**Глубокое обучение в задаче обнаружения объектов на спутниковых снимках высокого разрешения**

## Владимир Хрящев

*ЯрГУ им. П.Г. Демидова,Россия*

[*v.khryashchev@uniyar.ac.ru*](mailto:v.khryashchev@uniyar.ac.ru)

Владимир Павлов

*ЯрГУ им. П.Г Демидова, Россия*

[*vladimir@1pavlov.com*](mailto:vladimir@1pavlov.com)

Андрей Приоров

*ЯрГУ им. П.Г. Демидова,Россия* [*andcat@yandex.ru*](mailto:andcat@yandex.ru)

Анна Островская

*РУДН, Россия*

[*ostrovskaya\_aa@rudn.university,*](mailto:ostrovskaya_aa@rudn.university)

**Аннотация**

*Цель нашего исследования состоит в разработке сверточной нейронной сети для автоматического обнаружения объектов, таких как вода, лес и жилые постройки, на спутниковых снимках высокого разрешения. В данной работе показано, как глубокие нейронные сети, реализованные на современных графических процессорах, могут быть эффективно использованы для извлечения необходимых признаков на картинках. Для обучения модели на суперкомпьютере NVIDIA DGX-1, была использована размеченная база данных UrbanAtlas, которая содержит изображения 21 класса. Изображения, полученные со спутников Landsat-8, используются также для оценки качества автоматического детектирования объектов. Задача обнаружения объектов на спутниковых снимках находит свое применение в области градостроительства, для моделирования климата и т.д.*

# Введение

Сверточная нейронная сеть, используемая в этом исследовании, основана на сочетании алгоритмов глубокого обучения и передовой технологии GPU [1]. Основным преимуществом такого подхода является возможность детектирования и классификации объектов в режиме реального времени, будучи в то же время менее дорогостоящим и превосходящим по производительности в сравнении с другими методами машинного обучения [2-7]. И хотя алгоритмы сверточных нейронных сетей известны уже много десятилетий, недавние достижения в области создания графических процессоров сделали возможной одновременную обработку данных [8, 9].

Лежащая в основе математическая структура нейронных сетей параллельна и идеально подходит для архитектуры графического процессора (GPU), состоящего из тысяч ядер, предназначенных для одновременного выполнения нескольких задач [10-14]. Такая архитектура вычислений используется для значительного сокращения времени обучения алгоритма и повышения точности обнаружения и классификации объектов.

Как и для других задач, решаемых с помощью методов глубокого обучения, важно наличие большой базы аннотированных данных. Существует несколько общедоступных наборов размеченных спутниковых изображений.

**DeepSat.** База снимков DeepSat [15] была опубликована в 2015 году. Она содержит два набора аннотированных изображений: 500000 картинок со спутника Sat-4, разделенных на 4 класса («земля», «деревья», «поля» и класс который состоит объектов, отличных от указанных трех) и 405000 картинок со спутника Sat-6, разделенных на более чем 6 классов («земля», «деревья», «поля», «дороги», «здания» и «водные объекты»). Все образцы имеют размер 28×28 пикселей с пространственным разрешением 1 м/пиксель и содержат 4 цветовых канала (красный, зеленый, синий и инфракрасное излучение). В работе [16] было показано, что на основе такого рода изображений можно создать классификатор на основе сверточных сетей с точностью классификации около 99%. Хотя этот набор данных полезен для предварительной подготовки моделей, решающих более сложные задачи (например, сегментации изображений), он не позволяет предпринять дальнейшие шаги для детального анализа детектированных объектов.

**UCMerced.** Этот набор размеченных снимков был опубликован в 2010 году [17]. Он содержит 2100 изображений 21 класса размером 256 × 256 пикселей с разрешением 1 м/пиксель. Несмотря на малую выборку данных, нейронная сеть, обученный алгоритм из статьи [18], достигает более 95% точности.

978-1-5386-5710-2/18/$31.00 ©2018 IEEE

**UrbanAtlas** предоставляет карты с высоким разрешением (2,5 м/пиксель) 305 крупных городских зон и их окрестностей (более 100000 жителей) стран ЕС за 2006 г. и 695 городских районов и их окрестностей стран ЕАСТ за 2012 г.

Наша работа посвящена анализу работы сверточных нейронных сетей и оценке качества автоматического обнаружения объектов на спутниковых снимках. Для этого была использована открытая для общего пользования, размеченная база изображений UrbanAtlas. UrbanAtlas содержит снимки 21 класса со спутника Landsat-8 [20].

