

О. Г. Гвоздев^{1, 3}, канд. техн. наук, gvozdev@miigaik.ru,

В. А. Козуб¹, postbox-kozub@ya.ru, Н. В. Кошелева¹, antipova@phystech.edu,

А. Б. Мурынин^{1, 2}, канд. техн. наук, amuryinin@bk.ru, А. А. Рихтер¹, канд. техн. наук, urfin17@yandex.ru,

¹ Научно-исследовательский институт "АЭРОКОСМОС", Москва,

² Федеральный Исследовательский Центр "Информатика и Управление" Российской Академии Наук, Москва,

³ Московский государственный университет геодезии и картографии, Москва

Нейросетевой метод построения трехмерных моделей ригидных объектов по спутниковым изображениям¹

Разработан метод построения трехмерных моделей ригидных объектов на земной поверхности по одному спутниковому изображению на примере объектов железнодорожной инфраструктуры. Метод состоит в поэтапной обработке спутниковых изображений с последовательным применением двух сверточных нейронных сетей. На первом этапе обработки с помощью нейронной сети выполняется сегментация спутникового изображения для выделения совокупности объектов заданных классов. На втором этапе обработки с помощью нейронной сети выполняется локальный анализ областей изображения, выявленных по результатам первого этапа обработки. Результаты второго этапа обработки используются для оценки параметров трехмерной модели объекта. Возможности метода показаны на примере обработки спутникового изображения объектов железнодорожной инфраструктуры, причем рассмотрены такие информативные области объектов, как здание, тень здания, ребро стены, ребро крыши, вагон, рельсы, столбы. Показана возможность использования столбов и их теней в качестве эталонных объектов для оценки масштабирующих коэффициентов. Приведен пример применения разработанного метода выделения типичных объектов железнодорожной инфраструктуры для последующей оценки параметров трехмерной модели здания, частично заслоненного деревьями.

Ключевые слова: спутниковые изображения, трехмерная модель, растровая область, искусственная нейронная сеть, сверточная нейронная сеть, машинное обучение, объекты инфраструктуры, обучающая выборка

Введение

Современный уровень развития вычислительной техники позволяет с высоким качеством визуально воспроизводить картину местности по спутниковым изображениям на широкодоступных компьютерах. Однако одна из проблем воспроизведения заключается в построении трехмерных моделей наблюдаемых объектов.

Трехмерная цифровая модель местности включает в себя цифровую модель собственно рельефа и модели других объектов. Пространственная модель местности — ее наглядное и измеримое трехмерное изображение на электронных средствах отображения информации.

Цель настоящей работы заключается в разработке метода построения трехмерных моделей ригидных объектов на примере объектов железнодорожной инфраструктуры по двумерным спутниковым изображениям, основанного на использовании сверточных искусственных нейронных сетей (ИНС). Здесь под трехмерной моделью объекта понимается совокупность геометрических фигур, описывающих объект

земной поверхности, имеющих физические размеры, соответствующие размерам объекта, и координаты узловых точек, соответствующие координатам на поверхности объекта.

Для построения трехмерной модели объекта по спутниковому изображению должна быть решена совокупность следующих задач:

- сегментация спутникового изображения для выделения объектов;
- оценка высоты объектов по данным сегментированного спутникового изображения;
- построение объемной трехмерной модели по выделенным элементам объекта.

Существующие методы извлечения трехмерной информации из изображений

Современные подходы к извлечению трехмерной информации из изображений местности можно условно разделить на *традиционные методы*, основанные на классических подходах к обработке изображений, и *нейросетевые методы*, основанные на интеллектуальном анализе данных. Классический метод извлечения трехмерной информации из двумерных изображений основан на использовании стереопар изображений и построении по ним

¹ Исследования выполнены при поддержке Министерства науки и высшего образования РФ (уникальный идентификатор проекта RFMEFI160719X0312).

карт диспаратности. Поскольку получение стереопар космических изображений трудоемко и относительно дорого, активно развиваются методы извлечения 3D-информации из одиночных спутниковых изображений, которые предполагают как привлечение вспомогательных данных, так и дополнительную операторскую обработку данных спутниковой съемки.

В последнее время интенсивно развиваются нейросетевые методы, основанные на машинном обучении и работающие по одному изображению.

Совокупность задач высокоточной локализации объектов на изображении и их классификации с помощью искусственных нейронных сетей (ИНС) чаще всего упоминаются под названием *сегментация образца* (instance segmentation). Современные методы решения этих задач можно разделить на три класса:

- двухпроходные методы на основе якорей: Faster R-CNN [1], Mask R-CNN [2];
- однопроходные методы на основе якорей: YOLO v3 [3], SSD [4];
- однопроходные методы на основе встраивания объектов в латентное пространство: DeepWatershed [5], RSIS [6].

Методы первого и второго классов основаны на "якорях" — опорных точках на изображении, окрестность которых оценивается ИНС на предмет принадлежности объекту искомого класса. На основании совокупности таких оценок для каждого из якорей формируются, а затем отсеиваются кандидаты объектов. Данные методы ограничены плотностью расположения якорей: ее увеличение снижает производительность, уменьшение — снижает способность распознавать компактные и близкорасположенные объекты.

