hoerl, Kennard, 1970). Авторы употребили несколько наборов в качестве обучающих: набор, созданный ими экспертно по снимкам PlanetScope и Sentinel-2, а также набор, доступный на площадке Kaggle (https://www.kaggle.com/c/planet-understanding-the-amazon-from-space). Заявленная точность алгоритма — 78–80 %.

Обработка снимков и предобработка результатов

Для расчётов были выбраны 12 разносезонных снимков Sentinel-2 на территорию Пермского края (*рис. 1*, *табл. 2*). Процент облачности на выбранных снимках составлял от 0 до 60. Запуск алгоритмов выполнялся со стандартными параметрами.

Результатом работы алгоритмов стали классифицированные растры. Число классов варьируется, поскольку алгоритм Sen2Cor выделяет 9 классов, Fmask — 4 класса, S2cloudless — 2 класса. Для запуска алгоритмов Fmask и Sen2Cor требуется только загрузить и разархивировать данные со спутника Sentinel-2 уровня обработки L1C. Для алгоритма S2cloudless необходимо перевести данные Sentinel-2 в массив размером $n \times m \times k$, где n — число каналов, а m и k — размеры сцены Sentinel-2. Такое преобразование было выполнено с помощью библиотеки gdal (https://gdal.org/). Полученные результаты классификации переводились в векторный вид (формат shp), параллельно удалялись все классы, которые не относятся к облачности.

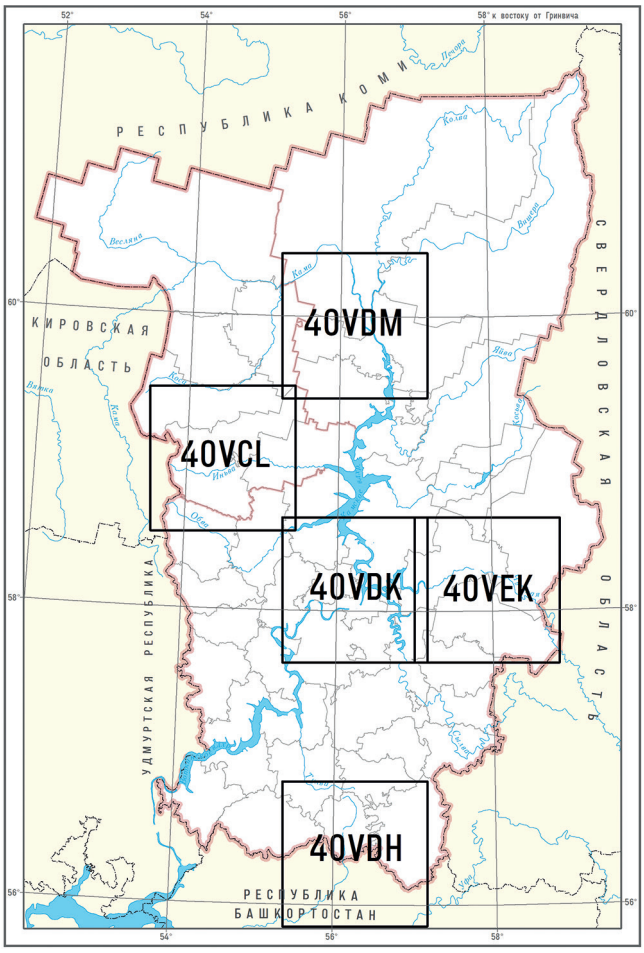
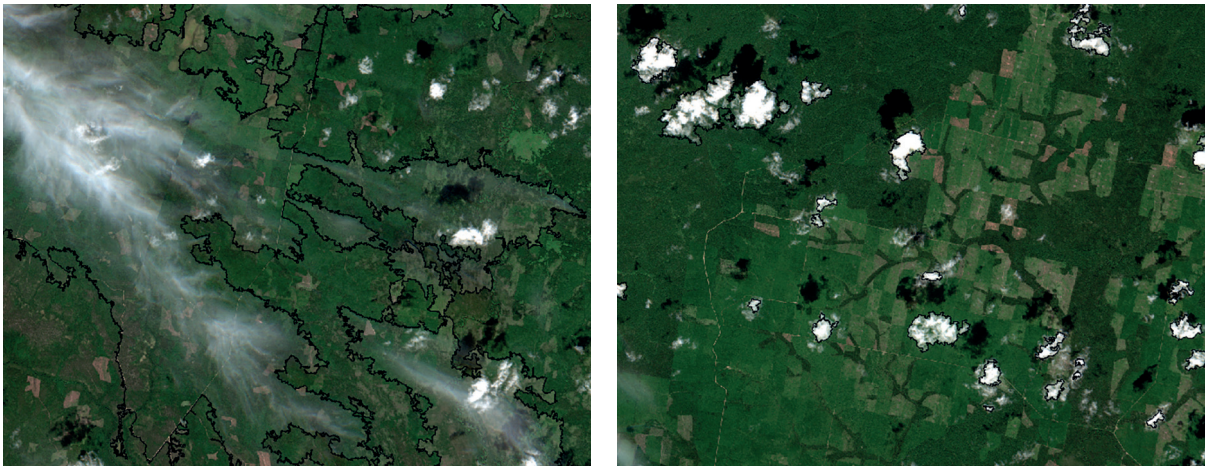

