Семантическая сегментация спутниковых снимков аэропортов с помощью свёрточных нейронных сетей

В.А. Горбачёв [1], И.А. Криворотов [1,2], А.О. Маркелов [1,2], Е.В. Котлярова [2]
[1] Государственный научно-исследовательский институт авиационных систем (ГНЦ РФ), Москва, Россия,
[2] МФТИ, Москва, Россия

Аннотация

Статья посвящена разработке эффективного алгоритма семантической сегментации для разметки элементов аэропортовой инфраструктуры на космических снимках оптического диапазона. В данной работе применены алгоритмы сегментации на основе глубоких свёрточных нейронных сетей. Они зарекомендовали себя в широком ряде задач, в том числе сегментации изображений наземной съёмки, где они показывают стабильно высокие результаты. В ходе работы были вручную размечены обучающие и тестовые изображения. Был произведён поиск оптимальной для данной задачи архитектуры нейронной сети. Исследованы различные комбинации энкодеров и декодеров. Для постобработки и учёта контекстной информации и соседства объектов различных классов с целью устранения выбросов применена модель условных случайных полей. Описаны особенности применённых решений на всех этапах подготовки алгоритма: подготовка данных, обучение нейронной сети и постобработка её результатов.

Ключевые слова: семантическая сегментация, искусственные нейронные сети, глубокое обучение, обработка изображений.

Цитирование: **Горбачёв, В.А**. Семантическая сегментация спутниковых снимков аэропортов с помощью свёрточных нейронных сетей / В.А. Горбачёв, И.А. Криворотов, А.О. Маркелов, Е.В. Котлярова // Компьютерная оптика. – 2020. – Т. 44, № 4. – С. 636-645. – DOI: 10.18287/2412-6179-CO-636.

Citation: Gorbachev VA, Krivorotov IA, Markelov AO, Kotlyarova EV. Semantic segmentation of satellite images of airports using convolutional neural networks. Computer Optics 2020; 44(4): 636-645. DOI: 10.18287/2412-6179-CO-636.

Введение

Сегодня в разных областях науки и техники широко востребованы цифровые карты местности и геоинформационные системы. На современном этапе карты должны содержать не только пространственное расположение объектов и высоты точек рельефа, но и подробную информацию об объектовом составе. Эта информация необходима в широком круге задач от планирования и администрирования территорий до экологического и кадастрового мониторинга. Отдельную роль электронные карты и планы играют в авиации. Не только маршрутизация воздушного и наземного аэропортового транспорта, но и системы повышения ситуационной осведомлённости и дополненного или синтезированного видения существенно опираются на детальные карты местности и информацию об объектовом составе.

При этом важны не только состав и детальность электронных карт, но и сроки их изготовления и обновления. Процесс их создания является сложной и трудоёмкой задачей, требующей значительного количества ручного труда. Алгоритмы семантической сегментации позволяют в существенной степени автоматизировать этот процесс. Полученная алгоритмами информация потребует обработки оператором-картографом, но существенно снизит его нагрузку.

В связи с этим в работе рассмотрен вопрос применения алгоритмов семантической сегментации к задаче автоматизации обработки аэрофото- и космических снимков аэропортов для выделения границ объектов и объектового состава.

Семантическая сегментация изображений – это разделение изображения на отдельные группы пикселей, области, соответствующие одному классу объектов с одновременным определением типа объекта в каждой области. Задача семантической сегментации является высокоуровневой задачей обработки изображений, относящейся к группе задач так называемого слабого искусственного интеллекта. Она является более сложной, чем задача классификации изображений и детектирования объектов, так как необходимо не только определять классы объектов, но и правильно выделять их границы на изображении. В то же время задача семантической сегментации заметно отличается от обычной сегментации, когда области объединяются по принципу цветового или текстурного сходства. Объекты могут иметь существенно различающиеся по фотометрическим характеристикам элементы и иметь значительный разброс показателей объектов внутри одного класса. В данной работе семантическая сегментация изображений применяется к спутниковым снимкам аэропортов в целях автоматизации процесса обработки снимков для обновления

карт и извлечения информации о расположении элементов аэропортовой инфраструктуры, зелёных зон, зданий и проч. Задачей работы является исследование влияния различных элементов алгоритма на качество сегментации.

Обзор существующих работ

Для задачи семантической сегментации исторически существует большое количество методов решения, однако результаты сравнения алгоритмов на открытых наборах данных, например, ISPRS Semantic Labeling Contest [1], показывают значительное превосходство алгоритмов, основанных на свёрточных нейронных сетях в комбинации с различными подходами к предобработке и постобработке изображений.

Подобные алгоритмы были разработаны относительно недавно, статья о первой успешной нейросетевой архитектуре для сегментации FCN-8s вышла в 2014 году [2], однако сейчас именно они показывают наилучшую точность работы. Большинство нейросетевых алгоритмов семантической сегментации имеют аналогичную этой сети архитектуру: сначала для выделения семантической информации изображение преобразуется в вектор признаков с помощью сети-шифровальщика (*encoder*), затем вектор обратно разворачивается в матрицу изображения с помощью сети-дешифровальщика (*decoder*). В качестве сети-шифровальщика часто используют различные заранее обученные свёрточные сети, такие как VGG [3] или ResNet [4]. Построение сети-дешифровальщика – задача более открытая, так как необходимо по семантической карте низкого разрешения построить попиксельную карту разметки высокого разрешения, восстановив пространственную информацию. Различные архитектуры используют разные механизмы для решения этой проблемы.

