Оптимизация сверточной нейронной сети для распознавания объектов на спутниковых снимках

Владимир Хрящев, Владимир Павлов,

ЯрГУ им. П.Г Демидова, ул. Советская 14, Ярославль, Россия

[v.khryashchev@uniyar.ac.ru,](mailto:v.khryashchev@uniyar.ac.ru) [i@yajon.ru](mailto:i@yajon.ru)

***Аннотация***— **в данной статье представлены результаты исследования работы сверточной нейронной сети для обнаружения объектов на изображениях со спутника Landsat-8. В качестве алгоритма распознавания использовалась архитектура U-Net. Нейронная сеть была обучена на размеченной базе данных UrbanAtlas. Она содержит изображения 21 класса, но в рамках проведенного исследования учитывались только 3 класса: «лес», «сельскохозяйственные угодья» и «вода». Изображения, полученные со спутника Landsat-8, использовались также и для оценки качества детектирования объектов. Для анализа точности алгоритма обнаружения объектов, детектированные регионы были сопоставлены с областями, ранее размеченными экспертами. Помимо этого, в ходе исследования была проанализирована работа двух модифицированных детекторов.**

***Ключевые слова — сверточные нейронные сети, обнаружение объектов, спутниковые снимки, глубокое машинное обучение.***

1. ВВЕДЕНИЕ

В приложениях машинного обучения задача обнаружения объектов на спутниковых снимках обычно формулируется как задача маркировки пикселей. Учитывая такую постановку, для изображения необходимо произвести полную его сегментацию на такие классы, как «здания», «дороги», «деревья», «поля» или «вода» [1-5].

Сверточные нейронные сети стали повсеместными в области компьютерного зрения с тех пор, как архитектура глубокой сети AlexNet [6] выиграла в конкурсе ImageNet Challenge: ILSVRC 2012 [7]. Пример реализации сверточной нейронной сети для сегментации спутниковых снимков (для обнаружения оврагов, лесов и т.д.) можно найти в статье [8]. Модель была обучена на открытом наборе данных UC Merced [9] (2100 изображений из 21 класса). Решение задачи детектирования полей представлено в статье [10]. Авторы этого исследования адаптировали архитектуру нейронной сети VGG16 [11], предварительно обученной на наборе данных ImageNet для распознавания изображений.

Обнаружение объектов на снимках является распространенной задачей компьютерного зрения, суть которой заключается в определении наличия или отсутствия конкретных признаков на изображениях. После обнаружения этих признаков объект может быть дополнительно классифицирован как принадлежащий к одному из заранее определенных классов. Такого рода задача называется классификацией объектов. Детектирование и классификация объектов являются фундаментальными задачами области искусственного интеллекта. Основная проблема в применении машинного обучения в аэрофотосъемке заключается в том, что сегментацию снимков довольно трудно выполнить в режиме реального времени в связи со сложностями поставленной задачи и ее вычислительными затратами.

Одним из предлагаемых решений является использование глубокого машинного обучения, использующего алгоритмы сверточной нейронной сети для отслеживания, обнаружения и классификации объектов из необработанных данных в режиме реального времени.

Анна Островская, Александр Семенов

РУДН, ул. Миклухо-Маклая 6, Москва, Россия

[ostrovskaya\_aa@rudn.university,](mailto:ostrovskaya_aa@rudn.university) [semenov.venture@mail.ru](mailto:semenov.venture@mail.ru)

В последние несколько лет глубокие нейронные сети показали выдающиеся результаты в задачах обнаружения и классификации объектов на изображениях из-за их относительно высокой точности и скорости работы.

Данные дистанционного зондирования и методы глубокого обучения используются для решения задач, связанных с обнаружением объектов. Например, в статье [12] геолокация снимка определяется визуальной корреляцией содержимого картинки и спутникового изображения с использованием сверточных нейронных сетей. Другие приложения в этой области касаются прогнозирования уровня бедности в некоторых африканских странах. Так, для решения этой задачи, в статье [13] используются ночные снимки с целью выявления связи между уровнем городской освещенности и экономической активности региона.