Примеры изображений из этой базы показаны на рис. 1. Изображения со спутника Landsat 8 имеют разрешение 30 м/пиксель - это самое высокое разрешение изображений такого рода.

Во второй части этой статьи описывается принцип оптимизации сверточной нейронной сети U-Net для анализа спутниковых изображений. Третья часть содержит результаты численного эксперимента по обнаружению объектов 3 классов («вода», «лес» и «сельскохозяйственные угодья»). В четвертой части строятся выводы по проделанной работе. Пятая часть содержит слова благодарности.

а) б)

в) г)

### Рис. 1. Примеры изображений со спутника Landsat-8: а) картинка с объектами класса «лес», б) картинка с объектами классов «лес» и «вода», в) картинка с объектами класса «сельскохозяйственные угодья», г) картинка с объектами классов «лес», «вода» и «сельскохозяйственные угодья»

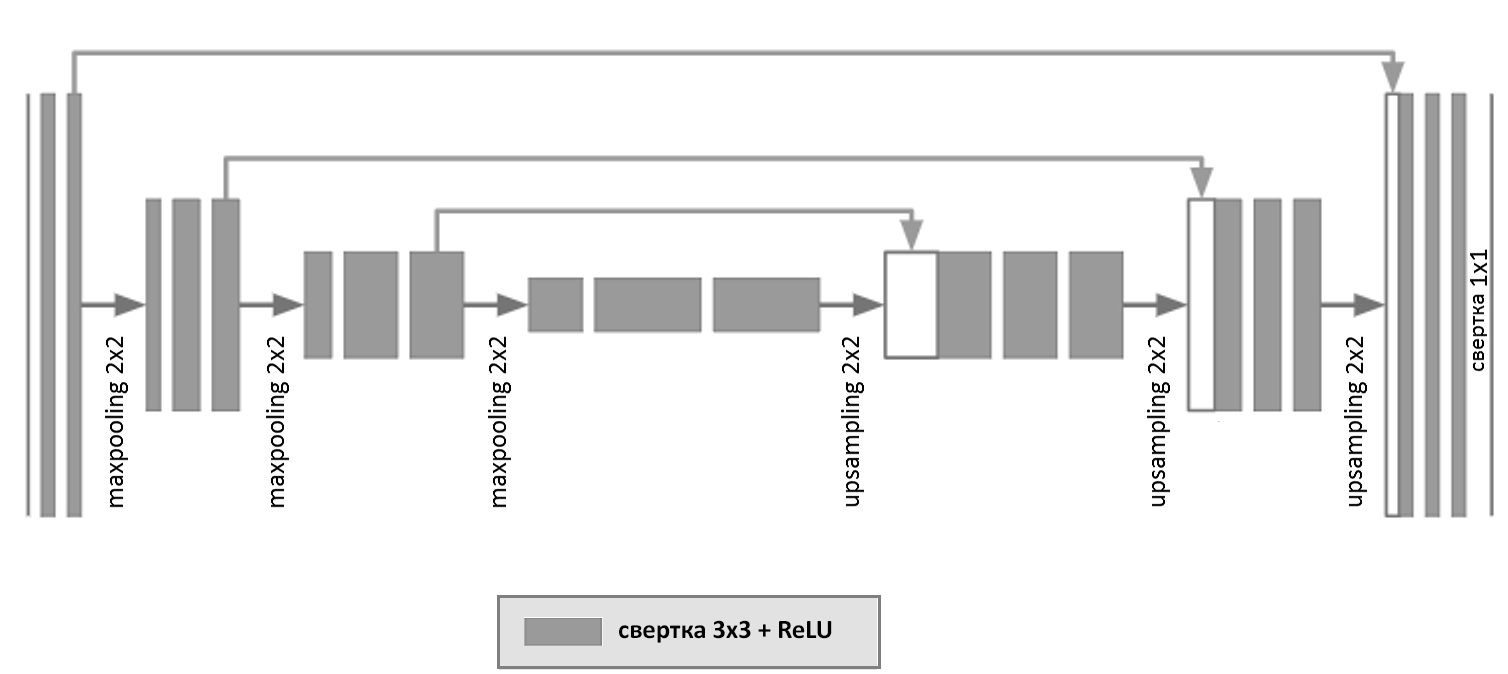


Рис. 2. Архитектура сети U-Net

1. **Обучение, тестирование и оптимизация сверточной нейронной сети**

Схема исследования работы содержит 3 основных этапа: обучение, тестирование и оптимизация. Выполнение этих шагов приводит к построению окончательного алгоритма для обнаружения объектов на изображениях со спутника Landsat-8.

