УДК 004.855.5

РАСПОЗНАВАНИЕ ОСНОВНЫХ ОБЪЕКТОВ ИНФРАСТРУКТУРЫ ГОРОДСКОЙ МЕСТНОСТИ ПРИ ПОМОЩИ БПЛА И НЕЙРОСЕТИ U-NET

В.А. Михайлов (*vladislav.mikhailov@phystech.edu*)
Московский Физико-Технический Институт, Долгопрудный

О.Г. Пилипенко (*pilipenko@phystech.edu*)
Московский Физико-Технический Институт, Долгопрудный

Аннотация. В статье рассматривается способ распознавания и кластеризации основных объектов жилого кампуса МФТИ посредством нейронной сверточной сети квадрокоптером. Сперва была написана нейронная сеть, задача которой распознавание уличных объектов по фотографиям, полученным с высоты 30 метров квадрокоптером. Обученная нейросеть была использована с БПЛА, который летал надо кампусом МФТИ и отслеживал изменение положений, распознанных объектов.

Ключевые слова: машинное обучение, компьютерное зрение, сверточная нейронная сеть, беспилотный летающий аппарат (БПЛА), Open CV, распознавание подвижных объектов.

Введение

Машинное распознавание объектов является сложной задачей, которую не всегда можно решить привычными методами. Большое количество алгоритмов и реализованы в библиотеке Open CV. Однако для некоторых проблем, результат работы Open CV и других инструментов компьютерного зрения является неприемлемым. Так же значительная часть ныне существующих методов сложно применить на переносимых устройствах, ввиду ограниченности вычислительных мощностей.

В данной статье рассматривается БПЛА со сверточной нейронной сетью, обученной на снимках кампуса МФТИ. Нейронная сеть распознает объекты по принципу от простого к сложному:

1. Предсказывает крупномасштабное распределение объектов;
2. Таргетирует объекты и присваивает им метки;
3. Уточняет и распознает объект.

Такая многослойная структура нейросети позволяет избежать ошибки переобучения и достигнуть приемлемых результатов распознавания.

Целью данной работы является демонстрация возможностей сверточных нейросетей применительно к задачам распознавания объектов камерой БПЛА.

1. Описание модели эксперимента

Основными методами, используемыми в современных устройствах компьютерного зрения являются, как правило, Open CV. Зачастую их низкое качество работы, плохая обучаемость и отсутствие вычислительных мощностей у БПЛА, приводят к неудовлетворительным результатам, однако быстродействие данных алгоритмов достаточно высоко. Ввиду этих проблем, в качестве основного инструмента распознавания изображений БПЛА, была выбранная сверточная нейронная сеть с многоуровневой архитектурой, для улучшения обобщающей способности.

Так как БПЛА должен производить сьемку местности на высоте от 30 до 500 метров, чтобы снимки с имеющейся камеры были в хорошем качестве. Так же аппарат должен быть устойчивым к сильным порывам ветра, которые происходят на данных высотах, чтобы повысить качество съемки и безопасность.

Ввиду причин, описанных выше создать такой БПЛА в короткие сроки не представлялось возможным, поэтому выбор арендуемого БПЛА был остановлен на Phantom 4 (рисунок 1).

Рис. 1. Квадрокоптер Phantom 4

Ниже будут приведены основные интересующие характеристики данного аппарата, а также его камеры.

Как видно из таблицы 1 все характеристики БПЛА позволяют справиться со всеми техническими проблемами. Так же данный БПЛА оснащен камерой и стабилизатором, характеристики которых приведены в таблице 2.

Таблица 1

Характеристики Phantom 4

Масса (г)	**Высота полета (м)**	**Время полета (мин)**	**Макс. Допустимая скорость ветра (м/с)**	**Макс. Скорость (м/с)**
1388	До 6000	30	10	20

Таблица 2

Характеристики камеры и стабилизатора Phantom 4

Матрица	**Угол обзора объектива в (градусах)**	**Макс размер изображ.**	**Стабилизация по 3 осям**	**Точность работы стабилизатора в (градусах)**
1’’CMOS, 20 Мп.	84	5472 × 3078	Да	0.03

Основными процессами обработки изображения являются:

1. Захват изображения
2. Обработка изображения
3. Определение объектов и разбиение на группы

Первый пункт обработки реализуется при помощи камеры Phantom 4 и программного кода, использующего возможности библиотеки Open CV [Kevin Yu et al.]. Так мы получаем видео и при помощи данной библиотеки вырезаем из видео интересующие нас кадры таким образом, чтобы не было потеряно стыков краев каждого, вырезанного кадра. Далее происходит сама обработка изображения, попадающего на ПК по каналу Wi-Fi. Мы арендовали сервера компании Amazon, чтобы вся работа нейросети происходила в облаке. Данные передавались на сервер при помощи LTE. Так же хочется отметить, что несмотря на обработку данных в облаке, бортовой комплекс БПЛА не загружен, что позволяет организовать

высококачественное распознавание изображения. Данный подход позволяет иметь хорошие вычислительные мощности, которые требует нейросеть u-net на небольших маломощных ПК.