Третий класс методов основан на оценке положения каждого пикселя в латентном пространстве объектов сцены, после чего с помощью методов кластеризации выполняется локализация отдельных объектов. Семейство методов данного класса является наиболее перспективным для решения задачи instance segmentation на космических изображениях.

Подход к обработке каждого объекта зависит от класса этого объекта. Для типовых объектов, имеющих форму геометрического примитива, возможно непосредственное определение характеристик по спутниковому изображению, снабженному метаданными. В более сложных случаях трехмерная модель объектов может быть

построена с помощью подходов, основанных на облаках точек и их проекциях [7] и методах геометрического машинного обучения [8, 9].

Также для повышения качества обучения целесообразно повышать пространственное разрешение изображений. Основным подходом к применению ИНС для повышения разрешения космических изображений является обучение модели сопоставлению фрагментов изображения с высоким и низким разрешением, что концептуально соответствует процессам обработки визуальной информации человеческим мозгом [10–12].

При выборе способа построения трехмерной модели объекта учитываются [13–16]: уровень детальности объектов сцены (LoD0, LoD1, LoD2 и др.); характер входных данных; степень автоматизации; размер обрабатываемой области; привязка модели к конкретной местности.

Впоследствии построение трехмерных моделей объектов может быть реализовано в интерактивном режиме, при этом в процессе детектирования объектов земной поверхности используются методы машинного обучения.

В последнее время было разработано большое число методов построения трехмерных моделей местности [17–32], однако все они имеют ряд недостатков. Таким образом, остается потребность в разработке универсального автоматизированного метода с невысокими затратами на реализацию и входные данные.

Разработанный метод построения трехмерных моделей

Авторами данной работы разработан метод построения трехмерных моделей, который заключается в обработке изображения с помощью ИНС в несколько последовательных этапов:

- интегральный нейросетевой анализ для сегментации изображения;
- локальный нейросетевой анализ выделенных при сегментации областей в целях выделения характерных информативных признаков и оценки геометрической формы и размеров изучаемых объектов;
- построение трехмерной модели объекта по выделенным информативным признакам и полученным оценкам размеров.

При подготовке данных для обучения ИНС на каждом снимке проводится разметка информативных классов c_i (рис. 1, см. третью

сторону обложки): здание (полигон, c_1); крыша (полигон, c_2); ребро крыши (полилиния, c_3); ребро стены (полилиния, c_4); тень здания (полигон, c_5); железнодорожная инфраструктура (полигон, c_6); рельсы (полигон, c_7); столб (полилиния, c_8); тень столба (полилиния, c_9); вагон (полигон, c_{10}); автомобильная дорога (полигон, c_{11}); исключенные объекты (полигон, c); фон (класс c_0).

В данной работе рассматриваются типовые инвариантные объекты (стандартизированные) и типовые параметрические объекты (задаются одинаковым конечным набором параметров, имеющих разные значения).

На первом этапе обработки изображения выполняется интегральный нейросетевой анализ. Целью первого этапа является решение задачи пообъектной сегментации изображения (задача instance segmentation). В данном случае применяется ИНС, обученная для сегментации объектов из классов $c_1, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}$ на спутниковых изображениях. Результатом данного этапа является выделение областей на изображении, где каждая отдельная выделенная область соответствует одному объекту (столбу, тени, зданию и др.). Области разных классов могут пересекаться, а области одного класса не пересекаются. В процессе машинного обучения возможно использование для обучения других вариантов классификации частей объекта (выявление двумерных или трехмерных геометрических примитивов, а также углов зданий и крыш и т.д.).

Поскольку среди классов $c_i, i = 1, \dots, 11$, встречаются как линейные, так и площадные (в том числе вытянутые), то для реализации данного этапа был выбран метод пообъектной сегментации, использующий концепцию Deep Watershed Transform (метод водоразделов) [30]. Его работа состоит из двух шагов. На первом шаге обученная ИНС предсказывает энергии для метода водоразделов. Под энергией в данном случае понимается некоторая строго монотонная функция, подсчитываемая для каждого пикселя изображения и зависящая от расстояния между пикселем и границей ближайшего объекта из сег-

ментируемого класса. В простейшем случае такая функция возвращает 0 для пикселей вне объекта и расстояние до края объекта для внутренних пикселей. После нахождения энергии применяется метод водоразделов и восстанавливаются реальные контуры объектов.

Для интегрального анализа наиболее важно обеспечить пиксельную точность локализации и классификации объектов, а также аппроксимации полей энергии для применения метода водораздела [30], что успешно обеспечивают топологии ИНС семантической сегментации. В рамках данного метода предлагается применение топологии ИНС на основе архитектуры U-Net, расширенной и дополненной авторами для задач интерпретации аэрокосмических изображений [33, 34].

Ввиду необходимости решения в рамках одной модели нескольких задач для сокращения времени обучения, потребления памяти и обеспечения согласованности результатов для каждой из подзадач реализованы отдельные блоки декодеровщика и функции потерь, для каждого из которых выполняется отдельная итерация градиентного спуска.

Для блока полей энергии для водораздела выходная активационная функция — Sigmoid, функция потерь — Mean Square Error Loss (среднеквадратичное отклонение); для тематических классов задействована активационная функция SoftMax и функции потерь — Cross Entropy Loss [32]; для теней — Sigmoid и Focal Loss [32] соответственно.