52°
54°
56°
58° к востоку от Гринвича
РЕСПУБЛИКА КОМИ
Печора
Колва
Вишера
Вишера
Вишера
Кама
Кама
Кама
Вятка
Вятка
Вятка

Таблица 2. Даты выбранных снимков Sentinel-2

Идентификатор тайла Sentinel-2	Сезон съёмки			
	Зима	Весна	Лето	Осень
40VDM	22.01.2019	–	05.08.2020	–
40VCL	17.01.2020	12.05.2019	26.07.2019	04.10.2019
40VDK	05.12.2019	–	–	–
40VEK	–	26.05.2018	–	–
40VDH	29.01.2019	09.04.2019	25.06.2018	03.09.2018

Оценка точности выделения маски облачности

Для оценки точности работы алгоритмов по уже загруженным снимкам экспертно были выделены контуры облачности. При создании объектов каждому присваивался один из классов: сплошная облачность (1-й класс) и полупрозрачная облачность (2-й класс) (*рис. 2*). Критерием разделения стала видимость поверхности Земли и наличие тени от объекта. Всего было оцифровано 5340 объектов, включая 5100 объектов класса «сплошная облачность» и 240 объектов класса «полупрозрачная облачность».



а *б*

Рис. 2. Примеры типов выделяемой облачности в контрольном наборе:
а — полупрозрачная; *б* — сплошная. Границы объектов выделены чёрным цветом

Для оценки точности была построена матрица ошибок результатов автоматизированного выделения облачности в сравнении с результатами экспертного выделения. На её основе рассчитана основная метрика для оценки точности — F-мера (Fmeasure), а также её составляющие — точность (precision) и полнота (recall). Метрика precision позволяет оценить, какая доля объектов, распознанных как объекты положительного класса (в данном случае — облачности), в действительности является таковыми, а recall показывает, какую долю объектов положительного класса из всех таких объектов нашёл алгоритм.

$$\text{Fmeasure}=\frac{2\text{precision}\times\text{recall}}{\text{precision}+\text{recall}},$$

$$\text{precision}=\frac{\text{TP}}{\text{TP}+\text{FP}},$$

$$\text{recall}=\frac{\text{TP}}{\text{TP}+\text{FN}},$$

где TP — объекты, правильно отнесённые к нужному классу; FP — неправильно отнесённые к нужному классу (ошибка I рода); FN — не отнесённые к нужному классу (ошибка II рода) (Powers, 2011).

На первом этапе оценивалась точность выделения всей облачности без разделения на типы. Расчёты производились с помощью инструментов пространственного анализа. Использование этих методов позволило разделить объекты на несколько классов (TP, FP, FN). Для каждого полученного объекта была рассчитана площадь, а потом она суммировалась для каждого класса (TP, FP, FN). В качестве итоговой оценки использовались среднее, медиана, минимальные и максимальные значения F-меры по всем сценам (*рис. 3*).

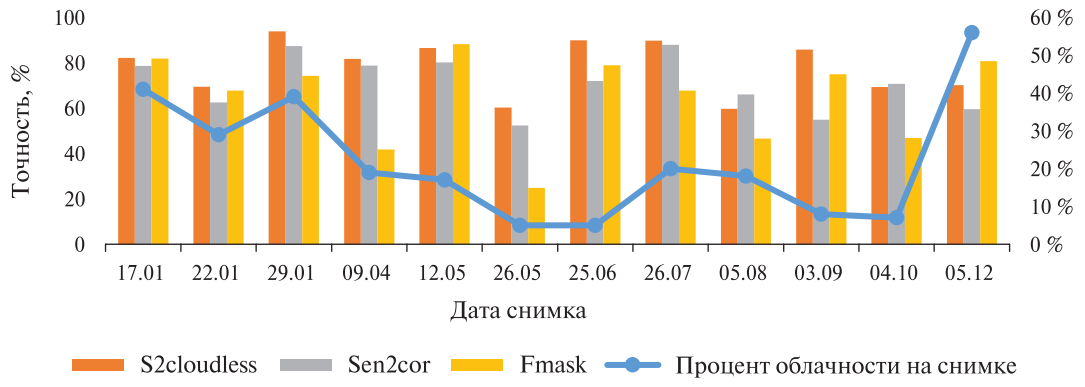


Рис. 3. Точность маскирования облачности с учётом процента облачности на снимках

Наилучший результат показал алгоритм S2cloudless (средняя точность — 83 %), на втором месте — алгоритм Sen2Cor (средняя точность — 75 %), а самый низкий показатель точности у алгоритма Fmask — 70 %. Низкая точность алгоритма Fmask в основном получается за счёт большого объёма ложного детектирования. Во-первых, на это оказывают влияние настройки буфера вокруг облачности, поскольку основная часть ошибочно выделенных объектов относится именно к буферным зонам. Во-вторых, точность данного алгоритма снижается до 40–50 % по снимкам осеннего и весеннего сезонов (апрель, октябрь) за счёт большого числа ошибочно выделенных объектов.

