

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы
направления 02.03.02 — Фундаментальная информатика и информационные
технологии
факультета КНиИТ
Кузнецова Ярослава Алексеевича

Научный руководитель
доцент, к. ф.-м. н.

А. А. Кузнецов

Заведующий кафедрой
к. ф.-м. н.

С. В. Миронов

ВВЕДЕНИЕ

Задача сегментации изображений заключается в поиске и выделении пикселей имеющих принадлежность к какому-либо описанному классу объектов. В статистике эта проблема известна как кластерный анализ и является широко изученной областью с сотнями различных алгоритмов. Сегментация применяется во многих областях, например, в производстве для индикации дефектов при сборке деталей, в медицине для первичной обработки снимков, также для составления карт местности по снимкам со спутников. В компьютерном зрении сегментация изображения является одной из старейших и широко изучаемых проблем.

Актуальность выбранной темы дипломной работы неоспоримо важна. В настоящее время обучение нейронных сетей для обработки данных и автоматизации работы с ними обрело новые масштабы. С развитием технологий начали появляться новые архитектурные решения и ресурсы для выполнения поставленных задач. Это дало возможность обрабатывать огромное количество данных в считанные секунды, что позволило значительно упростить жизнь людей.

Предмет исследования. В ходе написания бакалаврской работы были рассмотрены основные стандартные задачи машинного обучения. Показаны различные архитектуры нейронных сетей для решения задач анализа данных и компьютерного зрения. Была обучена и продемонстрирована работа нескольких обученных моделей нейронной сети на изображениях разного разрешения. Показана важность контроля обучаемой сети для избежания проблем недоученной и переобученной моделью во время ее предсказаний.

Цель бакалаврской работы — обучение модели для сегментации изображений при помощи библиотеки глубокого машинного обучения PyTorch на языке Python. Демонстрация важности контроля обучаемой нейронной сети.

Для достижения данной цели были определены **следующие задачи**:

- подготовка набора данных для обучения модели сегментации;
- обучение нейронной сети на тренировочных данных;
- демонстрация работы обученной модели для тестовом наборе данных;
- демонстрация погрешности предсказаний на недоученных и переобученных моделях.

Методологические основы распознавания образов представлены в кни-

ге В. Н. Ватника [1], архитектура нейронной сети U-Net в статье О. Ронненберга [2] и использование машинного обучения при помощи языка Python в книге С. Рашка [3].

Практическая значимость бакалаврской работы. В ходе этой работы была обучена нейронная сеть для сегментации изображений решающая соревновательную задачу Carvana Image Masking Challenge представленную сайтом Kaggle.com в 2017 году. Суть соревнования заключалась в создании алгоритма, который без вмешательства человека вырезал изображение автомобиля по предсказанной маске для дальнейшего его накладывания на самые разные фоны.

Структура и объем работы. Бакалаврская работа состоит из введения, двух разделов, заключения, списка использованных источников и трех приложений. Общий объем работы — 46 страниц, из них 34 страницы — основное содержание, включая 20 рисунков, цифровой носитель в качестве приложения, список использованных источников информации — 23 наименования.

1 Сегментация изображений с помощью нейронных сетей

Машинное обучение [1] — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Различают два типа обучения. Обучение по прецедентам, или индуктивное обучение, основано на выявлении общих закономерностей по частным эмпирическим данным. Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Дедуктивное обучение принято относить к области экспертных систем, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами.

В работе были описаны следующие варианты обучения нейронных сетей:

- обучение с учителем — наиболее распространенный случай. Каждый прецедент представляет собой пару «объект, ответ». Требуется найти функциональную зависимость ответов от описаний объектов и построить алгоритм, принимающий на входе описание объекта и выдающий на выходе ответ. Функционал качества обычно определяется как средняя ошибка ответов, выданных алгоритмом, по всем объектам выборки;
- обучение без учителя — в этом случае ответы не задаются, и требуется искать зависимости между объектами;
- частичное обучение — занимает промежуточное положение между обучением с учителем и без учителя. Каждый прецедент представляет собой пару «объект, ответ», но ответы известны только на части прецедентов. Пример прикладной задачи — автоматическая рубрикация большого количества текстов при условии, что некоторые из них уже отнесены к каким-то рубрикам;
- трансдуктивное обучение. Дана конечная обучающая выборка прецедентов. Требуется по этим частным данным сделать предсказания относительно других частных данных — тестовой выборки. В отличие от стандартной постановки, здесь не требуется выявлять общую закономерность, поскольку известно, что новых тестовых прецедентов не будет. С другой стороны, появляется возможность улучшить качество предсказаний за счет анализа всей тестовой выборки целиком, например, путем её кластеризации. Во многих приложениях трансдуктивное обучение практически не отличается от частичного обучения;

