

5. Для оценки качества изображений с **изменением контраста** лучше использовать значение параметра ФОРМА распределения Вейбулла для локальных оценок меры HELM, так как в данном случае коэффициент корреляции Спирмена максимален и равен 0,83. Нецелесообразно использовать локальные оценки мер BREN, CON3, LAPD, LARM, LOCC, LOEN, SVDB потому, что для них отсутствует корреляция с визуальными оценками экспертов.

Исследования выполнены в рамках проекта БРФФИ Ф18МС-028.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

- Xu, S. No-reference/blind image quality assessment: a survey / S. Xu, S. Jiang, W. Min // IETE Technical Review. – 2017. – Vol. 34. – № 3. – С. 223–245.
- Старовойтов, В. В. Уточнение индекса SSIM структурного сходства изображений / В. В. Старовойтов // Информатика. – 2018. – Т. 15. – № 3. – С. 41–55.
- Голуб, Ю. И. Сравнительный анализ безэталонных оценок резкости цифровых изображений / Ю. И. Голуб, Ф. В. Старовойтов, В. В. Старовойтов // Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2019. – № 7(125). – С. 113–120.
- Голуб, Ю. И. Исследование локальных оценок контраста цифровых изображений при отсутствии эталона / Ю. И. Голуб, Ф. В. Старовойтов // Системный анализ и прикладная информатика. – 2019. – № 2(22). – С. 4–11.
- Ponomarenko, N. Image database TID2013: Peculiarities, results and perspectives / N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo // Signal Processing: Image Communication. – 2015. – V. 30. – P. 57–77.
- Larson, E. C. Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy / E. C. Larson, D. M. Chandler // Journal of Electronic Imaging. – 2010. – Vol. 19. – № 1. – P. 1–21.
- Gu, K. Subjective and objective quality assessment for images with contrast change / K. Gu, G. Zhai, X. Yang, W. Zhang, M. Liu // Proc. IEEE Int. Conf. on Image Processing. – Melbourne, VIC, Australia. – Sep. 2013. – P. 383–387.
- Univariate Distribution Relationships [Электронный ресурс] / The College of William and Mary in Virginia. – Режим доступа : <http://www.math.wm.edu/~leemis/chart/UDR/links.html>.
- Старовойтов, Ф. В. Параметры кривой распределения локальных оценок как меры качества изображений / Ф. В. Старовойтов, В. В. Старовойтов // Системный анализ и прикладная информатика. – 2018. – № 3. – С. 26–41.
- Guan, J. No-reference Blur Assessment Based on Edge Modeling / J. Guan, W. Zhang, J. Gu, H. Ren // Journal of Visual Communication and Image Representation. – 2015. – T. 29. – C. 1–7.

03.02.2020

GOLUB Y. I., STAROVOITOV F. V., STAROVOITOV V. V. Study of no-reference local image quality assessments

The article presents results of new generalizing for the set of local no-reference image quality estimates. Currently, for this, the arithmetic mean value is mainly used. It is one of the parameters of the normal distribution. We propose to use the parameters of distributions of the logarithmic type as new quality estimates for images with different types of distortions. Using the maximum likelihood method, we proved that they better generalize properties of sets of local estimates. To compare the expert mean opinion scores of image quality with the proposed assessments, the Spearman rank correlation coefficient was used. Digital images from public databases were used in our experiments.

004.89

• • • • •

Введение. Автоматическое распознавание спутниковых или аэроснимков – это наиболее перспективный способ получения информации о расположении различных объектов на местности. Отказ от ручной разметки снимков особенно актуален, когда речь заходит об обработке больших участков земной поверхности в сжатые сроки.

Распознавание изображений – это технология, которая заключается в обработке, анализе и понимании изображений реального мира: идентификации целевых объектов посредством использования алгоритмов и концепций машинного обучения.

Задача распознавания крон деревьев на спутниковых изображениях заключается в выделении на них областей, соответствующих реальному расположению выбранных объектов исследования. Для человека не составляет особого труда решить данную задачу, однако это довольно нетривиальная задача для вычислительной машины, так как практически любое её действие – это чётко определённый алгоритм. В общем виде задачи распознавания трудно алгоритмируются, потому для их решения в настоящее время широко используются методы искусственного интеллекта, например, нейронные сети.

Наибольшей популярностью при решении задач такого типа пользуются свёрточные нейронные сети, которые входят в состав технологий глубокого обучения, – попытка научить компьютеры самостоятельно обучаться на большом количестве данных.

Спутниковые изображения находят применение во многих отраслях деятельности – сельском хозяйстве, геологических и гидрологических исследованиях, лесоводстве, охране окружающей среды, планировке территорий, образовательных, разведывательных и военных целях. Такие изображения могут быть выполнены как в видимой части спектра, так и в ультрафиолетовой, инфракрасной и других частях диапазона.