Так, например, в архитектуре сети SegNet [5] используется операция рассоединения (*unpooling*). Её новшество заключается в том, что при операции объединения по максимуму (*max-pooling*), на этапе свёртки в сети-шифровальщике индексы максимальных значений сохраняются и позже используются, чтобы повысить дискретизацию соответствующей карты признаков в сети-дешифровальщике, совершив операцию рассоединения (*unpooling*) с использованием сохраненных индексов. Модель U-net [6] использует идею сквозных соединений (*skip-connection*) для сохранения пространственной информации. Карты признаков из сети-шифровальщика напрямую передаются и конкатенируются с картами признаков на соответствующих слоях сети-дешифровальщика, параллельно с обычными свёрточными слоями. В LinkNet [7] вместо конкатенации применяется сложение карт признаков. Архитектура DeepLab [8] привнесла три новшества. Во-первых, это свёртка фильтрами с повышенной дискретизацией (*atrous convolution, dilated convolution*). Во-вторых, авторы первыми предложили пространственное пирамидальное объединение (ASPP) таких фильтров для сегментирования объектов в разных масштабах. В-третьих, была улучшена локализация границ объектов с помощью комбинирования методов из глубоких свёрточных нейронных сетей и вероятностных графических моделей (CRF) для учёта контекстной информации. В CFNet [9] для учёта контекста предлагается использовать специальный модуль Aggregated Co-Occurrent Feature Module, оценивающий вероятности совместного проявления различных признаков. Приём объединения карт признаков spatial pyramid pooling, с помощью которого сеть-дешифровальщик получает информацию о глобальном контексте, предложен в архитектуре PSPNet [10]. RefineNet [11] подходит к проблеме потери контекста другим, более производительным способом, заметно отличающимся от PSPNet. Предложенная авторами архитектура итеративно объединяет повышающие разрешение векторы признаков с помощью специальных блоков RefineNet для нескольких диапазонов разрешений, и, наконец, создаёт карту сегментации с высоким разрешением. В DANet [12] авторы используют механизм внимания (*self-attention*) для моделирования зависимостей как внутри каждого канала, так и между каналами. Модель Mask R-CNN [13] явилась развитием методов детектирования и решает одновременно две задачи: строит ограничивающий прямоугольник (*bounding box*) объекта для решения задачи детектирования и одновременно в этом прямоугольнике производит сегментацию. Наконец, принципиально другим подходом к задачам семантической сегментации стало использование генеративных состязательных сетей (*generative adversarial networks*), работающих без прямого использования функции потерь для сегментации [14].

План работы

Решение задачи сегментации включает в себя следующие этапы:

- предварительная обработка данных;
- создание обучающих выборок;
- выбор архитектуры алгоритма;
- выбор наиболее подходящей функции потерь;
- обучение и выполнение алгоритма;
- постобработка полученных карт разметки.

Можно заметить, что алгоритм имеет модульную структуру и допускает выбор различных методов на каждом этапе и их комбинирование. В работе было приведено исследование каждого этапа, предложены различные подходы по оптимизации алгоритма на каждом этапе и их экспериментальное сравнение.

Исходные данные

В качестве исходных данных были выбраны спутниковые снимки оптического диапазона аэропортов Домодедово, Шереметьево, Пулково, Внуково и

Хельсинки, с разрешением 1/3 метра на пиксел, полученных спутником WorldView-3. Разметка была произведена вручную по семи классам, представленным в табл. 1. Всего в обучающей коллекции было 31 изображение высокого разрешения для обучения нейросети, 3 для валидации и 4 для тестирования. Пример сегментированного человеком изображения (*ground truth*) представлен на рис. 1.

Табл. 1. Кодирование объектов на снимках

	Тип объекта	Hex-код цвета	Цвет
0	Здания	0000ff	синий
1	Растительность	00ff00	зелёный
2	Земля, стройка	ffff00	жёлтый
3	«Бетон» (ВПП, рулёжки)	ffffff	белый
4	«Асфальт» (автодороги)	00ffff	бирюзовый
5	Нестационарные объекты	ff00ff	фиолетовый
6	Другое	ff0000	красный

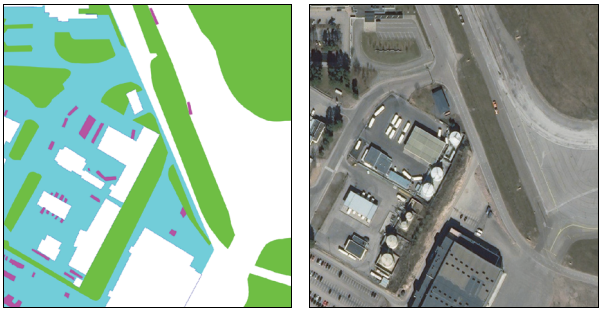

Рис. 1. Пример пары изображений из обучающей выборки

При обучении разметка классов была редуцирована. Отличить покрытие автомобильных дорог от покрытия аэродромов можно только по глобальному контексту, визуально они идентичны, а переходы между ними носят крайне условный характер. Поэтому они были объединены в один класс. Класс «объекты» содержал самолёты, небольшие аэропортовые объекты, автомобили. Он крайне мало представлен в выборке и недостаточно точно размечен (оператор разметки, как правило, объединял рядом стоящие объекты в целые области), поэтому был объединён с предыдущим классом, так как на собранной выборке выучить его сегментацию оказалось невозможно. Он только ухудшал точность остальных классов. Классы «земля», «стройка» и «прочее» тоже были объединены, так как визуально очень похожи. В итоге использовалось для обучения 4 класса: 1 – земля, стройка, прочее; 2 – растительность; 3 – асфальто-бетонное покрытие; 4 – здания.

Аугментация данных

Отметим, что изображения различных использованных аэропортов имеют существенные различия. Снимки были произведены с разных спутников и в разное время, имеют различные цветопередачу, условия освещённости, на снимках могут присутствовать тени от облаков и т.д. Например, из рис. 2 видно, что цвета травы, асфальта и бетона сильно отличаются у двух разных изображений. Искусственная имитация таких фотометрических особенностей на этапе обучения необходима для повышения обобщающей способности алгоритма.

Рис. 2. Пример различия цветового баланса между изображениями различных аэропортов

Изображения в высоком разрешении были нарезаны на фрагменты размером 440 × 440 и 660 × 660 с перекрытием в половину размера фрагмента. Затем фрагменты масштабировались к единому размеру. Всего было получено 3000 фрагментов. Использование масштабирования позволяет моделировать возможную разницу в размерах объектов, а нарезка с перекрытием позволяет использовать большее количество контекстов. Для того чтобы дополнительно увеличить обучающую выборку и имитировать различия между снимками, были проведены следующие операции: случайные повороты изображений, случайные изменения масштаба в небольшом диапазоне (15 %) и случайные мультипликативные изменения яркости (до 30 %). Примеры аргументированных снимков приведены на рис. 3.