Принципиальная схема ИНС приведена на рис. 2. Подробный состав блоков и их характеристики приведены в таблице.

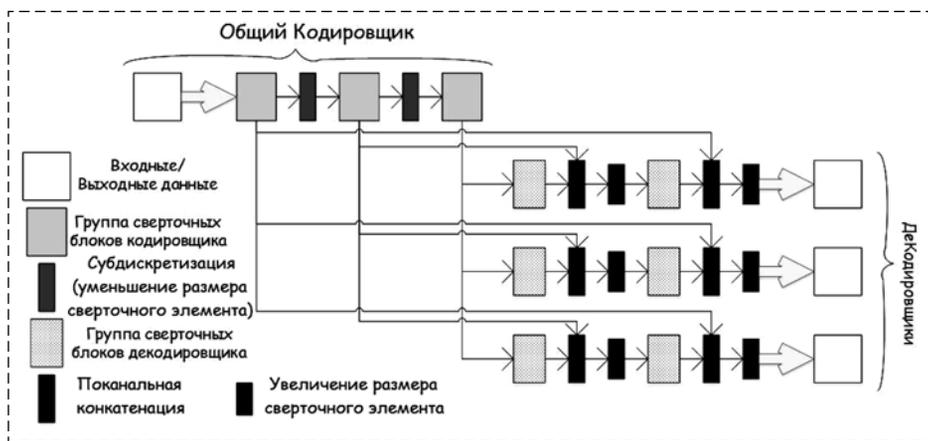


Рис. 2. Принципиальная схема топологии ИНС на базе U-Net с несколькими декодировщиками.

Fig. 2. Schematic diagram of the topology of an ANN based on U-Net with several decoders

Состав блоков нейросетевой модели интегральной оценки

The composition of the blocks of the neural network model of integral assessment

Тип блока	Пространственное окно	Выходных каналов	Число экземпляров блоков
Входная группа			
Convolve2D	7x7	32	1
ELU		32	1
Кодировщик			
ResNet Basic	3x3	32	3
MaxPool	2x2	32	1
ResNet Basic	3x3	64	3
MaxPool	2x2	64	1
ResNet Basic	3x3	128	4
MaxPool	2x2	128	1
ResNet Basic	3x3	256	5
Декодировщик			
Upsample 2x		256	1
Concatenate		384	1
ResNet Basic	3x3	128	2
Upsample 2x		128	1
Concatenate		192	1
ResNet Basic	3x3	64	2
Upsample 2x		64	1
Concatenate		96	1
ResNet Basic	3x3	32	2
Выходная группа			
Convolve2D	3x3	32	1
ELU		32	1
Convolve2D	1x1	8	1
Sigmoid		8	1

Результат интегрального анализа не позволяет сам по себе построить трехмерную модель объекта, однако достаточен для определения класса этого объекта. Для каждого выявленного объекта выполняется извлечение соответствующего фрагмента изображения и всей релевантной информации. Совокупность этих данных направляется на процедуру локального нейросетевого анализа (второй этап обработки).

Дальнейшая обработка при локальном анализе зависит от класса объекта. Наиболее значимыми являются два частных случая: распознавание объектов типовых инвариантных и параметрических классов и распознавание уникальных объектов.

Первый частный случай является тривиальным и выполняется с помощью специализированной для данного класса объектов ИНС, вы-

полняющей предсказание пространственной ориентации, а также конкретных значений дискретных и непрерывных параметров объекта. Дальнейшая интерпретация этих параметров, как правило, заключающаяся в переводе условных величин измерения и представлений дискретных признаков, возвращаемых ИНС, в физические параметры объекта, зависит от определенного ранее класса объекта.

Второй частный случай сводится к насыщению исходных данных информацией, необходимой для построения трехмерной модели, и передаче полученной совокупности данных в процедуру определения геометрической формы и размеров, описанной в следующем разделе. Насыщение данных проводится с помощью специализированной ИНС и заключается в локализации на изображении объектов его ключевых пространственных компонентов (классов c_2, c_3, c_4 как компонентов зданий).

Концептуально эта ИНС аналогична используемой для интегрального анализа. Она снабжена единственным декодером, возвращающим независимо обучаемые каналы для каждого из классов c_2, c_3, c_4 . Используется активационная функция Sigmoid и Focal Loss [32] с $\gamma = 4$ (γ отвечает за степень компенсации дисбаланса классов). Последнее особенно важно ввиду необходимости пиксельной точности локализации элементов объектов и их корректных пространственных соотношений.

Для обучения ИНС используется векторная разметка классов, подвергнутая растризации. Результатом работы ИНС являются одноканальные растры для каждого из распознаваемых классов.

Завершающим этапом обработки изображения является собственно построение трехмерной модели объекта.

Выбор классов c_7, c_8, c_9 обусловлен возможностью построения моделей объектов (зданий) по одному изображению при отсутствии метаданных. Данные классы описывают эталонные размеры на каждом изображении для оценки направляющих и масштабирующих коэффициентов [25, 26].