В ходе исследования была использована архитектура сверточной нейронной сети U-Net, представленной на рис. 2. Сеть состоит из двух частей: кодировщика и декодировщика (слева и справа соответственно), представленные 23 сверточными слоями [21-22].

Исследование алгоритма проводилось на суперкомпьютере NVIDIA DGX-1 в Центре искусственного интеллекта Ярославского университета им. П.Г. Демидова. Общее время обучения сети составило около 5 ч. Увеличение количества классов обнаружения будет линейно увеличивать время обучения на том же наборе данных, без увеличения количества эпох обученеия.

Функция ошибки, которая минимизируется во время обучения алгоритма необходима для оценки этапов обучения и тестирования нейронной сети. Для задач многоклассовой классификации обычно используется функция кросс-энтропии, значение которой вычисляется по формуле

(1)

где - истинная вероятность принадлежности -ому классу, – предсказанное алгоритмом значение вероятности принадлежности -ому классу, а – общее количество классов (в нашем случае их три: «лес», «вода» и «сельскохозяйственные угодья»).

Для поставленной задачи классы не являются взаимоисключающими, поэтому использовать формулу (1) представляется не совсем корректным,

поэтому в нашем случае, в ходе численного эксперимента, использовалась функция, представляющая собой соединение бинарной кросс-энтропии и индекса Джаккарда:

, (2)

где – функция бинарной кросс-энтропии, рассчитываемая по формуле

,

а - индекс Джаккарда, рассчитываемый, как

Функция (2) позволяет меру сходства между классами (тремя, в нашем случае).

# 3. Численные результаты

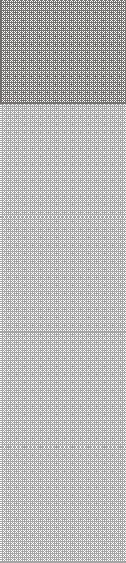
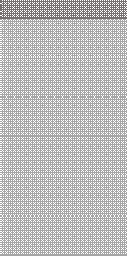
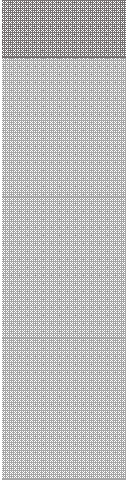
Для оценки качества работы алгоритма было произведено сравнение результатов алгоритма обнаружения объектов с экспертной маркировкой на тестовом наборе данных. Тестовая выборка содержала 100 спутниковых снимков. Каждая такая фотография была отмечена 10 экспертами, чтобы исключить субъективный фактор. Итоговая разметка представляла собой «средний» контур экспертной маркировки. Точность работы алгоритма детектирования определялась исходя из процента пересечения области, выделенной детектором и экспертной маркировки

.

Результаты эксперимента показаны на рис. 3. Каждая диаграмма состоит из двух частей: светлая часть отражает точность детектирования алгоритма, обучаемого с помощью функции кросс-энтропии, а темная – увеличение точности обнаружения при запуске модели, обученной с помощью преобразованной функции ошибки.

i

Рис. 3. Диаграмма точности обнаружения трехклассов: «лес», «вода» и «сельскохозяйственные угодья»



100.00%

95.00%

90.00%

85.00%

80.00%

75.00%

70.00%

Forest

Water

Agriculture

Как видно из результатов численного эксперимента, изменение функции потерь дало значительное улучшение: точность детектирования каждого класса повысилась не менее чем на 2%. Лучше всего алгоритму удается обнаруживать объекты класса «сельскохозяйственные угодья». Такой высокий процент точности связан с четкостью границ и видимым визуальным разделением окружающих объектов. При детектировании объектов класса «вода», выделяется много разделенных объектов. Это связано с наличием льда и других инородных объектов на поверхности рек или озер. Лес имеет среднюю величину обнаружения 92,3%. Ошибки в детектировании здесь объясняются нечетким распределением границ леса.

# Заключение

Разработанный алгоритм обнаружения объектов на спутниковых снимках основан на реализации относительно нового подхода - глубокого обучения и сверточной нейронной сети. В статье было показано, как глубокие нейронные сети, реализованные на современных графических процессорах, могут быть эффективно использованы для извлечения необходимых признаков на картинках. Разработанный алгоритм может применяться для сегментации спутниковых снимков: выделения территорий городов, контроля строительства, моделирования климата и проч.