Рис. 3. Пример аугментации изображения: оригинальный снимок, поворот, изменение яркости и приближение

Выбор архитектуры нейросети

Было исследовано 3 разных базовых архитектуры для сегментации изображений: Unet, PSPNet, LinkNet. Каждая архитектура обучалась с несколькими различными энкодерами, такими как: VGG16 [3], ResNet34 [4], InceptionV3 [15], MobileNetV2 [16], EfficientNetB0 [17]. Всего для выявления лучшего подхода и особенностей для данной задачи было обучено 15 различных моделей. Результаты сравнения приведены в параграфе Эксперименты. Пример комбинирования различных энкодеров и декодеров проиллюстрируем на примере сети Unet-VGG16.

По ряду причин U-net [5] является хорошей базовой архитектурой. U-net был создан для семантической сегментации медицинских изображений, для которых характерен постоянный ракурс и масштаб объектов, что соответствует постановке нашей задачи. U-net существенно использует идею сквозных соединений (*skip-connection*), которая даёт очень хорошие результаты по сравнению с обычными автоэнкодерами. Благодаря этому U-net не требует большого количества изображений для обучения, так как имеет сравнительно небольшое количество параметров.

Архитектура U-Net состоит из двух соединённых между собой сетей: сети-шифровальщика (энкодера) для извлечения из изображения семантической информации в виде вектора признаков и сети-дешифровальщика (декодера) для превращения вектора признаков в матрицу нового изображения – маски классов (рис. 4). Чтобы сохранить пространственную информацию (контекст), карты признаков из энкодера с помощью сквозных соединений напрямую передаются в декодер и конкатенируются с картами признаков соответствующего разрешения декодера.

Для повышения ёмкости модели и точности работы сети в качестве сети-шифровальщика вместо исходного энкодера можно использовать сеть VGG-16 [3]. Получившуюся архитектуру VGG-Unet можно увидеть на рис. 6.

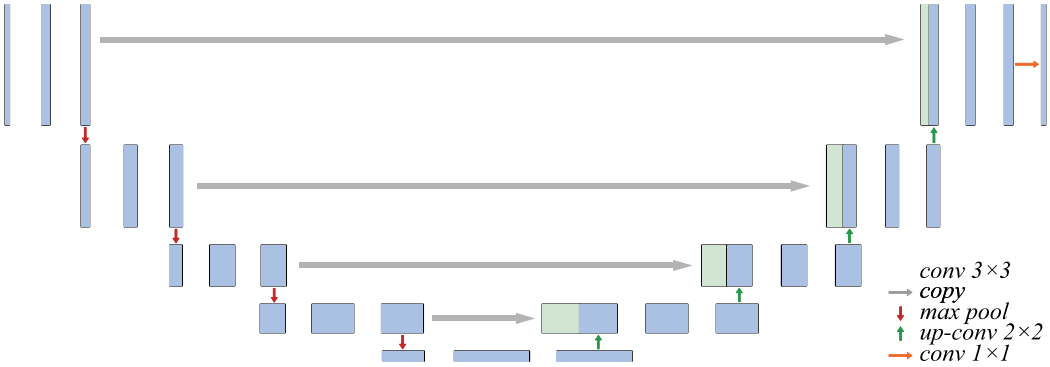

Рис. 4. Архитектура U-net

Рис. 5. Архитектура VGG-Unet

Функция потерь

Распространённой проблемой в области анализа изображений является детектирование или сегментирование очень маленького «аномального» региона большого изображения (например, выделение небольшого здания на фоне большого количества зелёных насаждений). Такие данные называются несбалансированными. Легко классифицируемые примеры составляют большую часть обучающей выборки и доминируют при расчёте функции потерь. Сложные примеры, на которых сеть ошибается, практически игнорируются при обучении, так как их количество относительно мало. Проблему несбалансированности данных можно решать, например, с помощью изменения обучающей выборки либо с помощью выбора функции потерь (*loss function*). В данной работе были исследованы три функции потерь: кросс-энтропия, обобщённая функция потерь Дайса и фокальная функция потерь.

Кросс-энтропия

Кросс-энтропийная (CE) функция потерь часто используется в задачах семантической сегментации.

Её выходной сигнал представляет собой значение вероятности в диапазоне от 0 до 1. Величина кросс-энтропийной функции потерь увеличивается, когда прогнозируемая вероятность отклоняется от целевой метки. В бинарной классификации, где количество классов равно двум, кросс-энтропия может быть посчитана так:

$$CE\left(p,y\right)=-\left(y\ \ln p+\left(1-y\right)\ln(1-p)\right),$$

где $y=0$ для объекта первого класса и $y=1$ для второго, p – вероятность того, что объект принадлежит ко второму классу.

Если классов больше двух, нужно рассчитать отдельные значения для каждого класса и просуммировать результат:

$$CE\left(p,y\right)=-\sum_i y_i\ \ln p_i\ .$$

Здесь и далее $y_i=1$, когда объект принадлежит к классу i, и $y_i=0$ иначе, а p_i – предсказанная вероятность принадлежности объекта к классу i.

Обобщённая функция потерь Дайса

На основе известной меры сходства между множествами такой, как коэффициент Дайса–Сёренсена, можно построить функцию потерь (т.н. *Dice loss*):

$$DL=1-2\frac{\sum y_i\ p}{\sum\left(p_i+y_i\right)}\ .$$

Чтобы избежать проблемы влияния несбалансированности классов на функцию потерь, в статье [18] авторы предложили использовать функцию потерь следующего вида (*Generalized Dice loss*):

$$GDL=1-2\frac{w_i\sum y_i\ p}{w_i\sum\left(p_i+y_i\right)},$$

где $w_i=(\Sigma y_i)^{-2}$, а Σy_i – это сумма y_i по всему изображению, то есть количество пикселей класса i. Благодаря такому взвешиванию сеть лучше обучается на объектах редко встречающихся классов, так как вес ошибок на них в функции потерь увеличивается.