Расстояние между рельсами $l_3 = 1,52$ м задает эталон горизонтальных размеров. Стандартная высота опорных столбов (промежуточные опоры с нормальными габаритными размерами на двухпутном участке) $h_3 = 11,35$ м задает эталон вертикальных размеров. Тогда масштабирующие коэффициенты можно оценить по следующим формулам:

$$m_h = h_s / N_{h_s}, m_l = l_s / N_{l_s}, m_s = h_s / N_{s_s}, \quad (1)$$

где m_l и m_h — масштабирующие коэффициенты для оценки размеров объектов вдоль горизонтальных r_1, r_2 (горизонтальных размеров) и вертикальной r_h (вертикальных размеров) направляющих; m_s — масштабирующий коэффициент для оценки вертикальных размеров по теням вдоль направляющей теней r_s ; r_1 и r_2 могут быть найдены для каждого здания как линии полуосей эквивалентного эллипса области c_2 для этого здания; r_h находится как средняя линия полуосей эквивалентных эллипсов областей c_8 для данного изображения, а r_s — для областей c_9 ; N_{l_s} — средняя толщина областей c_7 на изображении в пикселях; N_{h_s} — средняя длина областей c_8 ; N_{s_s} — средняя длина областей c_9 .

Выделенные при локальном анализе характеристики и значимые области объекта проходят постобработку и используются для оценки формы и размеров объекта.

Здания описываются областями c_1, c_2, c_3, c_4, c_5 . В зависимости от полноты описания имеют место разные уровни детализации LoD геометрической модели здания. Так, при наличии c_2 может быть получена только двумерная модель (уровень LoD0). При наличии набора $\{c_2, c_4\}$ или $\{c_2, c_5\}$ вырабатывается простейшая трехмерная модель с описанием формы основания и средней высоты объекта (уровень LoD1). При наличии набора $\{c_1, c_2, c_3, c_4\}$ вырабатывается трехмерная модель с описанием формы основания и высоты на разных участках объекта (уровень LoD2).

Для отнесения областей к одному объекту (зданию) предполагается: в класс c_1 входят все крыши и стены, связанные в одну непрерывную область (в том числе надстройки и пристройки); c_2, c_3 и c_4 полностью включены в c_1 , стены здания составляют разность множеств c_1 и c_2 ($c_1 \setminus c_2$) и включают c_4 , а c_3 включается в c_2 ; область c_5 , относящаяся к объекту, лежит в непосредственной близости от c_1 .

Пример применения разработанного метода построения трехмерных моделей. На рис. 3 (см. третью сторону обложки) приведены иллюстрации этапов обработки на примере изображения I объекта Z — здания с видимой двухскатной крышей, частично заслоненной деревьями. Трехмерная модель объекта оценивается по классам $\{c_7, c_8, c_9\}$, отвечающим за эквивалентные размеры изображения I ,

а также по классам $\{c_1, c_2\}$, описывающим объект. Метаданные изображения считаются неизвестными и не используются в примере.

На первом этапе (интегральный нейросетевой анализ) выявлены области: рельсы (c_7), столб (c_8), тень от столба (c_9), а также выявлена область всего объекта c_1 . На втором этапе (локальный нейросетевой анализ) выявлена область элемента объекта c_2 , "крыша" (растр A), а также область элемента объекта "стена" — $c_1 \setminus c_2$ (растр B). На третьем этапе (оценка геометрической формы и размеров) получены оценки направляющих и масштабирующих коэффициентов.

Для изображения находится направляющая r_h большой полуоси эквивалентного эллипса области c_8 и направляющая r_s большой полуоси эквивалентного эллипса области c_9 (при наличии видимой части стен тени c_9 не применяются для оценки вертикальных размеров). Также для изображения I по формулам (1) по c_7 находится масштабирующий коэффициент $m_l \approx 0,26$, по c_8 — $m_h \approx 1,09$, по c_9 — $m_s \approx 0,49$.

Для объекта по растру A рассчитываются ортогональные направляющие r_1 и r_2 полуосей эквивалентного эллипса.

По результатам линеаризации и поляризации A и B получаются линии L_A (построена по направляющим r_1 и r_2) и L_B (построена по направляющим r_1 и r_h) соответственно.

По формуле (2) рассчитываются горизонтальные размеры l элементов крыши для линии L_A вдоль направляющих r_1 и r_2 , а также вертикальные размеры h элементов стены L_B вдоль направляющей r_h .

Размеры l дают основание здания (I), среднее значение вертикальных размеров h дает высоту здания до карниза (II), максимальное значение вертикальных размеров и его горизонтальное положение на карнизе дают высоту здания до крыши (III).

Заключение

Разработан метод определения формы и размеров объектов железнодорожной инфраструктуры по одному спутниковому изображению. Определены правила разметки классов для обучения в соответствии с особенностями изображений. Работа метода продемонстрирована на примере.

Предлагаемый подход позволяет работать с областями "неправильной" формы, с частичным заслонением областей, в условиях дефицита данных для обучения, неполноты или нечеткости разметки, с одним изображением при отсутствии метаданных.

Размер обучающей выборки ограничен в связи с тем, что последняя строится вручную. Ввиду этого целесообразна разработка методов и алгоритмов автоматической генерации обучающей выборки с применением алгебраических операций над геометрическими объектами для ускорения ее построения.