И итоговым пунктом обработки изображения является обработка его нейросетью, которая была уже предварительно обучена на других похожих снимках местности. Она разбивает объекты карты на пять основных групп: люди, машины, дороги, деревья и здания. Результатом работы нейросети является карта, с четырьмя основными, распознанными классами объектов. Критерии качества работы нейросети приведены в таблице 3 в виде четырех параметров оценки каждого алгоритма.

2. Параметры и архитектура нейросети

Прежде чем проектировать структуру нейросети была проведена проверка существующих алгоритмов классификации на данном датасете. Ниже приведены результаты работы алгоритмов: random-forest [Anna Bosch et al. 2007], линейная регрессия, логистическая регрессия и XGboost [Chen et al., 2016]. Ниже на рисунке 2 приведены результаты распознавания каждым из этих алгоритмов примеров из созданного датасета.

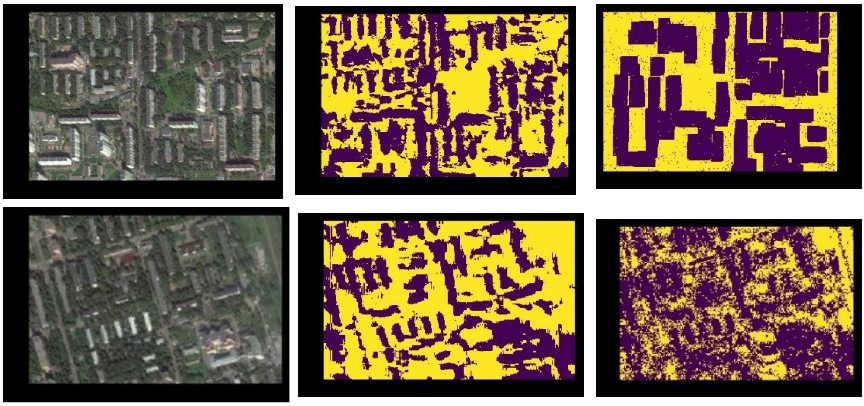

Рис. 2. Примеры обучения нейросети распознаванию строений известными алгоритмами (слева – исходные фото, по центру – линейная регрессия, справа – random-forest)

Из рисунка 2 видно, что ни один алгоритм не может предоставить качественные результаты. Так в случае использования линейной регрессии строения распознаются неточно с границами меньше чем это есть на картинке с БПЛА. В то же время реализация алгоритма random-forest определяет каждый дом больше, чем он есть, что приводит к слипанию

домов (рис. 2 справа). На кросс-валидации Random-forest показывает сильное зашумление.

В таблице три изложены основные параметры, которые были получены при сравнении параметров работы известных алгоритмов, с проектируемой сетью u-net, в применении к данной задаче.

Таблица 3

Численные характеристики параметров, полученные в результате экспериментов

Название алгоритма	**Train accuracy**	**Validation accuracy**	**Скорость обучения(с)**	**Скорость предсказания(мс)**
Random-forest	0,994	0,583	2,98	247
Linear reg.	0,787	0,649	1,344	68
XGboost	0,973	0,785	12,75	137
U-Net	0,989	0,954	>> 1 мин.	322

Как можно видеть из таблицы 3, по подавляющему большинству параметров лучшим, для решения этой задачи, является алгоритм U-Net. Эти результаты предопределили наш выбор в пользу данной архитектуры.

2.1 Архитектура нейросети

Главной задачей, проектируемой сверточной нейросети является распознавание четырех основных классов объектов: строения, дороги, зеленые массивы, вода.

Основой архитектуры, данной нейросети является нейросеть u-net [Chen et al., 2016]. На рисунке 3 приведена блок-схема нейросети.

Для предсказания шаблона- маски исползовалось обычное скользящее окно 3 на 3. Каждая из картинок нарезалась на патчи, которые предсказываются и собираются в исходную. Итогом работы данного этапа предсказания являются вероятности того, что в конкретном пикселе находится каждый из шести классов. Чтобы получить бинарную маску обучения мы вводим фильтр с уровнем пропускания 0,42, который был подобран экспериментально.