В последнее время многие команды разработчиков ставят перед собой задачу преобразования спутниковых снимков в полноценные размеченные карты. Чтобы с картой было удобно работать, на снимках многоугольниками размечаются объекты: леса, водоёмы, улицы, дома и т. п. Обычно разметкой занимаются специалисты-картографы. Задача разметки карт вручную является очень трудоёмкой, ресурсоёмкой и требующей много времени для выполнения, потому задача автоматизации этого процесса является очень актуальной.

За операции с изображениями отвечает область информационных технологий, которая называется компьютерным зрением. Одна из областей, которая описывается как принадлежащая к искусственному интеллекту и одновременно имеющая связь с компьютерным зрением, – это распознавание образов. Распознавание образов – направление, связанное с разработкой принципов и построением систем, предназначенных для определения принадлежности данного объекта к одному из заранее выделенных классов объектов.

Хацкевич Мария Викторовна, старший преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Михняев А. Л., старший преподаватель кафедры интеллектуальных информационных технологий Брестского государственного технического университета.

Беларусь, БрГТУ, 224017, г. Брест, ул. Московская, 267.

Задачи распознавания – это задачи, включающие в себя два крупных этапа:

- преобразование исходных данных к виду, удобному для распознавания (может включать в себя целый ряд этапов и действий – например, детектирование, сегментацию, локализацию и т. д.);
- собственно распознавание (указание принадлежности объекта определенному классу).

Успешное решение задачи распознавания прямо пропорционально зависит от успешного выполнения этапа подготовки данных, потому не стоит пренебрегать возможностями улучшить исходные данные.

Классическая задача в компьютерном зрении, обработке изображений и машинном зрении – это определение, содержат ли данные (изображение, видеоряд) некоторый характерный объект, особенность или активность.

Существующие методы решения этой задачи эффективны только для отдельных объектов, таких как простые геометрические объекты (например, многогранники), человеческие лица, печатные или рукописные символы, автомобили в определённых условиях – обычно это определённое освещение, фон и положение объекта относительно камеры.

Задача распознавания крон деревьев на изображении с применением нейросетевых методов ориентирована на распознавание изображения по семантической составляющей, то есть по содержанию – выделение крон деревьев среди всех объектов, присутствующих на изображении.

В результате решения задачи распознавания крон деревьев на изображении с применением нейросетевых методов получаем бинарное изображение, отображающее кроны как единственный значимый объект, а все остальные объекты рассматриваем в качестве фона. Данный процесс заключается в выделении на изображении связанных областей по критерию однородности, а затем использование метода порогового отсечения для обработки цветных изображений. Такой выбор обусловлен возможностью значительно сократить исходный объём данных в процессе разметки изображения, ведь файлы изображений карты в высоком разрешении могут занимать гигабайты памяти системы. Важным будет упомянуть, что не предусматривается распознавание вообще всех зелёных насаждений (например, газон, трава, поля).

Решение данной задачи может быть использовано в реальных прикладных задачах, например, в решении задачи подсчёта плотности лесных насаждений. Также возможен вариант использования работниками экологических служб или лесничеств, или для работников сельского хозяйства, работников охраны окружающей среды. Практическую ценность поставленная задача представляет и для инженеров-строителей, перед которыми ставится задача планировки, то есть использования земельных участков под строительство зданий.

Интеллектуальное ядро поставленной задачи потенциально может быть полезным для решения задачи распознавания других объектов, а, следовательно, оно может быть полезно для многих сфер деятельности. Путём замены входных данных, а именно масок (например, если необходимо распознать строения, дороги, водоёмы и т. п.), можно получить размеченную карту по интересующему критерию, переобучив сеть.

Для реализации поставленной задачи распознавания изображения по семантической составляющей, то есть по содержанию – выделение крон деревьев среди всех объектов, присутствующих на изображении, необходимо:

- сформировать входные данные;
- провести анализ, на основе которого выбирается архитектура нейронной сети;
- выполнить процесс Transfer learning;
- проанализировать полученные результаты.

Задача распознавания крон деревьев является высокоуровневой задачей обработки изображений, относящейся к группе задач т. н. слабого искусственного интеллекта. Она является даже более сложной, чем задача классификации изображений и детектирования объектов, что обусловлено не только необходимостью определения класса объекта, но и выявления его структуры, правильного выделения частей объекта на изображении.

Решение задачи анализа аэрофотоснимков высокого разрешения и распознавания, выбранных для исследования объектов, приводит к возможности автоматического построения детальных карт местности или города. При создании карт как раз необходимо точно определить границы объектов на снимке поверхности земли.

1. Исходные данные. Источником исходных данных для реализации поставленной задачи являются картографические сервисы и непосредственно оцифрованные карты. Ключевым понятием предметной области является тайл.

Так как существуют различные карты по назначению – ROADMAP (обычная 2D-карта), SATELLITE (фотографические карты), HYBRID (фотографическая карта в совокупности с дорогами и нанесёнными названиями городов), TERRAIN (карта с горами, реками и т. д.) – тайлы тоже выглядят по-разному. Примеры тайлов приведены на рисунке 1.