Список литературы

1. **Тусикова А. А., Вихтенко Э. М.** О распознавании автомобильных дорог на спутниковых снимках с использованием сверточных сетей mask-rcnn // V Международная конференция "Информационные технологии и высокопроизводительные вычисления" (ИТНС-2019), Хабаровск, Россия, 16—19 сентября 2019 г. Сборник трудов. Под ред. Смагин С. И., Зацаринный А. А., Россия, 2019. С. 308—314.
2. **Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun.** Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497v3 [cs.CV] 6 Jan 2016. URL: <https://arxiv.org/abs/1506.01497>; 56 p.
3. **Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick.** Mask R-CNN / Facebook AI Research (FAIR). arXiv:1703.06870v3 [cs.CV] 24 Jan 2018. URL: <https://arxiv.org/abs/1703.06870>, 10 p.
4. **Joseph Redmon, Ali Farhadi.** YOLOv3: An Incremental Improvement. University of Washington. arXiv:1804.02767v1 [cs.CV] 8 Apr 2018. URL: <https://arxiv.org/abs/1804.02767>, 6 p.
5. **Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg.** SSD: Single Shot MultiBox Detector. arXiv:1512.02325v5 [cs.CV] 29 Dec 2016, doi 10.1007/978-3-319-46448-0_2. URL: <https://arxiv.org/abs/1512.02325>.
6. **Amaia Salvador, Miriam Bellver, Manel Baradad, Xavier Giró, Jordi Torres, Ferran Marqués.** Recurrent Neural Networks for Semantic Instance Segmentation. Universitat Politècnica de Catalunya Barcelona Supercomputing Center, 2012. URL: <https://imatge-upc.github.io/rsis>.
7. **Adrian Bulat, Jing Yang, Georgios Tzimiropoulos.** To learn image super-resolution, use a GAN to learn how to do image degradation first. arXiv:1807.11458v1 [cs.CV] 30 Jul 2018. URL: <https://arxiv.org/abs/1807.11458>, 16 p.
8. **Chen-Hsuan Lin, Chen Kong, Simon Lucey.** Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction / The Robotics Institute, Carnegie Mellon University. arXiv:1706.07036v1 [cs.CV] 21 Jun 2017. URL: <https://arxiv.org/abs/1706.07036>, 10 p.
9. **Davide Boscaini, Jonathan Masci, Emanuele Rodola, Michael Bronstein.** Learning shape correspondence with anisotropic convolutional neural networks // 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. URL: <https://arxiv.org/abs/1605.06437>, 13 p.
10. **Li Yi, Hao Su, Xingwen Guo, Leonidas Guibas.** Sync-SpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation. arXiv:1612.00606v1 [cs.CV] 2 Dec 2016. URL: <https://arxiv.org/pdf/1612.00606.pdf>, 11 p.
11. **Marte Otten, Yair Pinto, Chris L. E. Paffen, Anil K. Seth, Ryota Kanai.** The Uniformity Illusion: Central Stimuli Can Determine Peripheral Perception // Psychological Science. 2017. Vol. 28(1). P. 56—68. URL: <https://journals.sagepub.com/doi/abs/10.1177/0956797616672270>.

12. **Wen Ma, Zongxu Pan, Feng Yuan, Bin Lei.** Super-Resolution of Remote Sensing Images via a Dense Residual Generative Adversarial Network // Remote Sens. 2019. Vol. 11. 2578. doi:10.3390/rs11212578. URL: <https://www.mdpi.com/2072-4292/11/21/2578/htm>, 24 p.
13. **Fatemeh Alidoost, Hossein Arefi, Federico Tombari.** 2D Image-To-3D Model: Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs) // Remote Sens. USA. 2019. Vol. 11, 2219. doi:10.3390/rs11192219. 25 p.
14. **Fatemeh Alidoost, Hossein Arefi.** A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image // PFG. Germany. 2018. Vol. 86. P. 235—248.
15. **Park J. S.** Interactive 3D reconstruction from multiple images: A primitivebased approach // Pattern recognition letters. 2005. N. 26(16). P. 2558—2571.
16. **Sherrie Wang, William Chen, Sang Michael Xie, George Azzari, David B. Lobell.** Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery // Remote Sens. USA. 2020. Vol. 12, 207. doi:10.3390/rs12020207. 25 p.
17. **Arsalan Mousavian, Dragomir Anguelov John Flynn, Jana Kořecká.** 3D bounding box estimation using deep learning and geometry. arXiv:1612.00496v2 [cs.CV] 10 Apr 2017. URL: <https://arxiv.org/pdf/1612.00496.pdf>, 10 p.
18. **Olaf Ronneberger, Philipp Fischer, Thomas Brox.** U-Net: Convolutional Networks for Biomedical Image Segmentation // Computer Vision and Pattern Recognition (cs.CV) 18 May 2015. URL: <https://arxiv.org/abs/1505.04597>.
19. **Smith, S. M., Brady J. M.** SUSAN — A New Approach to Low Level Image Processing // International Journal of Computer Vision. 1997. Vol. 23, N. 1. P. 45—78.
20. **Trevor Standley, Ozan Sener, Dawn Chen, Silvio Savarese.** Image2mass: estimating the mass of an object from its image // 1st Conference on Robot Learning (CoRL 2017). Mountain View, United States. URL: <http://proceedings.mlr.press/v78/standley17a/standley17a.pdf>, 10 p.
21. **Umut Gunes Sefercik, Serkan Karakis, Caglar Bayik, Mehmet Alkan & Naci Yastikli.** Contribution of Normalized DSM to Automatic Building Extraction from HR Mono Optical Satellite Imagery // European Journal of Remote Sensing, Turkey, 2014. Vol. 47. P. 575—591. doi: 10.5721/EuJRS20144732.
22. **Yaning Yi, Zhijie Zhang, Wanchang Zhang, Chuanrong Zhang, Weidong Li, Tian Zhao.** Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network // Remote Sens. USA. 2019. Vol. 11, 1774. doi:10.3390/rs11151774. 19 p.
23. **Zhongyuan Wang, Kui Jiang, Peng Yi, Zhen Han, Zheng He.** Ultra-dense GAN for satellite imagery super-resolution // Neurocomputing Available online 21 October 2019. URL: <https://www.sciencedirect.com/science/article/pii/S0925231219314602>, p. 328—337.
24. **Бондур В. Г.** Аэрокосмические методы и технологии мониторинга нефтегазовых территорий и объектов нефтегазового комплекса // Исследование Земли из космоса. 2010. № 6. С. 3—17.
25. **Казарян М. Л., Шахраманьян М. А., Мурынин А. Б., Рихтер А. А.** Построение 3D-моделей ригидных объектов по одному изображению методом пространственных сеток // Информатика и космос. 2018. № 4. С. 90—102.
26. **Мурынин А. Б., Рихтер А. А.** Особенности применения методов и алгоритмов реконструкции трехмерной формы ригидных объектов по данным панорамной съемки // Машинное обучение и анализ данных, Россия. 2018. Т. 4, № 4. С. 235—247. DOI:10.21469/22233792.4.4.02.
27. **Мурынин А. Б., Трекин А. Н., Игнатьев В. Ю., Кульченкова В. Г., Ракова К. О.** Метод повышения разрешения космических изображений ригидных объектов // Машинное обучение и анализ данных. 2019. Т. 4, № 5. С. 296—308.
28. **Соколова Н. О.** Распознавание контуров зданий на спутниковых изображениях высокого пространственного разрешения // Вестник ХНТУ. 2015, № 3 (54), С. 610—614.