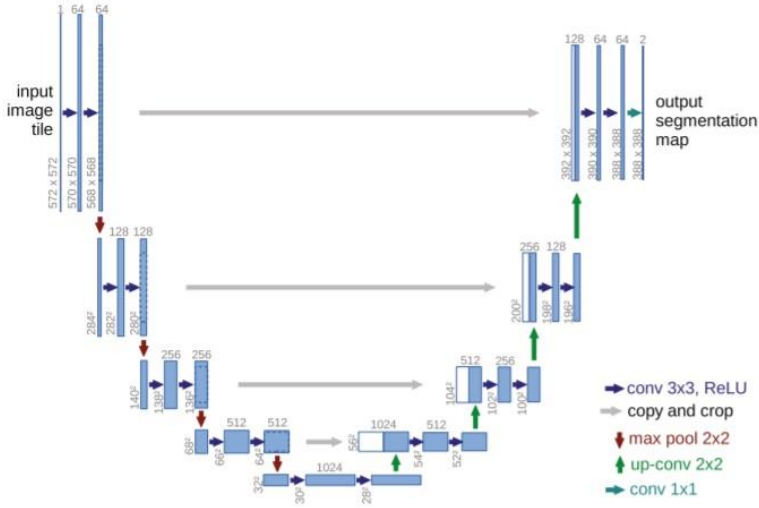

Рис. 3. Схема, реализуемой нейросети u-net

Архитектура сети показана на рисунке 3. Она состоит из контрактного пути (левая сторона) и расширенного пути (правая сторона). Контрактный путь является типичной архитектурой сверточной сети. Он состоит из повторяющихся применений двух сверток 3 × 3 (неуплотненные свертки), с активатором ReLU. Далее производится операция объединения 2x2 с максимальным шагом 2, для понижающей дискретизации. На каждом понижающем шаге дискретизации мы удваиваем число признаков в каналах. Каждый шаг расширенного пути состоит из повышающей дискретизации, после чего следует свертка 2x2 («up-convolution»), которая уменьшает вдвое количество каналов функции, далее следует склейка с соответствующим обрезанным образом карты признаков из левой части сети и две свертки 3x3, с активатором ReLU. Обрезка необходима из-за потери граничных пикселей в каждой свертке. На конечном слое свертка 1x1 используется для отображения каждого 64 - компонентного вектора признаков в необходимое количество классов. Всего сеть имеет 23 сверточных слоя. Чтобы обеспечить бесшовную разбивку карты сегментации вывода (см. Рисунок 2), важно выбрать размер входного фрагмента так, чтобы все операции с максимальным пулом 2x2 применялись к слою с четным размером по x и y.

2.2. Обучение нейросети

Для обучения нейросети было использовано 20 снимков территории города Долгопрудный, которые были размечены на 4 основных распознаваемых классов (см. п. 2.1) вручную. Каждое изображение было отмасштабировано к одному размеру 900 на 900 пикселей.

Входные изображения и соответствующие им карты сегментирования используются для обучения сети с реализацией стохастического градиентного спуска Theano [Yichuan Tang, 2013]. Из-за особенностей сверток, выходное изображение меньше входного на постоянную ширину границы. Чтобы минимизировать накладные расходы и максимально использовать памяти GPU, мы отдаем предпочтение большим входным фрагментам за большой размер партии и, следовательно, уменьшите партию до одного изображения. Соответственно, мы используем большой импульс (0,99) так, чтобы большое количество ранее рассмотренных учебных образцов определяло обновление в текущем шаге оптимизации.

Для оценки функции потерь проектируемой нейросети используется логистическая функция потерь (log loss multiclass). Подсчет ошибки производится по формуле:

$$F_{loss} = -\frac{1}{N}\sum_{i}^{N}\sum_{J}^{M} y_{ij}\ln(p_{ij})$$

где N – количество экземпляров, M – количество разных меток, y_{ij} – двоичная переменная с ожидаемыми метками, p_{ij} – вероятность классификации, выдаваемая классификатором для i-го экземпляра и j-ой метки.

Мы предварительно вычисляем карту весов для каждой основной сегментации, чтобы компенсировать различную частоту пикселей от определенного класса в наборе данных обучаемых и принудить сеть к изучению малых классов.

Граница раздела вычисляется с использованием морфологических операций. Затем карта весов вычисляется по формуле:

$$w(x) = w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right)$$

где $w_c(x)$ – карта весов для компенсирования частот классов, w_0 = 5 в нашем эксперименте, σ=5 пикселей, $d_1(x)$ - расстояние до границы ближайшей ячейки, $d_2(x)$ – расстояние до границы второй ближайшей ячейки.