Рисунок 1 – Примеры тайлов

В данном случае интерес вызывают тайлы фотографической карты, то есть изображения, снятые спутниками. Каждый тайл представляет собой изображение формата jpeg (спутниковые снимки, некоторые отсканированные карты) или png (карты, слои) и хранится в файле с уникальным именем, которое определяется координатами этого тайла по осям X и Y (существуют также альтернативные варианты названий, например, только по y – координате и т. п.).

Большинство картографических сервисов предоставляют тайлы размерами 256x256 пикселей. Количество тайлов, из которого состоит изображение, зависит от масштаба. Масштаб (от немецкого messen – мера и stab – палка) – отношение длины отрезка на карте, плане, аэро- или космическом снимке к его действительной длине на местности. Например, на сервисе Google Maps изображение в масштабе Z1 (самом мелком) состоит всего из 4-х тайлов. Обозначение Z1 указывает на уровень детализации. Всего их 24: Z1, Z2, Z3... Z24. Z1 – наименьшая детализация, а Z24 – наибольшая. Собственно буква Z означает Zoom (с английского – увеличение изображения).

Использование тайловой структуры позволяет загружать не всё изображение целиком (которое может быть действительно огромным), а только некоторую его часть.

Не менее важной сущностью, является маска. Маска представляет собой размеченную бинарным образом копию исходного изображения. Для решения поставленной задачи исходный тайл и маска имеют размерность 256x256 пикселей. Под словом «размеченная» подразумевается выделение областей, представляющих собой кроны деревьев, чёрным цветом и фона – белым цветом. Маска является составляющей обучающей выборки наряду с исходным изображением (тайлом).

Пример маски представлен на рисунке 2.

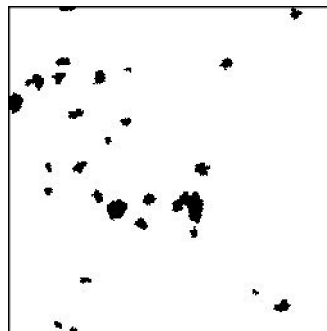


Рисунок 2 – Маска для исходного изображения (тайла)

Огромная часть времени для решения задачи распознавания крон деревьев на спутниковых изображениях отведена на создание обучающей и валидационной выборки для тренировки сети. Проблем с нахождением непосредственно спутниковых снимков совершенно не возникало, так как большинство картографических сервисов предоставляют возможность скачать данные в любом удобном формате – tiff, jpeg, png, json, csv и др. Однако масок, необходимых для обучения сети, недостаточно. Каждое из изображений датасета проходило через следующие этапы обработки:

- наложение прозрачного слоя поверх пикселей картинки;
- вырисовывание границ крон деревьев на добавленном слое;
- закрашивание областей с деревьями чёрным цветом, фона – белым цветом;
- удаление слоя основной картинки;
- сохранение верхнего слоя (непосредственно маска).

Процесс создания маски показан на рисунке 3. Конечно, такая обработка не могла не сказаться отрицательно на точности сети. Во-первых, рисовать вручную точно по границам очень сложно. Во-вторых, визуально даже человеку не всегда понятно, являются ли некоторые области кронами или всё же чем-то другим. В-третьих, некоторые кроны, в особенности одиночные деревья, при таком масштабе занимают настолько малое количество пикселей на изображении, что минимальная область захвата инструмента для выделения больше, чем участок, который должен быть выделен. В-четвертых, на изображениях иногда присутствуют искажения из-за условий съемки – облака, тени и т. п.

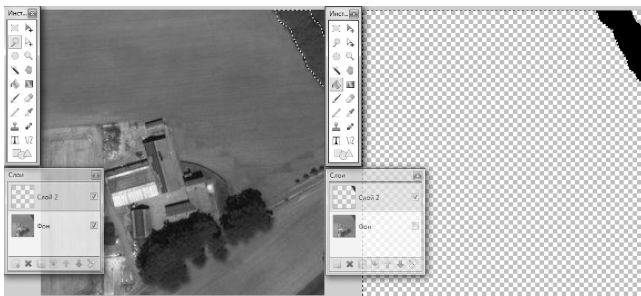


Рисунок 3 – Процесс создания маски для тайла

Таким образом, сформированы данные для обучающей, тестовой и валидационной выборки (рисунок 4).