29. **Тужилкин А. Ю.** Распознавание и реконструкция 3D-объектов по спутниковым изображениям на основе сравнения спектров графов // *Фундаментальные исследования*. 2015, № 2. С. 3727–3732.

30. **Min Bai, Raquel Urtasun.** Deep Watershed Transform for Instance Segmentation // Department of Computer Science, University of Toronto, arXiv:1611.08303v2 [cs.CV]. 4 May 2017. 9 p. URL: <https://arxiv.org/abs/1611.08303>.

31. **Кудряшов А. П., Соловьев И. В.** Реконструкция текстурированной модели городского пространства по топографическому плану и фотоснимкам // Программные продукты и системы // *Software & Systems*. 2019. Vol. 3 (32). DOI: 10.15827/0236-235X.127.496-503.

32. **Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar.** Focal Loss for Dense Object Detection. arXiv:1708.02002v2 [cs.CV] 7 Feb 2018. URL: <https://arxiv.org/pdf/1708.02002.pdf>.

33. **Рихтер А. А., Мурынин А. Б., Шлюпиков В. А.** Особенности комплексирования данных дистанционного зондирования для восстановления трехмерных моделей ригидных объектов по спутниковым снимкам // Матер. Междунар. науч.-практ. Конф. "Достижения науки в 2019 году", 27 декабря 2019 г. Центр научного развития "Большая книга".

34. **Гвоздев О. Г., Мурынин А. Б., Рихтер А. А.** Комплекс прикладных решений по построению и обучению искусственных нейронных сетей для семантической сегментации аэрокосмических изображений произвольной канальности спектральной структуры в условиях дефицита обучающих данных // Матер. 19-й Всеросс. конф. С междунар. участием: Математические методы распознавания образов (ММРО-2019, г. Москва). Российская академия наук. 2019. С. 344–348.

35. **Pytorch.** URL: <https://pytorch.org>.

Neural Network Method for Constructing Three-Dimensional Models of Rigid Objects from Satellite Images

O. G. Gvozdev^{1,3}, gvozdev@miigaik.ru, **V. A. Kozub**¹, postbox-kozub@ya.ru,
N. V. Kosheleva¹, antipova@phystech.edu, **A. B. Murynin**^{1,2}, PhD, amurynin@bk.ru,
A. A. Richter¹, PhD, urfin17@yandex.ru,

¹ Institute for Scientific Research of Aerospace Monitoring AEROCOSMOS, Moscow, 105064, Russian Federation,

² Federal Research Center "Computer Science and Control" of RAS, Moscow, 119333, Russian Federation,

³ State University of Geodesy and Cartography, Moscow, 105064, Russian Federation

Corresponding author: Murynin A. B., Ph.D., Leading Researcher, Institute for Scientific Research of Aerospace Monitoring AEROCOSMOS, Moscow, 105064, Russian Federation, e-mail: amurynin@bk.ru

Accepted on September 21, 2020

Abstract

A method has been developed for constructing three-dimensional models of rigid objects on the earth's surface using one satellite image using the example of railway infrastructure. The method consists in step-by-step processing of satellite images with sequential application of two convolutional neural networks. In the first processing step, a satellite image is segmented by a neural network to select a plurality of objects of predetermined classes. At the second stage of processing with the help of neural network local analysis of image areas detected by results of the first stage of processing is performed. The results of the second processing step are used to estimate the parameters of the 3D model of the object. The possibilities of the method are shown by the example of processing a satellite image of the railway infrastructure. The following classes of informative areas are considered: building, wall edge, roof edge, building shadow, railway infrastructure, car, highway; rails, poles and shadows from poles (taken as reference objects for estimating scaling coefficients in certain directions). An example is given of using the developed method of highlighting typical railway infrastructure objects and for subsequent evaluation of the parameters of a three-dimensional building model partially obscured by trees.