Фокальная функция потерь

Авторы статьи [19] предлагают изменить форму функции потери таким образом, чтобы сосредоточить её внимание на сложных редко встречающихся примерах. В случае доминирования объектов одного из классов сеть при обучении будет пытаться устранить даже небольшие ошибки на объектах доминирующих классов, а существенные ошибки редких классов будет игнорировать. Авторами предлагается добавить модулирующий фактор $(1-p_t)^\gamma$ к кросс-энтропийной функции потерь с настраиваемым фокусирующим параметром $\gamma\geq0$.

$$FL\left(p_t\right)=-\sum_i y_i\left(1-p_i\right)^\gamma\ln p_i\ ,$$

где $y_i=1$, если объект принадлежит к классу i, и 0 иначе, а p_i – предсказанная вероятность принадлежности объекта к классу i.

Когда пример классифицирован неправильно и вероятность p_i_x0001_ небольшая, модулирующий фактор близок к единице и значение функции потерь не изменяется. При вероятности p_i_x0001_, близкой к 1, коэффициент модуляции становится равным 0, а значение функции потери для уверенно классифицированных примеров снижаются. Параметр фокусировки γ плавно регулирует скорость, с которой у «простых» примеров понижаются веса. Когда $\gamma=0$, фокальная функция потерь становится тождественна кросс-энтропийной.

Постобработка

Для улучшения точности классификации необходимо учитывать пространственные зависимости между целевыми переменными для улучшения пространственной поддержки разметки, при этом оставляя задачу эффективно вычислимой. Был использован структурный подход на основе модели условных случайных полей (CRF). Условные случайные поля являются методом статистического моделирования, который часто применяется в машинном обучении. В то время как дискретный классификатор предсказывает метку для одного образца без учёта меток соседних объектов, CRF может учитывать контекст, что является важным для данной задачи, что было показано в статье [20]. Учёт контекста заключается в гипотезе о том, что соседний сегмент для данного сегмента имеет тот же класс (то есть границы между классами относительно редки), а вероятность встретить по соседству сегменты различных классов соответствует вероятности на обучающей выборке. Математически такая гипотеза формулируется в виде модели условного случайного поля. Вероятность того, что изображению I соответствует разметка Y, можно описать следующим образом:

$$P(Y\mid I)=\frac{1}{Z(I)}\prod_i\Phi(y_i\mid I)\prod_i\prod_{j\in N(i)}\Psi(y_i,y_j\mid I)\ ,$$

где $\Phi(y_i\mid I)$ – фактор, выражающий вероятность того, что сегмент i изображения будет иметь класс y_i, $\Psi(y_i,y_i\mid I)$ – фактор, выражающий вероятность того, что сегмент i и его сосед j из окрестности $N(i)$ будут иметь классы y_i и y_j одновременно. Перемножение производится по всем сегментам i изображения I. $Z(i)$ – константа нормализации, равная сумме по всем возможным разметкам произведений факторов. Максимизацию вероятности можно заменить минимизацией логарифма вероятности, т.н. «энергии» карты разметки:

$$E(Y \mid I) = \sum_i \varphi(y_i \mid I) + \sum_i \sum_{j \in N(i)} \psi(y_i, y_j \mid I),$$

где $\varphi = ln\Phi$ – «унарный потенциал», $\psi = ln\Psi$ – «бинарный потенциал». При этом вычислять константу нормализации не требуется. В результате постобработки карта разметки, полученная нейростетью, заменяется картой разметки, имеющей наименьшую энергию (то есть наибольшую совместную правдоподобность в смысле указанного функционала). Результаты экспериментов показывают, что условные случайные поля (CRF) являются эффективным инструментом для постобработки изображений. Она позволяет заметно устранить выбросы и сгладить результаты сегментации.

Результаты экспериментов

При обучении в качестве оптимизатора выбран Adadelta, функция потерь – Focal Loss, размер батча – 32. Для инициализации параметров использовались предобученные на коллекции данных ImageNet [21] веса энкодеров. Время обучения составляло от 4 до 8 часов на видеокарте Tesla K80.

По графикам обучения (рис. 6) видно, что для всех моделей значение 0,9 для F1-score является предельным значением. Так как среди моделей присутствуют разного рода архитектуры, как более лёгкие (MobileNetV2, EfficientNetB0), так и более тяжёлые (InceptionV3, VGG16), и все они запоминают обучающую выборку одинаково, то можно предположить, что эта граница обусловлена качеством разметки обучающей выборки. Однако модели имеют различную обобщающую способность, что видно на графиках валидации (рис. 7).

На рис. 8 можно увидеть результаты сегментации по изначально указанным в разметке 7 классам. У исходно обученной сети плохо распознаётся класс земля/стройка, потому что он очень разнообразен внешне (имеет большую внутриклассовую дисперсию) и недостаточно обширно представлен на изображениях из обучающей выборки. Также плохой результат у класса дорог, так как отличить взлётно-посадочную полосу и рулёжные дорожки от автомобильной дороги можно только по контексту или по специфическим линиям разметки. Это естественно, так как классы локально визуально практически неразличимы, имеют почти идентичную текстуру и различаются только функционально. На данной выборке сеть оказалась не в состоянии выполнить такой высокоуровневый анализ. Помимо этого, довольно редко встречается класс «Нестационарные объекты», его удельная площадь на изображениях крайне низка. Это объясняет предпринятое далее перераспределение классов и редуцирование выборки до 4 классов на обучении. Сеть обучалась на тех же исходных изображениях, но с изменёнными метками.

Тестирование алгоритмов производилось на 4 отложенных снимках высокого разрешения, поделённых на 512 тестовых изображений. Точность вычислялась как доля правильно классифицированных пикселей изображения. При подсчёте сравнивались две карты разметки: полученное нейросетью и размеченное человеком. Результаты по разным изображениям усреднялись. Для оценки результатов использовались метрики F1-score по каждому классу, их среднее по классам (*Avg*) и общая точность (*Accuracy*) по всему изображению, они приведены в табл. 1. В целом видно, что лучше всего себя показало семейство LinkNet. С лучшей в среднем сетью LinkNet-EfficientNetB0 была проведена серия экспериментов с различными вариациями условий (табл. 2).