Полученные оценки точности также были агрегированы по сезонам года. Как и ожидалось, самые низкие показатели получены в переходные сезоны: весной и осенью. Это связано с наличием частичного снежного покрова на апрельском снимке и с частичным покрытием осеннего снимка полупрозрачной облачностью. В среднем снижение точности составляет не более 6 %, а характер её изменения по сезонам года совпадает у всех алгоритмов.

После общей оценки точности была проведена оценка в разрезе типов облачности (сплошной и полупрозрачной). Оценка точности для каждого типа облаков производилась только с помощью расчёта показателя recall (процент совпадения фактической и выделенной площади облачности). Сначала из контрольного набора выбирались объекты одного из типов, после этого анализировалось, насколько полно они покрываются масками облачности, созданными с помощью алгоритмов. Сплошная облачность была представлена на всех снимках. Для полупрозрачной облачности такой расчёт проводился только для трёх снимков, где присутствовал этот тип, а площадь облачности превышала 1 км^2. В качестве общей оценки были вычислены средние значения для 12 сцен по каждому типу облачности.

Самый высокий процент совпадения для сплошной облачности (98 %) показал алгоритм Fmask. Таким образом, этот алгоритм практически не пропускает объекты данного класса. Алгоритмы S2cloudless и Sen2Cor дают значительно более низкий процент совпадения — 82 и 72 % соответственно.

При выделении полупрозрачной облачности самый высокий процент показал также алгоритм Fmask, однако он даёт большой объём ложных срабатываний. Алгоритм S2cloudless выделяет 55 % полупрозрачной облачности, а Sen2Cor — только 1 %, поскольку полупрозрачная

облачность попадает в класс «облачность с низкой вероятностью» (clouds low probability), который не учитывался при оценке. Данный класс содержит много ложных объектов (участков дорог и открытого грунта).

Применение алгоритмов выделения маски облачности для снимков PlanetScope

При анализе литературы было выделено относительно немного работ, которые посвящены маскированию облачности на снимках PlanetScope и при этом находятся в открытом доступе (Dagobert et al., 2019; Xu et al., 2019). Вероятно, это связано с тем, что период активной эксплуатации спутников PlanetScope составляет всего около трёх лет.

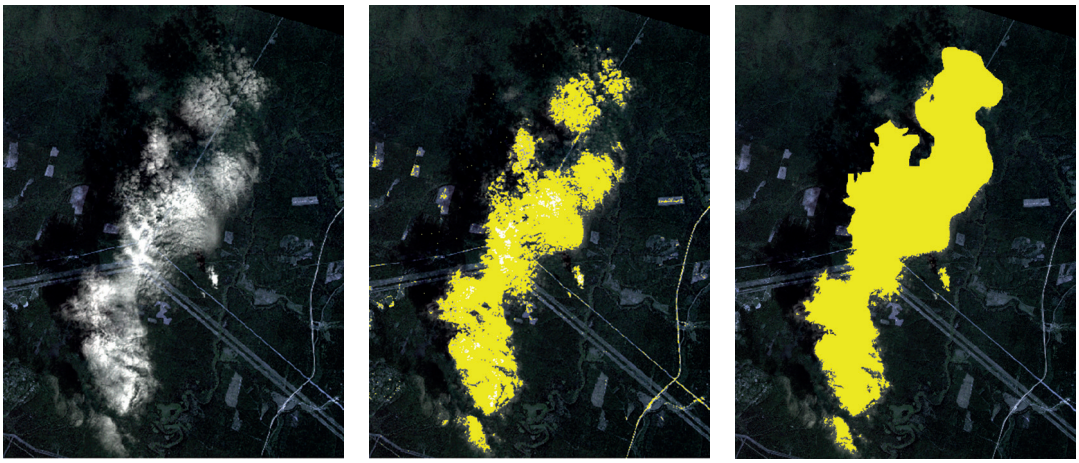
Традиционные алгоритмы маскирования облачности Fmask и Sen2Cor не могут быть адаптированы под данные PlanetScope, поскольку эти данные не содержат необходимых спектральных каналов. Также при их запуске требуется использовать метаданные съёмки Sentinel-2. К данным PlanetScope может быть адаптирован только алгоритм S2cloudless. Такая адаптация возможна благодаря способности алгоритмов машинного обучения переносить полученный опыт на новые задачи, которые связаны с исходной задачей (transfer learning) (Chuanqi et al., 2018). Для этого средствами языка программирования Python и библиотеки gdal (https://gdal.org/) снимки PlanetScope были конвертированы в формат 12-канальных растров, причём на место соответствующих каналов Sentinel были подставлены каналы PlanetScope. Значения яркости пикселей в остальных каналах были заполнены нулями.