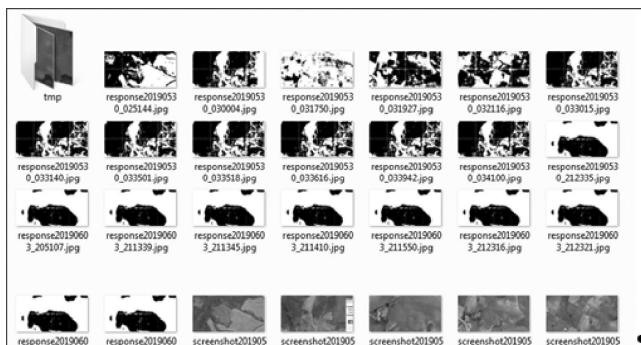


Рисунок 4 – Данные для обучающей, тестовой и валидационной выборки

2. Архитектура нейронной сети. Сверточная нейронная сеть – одна из наиболее популярных архитектур на сегодняшний день. Данный вид сети сильно отличается от остальных и в основном используется для работы с графической и аудиоинформацией. Данная архитектура обрабатывает данные не целиком, а фрагментами, но при этом данные не дробятся на части, а осуществляется своего рода последовательный прогон. Далее данные передаются дальше по слоям. Кроме сверточных слоёв (С) в сверточных нейронных сетях используются также слои объединения (P – pooling). Pooling – слои объединения сжимаются с глубиной (обычно степенью двойки).

К конечным слоям добавляются несколько перцептронов (сеть прямого распространения) для последующей обработки данных.

Архитектура сверточной нейронной сети требует настройки большого количества параметров, например, настраивается количество слоев и количество ядер, выбираются функции активации и т. д. Поэтому выбор архитектуры сверточной нейронной сети основывается главным образом на типе решаемой задачи (аппроксимация, прогнозирование, кластеризация) и входных данных (размерности, ключевых факторах). Так, в данном случае делается упор на обработку графической информации с целью распознавания (для реализации задачи распознавания крон деревьев на спутниковых изображениях) была выбрана архитектура, которая относится к сверточным нейронным сетям – U-Net.

Модель U-Net – успешная и популярная модель для подобных задач. Успех обусловлен возможностью учета двумерной топологии изображения в отличие от многослойного перцептрона. U-Net считается одной из стандартных архитектур сверточных нейронных сетей для задач распознавания изображений, когда нужно не только определить класс изображения целиком, но и создать маску, которая будет разделять изображение на несколько классов. Архитектура состоит из сжимающего пути для захвата контекста и симметричного расширяющегося пути, который позволяет осуществить точную локализацию.

Для U-Net характерно достижение высоких результатов в различных реальных задачах, особенно для биомедицинских приложений и использование небольшого количества данных для достижения хороших результатов.

Архитектура U-Net состоит из сокращающегося пути (левая сторона) и расширяющегося пути (правая сторона). Архитектура сети показана на рисунке 5.

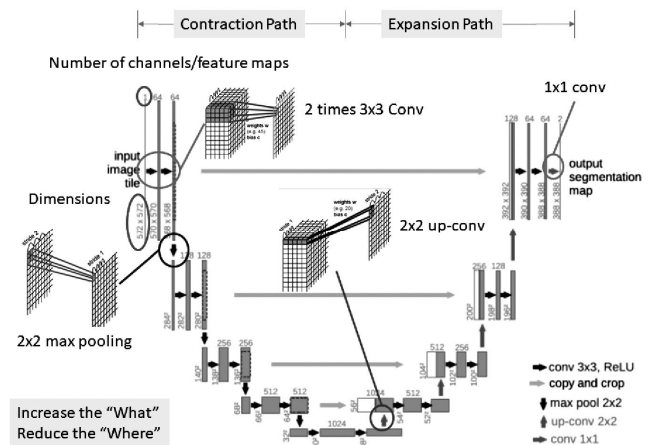


Рисунок 5 – Пример архитектуры сети U-Net

Основные этапы, которые проходит изображение в процессе обработки:

- операция свёртки;
- активация;
- batch-нормализация;
- операция субдискретизации (например, max pooling);
- dropout-регуляризация (процедура временного «выбрасывания»).

В сверточной нейронной сети обработка участка изображения должна происходить без учёта конкретной локации рассматриваемого участка. Конечно, нельзя отрицать, что взаимное расположение предметов на изображении имеет значение, но, в любом случае, первостепенной задачей является распознавание этих объектов, которое является локальной и независимой задачей относительно конкретных координат объекта на всём изображении.

Для решения такой задачи в сверточной нейронной сети производится анализ небольших «окон», которыми покрывается всё изображение, с целью выделения признаков в каждом таком «окне». Затем выходные данные такого преобразования можно представить,

например, в виде изображения, состоящего только из центральных пикселей каждого «окна», и передвигать «окно» уже по нему.

Значения каждого признака, которые выделяются из «окон» в исходном изображении, будут представлять собой целую матрицу. Каждая такая матрица называется картой признаков (feature map). Зачастую понятие карты признаков заменяют понятием каналов.

Такой метод сканирования изображения называется методом скользящего окна, или локального рецептивного поля (receptive field) для соответствующего ему нейрона карты признаков. Поэтому если размер скользящего окна, которое называется ядром, равняется \times (рецептивное поле), то каждый нейрон свёрточного слоя связан с \times^2 элементами соответствующей области рецептивного поля изображения. Каждое рецептивное поле (окно) во входном пространстве образов отображается на специальный (отдельный) нейрон в каждой карте признаков. Количество нейронов в каждой карте признаков в общем случае вычисляется следующим образом:

$$D(C_1) = \left(\frac{n-p}{s} + 1 \right) \left(\frac{n-p}{s} + 1 \right),$$

где $n \times n$ – размерность исходного изображения,

\times – размерность ядра,

s – шаг скольжения окна.