База данных European Urban Atlas [14] 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.

Данная работа посвящена анализу сверточных нейронных сетей для обнаружения объектов на спутниковых снимках. Для изучения работы алгоритмов глубокого машинного обучения была использована размеченная база данных UrbanAtlas. Она содержит изображения 21 класса, но в рамках проведенного исследования использовались только 3 класса: «лес», «сельскохозяйственные угодья» и «вода». Изображения, полученные со спутника Landsat-8 [15], использовались также и для оценки качества автоматического обнаружения объектов.

2. ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ

Для обучения нейронной сети, решающей задачу обнаружения объектов на спутниковых снимках, использовались 500 уникальных изображений. Каждая фотография со спутника Landsat-8, снятая не раньше лета 2017 г., включает в себя территорию площадью в 1 км². Подготовленные изображения были разделены на 2 выборки, обучающую и тестовую, в отношении 70/30.

Перед обучением сети U-Net для подготовки набора данных каждому пикселю изображения присваивался определенный класс. Из имеющегося набора классов, были выбраны 3 наиболее сбалансированные. Процентное соотношение каждого из 3 классов на снимках из обучающего множества представлено на рис. 1.

Для оценки качества работы алгоритма, был использован индекс Джаккарда, который определяет степень сходства сравниваемых объектов и вычисляется по формуле (1):

(1)

Из формулы (1) следует, что значение индекса Джаккарда всегда лежит в отрезке . Если значение этого коэффициента стремится к 1, то с геометрической точки зрения, объекты идентичны. Стремление индекса Джаккарда к 0 означает, что объекты значительно отличаются друг от друга.

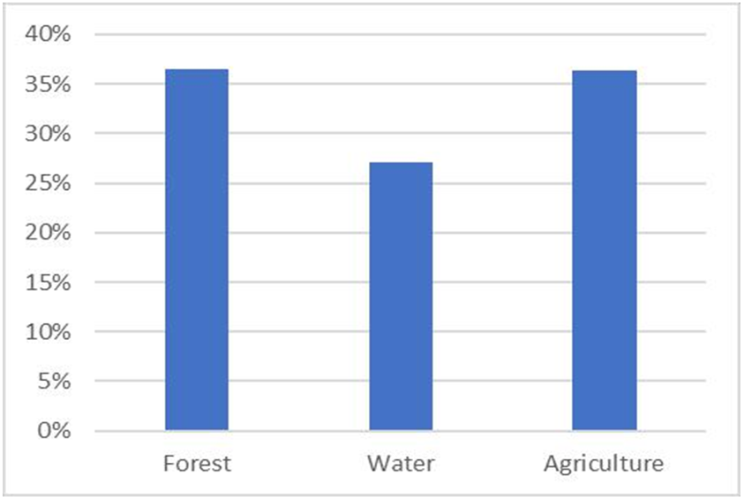
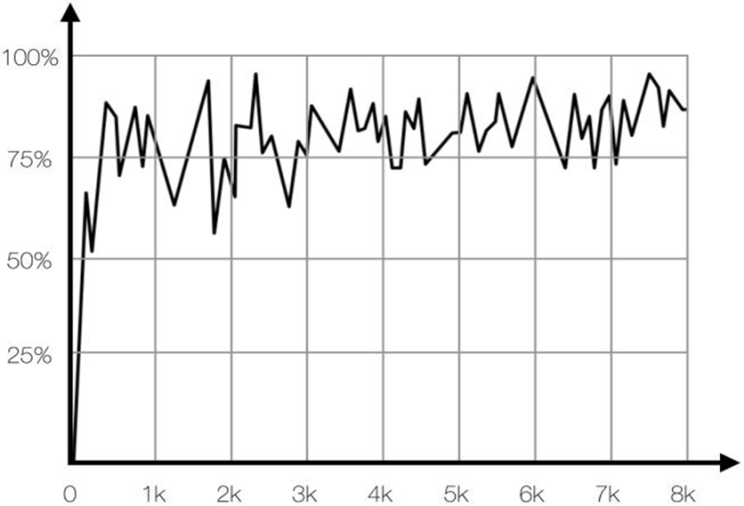
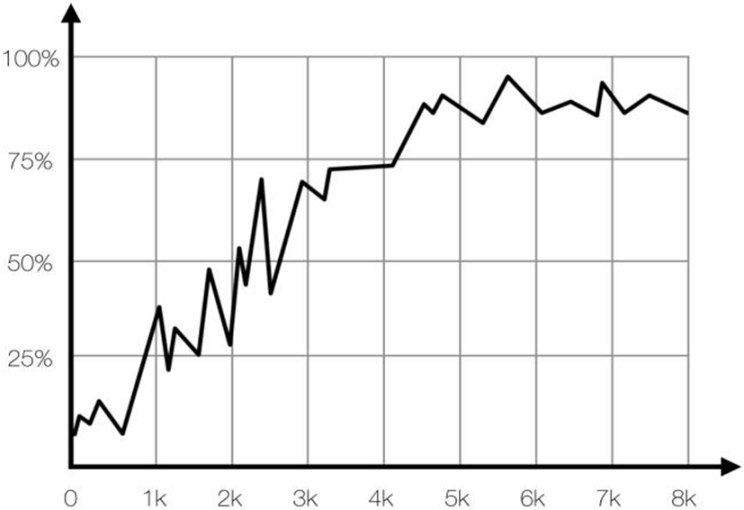