Таблица 3. Соответствующие каналы съёмочной системы Sentinel-2 и PlanetScope

Канал	Спектральный интервал Sentinel-2, нм	Спектральный интервал PlanetScope, нм
Голубой (Blue)	448–546	455–515
Зелёный (Green)	537–583	500–590
Красный (Red)	545–583	590–670
Ближний инфракрасный (NIR)	763–909	780–860

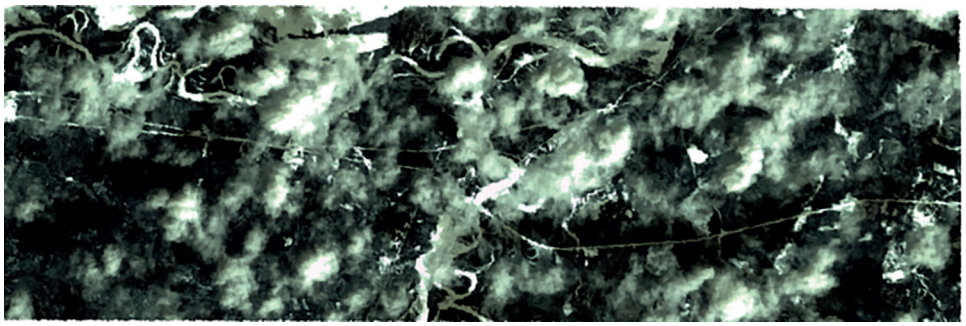
Следующим этапом подготовки данных был пересчёт исходных значений яркости в диапазон значений от 0 до 1, для чего производилось нормирование. Реализованы три способа нормирования: деление на постоянное значение 10 000 (как указано в документации алгоритма S2cloudless), на максимальное значение в канале и на среднее значение яркости в канале. Полученный результат передавался на вход алгоритма S2cloudless. Первый запуск алгоритма показал, что точность маскирования облачности по снимкам, не прошедшим предварительную атмосферную коррекцию (без префикса SR в имени файла), была значительно ниже, чем по откорректированным изображениям. Поэтому далее использовались только снимки, предварительно прошедшие атмосферную коррекцию.

Алгоритм показал хороший результат на снимках PlanetScope с нормировкой по максимальному значению (*рис. 4*, *5*). К сожалению, провести оценку по сезонам года или получить большую выборку по данным PlanetScope не удалось, так как большинство загруженных снимков были безоблачными, а в настоящее время данные распространяются на коммерческой основе. Для 10-летних снимков PlanetScope было выполнено сравнение точности между маской облачности, выделенной алгоритмом S2cloudless, и стандартной маской облаков PlanetScope. В качестве эталона использовались результаты экспертного (ручного) выделения облачности на этих снимках. В результате установлено, что процент совпадения с фактически выделенной облачностью для результатов, полученных с помощью алгоритма S2cloudless (F-мера — 75 %), оказался ниже, чем для стандартной маски (90 %), при этом на отдельных снимках результат сопоставим со стандартной маской (см. *рис. 4*, *5*, см. с. 33).

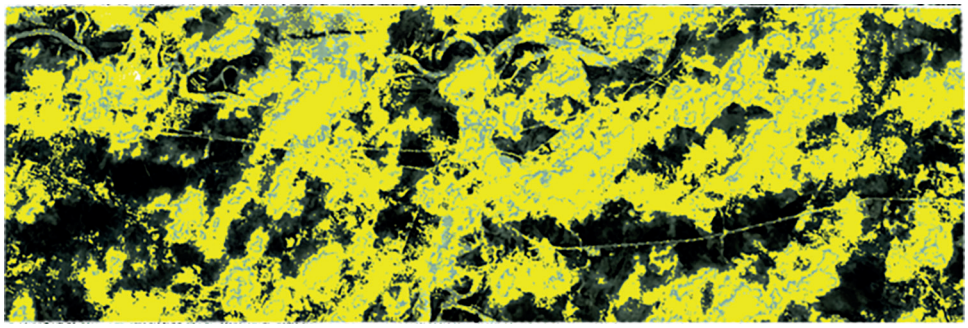


а *б* *в*

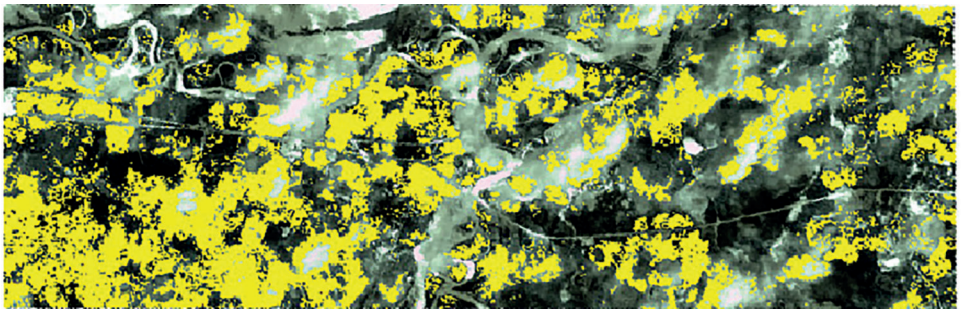
Рис. 4. Пример выделения маски облачности на летнем снимке PlanetScope с помощью алгоритма S2cloudless: *а* — исходное изображение в синтезе RGB; *б* — маска облачности (жёлтый цвет); *в* — стандартная маска облачности (жёлтый цвет)



а



б



в

Рис. 5. Пример выделения маски облачности на зимнем снимке PlanetScope алгоритмом S2cloudless: *а* — исходное изображение в синтезе RGB; *б* — маска облачности, выделенная алгоритмом S2cloudless (жёлтый цвет); *в* — стандартная маска облачности (жёлтый цвет)

Также апробация работы алгоритма была произведена и на зимних снимках PlanetScope. На представленном зимнем снимке точность (F-мера) составила 70 %, в то время как стандартная маска облачности, которую можно получить вместе со снимком, показала точность чуть выше 50 %.