В глубоких сетях со многими сверточными слоями и разными путями в сети, очень важна инициализация весов. В противном случае, части сети могут иметь чрезмерную активацию, в то время как другие части никогда не вносят свой вклад. В идеале первоначальные веса должны быть

адаптированы таким образом, чтобы каждая карта функции в сети имела единичную дисперсию. Для сети с нашей архитектурой (чередующиеся свертки и слои ReLU), это может быть достигнуто путем отрисовки начальных весов из гауссовского распределения со стандартным отклонением $\sqrt{2/N}$, где N обозначает количество входящих узлов одного нейрона[5]. Например, для свертки 3x3 и 64 функциональных каналов на предыдущем уровне N = 9 · 64 = 576.

Лучшие показатели нейросети U-Net во время тренировки были при реализации 64 эпох обучения (при данном количестве появлялся нормальный результат):

- Качество тренировки - 0,9889
- Train loss - 0,0104
- Качество на валидации 0,954
- Validation loss 0,117

3. Примеры распознавания и сегментации объектов инфраструктуры

На рисунках 4 - 7 приведены результаты работы сверточной нейросети. Каждый результат представляет собой три картинки, снимок с камеры, распознавание человеком и распознавание какого-либо класса компьютером. В результате получили качественные предсказания для четырех классов из пяти.

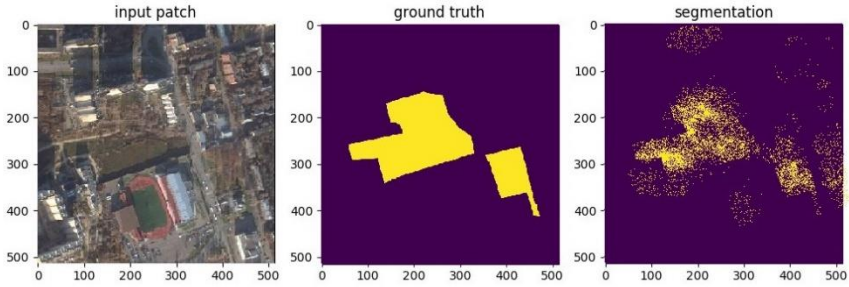

Рис. 4. Распознавание зеленых массивов

Одними из самых неточных результатов получился при распознавании зеленых массивов (см. рис.4). Видно, что сеть способна вычленять основные скопления деревьев, а также давать положения более мелких скоплений. Однако следует отметить большую зашумленность результата случайными точками, из-за неточности разметки и большого разброса деревьев.

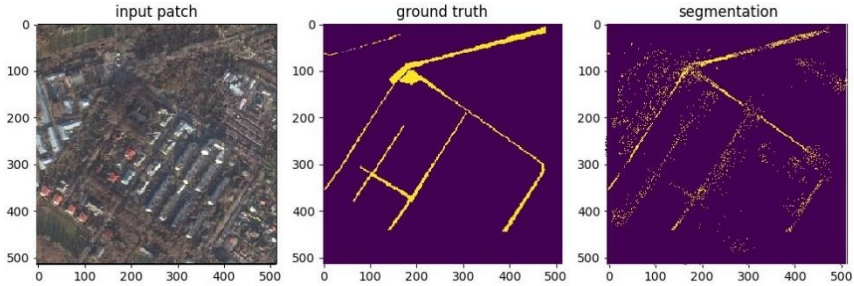

Рис. 5. Распознавание дорог

Вторым классом по качеству распознавания мы отнесли класс дорог (см. рис.5). Здесь присутствует большая зашумленность и периодически непрерывность дорог пропадает, мы считаем данный результат нормальным, данные результаты связываем с тем, что градиентные фильтры, которые присутствуют в нейросети не могут четко определять перепад между жилой застройкой и проезжей частью.

Самые лучшие результаты были продемонстрированы при распознавании жилых построек (см рис. 6-7). Спроектированная нейросеть смогла находить некоторые дома, которые не получалось найти изначально человеческим глазом. Несмотря на большой уровень зашумления основные дома были распознаны, за исключением центрального дома в левом верхнем углу рисунка 6. Связываем данный результат с тем, что в обучающей выборке был всего лишь один дом с оранжевым цветом, что не достаточно для нейросети, чтобы обучиться.