# Благодарности

Статья была подготовлена при финансовой поддержке Министерства образования Российской Федерации в рамках научного проекта № 14.575.21.0167 (идентификатор RFMEFI57517X0167).

Авторы также благодарны Центру Искусственного Интеллекта Ярославского государственного университета им. П.Г. Демидова за обеспечение доступа к суперкомпьютеру NVIDIA DGX-1.

# Список литературы

1. T. Qu, Q. Zhang, S. Sun, “Vehicle detection from high- resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks”, Multimedia Tools and Applications, Volume 76, Issue 20, pp. 21651–21663, October 2017.
2. E. P. Baltsavias, “Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems”, ISPRS Journal of Photogrammetry and Remote Sensing, 58(3-4): pp.129-151, January, 2004.
3. S. Kluckner, H. Bischof, “Semantic classification by covariance descriptors within a randomized forest”, In Computer Vision Workshops (ICCV), pp. 665-672. IEEE, 2009.
4. S. Kluckner, T. Mauthner, P. M. Roth, H. Bischof, “Semantic classification in aerial imagery by integrating appearance and height information”, In ACCV, volume 5995 of Lecture Notes in Computer Science, pp. 477-488. Springer, 2009.
5. V. Mnih, G. Hinton, “Learning to detect roads in high- resolution aerial images”, In Proceedings of the 11th European Conference on Computer Vision (ECCV), September 2010.
6. V. Mnih, G. Hinton, “Learning to label aerial images from noisy data”, In A. McCallum and S. Roweis, editors, Proceedings of the 29th Annual International Conference on Machine Learning (ICML 2012), June 2012.
7. A. Krizhevsky, I. Sutskever, G. Hinton, “Imagenet classification with deep convolutional neural networks”, In Advances in neural information processing systems, pp. 1097–1105, 2012.
8. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, “Overfeat: integrated recognition, localization and detection using convolutional networks”, International Conference on Learning Representations, 2013.
9. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang,

E. Tzeng, T. Darrell, “DeCAF: a deep convolutional activation feature for generic visual recognition”, ICML’14, Proceedings of the 31st International Conference on Machine Learning, volume 32, pp. I-647-I-655, Beijing, China 2014.

1. H. Mayer, “Object extraction in photogrammetric computer vision”, ISPRS Journal of Photogrammetry and Remote Sensing, 63(2):213-222, March 2008.
2. O. A. B. Penai, K. Nogueira, J. A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?”, IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44–51, 2015.
3. K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, Web: https://arxiv.org/abs/1409.1556, 2014.
4. M. Zhai, Z. Bessinger, S. Workman, N. Jacobs, “Predicting ground-level scene layout from aerial imager”, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, pp. 867-875, 2017.
5. N. Jean, M. Burke, M. Xie, W. Davis, D. Lobell, S. Ermon, “Combining satellite imagery and machine learning to predict poverty”, Science 353, 6301, pp. 790–794, 2016.
6. S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano,

M. Karki, R. Nemani, “DeepSat - A learning framework for Satellite Imagery”, Proceedings of SIGSPATIAL’15, Bellevue, WA, USA, 2015.

1. M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko,

K. Karantzalos, “Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data”, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 83– 88, June 2016.

1. K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition”, In 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778, 2016.
2. M. Castelluccio, G. Poggi, C. Sansone, L. Verdoliva, “Land use classification in remote sensing images by convolutional neural networks”, Web: https://arxiv.org/abs/1508.00092, 2015.
3. “European Union. 2011. Urban Atlas”, Web: https:/[/www.eea.europa.eu/data-and-maps/data/urban-atlas.](http://www.eea.europa.eu/data-and-maps/data/urban-atlas)
4. “Landsat8”, Web: https://landsat.usgs.gov/landsat-8.
5. O. Ronneberger, P. Fischer, T. Brox “U-Net: convolutional networks for biomedical image segmentation”, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol. 9351: pp. 234-241, 2015.
6. V. Khryashchev, V. Pavlov, A. Priorov, E. Kazina, “Convolutional Neural Network for Satellite Imagery”, Proceedings of the 22th Conference of Open Innovations Association FRUCT'22. Jyvaskyla, Finland, pp. 344-347, 2018.