Рис. 1. Процентное соотношение классов на снимках обучающего множества



а)



б)

Рис. 2. Точность обнаружения объектов класса «вода»: во время обучения (а) и во время тестирования (б)

Обучение сверточной нейронной сети проводилось в 60 эпох. Такой подход позволил избежать переобучения, а также получить максимальную точность обнаружения объектов каждого класса. Во время выполнения эпохи проводилось 450 тренировочных итераций. За каждую итерацию, на вход сверточной нейронной сети подавалось 200 изображений, случайным образом выбранных из тренировочной выборки. На вход, модель принимала спутниковые снимки размером 224224 пикселей. Такой размер оказался наиболее оптимальным для достижения высокой скорости обучения, с учетом отсутствия необходимости выделения слишком мелких признаков. Увеличение размера изображения будет иметь смысл при рассмотрении дополнительных классов для обнаружения объектов, таких как «транспорт», «здания» и «самолеты».

На рис. 2 отражены изменения в точности обнаружения объектов класса «вода» во время обучения (а) и во время тестирования (б).

График функции потерь показан на рис. 3.

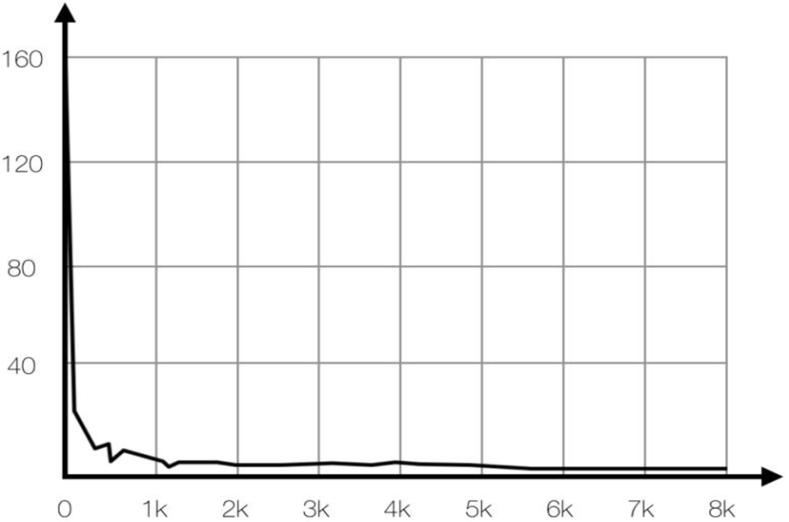


Рис. 3. График функции потерь во время обучения сверточной нейронной сети с архитектурой U-Net для обнаружения объектов класса «вода»

3. ОПТИМИЗАЦИЯ РАБОТЫ АЛГОРИТМА

Классический метод обнаружения объектов на спутниковых снимках предполагает подачу на вход уже обученной нейронной сети изображений необходимого размера. В рамках разработанной модели это спутниковые снимки размера 224224 пикселей. Поскольку скорость обнаружения объектов с помощью этого алгоритма достаточно низкая, для максимальной скорости обхода спутникового изображения детектор последовательно разрезает изображение на части, как показано на рис. 4.