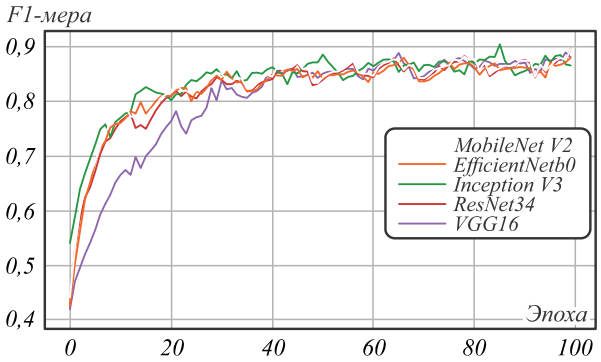

Рис. 6. График функции потерь на обучении

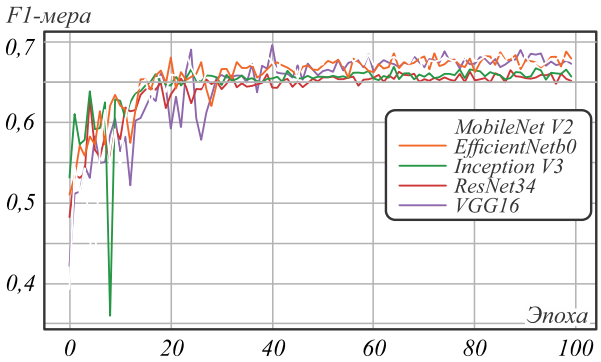

Рис. 7. График функции потерь на валидации

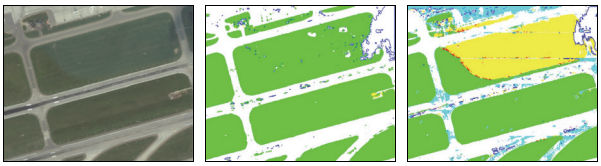

Рис. 8. Оригинальное изображение, сегментированное нейронной сетью на 4 класса, сегментированное нейронной сетью на 7 классов

Из анализа выборки можно заключить, что, помимо низкого качества разметки, основной проблемой является дисбаланс классов. Класс «Стройка, земля, мусор» представлен в выборке крайне незначительно по сравнению с другими классами, вследствие чего точность распознавания на этом классе довольно низка. Это подтолкнуло к попытке искусственно изменить баланс классов в обучающей выборке при обучении, включая больше изображений, где данный класс занимает большую точность. Для одной из се-

тей (Unet-MobileNetV2), которая была не самой лучшей, была проведена серия экспериментов на таких перебалансированных данных (табл. 3). Результаты в целом были улучшены. На классе «здания» был достигнут лучший среди всех результат. Неожиданно балансировка отрицательно сказалась на результатах сети LinkNet-EfficientNetB0.

Применение искусственного расширения обучающих данных привело к большей устойчивости алгоритма при появлении разных условий освещённости, изменении цвета травы, также стали лучше выделяться здания на снимках. К сожалению, в случае с тенями облаков использование данного приёма не всегда приводило к существенному результату (рис. 9). Предположительно, это вызвано тем, что тень от облака имеет ярко выраженную границу, в то время как аугментированные снимки затемнены целиком.

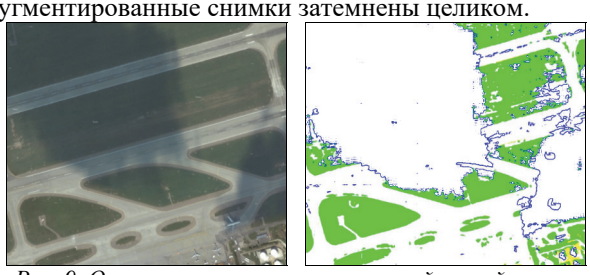

Рис. 9. Оригинальное и распознанное нейронной сетью изображение

Табл. 2. Сравнение различных архитектур. В ячейках указана F1-мера, её среднее значение по классам (Avg) и точность по всем пикселам в столбце Accuracy

Модель	**1**	**2**	**3**	**4**	**Avg**	**Accuracy**
Unet-EfficientNetB0	30,65	97,1	95,08	64,11	71,73	94,53
Unet-ResNet34	19,10	97,25	94,15	75,00	71,37	94,35
Unet-MobileNetV2	37,21	96,89	94,57	72,47	75,29	94,05
Unet-VGG16	29,17	96,61	93,52	71,26	72,64	92,95
Unet-InceptionV3	20,95	97,25	94,50	76,75	72,36	94,20
Linknet-VGG16	31,51	97,32	95,33	78,83	**75,75**	95,14
LinkNet-MobileNetV2	23,65	97,05	94,60	65,30	70,15	94,10
LinkNet-ResNet34	25,65	96,90	93,85	**80,45**	74,21	94,10
LinkNet-EfficientNetB0	35,84	**97,49**	95,34	73,56	75,56	**95,22**
LinkNet-Inceptionv3	23,50	97,42	**95,46**	75,99	73,09	94,79
PSPNet-EfficientNetB0	18,85	96,70	93,35	38,75	61,91	92,40
PSPNet-MobileNetV2	9,10	96,25	92,00	40,75	59,52	91,50
PSPNet-VGG16	**45,35**	96,45	94,2	51,05	71,76	93,70
PSPNet-ResNet34	25,76	96,72	93,31	39,07	63,71	93,08
PSPNet-Inceptionv3	32,31	96,38	93,52	47,19	67,35	92,17

Табл. 3. Точности алгоритма LinkNet-EfficientNetB0 для различных условий обучения. В ячейках указана F1-мера, её среднее значение по классам (Avg) и точность по всем пикселам в столбце Accuracy

Balance	**CCE**	**Focal**	**Dice**	**CRF**	**1**	**2**	**3**	**4**	**Avg**	**Accuracy**
V		V			29,12	97,22	94,35	59,11	69,95	94,28
	V				26,81	96,48	94,9	60,07	69,57	94,36
		V			35,84	97,49	95,34	73,56	75,56	95,22
			V		32,74	96,29	94,19	72,51	73,93	93,86
	V			V	26,93	96,83	95,07	59,16	69,49	94,66
		V		V	**37,04**	**97,65**	**95,46**	**74,70**	**76,21**	**95,41**
			V	V	33,66	96,90	94,49	73,07	74,53	94,25