Также были проведены эксперименты с алгоритмом CCNN, который основан на глубоком обучении. Данный алгоритм позволяет успешно выделять облачность на летних снимках, однако по зимним снимкам получен неудовлетворительный результат.

Метод CCNN работает на летних снимках, однако он не справляется со снимками, на которых присутствует снежный покров, что было выявлено в рамках настоящего исследования. Это объясняется отсутствием в обучающих выборках изображений со снегом, поэтому было предпринято несколько попыток подстроить алгоритм под зимние снимки PlanetScope. Однако такие методы оказались хуже, чем предыдущий. Суть данных подходов представлена на *рис. 6*. Отдельно стоит отметить первый из них, где на первом этапе модель обучалась трансформировать летние фотографии в зимние, а после была проведена попытка перенести это преобразование на спутниковые снимки.

Рис. 6. Подходы к адаптированию алгоритма CCNN для зимних снимков

Результаты и обсуждение

С появлением данных с высоким пространственным разрешением (Sentinel-2, PlanetScope) становится всё более актуальной задача автоматического маскирования облачности для последующего использования снимков в системах мониторинга. В настоящем исследовании проанализирована работа нескольких традиционных алгоритмов маскирования облачности (Fmask, Sen2Cor) и методов, основанных на применении машинного обучения (S2cloudless). В качестве контрольного набора было выбрано 12 сцен разных сезонов года на территории Пермского края. По ним была проведена оценка в разрезе сезонов года и по разным типам облачности.

Наилучший результат при оценке общей точности (F-меры) показал алгоритм S2cloudless (средняя F-мера — 83%), на втором месте — Sen2Cor (75 %), а самый низкий процент точности — у алгоритма Fmask (70 %). Самый высокий процент точности для сплошной облачности продемонстрировал алгоритм Fmask — 98 %. На втором месте — S2cloudless (82 %), на третьем — Sen2Cor (72 %). Для полупрозрачной облачности самую высокую точность (F-мера) показал также алгоритм Fmask — 87 %. На втором месте снова оказался алгоритм S2cloudless — 55 %, на третьем — Sen2Cor (1 %).

Для алгоритма с самыми высокими значениями F-меры (S2cloudless) была изучена возможность его применения для снимков PlanetScope. Такой эксперимент оказался удачным для зимних снимков — S2cloudless успешно выделяет на них облачность (Fmeasure ~70 %, что на 20 % лучше, чем стандартная маска облачности). Что касается летних снимков, то здесь точность выделения облачности оказалась меньше, чем у стандартной маски облачности. Поскольку снимки PlanetScope в настоящее время имеются только в коммерческом доступе, было невозможно оценить вариацию точности по сезонам года и использовать большой набор проверочных данных.

Таким образом, методы машинного обучения имеют высокие перспективы применения при обработке данных дистанционного зондирования Земли, в том числе при маскировании облачности за счёт использования спектральных и текстурных признаков, важность которых возрастает на снимках высокого пространственного разрешения (PlanetScope), что позволяет увеличивать точность выделения облачности, в частности на зимних снимках. Методы машинного обучения позволяют разрабатывать алгоритмы для нескольких съёмочных систем без существенной модификации алгоритма, поэтому такие методы можно эффективно применять в системах мониторинга, где используются снимки Sentinel-2 и PlanetScope.

Исследование проведено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 19-35-90005).