Рис. 4. Разделение изображения на части для подачи на вход детектору

Как показывают результаты численных экспериментов, U-Net хуже работает с крайними пикселями изображения, чем с расположенными в центре. В результате была выполнена оптимизация обхода изображения с помощью маски: каждое изображение проходилось дважды со смещением на ½ фрагмента (рис.5).

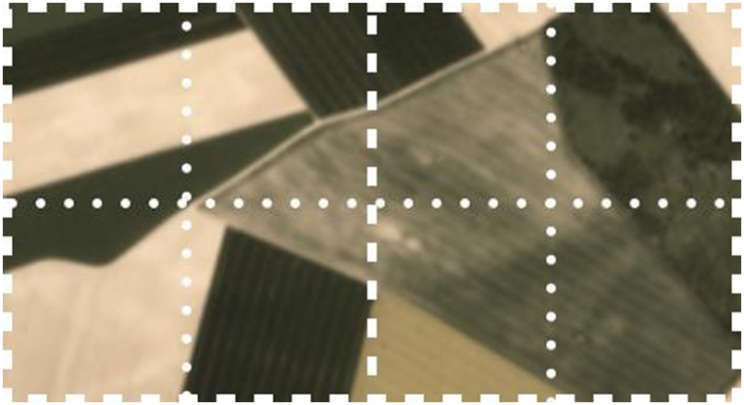


Рис. 5. Пример обхода фрагментов снимка ячеек со смещением

Вероятность принадлежности к определенному классу для пересекающихся частей определялась как среднее арифметическое вероятностей каждого класса. Результирующее значение вероятности и соответствующий класс определялся, исходя из наивысшего среднего значения процента вероятности.

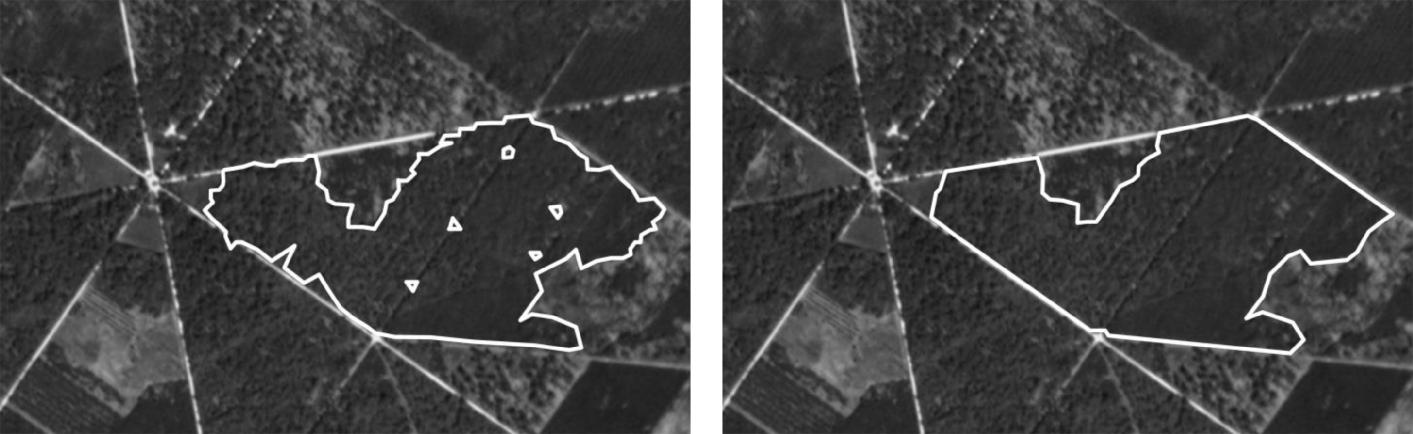
Результатом работы алгоритма является назначение каждому пикселю изображения одного из 3 классов («лес», «вода», «сельскохозяйственные угодья»). Для работы с этой информацией необходимо сгруппировать пиксели одного класса. Для этого решалась задача кластеризации. Было произведено сравнение двух алгоритмов такого рода: алгоритм k-средних и дискретизации, которые включены в библиотеку sklearn языка программирования Python. На рис. 6 показаны результаты работы этих алгоритмов.