Keywords: satellite images, 3D model, raster area, artificial neural network, convolutional network, machine learning, infrastructure, training sample

Acknowledgements: The research was carried out with the support of the Ministry of Science and Higher Education of the Russian Federation (unique project identifier RFMEFI60719X0312).

For citation:

Gvozdev O. G., Kozub V. A., Kosheleva N. V., Murynin A. B., Richter A. A. Neural Network Method for Constructing Three-Dimensional Models of Rigid Objects from Satellite Images, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2021, vol. 22, no. 1, pp. 48–55.

DOI: 10.17587/mau.22.48-55

References

1. **Tusikova E. M., Vikhtenko A. A.** On road recognition on satellite imagery using convolutional mask-rcnn networks,

V International Conference "Information Technologies and High Performance Computing" (ITHPC-2019), Khabarovsk, Russia, September 16–19, 2019, Proceedings, 2019, pp. 308–314 (in Russian).

2. **Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun.** Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv:1506.01497v3 [cs.CV] 6 Jan 2016, 56 p., available at: <https://arxiv.org/abs/1506.01497>.

3. **Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick.** Mask R-CNN, Facebook AI Research (FAIR), arXiv:1703.06870v3 [cs.CV] 24 Jan 2018, 10 p., available at: <https://arxiv.org/abs/1703.06870>.

4. **Joseph Redmon, Ali Farhadi.** YOLOv3: An Incremental Improvement / University of Washington, arXiv:1804.02767v1 [cs.CV] 8 Apr 2018, 6 p., available at: <https://arxiv.org/abs/1804.02767>.

5. **Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg.** SSD:

Single Shot MultiBox Detector, arXiv:1512.02325v5 [cs.CV] 29 Dec 2016, doi 10.1007/978-3-319-46448-0_2, available at: <https://arxiv.org/abs/1512.02325>.

6. **Amaia Salvador, Míriam Bellver, Manel Baradad, Xavier Giró, Jordi Torres, Ferran Marqués.** Recurrent Neural Networks for Semantic Instance Segmentation, Universitat Politècnica de Catalunya Barcelona Supercomputing Center, 2012, available at: <https://imatge-upc.github.io/rsis>.

7. **Adrian Bulat, Jing Yang, Georgios Tzimiropoulos.** To learn image super-resolution, use a GAN to learn how to do image degradation first, arXiv:1807.11458v1 [cs.CV], 30 Jul 2018, 16 p. available at: <https://arxiv.org/abs/1807.11458>.

8. **Chen-Hsuan Lin, Chen Kong, Simon Lucey.** Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction, The Robotics Institute, Carnegie Mellon University, arXiv:1706.07036v1 [cs.CV], 21 Jun 2017, 10 p., available at: <https://arxiv.org/abs/1706.07036>.

9. **Davide Boscaini, Jonathan Masci, Emanuele Rodola, Michael Bronstein.** Learning shape correspondence with anisotropic convolutional neural networks, Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 13 p., available at: <https://arxiv.org/abs/1605.06437>.

10. **Li Yi, Hao Su, Xingwen Guo, Leonidas Guibas.** Sync-SpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation, arXiv:1612.00606v1 [cs.CV] 2 Dec 2016, 11 p., available at: <https://arxiv.org/pdf/1612.00606.pdf>.

11. **Marte Otten, Yair Pinto, Chris L. E. Paffen, Anil K. Seth, Ryota Kanai.** The Uniformity Illusion: Central Stimuli Can Determine Peripheral Perception, *Psychological Science* 2017, Vol. 28(1) 56–68, available at: <https://journals.sagepub.com/doi/abs/10.1177/0956797616672270>.

12. **Wen Ma, Zongxu Pan, Feng Yuan, Bin Lei.** Super-Resolution of Remote Sensing Images via a Dense Residual Generative Adversarial Network / *Remote Sens.* 2019, 11, 2578; doi:10.3390/rs11212578, 24 p. available at: <https://www.mdpi.com/2072-4292/11/21/2578/htm>.

13. **Fatemeh Alidoost, Hossein Arefi, Federico Tombari.** 2D Image-To-3D Model: Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs), *Remote Sens.*, USA, 2019, vol.11, no. 2219, 25 p. doi:10.3390/rs11192219.

14. **Fatemeh Alidoost, Hossein Arefi.** A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image, PFG, Germany, 2018, 86, pp. 235–248.

15. **Park J. S.** Interactive 3D reconstruction from multiple images: A primitivebased approach, *Pattern Recognition Letters*, 2005, no. 26(16), pp. 2558–2571.

16. **Sherrie Wang, William Chen, Sang Michael Xie, George Azzari, David B. Lobell.** Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery, *Remote Sens.*, USA, 2020, vol.12, no. 207, 25 p., doi:10.3390/rs12020207.