Общее количество различных синаптических связей в свёрточном слое равно:

$$V(C_1) = M(p^2 + 1),$$

где – общее количество карт признаков в свёрточном слое.

Для упрощения математического описания свёрточного слоя пиксели входного изображения можно представить в одномерном пространстве. Тогда выходное значение ij -го нейрона для k -й карты признаков в конволюционном слое определяется по следующей формуле:

$$y_{ij}^k = F \left(\sum_c w_{cij}^k x_c - T_{ij}^k \right),$$

где F – функция активации;

– $1, p^2$;

– c -я компонента входного образа соответствующей карты признаков свёрточного слоя;

k – номер карты признаков,

w_{cij}^k – весовой коэффициент между c -м нейроном входного

слоя и ij -м нейроном в k -й карте признаков,

T_{ij}^k – пороговое значение ij -го нейрона в k -й карте признаков,

ij – номер нейронного элемента в карте признаков [7, 8].

Следующим важным этапом является batch-нормализация. Обучение сложных нейронных сетей иногда может занимать дни и недели только для одной конфигурации. А чтобы подобрать оптимальную конфигурацию для конкретной задачи, требуется запустить обучение несколько раз — это может занять месяцы вычислений даже на действительно мощной машине.

Формально можно описать слой batch-нормализации с помощью нескольких формул. Слой получает на вход мини-batch $\{x_1, \dots, x_m\}$, а затем последовательно:

- вычисляет данные статистики по мини-batch:

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i,$$

где μ – математическое ожидание,

m – размерность batch,

x_i – i -й элемент в batch.

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2,$$

где σ^2 – дисперсия;

- нормализует входы:

$$x_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}},$$

где x_i – нормализованная величина для i -го элемента batch,

ε – константа, добавляемая к дисперсии во избежание деления на ноль;

- вычисляет результат:

$$y_i = \gamma x_i + \beta,$$

где y_i – результирующее значение,

γ, β – дополнительные параметры для масштабирования и сдвига нормализованной активации по каждой компоненте.

В классическом свёрточном слое, кроме линейной свёртки и следующей за ней нелинейности, есть и еще одна операция – субдискретизация (pooling; по-русски ее иногда называют еще операцией «подвыборки», от альтернативного английского термина subsampling).

Смысл субдискретизации: в свёрточных сетях обычно исходят из предположения, что наличие или отсутствие того или иного признака гораздо важнее, чем его точные координаты. Возможно «обобщить» выделяемые признаки, потеряв часть информации об их местоположении, но зато сократив размерность.

Обычно в качестве операции субдискретизации к каждой локальной группе нейронов применяется операция взятия максимума (max-pooling).

После операции субдискретизации может следовать этап регуляризации. Очень важным подводным камнем глубокого обучения является проблема переобучения (англ. overfitting). Отрицательный эффект переобучения заметно проявляется на сетях, но существует очень простой метод предотвращения такой проблемы.

Переобучение – это излишне точное соответствие нейронной сети конкретному набору обучающих примеров, при котором сеть теряет способность к обобщению. Другими словами, модель могла выучить обучающее множество (вместе с шумом, который в нем присутствует), но не смогла распознать скрытые процессы, которые это множество породили.

У глубоких свёрточных нейронных сетей масса разнообразных параметров. Переобучение может проявить себя в следующей форме: если при обучении сети присутствует недостаточное количество обучающих примеров, маленькая группа нейронов может стать ответственной за большинство вычислений, а остальные нейроны станут избыточны; или наоборот, некоторые нейроны могут нанести ущерб производительности, при этом другие нейроны из слоя не будут заниматься ничем, кроме исправления ошибок первых.

Чтобы сеть не утратила способности к обобщению, используют приемы регуляризации: вместо сокращения количества параметров накладываются ограничения на параметры модели во время обучения, не позволяя нейронам изучать шум обучающих данных. Одним из важнейших методов регуляризации нейронных сетей для революции глубокого обучения стал метод dropout («выбрасывание»).

Его идея чрезвычайно проста. Для каждого нейрона существует некоторая вероятность, с которой он будет выброшен из сети. Алгоритм обучения меняется таким образом: на каждом новом тренировочном примере сначала для каждого нейрона вычисляется такая вероятность, и затем, в зависимости от результата, выходное значение нейрона либо устанавливается в обычном порядке, либо устанавливается строго равным нулю. Нулевое значение на выходе приводит к тому, что нейрон фактически выпадает из графа вычислений: и прямое вычисление, и обратное распространение градиента останавливаются на этом нейроне и дальше не идут.