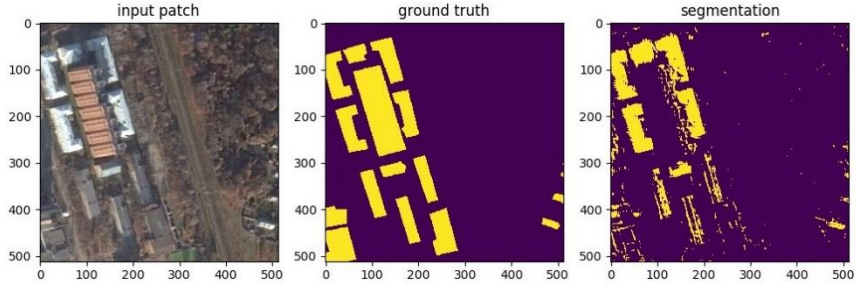

Рис. 6. Распознавание домов

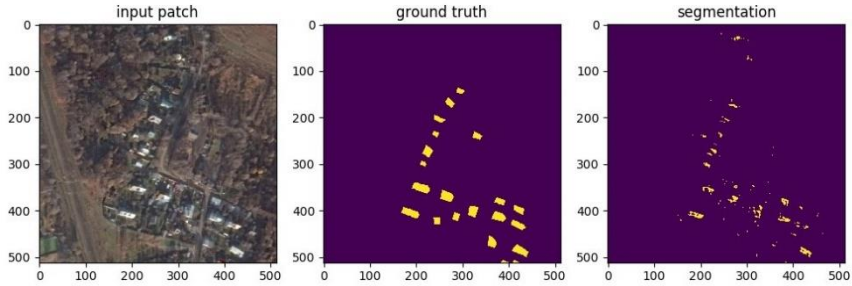

Рис. 7. Распознавание домов

Наилучшие результаты были достигнуты для класса строений, а наихудшие для класса дорог. Так, например, обученная нейросеть, распознавала все имеющиеся на фотографии дома, однако не всегда точно удавалось определять границы строений. Как можно видно из рисунка данный тип фильтрации нейросети не позволяет полностью избавиться от шумов, размеры которых могут выражаться в виде точек в 2-3 пикселя, что очень сильно затрудняет распознавание класса автомобилей.

4. Достигнутые результаты

Основным достижением авторы считают создание системы, которая в режиме реального времени способна распознавать основные элементы инфраструктуры местности (строения, зеленые массивы, дороги, вода). Основной принцип работы данной системы изложен в пункте 1 данной статьи.

Как было показано в пункте 2 данной статьи, нейросеть U-Net обладает наилучшим качеством распознавания, применительно к данной задаче. Поэтому авторы считают, что одной из возможностей для дальнейшего развития – расширение возможностей данной системы для распознавания объектов на различных типах местности.

После анализа полученных результатов распознавания (см. пункт 3) были сделаны выводы, позволяющие улучшить качество данного исследования. Условия улучшения результатов:

- увеличение качества изображения;
- увеличение размера обучающей выборки;
- улучшение качества разметки, увеличение числа слоев и фильтров в сверточных слоях, для чего необходимо более мощное оборудование;
- добавление дроп-аута (Dropout).

К сожалению, на данный момент, приведенное, решение вопроса распознавания объектов при помощи БПЛА, невозможно встроить в программно-аппаратный комплекс малых БПЛА, однако при использовании мощных вычислительных машин на земле или облачных технологий, данное решение имеет возможность для реализации.

При выполнении вышеперечисленных условий, применение данного вида нейросети возможно в картографических, розыскных мероприятиях,

Благодарности. Авторы считают своим приятным долгом поблагодарить д.ф.-м.н., профессора В. Е. Павловского, который направлял, консультировал и помогал данному исследованию.

Список литературы

[O Ronneberger et al., 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation.

[Anna Bosch et al. 2007] A. Bosch, A. Zisserman, X. Munoz. Image Classification using Random Forests and Ferns, 2007.

[Chen et al., 2016] Tianqi Chen, Carlos Guestrin., XGBoost: A Scalable Tree Boosting System

[Kevin Yu et al.] Kevin Yu Rahman Kadierding Umesh Dinkar., XGBoost: Capturing Webcam Images to Display on a Web Interface

[Yichuan Tang, 2013] Yichuan Tang, XGBoost: Deep Learning using Linear Support Vector Machines

[Martin Längkvist et al., 2016] Martin Längkvist, Andrey Kiselev, Marjan Alirezaie and Amy Loutfi: Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks

[Idris Kahraman et al., 2015] Idris Kahraman, Muhammed Kamil Turan, and Ismail Rakip Karas: Road Detection from High Satellite Images Using Neural Networks

http://lasagne.readthedocs.io/en/latest/index.html - documentation for Lasagne – good framework for neural network