Табл. 4. Точности алгоритма Unet-MobileNetV2 для различных условий обучения. В ячейках указана F1-мера, её среднее значение по классам (Avg) и точность по всем пикселам в столбце Accuracy

Balance	**CCE**	**Focal**	**Dice**	**CRF**	**1**	**2**	**3**	**4**	**Avg**	**Accuracy**
		V			37,21	96,89	94,57	72,47	75,29	94,05
V		V			54,72	97,40	95,43	69,01	79,14	94,83
V	V				28,46	96,61	94,52	63,36	70,74	93,81
V		V			54,72	97,40	95,43	69,01	79,14	94,83
V			V		37,70	96,98	95,22	81,32	77,80	94,64
V	V			V	28,43	96,77	94,64	63,40	70,81	93,95
V		V		V	**57,56**	**97,51**	**95,54**	69,30	**79,98**	**94,95**
V			V	V	37,79	97,10	95,21	**81,74**	77,96	94,74

Как видно из рис. 10, использование разных функции потерь для одной и той же архитектуры для одних и тех же данных приводит к получению сетей, которые по-разному выделяют классы. В целом, лучшие результаты показывает фокальная функция потерь. Две другие функции потерь позволяют в разных случаях получать более качественные результаты, в основном за счёт того, что они лучше работают при несбалансированных данных. Так, функция потерь Дайса хорошо себя показывает для выделения зданий, которые недостаточно представлены в выборке. В целом, разные архитектуры показывают в среднем сходные результаты (рис. 11), выбор должен производиться исходя из ценности выделения определённых классов в рамках конкретной практической задачи.

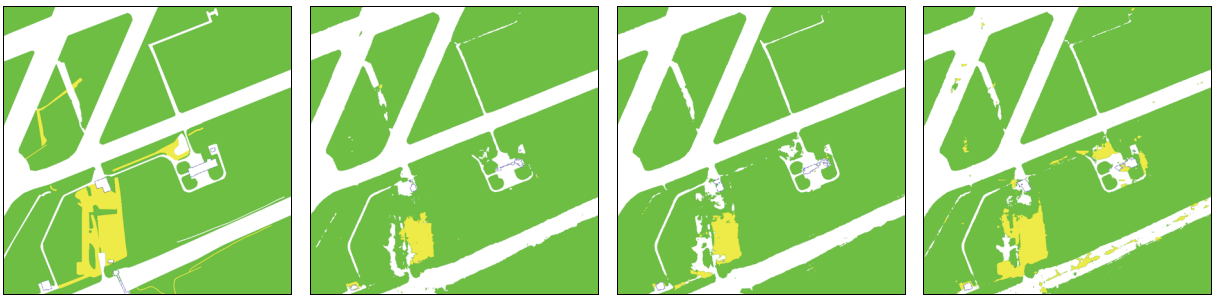

Рис. 10. Сравнение результатов работы нейронной сети с тремя разными функциями потерь, слева-направо и сверху-вниз: ручная разметка, кросс-энтропийная, фокальная функция потерь, функция потерь Дайса

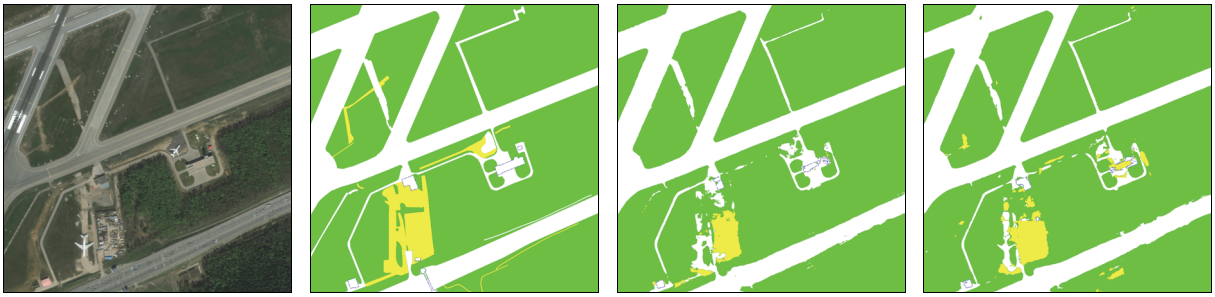

Рис. 11. Исходное изображение, разметка и результаты сегментации LinkNet-EfficientNetB0 и Unet-MobileNetV2

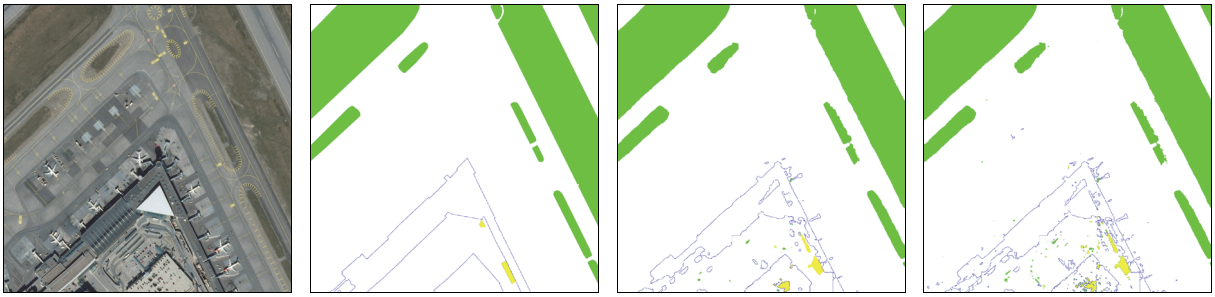

Рис. 12. Исходное изображение, разметка и результаты сегментации сетью LinkNet-EfficientNetB0 без постобработки, с постобработкой CRF

Из табл. 1 видно, что с использованием контекстной информации через модель CRF можно достичь более высокой точности, чем без неё. На рис. 12 видно, что подобная обработка визуально сглаживает результаты работы нейросети и устраняет выбросы.

В целом, доля правильно расставленных меток разнится в зависимости от класса, например, наиболее хорошо распознаются дорожное покрытие и растительность. Это объясняется тем, что для данных классов имеется более высокое количество обучающих примеров (дороги и растительность занимают на изображениях из обучающей выборки наибольшую площадь), к тому же объекты данных имеют более гладкие границы.