Решение поставленной задачи распознавания крон деревьев на спутниковых изображениях использует алгоритм раннего останова обучения и так называемый «обратный вызов» для скорости обучения в качестве методов обеспечения оптимальной тренировки сети. Ранняя остановка – это метод, который позволяет указать произвольно большое количество периодов обучения и прекратить обучение, как только производительность модели перестает улучшаться.

Что касается скорости, модели часто выигрывают от снижения скорости обучения в 2–10 раз после застоя обучения. Обратный вызов, предоставляемый API для обучения сетей, способен контролировать, не наблюдается ли улучшения для количества «терпения» эпох, и, в случае необходимости, скорость обучения снижается [7].

Для обучения сверточных нейронных сетей используется алгоритм обратного распространения ошибки, адаптированный к архитектуре сверточной сети. Целью обучения является минимизация суммарной квадратичной ошибки сети, которая характеризует разницу между реальными и эталонными выходными значениями сети. Значение суммарной квадратичной ошибки для L тренировочных наборов определяется следующим образом:

$$E_s = \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^m (y_j^k - e_j^k)^2, \quad (1)$$

где E_s – суммарная квадратичная ошибка;

L – количество тренировочных наборов;

m – количество нейронов в слое;

y_j^k – реальное выходное значение j -го нейрона выходного слоя

для k -го образа,

e_j^k – эталонное выходное значение j -го нейрона выходного

слоя для k -го образа.

Тогда согласно методу градиентного спуска для минимизации суммарной квадратичной ошибки сети синаптические связи ij -го нейрона карты признаков сверточного слоя в случае группового обучения (mini-batch) должны изменяться следующим образом:

$$w_{cij}(t+1) = w_{cij}(t) - \alpha \frac{\partial E(r)}{\partial w_{cij}(t)}, \quad (2)$$

где α – скорость обучения;

$E(r)$ – суммарная квадратичная ошибка для группы размером r -образов;

w_{cij} – весовой коэффициент между i -м и j -м нейроном соответственно предыдущего и сверточного слоя.

Поскольку нейроны каждой карты признаков сверточного слоя имеют одинаковые синаптические связи (веса и пороги), то частная

производная $\frac{\partial E(r)}{\partial w_{cij}(t)}$ определяется как сумма частных производ-

ных всех нейронов соответствующей карты признаков.

Таким образом, для вычисления определенного весового коэффициента необходимо взять производные по этой связи для всех нейронных элементов карты признаков и просуммировать их. Затем согласно выражению (2) полученное значение весового коэффициента присвоить одноименным связям для всех нейронных элементов карты признаков сверточного слоя.

Итоговое выражение для модификации весовых коэффициентов имеет следующий вид (пояснения некоторых элементов выражения приведены в формулах (1) и (2)):

$$w_{cij}(t+1) = w_{cij}(t) - \alpha(t) \sum_{ij} \sum_{k=1}^r \gamma_{ij}^k F'(S_{ij}^k) x_c^k,$$

где F' – производная функции активации для k -го образа;

S_{ij}^k – взвешенная сумма нейрона с номером ij в карте признаков;

γ_{ij}^k – ошибка ij -го нейрона для k -го образа [8].

На основе приведенных данных была выбрана следующая конфигурация сети U-Net: сжатие начинается с повторного применения двух сверток 3×3 , за которыми следуют слой batch-нормализации пачки изображений и ReLU активация, а также операции нелинейного преобразования (слой pooling) 2×2 с шагом 2 для понижающей дискретизации. На каждом шаге понижающей дискретизации удваивается количество функциональных каналов.

Расширение заключается в повышающей дискретизации карты объектов, за которой следует свертка 2×2 («свертка вверх»), которая

двое сокращает число каналов объектов, две 3×3 свертки, за которыми следуют batch-нормализация и ReLU. Целью этого расширения является точная локализация в сочетании с контекстной информацией из сжимающего пути.

На последнем слое свертка 1×1 используется для сопоставления каждого компонента с использованием сигмоидной функции активации.

Для анализа и сравнения было решено также исследовать другой вид сети – SegNet. Архитектура сети SegNet слабо отличается от ранее рассмотренной архитектуры сети U-Net по входящим в ее состав операциям. Самым существенным различием является наличие операции под названием upsampling в декодирующем слое, которая заключается в простом масштабировании изображения путём изменения его размера.

3. Процесс переноса знаний – Transfer learning. Экспериментальные результаты. Для решения поставленной задачи распознавания крон деревьев на спутниковых изображениях используем предобученную архитектуру сети U-Net.

Далее начинается процесс переноса знаний – Transfer learning. Создается модель сети с классификационными слоями и выходным слоем всей сети с сигмоидной функцией активации. Подаются на вход полученные ранее данные, и начинается обучение классификатора.

На вход сети поступают изображения - тайлы. Для более точной разметки исходное изображение представлено массивом только зеленого канала цветовой схемы RGB. Длина и ширина картинки равняется 256 пикселей; единица означает, что из всего изображения берется только составляющая зеленого канала (так как большинство крон деревьев зеленого цвета). Весь набор фотографий делится на тренировочную, валидационную, тестовую выборки, соотношение 70 – 10 – 20 %.