Литература

1. *Андреев А. И.*, *Шамилова Ю. А.*, *Холодов Е. И.* Применение сверточной нейронной сети для детектирования облачности по данным прибора МСУ-МР спутника «Метеор-М» № 2 // Метеорология и гидрология. 2019. № 7. С. 44–53.
2. *Aas C.*, *Jochemsen A.*, *Mantas V.*, *Lewyckyj N.*, *Jozefiak M.*, *Buchhorn M.* Maximizing forest value through using Sentinel-2 in combination with hyperspectral UAVs // Proc. 69th Intern. Astronautical Congress. Bremen, Germany. 2018. P. 4492–4498.
3. *Chuanqi T.*, *Fuchun S.*, *Tao K.*, *Wenchang Z.*, *Chao Y.*, *Chunfang L.* A Survey on Deep Transfer Learning. Artificial Neural Networks and Machine Learning // Proc. 27th Intern. Conf. Artificial Neural Networks (ICANN 2018). 4–7 Oct. 2018, Rhodes, Greece. 2018. V. 11141. P. 270–279. DOI: 10.1007/978-3-030-01424-7_27.
4. *Dagobert T.*, *Morel J.-M.*, *de Franchis C.*, *von Gioi R. G.* Visibility Detection in Time Series of Planetscope Images // Proc. International Geoscience and Remote Sensing Symposium (IGARSS). Yokohama, 2019. P. 1673–1676. DOI: 10.1109/IGARSS.2019.8898892.
5. *Drönner J.*, *Korfhage N.*, *Egli S.*, *Mühling M.*, *Thies B.*, *Bendix J.*, *Freisleben B.*, *Seeger B.* Fast Cloud Segmentation Using Convolutional Neural Networks // Remote Sensing. 2018. V. 10(11). Art. No. 1782. DOI: 10.3390/rs10111782.
6. *Feng Z.*, *Yang J.*, *Yao L.* Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation // IEEE Intern. Conf. Image Processing. Beijing. 2017. P. 1742–1746. DOI: 10.1109/ICIP.2017.8296580.
7. *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 // Proc. 31st Annual Conf. Neural Information Processing Systems (NIPS 2017). California. 2017. P. 3149–3157.
8. *Gursky J.* Boosting Showdown: Scikit-Learn vs XGBoost vs LightGBM vs CatBoost in Sentiment Classification // Towards Data Science. 2020. URL: https://towardsdatascience.com/boosting-showdown-scikit-learn-vs-xgboost-vs-lightgbm-vs-catboost-in-sentiment-classification-f7c7f46fd956 (accessed 20.07.2020).
9. *He K.*, *Zhang X.*, *Ren S.*, *Sun J.* Deep residual learning for image recognition // Proc. IEEE Conf. Computer Vision and Pattern Recognition. Las Vegas. 2016. P. 770–778. DOI: 10.1109/CVPR.2016.90.

10. *Hethcoat M. G.*, *Edwards D. P.*, *Carreiras J. M. B.*, *Bryant R. G.*, *França F. M.*, *Quegan S.* A machine learning approach to map tropical selective logging // Remote Sensing of Environment. 2019. V. 221. P. 569–582. DOI: 10.1016/j.rse.2018.11.044.
11. *Hoerl E.*, *Kennard W. R.* Ridge Regression: Biased Estimation for Nonorthogonal Problems // Technometrics. 1970. V. 12. P. 55–67. DOI: 10.1080/00401706.1970.10488634.
12. *Huang G.*, *Liu Z.*, *Van Der Maaten L.*, *Weinberger K. Q.* Densely connected convolutional networks // Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR). Honolulu. 2017. P. 4700–4708.
13. *Li Z.*, *Shen H.*, *Cheng Q.*, *Liu Y.*, *You S.*, *He Z.* Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors // ISPRS J. Photogrammetry and Remote Sensing. 2019. V. 150. P. 197–212. DOI: 10.1016/j.isprsjprs.2019.02.017.
14. *Liu C.-C.*, *Zhang Y.-C.*, *Chen P.-Y.*, *Lai C.-C.*, *Chen Y.-H.*, *Cheng J.-H.*, *Ko M.-H.* Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation // Remote Sensing. 2019. V. 11(2). Art. No. 119. DOI: 10.3390/rs11020119.
15. *Liu S.*, *Li M.*, *Zhang Z.*, *Xiao B.*, *Cao X.* Multimodal Ground-Based Cloud Classification Using Joint Fusion Convolutional Neural Network // Remote Sensing. 2018. V. 10. Art. No. 822. DOI: 10.3390/rs10060822.
16. *Lonjou V.*, *Desjardins C.*, *Hagolle O.*, *Petrucci B.*, *Thierry T.*, *Dejus M.*, *Makarau A.*, *Auer S.* MACCS-ATCOR joint algorithm (MAJA) // Proc. Conf. SPIE Remote Sensing. 28–29 Sept. 2016. Edinburgh. 2016. V. 1. P. 119.
17. *Main-Knorn M.*, *Pflug B.*, *Louis J.*, *Debaecker V.*, *Müller-Wilm U.*, *Gascon F.* Sen2Cor for Sentinel-2 // Image and Signal Processing for Remote Sensing XXIII: Proc. SPIE. Warsaw. 2017. V. 10427. P. 218.
18. Planet Imagery Product Specification. Planet Labs Inc., 2020. 97 p. URL: https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (accessed 20.07.2020).
19. *Powers D.* Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation // J. Machine Learning Technologies. 2011. V. 2(1). P. 37–63.
20. *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. V. 231. P. 83–94. DOI: 10.1016/j.rse.2019.05.024.
21. S2 MPS. Sen2Cor Configuration and User Manual. Ref.: S2-PDGS-MPC-L2A-SUM-V2.4. Iss. 1. ESA, 2017. 53 p. URL: https://step.esa.int/thirdparties/sen2cor/2.4.0/Sen2Cor_240_Documenation_PDF/S2-PDGS-MPC-L2A-SUM-V2.4.0.pdf (accessed 20.07.2020).
22. Sentinel-2 Products Specification Document. ESA, 2018. Ref.: S2-PDGS-TAS-DI-PSD. Iss. 14.5. 510 p. URL: https://sentinel.esa.int/documents/247904/685211/Sentinel-2-Products-Specification-Document (accessed 20.07.2020).
23. *Shendryk Y.*, *Rist Y.*, *Ticehurst C.*, *Thorbum P.* Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery// ISPRS J. Photogrammetry and Remote Sensing. 2019. V. 157. P. 124–136. DOI: 10.1016/j.isprsjprs.2019.08.018.
24. *Simonyan K.*, *Zisserman A.* Very deep convolutional networks for large-scale image recognition // Proc. 3rd Intern. Conf. Learning Representations. San Diego. 2014. arXiv: 1409.15556.
25. *Syrris V.*, *Hasenohr P.*, *Delipetrev B.*, *Kotsev A.*, *Kempeneers P.*, *Soille P.* Evaluation of the potential of convolutional neural networks and random forests for multi-class segmentation of Sentinel-2 imagery // Remote Sensing. 2019. V. 11(8). Art. No. 907. DOI: 10.3390/rs11080976.
26. *Xu K.*, *Guan K.*, *Peng J.*, *Luo Y.*, *Wang S.* DeepMask: an algorithm for cloud and cloud shadow detection in optical satellite remote sensing images using deep residual network // arXiv: 1911.03607. 2019. V. 1911. Article No. 3607.
27. *Zhe Z.*, *Curtis E. W.* Object-based cloud and cloud shadow detection in Landsat imagery. // Remote Sensing of Environment. 2012. V. 118. P. 83–94. DOI: 10.1016/j.rse.2011.10.028.