Заключение

В данной работе решалась задача семантической сегментации космических снимков аэропортов оптического диапазона с помощью аппарата свёрточных нейронных сетей. Было произведено исследование влияния на результат сегментации всех элементов алгоритма: предобработки данных, архитектуры сети, функции потерь, постобработки результатов. В качестве предобработки (аугментации) были применены повороты, отражения и изменения яркости. Было произведено сравнение различных архитектур энкодеров и декодеров. Наилучшим образом в данной задаче себя показала архитектура LinkNet-EfficientNetB0. Исследовано влияние на результат функции потерь: кросс-энропийной, фокальной и функции потерь Дайса.

Фокальная оказывается оптимальным выбором в условиях дисбаланса классов. Для повышения качества итоговой семантической разметки и подавления шумов была проведена постобработка с помощью модели условных случайных полей. Полученная точность составила около 95 % в среднем по всем классам. Проведён сравнительный анализ точностей различных подходов и их комбинаций.

Благодарности

Работа была поддержана Российский фондом фундаментальных исследований, грант № 17-08-00191.

Литература

1. ISPRS 2D semantic labeling contest [Electronical Resource]. – URL: http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html (request date 11.06.2019).
2. **Long, J.** Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39, Issue 4. – P. 640-651.
3. **Simonyan, K.** Very deep convolutional networks for large-scale image recognition [Electronical Resource] / K. Simonyan, A. Zisserman. – 2015. – URL: https://arxiv.org/pdf/1409.1556.pdf (request date 11.06.2019).
4. **Kaiming, H.** Deep residual learning for image recognition / H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 770-778.
5. **Badrinarayanan, V.** SegNet: A deep convolutional encoder-decoder architecture for image segmentation / V. Badrinarayanan, A. Kendall, R. Cipolla // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39, Issue 12. – P. 2481-2495.
6. **Ronneberger, O.** U-Net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical Image Computing and Computer-Assisted Intervention. – 2015. – Vol. 1, Issue 3. – P. 234-241.
7. **Chaurasia, A.** LinkNet: Exploiting encoder representations for efficient semantic segmentation / A. Chaurasia, E. Culurciello // IEEE Visual Communications and Image Processing (VCIP). – 2017. – P. 1-4.
8. **Chen, L.-Ch.** DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs / L.-Ch. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2018. – Vol. 40, Issue 4. – P. 834-848.
9. **Zhang, H.** Co-occurrent features in semantic segmentation / H. Zhang, H. Zhang, C. Wang, J. Xie // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 548-557.
10. **Hengshuang, Z.** Pyramid scene parsing network / Z. Hengshuang, S. Jianping, Q. Xiaojuan, W. Xiaogang, J. Jiaya // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 2881-2890.
11. **Lin, G.** RefineNet: Multi-path refinement networks for high-resolution semantic segmentation / G. Lin, A. Milan, Ch. Shen, I. Reid // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 5168-5177.
12. **Fu, J.** Dual attention network for scene segmentation / J. Fu, J. Liu, H. Tian, Z. Fang, H. Lu // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 3146-3154.
13. **Kaiming, H.** Mask R-CNN / H. Kaiming, G. Gkioxari, P. Dollar, R. Girshick // IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2980-2988.
14. **Goodfellow, I.** Generative adversarial nets / I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // Proceedings of the 27th International Conference on Neural Information Processing Systems. – 2014. – Vol. 2. – P. 2672-2680.
15. **Szegedy, C.** Rethinking the inception architecture for computer vision / C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna // IEEE Conference on Computer Vision and Pattern Recognition (CVPR): – 2015. – P. 2818-2826.
16. **Sandler, M.** MobileNetV2: Inverted residuals and linear bottlenecks / M. Sandler, A.G. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen // IEEE Conference on Computer Vision and Pattern Recognition (CVPR): – 2018. – P. 4510-4520.
17. **Tan, M.** EfficientNet: Rethinking model scaling for convolutional neural networks / M. Tan, Q.V. Le // International Conference on Machine Learning (ICML). – 2019. – P. 6105-6114.
18. **Carole, H.S.** Generalized Dice overlap as a deep learning loss function for highly unbalanced segmentations / H.S. Carole, L. Wenqi, T. Vercauteren, S. Ourselin, M.J. Cardoso // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. – 2017. – P. 240-248.
19. **Lin, T.** Focal loss for dense object detection / T. Lin, P. Goyal, R. Girshick, H. Kaiming, P. Dollar // IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2999-3007.
20. **Блохинов, Ю.Б.** Разработка алгоритма семантической сегментации аэрофотоснимков реального времени / Ю.Б. Блохинов, В.А. Горбачев, Ю.О. Ракутин, А.Д. Никитин // Компьютерная оптика. – 2018. – Т. 42, № 1. – С. 141-148. – DOI: 10.18287/2412-6179-2018-42-1-141-148.
21. ImageNet large scale visual recognition competition 2014 [Electronical Resource]. – URL: http://image-net.org/challenges/LSVRC/2014/ (request date 11.06.2019).

Сведения об авторах

Горбачев Вадим Александрович, 1988 года рождения, в 2011 году окончил Московский физико-технический институт (государственный университет) по направлению «Прикладные математика и физика», работает начальником сектора в ФГУП ГОСНИИАС (ГНЦ РФ). Область научных интересов: компьютерное зрение, машинное обучение, искусственный интеллект, обработка изображений.
E-mail: *vadim.gorbachev@gosniias.ru* .

Криворотов Иван Андреевич, 1997 года рождения, в 2019 году окончил бакалавриат Московского физико-технического института (государственного университета) по направлению «Прикладные математика и физика», поступил в магистратуру МФТИ, работает инженером в ФГУП ГОСНИИАС (ГНЦ РФ). Область научных интересов: компьютерное зрение, машинное обучение, искусственный интеллект, обработка изображений.
E-mail: *krivorotov.ia@phystech.edu* .