Далее объединяется сверточная часть нейронной сети с классификационной частью (см. рисунок 5), загружаются соответствующие веса и выполняется дообучение всей модели на небольшом количестве эпох для завершения процесса Transfer learning.

Для сравнения и анализа результатов использованы три архитектуры нейронных сетей:

- U-Net (kernel 3);
- U-Net (kernel 5);
- SegNet.

В таблице 1 приведено количество тренируемых и нетренируемых параметров сетей.

Таблица 1 – Параметры ИНС

Модель сети	Общее количество параметров	Тренируемые параметры	Нетренируемые параметры
U-Net (3)	2 164 305	2 161 361	2944
U-Net (5)	1 501 609	1 500 137	1472
SegNet	1 848 677	1 844 707	3970

Быстрее всего обучилась сеть U-Net с ядром свертки 5×5 , а медленнее всего такая же сеть с ядром 3×3 . Нетрудно заметить прямо пропорциональную зависимость времени обучения от количества тренируемых параметров сети.

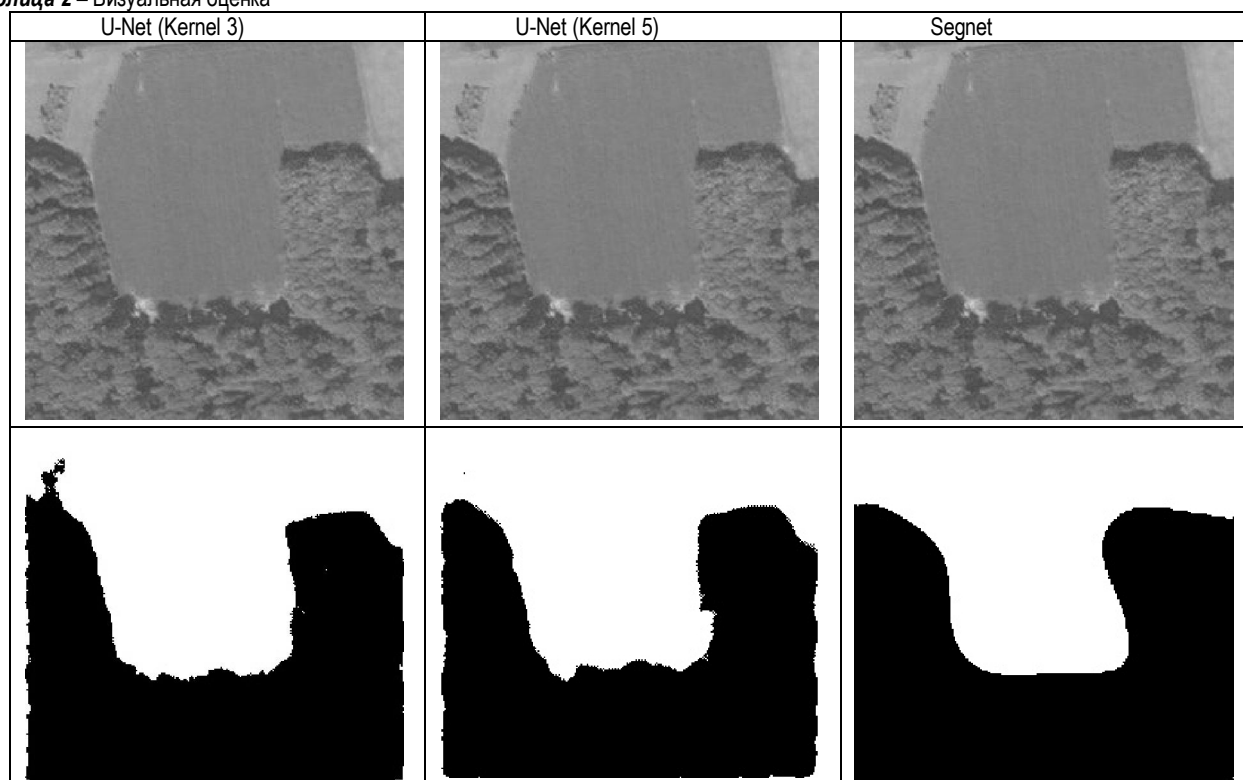
Результаты также можно проанализировать путём визуальной оценки изображений. В таблице 2 (см. стр. 23) хорошо видно, что первая сеть видит более детально, а последняя сеть делает границы слишком плавными.

В таблице 3 приводятся метрики сетей.

Таблица 3 – Метрики ИНС

Модель	Точность на обучающей выборке	Точность на валидационной выборке	Точность на тестовой выборке
U-Net (3)	0.917318	0.943089	0.945859
U-Net (5)	0.887473	0.940471	0.936079
SegNet	0.904689	0.925778	0.927248

Таблица 2 – Визуальная оценка



На рисунке 6 продемонстрированы графики изменения метрик во время обучения искусственных нейронных сетей.