Исходя из численных экспериментов, можно заметить, что алгоритм k-средних более точно определяет границы обнаруженных объектов, при этом дискретизация значительно снижает ошибку обнаружения, делая объекты геометрически цельными. В результате такого сравнения для системы детектирования, был выбран алгоритм дискретизации с параметрами по умолчанию.

Для оценки качества работы модели было произведено сравнение результатов детектирования с экспертной разметкой на тестовом наборе данных. Тестовая выборка содержала 100 спутниковых снимков. Точность работы алгоритма детектирования определялась исходя из процента пересечения области, выделенной моделью и экспертной разметки

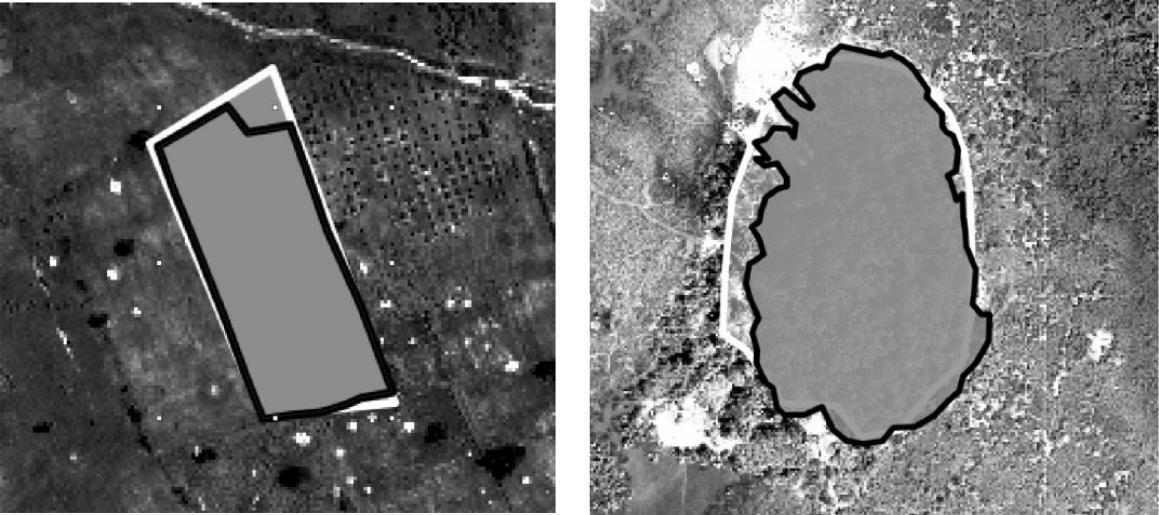
.

Примеры изображений с объектами, отмеченными алгоритмом и экспертом показаны на рис. 7. Белая линия определяет контур объекта, обнаруженного разработанным алгоритмом, а черная линия – выделенного экспертом.



а) б)

Рис. 6. Пример работы алгоритмов кластеризации для получения контура детектированного объекта a) алгоритм k-средних, b) дискретизация



а) б)

Рис. 7. Изображения со спутника Landsat-8 с автоматическим (белая кривая) и ручным (черная кривая) обнаружением объектов класса: а) «сельскохозяйственные угодья», б) «лес»

Для подтверждения гипотезы правильности выбора функции оценки точности алгоритма, были созданы два детектора с соответствующими функциями. Результаты их работы приведены в табл. 1 и 2. В приведенных таблицах общее количество объектов – общее число размеченных экспертами объектов соответствующего класса, а точность детектирования – величина, отражающая качество работы алгоритма и рассчитываемая по формуле:

,

где и - правильно и ошибочно детектированные объекты соответственно. Величины , и рассчитывались для каждого класса по отдельности.