Estimation of the accuracy of cloud masking algorithms using Sentinel-2 and PlanetScope data

A. V. Tarasov

Perm State University, Perm 614990, Russia
E-mail: gis@psu.ru

Nowadays, many automated satellite-based monitoring systems (e. g. for forestry or agriculture) widely use the images from Sentinel-2 and PlanetScope satellites, which combine high spatial and temporal resolution. An important challenge in the building of such monitoring systems is high-quality preprocessing of the images, including cloud masking. Traditional (threshold-based) cloud masking algorithms use only spectral reflectance in the visible and infrared (IR) bands. Also the machine learning methods (which take into account not only spectral characteristics but also geometry and texture) are successfully used for cloud masking on satellite images in recent decade. In this study, we estimated accuracy of cloud masking on Sentinel-2 images obtained at different seasons with the use of traditional Fmask and Sen2Cor algorithms and machine-learning-based S2cloudless algorithm, on the example of the Perm Region territory. For Sentinel-2 data, S2cloudless showed the best result (average accuracy 83 %), and the lowest accuracy was obtained with Fmask (70 %). The seasonal variability of cloud masking accuracy did not exceed 6 %. The applicability of the listed algorithms for PlanetScope data having only visible and near-IR spectral bands was also estimated. It was found that S2cloudless can be used for PlanetScope images obtained in winter, since the cloud masking accuracy is higher than that the standard cloud mask product delivered with the data.

Keywords: cloud masking, machine learning, PlanetScope, Sentinel-2

Accepted: 05.11.2020
DOI: 10.21046/2070-7401-2020-17-7-26-38

References

1. Andreev A. I., Shamilova Yu. A., Kholodov E. I., Using convolutional neural networks for cloud detection from “Meteor-M” No. 2 MSU-MR Data, *Russian Meteorology and Hydrology*, 2019, Vol. 44, No. 7, pp. 459–466.
2. Aas C., Jochemsen A., Mantas V., Lewyckyj N., Jozefiak M., Buchhorn M., Maximizing forest value through using Sentinel-2 in combination with hyperspectral UAVs, *Proc. 69th Intern. Astronautical Congress*, Bremen, Germany, 2018, pp. 4492–4498.
3. Chuanqi T., Fuchun S., Tao K., Wenchang Z., Chao Y., Chunfang L., A Survey on Deep Transfer Learning, *Proc. 27th Intern. Conf. Artificial Neural Networks* (*ICANN 2018*), Rhodes, Greece, 4–7 Oct. 2018, Vol. 11141, pp. 270–279, DOI: 10.1007/978-3-030-01424-7_27.
4. Dagobert T., Morel J.-M., Franchis C. D., von Gioi R. G., Visibility Detection in Time Series of Planetscope Images, *Proc. Intern. Geoscience and Remote Sensing Symp.* (*IGARSS*), Yokohama, 2019, pp. 1673–1676, DOI: 10.1109/IGARSS.2019.8898892.
5. Drönner J., Korfhage N., Egli S., Mühling M., Thies B., Bendix J., Freisleben B., Seeger B., Fast Cloud Segmentation Using Convolutional Neural Networks, *Remote Sensing*, 2018, Vol. 10(11), Art. No. 1782, DOI: 10.3390/rs10111782.
6. Feng Z., Yang J., Yao L., Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation, *IEEE Intern. Conf. Image Processing*, Beijing, 2017, pp. 1742–1746, DOI: 10.1109/ICIP.2017.8296580.
7. 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, *Proc. 31st Annual Conf. Neural Information Processing Systems* (*NIPS 2017*), California, 2017, pp. 3149–3157.
8. Gursky J., Boosting Showdown: Scikit-Learn vs XGBoost vs LightGBM vs CatBoost in Sentiment Classification, *Towards Data Science*, 2020, available at: https://towardsdatascience.com/boosting-showdown-scikit-learn-vs-xgboost-vs-lightgbm-vs-catboost-in-sentiment-classification-f7c7f46fd956 (accessed 20.07.2020).
9. He K., Zhang X., Ren S., Sun J., Deep residual learning for image recognition, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 770–778, DOI: 10.1109/CVPR.2016.90.