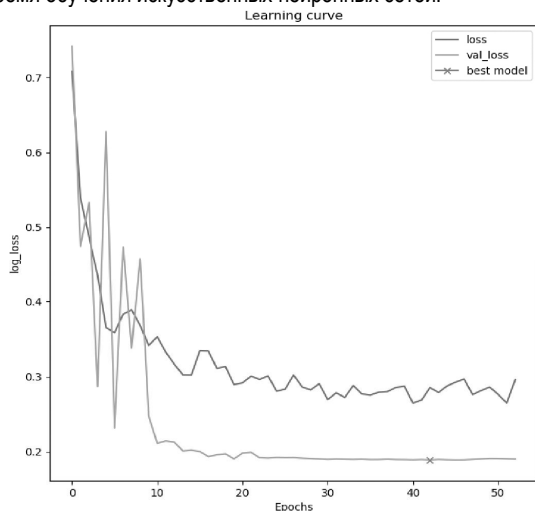


Рисунок 6 – График метрик для U-Net(3), Segnet

Исходя из приведенных выше данных, можно сделать вывод, что выбранная нейросетевая архитектура U-Net справляется с задачей распознавания немного лучше, чем архитектура SegNet. Обе конфигурации сети U-Net показали достаточно хорошие результаты.

Заключение. Поставленная задача является довольно узкоспециализированной, так как в качестве объекта распознавания выбраны лесные насаждения. В первую очередь, решение поставленной задачи позволит существенно упростить процесс разметки карт для картографов, сократить время на обработку данных, существенно уменьшить физический объем данных.

В данной статье предложен подход для решения задачи распознавания крон деревьев на спутниковых снимках. Исходя из экспериментальных данных, можно сделать вывод, что выбранная нейросетевая архитектура U-Net(3) показывает достаточно хорошие результаты для решения задачи распознавания крон деревьев на спутниковых снимках.

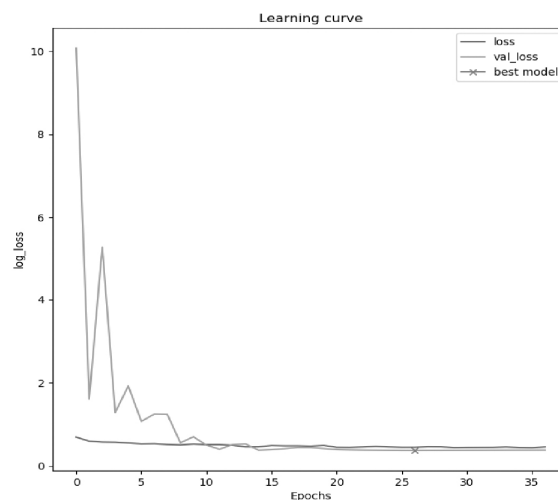


Рисунок 6 – График метрик для U-Net(3), Segnet

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Компьютерное зрение [Электронный ресурс]. – Режим доступа : https://ru.wikipedia.org/wiki/Компьютерное_зрение. – Дата доступа : 15.05.2019.
2. Зенин, А. В. Анализ методов распознавания образов // Молодой ученый. – 2017. – № 16. – С. 125-130. [Электронный ресурс]. – Режим доступа : <https://moluch.ru/archive/150/42393/>. – Дата доступа : 15.05.2019.
3. Dstl Satellite Imagery Feature Detection [Электронный ресурс]. – Режим доступа : <https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection>. – Дата доступа : 10.05.2019.
4. SegNet [Электронный ресурс]. – Режим доступа : https://quc.it.com/keras-segnet-implementation-of-segnet_en/. – Дата доступа : 20.05.2019.
5. Москалев, Н. С. Виды архитектур нейронных сетей // Молодой ученый. – 2016. – № 29. – С. 30-34. – [Электронный ресурс]. – Режим доступа : <https://moluch.ru/archive/133/37121/>. – Дата доступа : 19.05.2019.

-
6. [Электронный ресурс]. – Режим доступа : <http://www.deeplearning.net/tutorial/unet.html>. – Дата доступа : 17.05.2019.
7. Николенко, С. И. Глубокое обучение / С. И. Николенко, А. А. Ка-
дурин, Е. О. Архангельская – СПб. : Питер, 2018.
8. Головки, В. А. Нейросетевые технологии обработки данных :
учеб. пособие / В. А. Головки, В. В. Краснопоршин. – Минск :
БГУ, 2017.

31.01.2020

KHATSKEVICH M. V., GOLOVKO V. A., MIKHNYAEV L. P. Recognition of tree crowns on image using neural network methods

The article proposes the approach to solving the problem of tree crown recognition in satellite imagery. The main components of the architecture of U-Net neural network are given. The process of solving the problem of tree crown recognition in satellite imagery using U-Net neural network architecture. Based on experimental data, it can be concluded that the selected U-Net neural network architecture shows good enough results for solving the problem of tree crown recognition in satellite images.