Маркелов Александр Олегович, 1997 года рождения, в 2019 году окончил бакалавриат Московского физико-технического института (государственного университета) по направлению «Прикладные математика и физика», поступил в магистратуру МФТИ, работает инженером в ФГУП ГОСНИИАС (ГНЦ РФ). Область научных интересов: компьютерное зрение, машинное обучение, искусственный интеллект, обработка изображений.
E-mail: *markelov.ao@phystech.edu* .

Котлярова Екатерина Владимировна, 1995 года рождения, в 2019 году Московский физико-технический институт (государственный университет) по направлению «Прикладные математика и физика», в 2019 году поступила в аспирантуру МФТИ, работает младшим разработчиком в области машинного обучения в Huawei Labs RUS. Область научных интересов: компьютерное зрение, машинное обучение, искусственный интеллект, обработка изображений. E-mail: *kotlyarova.ev@phystech.edu* .

ГРНТИ: 28.23.15
Поступила в редакцию 20 сентября 2019 г. Окончательный вариант – 04 декабря 2019 г.

Semantic segmentation of satellite images of airports using convolutional neural networks

V.A. Gorbachev[1], I.A. Krivorotov[1,2], A.O. Markelov[1,2], E.V. Kotlyarova[2]
[1]State Research Institute of Aviation Systems (SSC of RF), Moscow, Russia,
[2]Moscow Institute of Physics and Technology (State University), Moscow, Russia

Abstract

The paper is devoted to the development of an effective semantic segmentation algorithm for automation of airport infrastructure labelling in RGB satellite images. This task is addressed using algorithms based on deep convolutional artificial neural networks, as they have proven themselves in a wide range of tasks, including the terrestrial imagery segmentation, where they show consistently high results. A new dataset was labelled for this particular task and a comparative analysis of different architectures and backbones was carried out. A conditional random field model (CRF) was used for postprocessing and accounting of contextual information and neighborhood of objects of different classes in order to eliminate outliers. Features of the solutions applied at all preparatory stages of the algorithm were described, including data preparation, neural network training and post-processing of the training results.

Keywords: semantic segmentation, artificial neural networks, deep learning, image processing.

Citation: Gorbachev VA, Krivorotov IA, Markelov AO, Kotlyarova EV. Semantic segmentation of satellite images of airports using convolutional neural networks. Computer Optics 2020; 44(4): 636-645. DOI: 10.18287/2412-6179-CO-636.

Acknowledgements The work was supported by the Russian Foundation of Basic Research under grant No. 17-08-00191.

References

[1] ISPRS 2D semantic labeling contest. Source: ⟨http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html⟩.

[2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 2017; 39(4): 640-651.

[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Source: ⟨https://arxiv.org/pdf/1409.1556.pdf⟩.

[4] Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. IEEE Conf Comp Vis Pattern Recogn (CVPR) 2016: 770-778.

[5] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017; 39(12): 2481-2495.

[6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Med Image Comput Comput Assist Interv 2015; 1(3): 234-241.

[7] Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. IEEE VCIP 2017: 1-4.

[8] Chen L-Ch, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deep Lab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018; 40(4): 834-848.

[9] Zhang H, Zhang H, Wang C, Xie J. Co-occurrent features in semantic segmentation. IEEE CVPR 2019: 548-557.

[10] Hengshuang Z, Jianping S, Xiaojuan Q, Xiaogang W, Jiaya J. Pyramid scene parsing network. IEEE CVPR 2017: 2881-2890.

[11] Lin G, Milan A, Shen Ch, Reid I. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. IEEE CVPR 2019: 5168-5177.

[12] Fu J, Liu H, Tian H, Fang Z, Lu H. Dual attention network for scene segmentation. IEEE CVPR 2019: 3146-3154.

[13] Kaiming H, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. IEEE ICCV 2017: 2980-2988.

[14] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Proc NIPS'14 2014: 2672-2680.

[15] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. IEEE CVPR 2015: 2818-2826.

[16] Sandler M, Howard AG, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted residuals and linear bottlenecks. IEEE CVPR 2018: 4510-4520.

[17] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. ICML 2019: 6105-6114.

[18] Carole HS, Wenqi L, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support 2017: 240-248.

[19] Lin T, Goyal P, Girshick R, Kaiming H, Dollar P. Focal loss for dense object detection. IEEE ICCV 2017: 2999-3007.

[20] Blokhinov YB, Gorbachev VA, Rakutin YO, Nikitin AD. A real-time semantic segmentation algorithm for aerial imagery. Computer Optics 2018; 42(1): 141-148. DOI: 10.18287/2412-6179-2018-42-1-141-148.

[21] ImageNet large scale visual recognition competition 2014. Source: ⟨http://image-net.org/challenges/ LSVRC/2014/⟩.

Authors' information

Vadim Aleksandrovich Gorbachev (b. 1988) graduated from Moscow Institute of Physics and Technology in 2011, majoring in Applied Mathematics and Physics. Currently he works as the head of sector at the FSUE State Research Institute of Aviation Systems. Research interests are currently focused on computer vision, machine learning, artificial intelligence and image analysis. E-mail: *vadim.gorbachev@gosniias.ru* .

Ivan Andreevich Krivorotov (b. 1997) graduated undergraduate from Moscow Institute of Physics and Technology in 2019, majoring in Applied Mathematics and Physics. Currently he works as the engineer at the FSUE State Research Institute of Aviation Systems. Research interests are currently focused on computer vision, machine learning, artificial intelligence and image analysis. E-mail: *krivorotov.ia@phystech.edu* .

Aleksandr Olegovich Markelov (b. 1997) graduated undergraduate from Moscow Institute of Physics and Technology in 2019, majoring in Applied Mathematics and Physics. Currently he works as the engineer at the FSUE State Research Institute of Aviation Systems. Research interests are currently focused on computer vision, machine learning, artificial intelligence and image analysis. E-mail: *markelov.ao@phystech.edu* .

Ekaterina Vladimirovna Kotlyarova (b. 1995) graduated from Moscow Institute of Physics and Technology in 2019. She works as junior machine learning developer at Huawei Labs RUS. Her research interests are currently focused on computer vision, machine learning, artificial intelligence and image analysis.
E-mail: *kotlyarova.ev@phystech.edu* .

Received September 20, 2019. The final version – December 04, 2019.