10. Hethcoat M. G., Edwards D. P., Carreiras J. M. B., Bryant R. G., França F. M., Quegan S., A machine learning approach to map tropical selective logging, *Remote Sensing of Environment*, 2019, Vol. 221, pp. 569–582, DOI: 10.1016/j.rse.2018.11.044.
11. Hoerl E., Kennard W. R., Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 1970, Vol. 12, pp. 55–67, DOI: 10.1080/00401706.1970.10488634.
12. Huang G., Liu Z., Van Der Maaten L., Weinberger K. Q., Densely connected convolutional networks, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 2017, pp. 4700–4708.
13. Li Z., Shen H., Cheng Q., Liu Y., You S., He Z., Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors, *ISPRS J. Photogrammetry and Remote Sensing*, 2019, Vol. 150, pp. 197–212, DOI: 10.1016/j.isprsjprs.2019.02.017.
14. Liu C.-C., Zhang Y.-C., Chen P.-Y., Lai C.-C., Chen Y.-H., Cheng J.-H., Ko M.-H., Clouds Classification from Sentinel-2 Imagery with Deep Residual Learning and Semantic Image Segmentation, *Remote Sensing*, 2019, Vol. 11(2), Art. No. 119, DOI: 10.3390/rs11020119.
15. Liu S., Li M., Zhang Z., Xiao B., Cao X., Multimodal Ground-Based Cloud Classification Using Joint Fusion Convolutional Neural Network, *Remote Sensing*, 2018, Vol. 10, Art. No. 822, DOI: 10.3390/rs10060822.
16. Lonjou V., Desjardins C., Hagolle O., Petrucci B., Thierry T., Dejus M., Makarau A., Auer S., MACCS-ATCOR joint algorithm (MAJA), *Proc. Conf. SPIE Remote Sensing*, 28–29 Sept. 2016, Edinburgh, 2016, Vol. 1, p. 119.
17. Main-Knorn M., Pflug B., Louis J., Debaecker V., Müller-Wilm U., Gascon F., Sen2Cor for Sentinel-2, *Image and Signal Processing for Remote Sensing XXIII: Proc. SPIE*, Warsaw, 2017, Vol. 10427, p. 218.
18. *Planet Imagery Product Specification*, Planet Labs Inc., 2020, 97 p., available at: https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (accessed 20.07.2020).
19. Powers D., Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, *J. Machine Learning Technologies*, 2011, Vol. 2(1), pp. 37–63.
20. 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, pp. 83–94, DOI: 10.1016/j.rse.2019.05.024.
21. *S2 MPS. Sen2Cor Configuration and User Manual*, Ref.: S2-PDGS-MPC-L2A-SUM-V2.4, Iss. 1, ESA, 2017, 53 p. available at: https://step.esa.int/thirdparties/sen2cor/2.4.0/Sen2Cor_240_Documenation_PDF/S2-PDGS-MPC-L2A-SUM-V2.4.0.pdf (accessed 20.07.2020).
22. *Sentinel-2 Products Specification Document*, ESA, 2018, Ref.: S2-PDGS-TAS-DI-PSD, Issue 14.5, 510 p., available at: https://sentinel.esa.int/documents/247904/685211/Sentinel-2-Products-Specification-Document (accessed 20.07.2020).
23. Shendryk Y., Rist Y., Ticehurst C., Thorbum P., Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery, *ISPRS J. Photogrammetry and Remote Sensing*, 2019, Vol. 157, pp. 124–136, DOI: 10.1016/j.isprsjprs.2019.08.018.
24. Simonyan K., Zisserman A., Very deep convolutional networks for large-scale image recognition, *Proc. 3rd Intern. Conf. Learning Representations*, San Diego, 2014, arXiv: 1409.15556.
25. Syrris V., Hasenohr P., Delipetrev B., Kotsev A., Kempeneers P., Soille P., Evaluation of the potential of convolutional neural networks and random forests for multi-class segmentation of Sentinel-2 imagery, *Remote Sensing*, 2019, Vol. 11(8), Art. No. 907, DOI: 10.3390/rs11080976.
26. Xu K., Guan K., Peng J., Luo Y., Wang S., DeepMask: an algorithm for cloud and cloud shadow detection in optical satellite remote sensing images using deep residual network, *ArXiv: 1911.03607*, 2019, Vol. 1911, Article No. 3607.
27. Zhe Z., Curtis E. W., Object-based cloud and cloud shadow detection in Landsat imagery, *Remote Sensing of Environment*, 2012. Vol. 118, pp. 83–94, DOI: 10.1016/j.rse.2011.10.028.