Таблица 1. Результаты обнаружения объектов простым детектором

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Метрики | | | | |
| Классы | Общее число объектов | TD | FD | A | «Средний» процент пересечения |
| Лес | 200 | 160 | 24 | 86,96% | 89,62% |
| Вода | 40 | 33 | 6 | 84,62% | 80,83% |
| СХ | 220 | 199 | 30 | 86,90% | 91,34% |
|  |  |  |  | ~86,16% | ~87,26% |

Таблица 2. Результаты обнаружения объектов

модифицированным детектором

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Метрики | | | | |
| Классы | Общее число объектов | TD | FD | A | «Средний» процент пересечения |
| Лес | 200 | 168 | 21 | 88,89% | 92,32% |
| Вода | 40 | 38 | 4 | 90,48% | 81,71% |
| СХ | 220 | 208 | 11 | 94,98% | 96,17% |
|  |  |  |  | ~92,73% | ~90,07% |

Согласно табл. 2 наибольший процент пересечения объектов, обнаруженного алгоритмом, с экспертной разметкой получился для класса «сельскохозяйственные угодья» (СХ). Такой высокий результат связан с четкостью границ и видимым визуальным разделением окружающих объектов этого класса. 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.

4. ЗАКЛЮЧЕНИЕ

Итоговая точность классификации составляет 81,7% для объектов класса «вода», 92,3% для объектов класса «лес» и 96,1% для объектов класса «сельскохозяйственные угодья». Разработанный алгоритм может применяться для сегментации спутниковых снимков: выделения территорий городов, контроля строительства, моделирования климата и проч.

5. 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. S. Kluckner and H. Bischof, “Semantic classication by covariance descriptors within a randomized forest”, In Computer Vision Workshops (ICCV), pp. 665-672, 2009.
2. S. Kluckner, T. Mauthner, P.M. Roth, and H. Bischof, “Semantic classication in aerial imagery by integrating appearance and height information”, In ACCV, volume 5995 of Lecture Notes in Computer Science, pp. 477-488. Springer, 2009.
3. V. Mnih and G. Hinton, “Learning to detect roads in high-resolution aerial images”, In Proceedings of the 11th European Conference on Computer Vision (ECCV), September 2010.
4. V. Mnih and G. Hinton, “Learning to label aerial images from noisy data”, In Proceedings of the 29th Annual International Conference on Machine Learning (ICML 2012), June 2012.
5. P. Dollar, Z. Tu, and S. Belongie, “Supervised learning of edges and object boundaries”, In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1964-1971, 2006.
6. A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks”, In Advances in neural information processing systems, pp. 1097–1105, 2012.
7. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., “Imagenet large scale visual recognition challenge”, International Journal of Computer Vision, 115(3), pp. 211–252, 2015.
8. M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva,

“Land Use Classication in Remote Sensing Images by Convolutional Neural Networks”, Web: https://arxiv.org/abs/1508.00092, 2015.

1. Y.Yang and S. Newsam, “Bag-of-visual-words and Spatial Extensions for Land-use Classication”, In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS ’10). ACM, NY, USA, pp. 270–279.
2. O. A. B. Penai, K. Nogueira, and J. A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?”, In 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44–51.
3. K. Simonyan and 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, and N. Jacobs, “Predicting Ground-Level Scene Layout from Aerial Imager”, Web: https://arxiv.org/abs/1612.02709 (2016).
5. N. Jean, M. Burke, M. Xie, W.M. Davis, D.B. Lobell, and S.Ermon, “Combining satellite imagery and machine learning to predict poverty”, Science 353, 6301 (2016), pp. 790–794.

[14] “European Union. 2011. Urban Atlas”, Web: https://www.eea.europa.eu/data-and-maps/data/urban-atlas.

[15] “Landsat8”,Web:https://en.wikipedia.org/wiki/Landsat\_8.