17. **Arsalan Mousavian, Dragomir Anguelov John Flynn, Jana Košecká.** 3D bounding box estimation using deep learning and geometry, arXiv:1612.00496v2 [cs.CV] 10 Apr 2017, 10 p. available at: <https://arxiv.org/pdf/1612.00496.pdf>.

18. **Olaf Ronneberger, Philipp Fischer, Thomas Brox.** U-Net: Convolutional Networks for Biomedical Image Segmentation, *Computer Vision and Pattern Recognition (cs.CV)* 18 May 2015, available at: <https://arxiv.org/abs/1505.04597>.

19. **Smith, S. M., Brady J. M.** SUSAN — A New Approach to Low Level Image Processing, *International Journal of Computer Vision*, 1997, vol. 23, no. 1, pp. 45–78.

20. **Trevor Standley, Ozan Sener, Dawn Chen, Silvio Savarese.** Image2mass: estimating the mass of an object from its image, *1st Conference on Robot Learning (CoRL 2017)*, Mountain View, United States, 10 p. available at: <http://proceedings.mlr.press/v78/standley17a/standley17a.pdf>.

21. **Umut Gunes Sefercik, Serkan Karakis, Caglar Bayik, Mehmet Alkan, Naci Yastikli.** Contribution of Normalized DSM to Automatic Building Extraction from HR Mono Optical Satellite Imagery, *European Journal of Remote Sensing*, Turkey, 2014, vol. 47, pp. 575–591. doi: 10.5721/EuJRS20144732.

22. **Yaning Yi, Zhijie Zhang, Wanchang Zhang, Chuanrong Zhang, Weidong Li, Tian Zhao.** Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network, *Remote Sens.*, USA, 2019, vol. 11, no. 1774, 19 p., doi:10.3390/rs11151774.

23. **Zhongyuan Wang, Kui Jiang, Peng Yi, Zhen Han, Zheng He.** Ultra-dense GAN for satellite imagery super-resolution, *Neuro-computing*, 21 October 2019, available at: <https://www.sciencedirect.com/science/article/pii/S0925231219314602>, p. 328–337.

24. **Bondur V. G.** Aerospace methods and technologies for monitoring oil and gas territories and oil and gas complex facilities, *Exploration of the Earth from space*, 2010, no. 6, pp. 3–17 (in Russian).

25. **Kazaryan M. L., Shakhramanyan M. A., Murynin A. B., Richter A. A.** Construction of 3D models of rigid objects according to one image by the method of spatial grids, *Information and Space*, 2018, no. 4, pp. 90–102 (in Russian).

26. **Murynin A. B., Richter A. A.** Features of application of methods and algorithms of reconstruction of three-dimensional shape of rigid objects according to panoramic survey data, *Machine Learning and Data Analysis*, Russia, 2018, vol. 4, no. 4, pp. 235–247, doi: 10.21469 / 22233792.4.4.02 (in Russian).

27. **Murynin A. B., Trekin A. N., Ignat'ev V. Yu., Kul'chenkova V. G., Rakova K. O.** Method for increasing the resolution of space images of rigid objects, *Machine Learning and Data Analysis*, 2019, vol. 4, no. 5, pp. 296–308 (in Russian).

28. **Sokolova N. O.** Recognition of building contours on satellite images of high spatial resolution, *Vestnik Kherson National Technical University*, 2015, no. 3 (54), pp. 610–614 (in Russian).

29. **Tuzhilkin A. Yu.** Recognition and reconstruction of 3D objects from satellite images based on comparison of graph spectra, *Fundamental research*, 2015, no. 2, pp. 3727–3732 (in Russian).

30. **Min Bai, Raquel Urtasun.** Deep Watershed Transform for Instance Segmentation, Department of Computer Science, University of Toronto, arXiv:1611.08303v2 [cs.CV] 4 May 2017, 9 p., available at: <https://arxiv.org/abs/1611.08303>.

31. **Kudryashov A. P., Soloviev I. V.** Reconstruction of the textured model of urban space according to the topographic plan and photographs, *Software products and systems*, Software & Systems, Russia, 2019, no. 3 (32), doi: 10.15827/0236-235X.127.496-503 (in Russian).

32. **Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar.** Focal Loss for Dense Object Detection, arXiv:1708.02002v2 [cs.CV] 7 Feb 2018, available at: <https://arxiv.org/pdf/1708.02002.pdf>.

33. **Richter A. A., Murynin A. B., Shlupikov V. A.** Features of remote sensing data integration for the restoration of 3D models of rigid objects from satellite images, *Materials of the International Scientific and Practical Conference "Achievements of Science in 2019"*, December 27, 2019, Center for Scientific Development "Big Book" (Moscow) (in Russian).

34. **Gvozdev O. G., Murynin A. B., Richter A. A.** Complex of applied solutions for the construction and training of artificial neural networks for semantic segmentation of aerospace images of an arbitrary channel-spectral structure in conditions of lack of training data, *Materials of the 19th All-Russia conf. with international participation: Mathematical methods of pattern recognition (MMPO-2019, Moscow)*, Russian Academy of Sciences, 2019, pp. 344–348 (in Russian).

35. **Pytorch**, available at: <https://pytorch.org>.