- обучение с подкреплением. Роль объектов играют пары «ситуация, принятое решение», ответами являются значения функционала качества, характеризующего правильность принятых решений (реакцию среды). Как и в задачах прогнозирования, здесь существенную роль играет фактор времени. Примеры прикладных задач: формирование инвестиционных стратегий, автоматическое управление технологическими процессами, самообучение роботов, и т.д;
- динамическое обучение может быть как обучением с учителем, так и без учителя. Специфика в том, что прецеденты поступают потоком. Требуется немедленно принимать решение по каждому прецеденту и одновременно доучивать модель зависимости с учетом новых прецедентов. Как и в задачах прогнозирования, здесь существенную роль играет фактор времени;
- активное обучение отличается тем, что обучаемый имеет возможность самостоятельно назначать следующий прецедент, который станет известен;
- мета-обучение отличается тем, что прецедентами являются ранее решенные задачи обучения. Требуется определить, какие из используемых в них эвристик работают более эффективно. Конечная цель — обеспечить постоянное автоматическое совершенствование алгоритма обучения с течением времени.

В следующем подразделе была описана искусственная нейронная сеть (ИНС) — вычислительная нелинейная модель, в основе которой лежит нейронная структура мозга, способная обучаться выполнению задач классификации, предсказания, принятия решений, визуализации и некоторых других только благодаря рассмотрению примеров.

Описан процесс обучения нейронной сети — процесс оптимизации весов, в котором минимизируется ошибка предсказания, и сеть достигает требуемого уровня точности. Наиболее используемый метод для определения вклада в ошибку каждого нейрона — обратное распространение ошибки, с помощью которого вычисляют градиент. Это одна из модификаций метода градиентного спуска.

Далее в работе представлены и описаны подразделы с различными архитектурами нейронных сетей, такие как:

- многослойный перцептрон — использует нелинейную функцию активации, часто тангенциальную или логистическую, которая позволяет классифицировать линейно неразделимые данные. Каждый узел в слое соединен с каждым узлом в последующем слое, что делает сеть полностью связанной;
- сверточная нейронная сеть — содержит один или более объединенных или соединенных сверточных слоев;
- рекурсивная нейронная сеть — тип глубокой нейронной сети, сформированный при применении одних и тех же наборов весов рекурсивно над структурой, чтобы сделать скалярное или структурированное предсказание над входной структурой переменного размера через активацию структуры в топологическом порядке;
- рекуррентная нейронная сеть — вид нейронных сетей, где связи между элементами образуют направленную последовательность;
- сеть долгой краткосрочной памяти — разновидность архитектуры рекуррентной нейросети, созданная для более точного моделирования временных последовательностей и их долгосрочных зависимостей, чем традиционная рекуррентная сеть;
- sequence-to-sequence модель — одна из наиболее популярных архитектур используемая в машинном переводе;
- неглубокие нейронные сети — в неглубоких нейронных сетях используется один скрытый слой.

Следующим подразделом является формулировка задачи сегментации изображений [4] — задача заключается в поиске и выделении пикселей имеющих принадлежность к какому-либо описанному классу объектов. Сегментация изображений применяется в различных областях, таких как, биомедицинских исследованиях, робототехника, картография и другие.

Обычно современные задачи компьютерного зрения разделяют на четыре вида:

- Classification — классификация изображения по типу объекта, которое оно содержит;
- Semantic segmentation — определение всех пикселей объектов определённого класса или фона на изображении. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от

друга;

- Object detection — обнаружение всех объектов указанных классов и определение охватывающей рамки для каждого из них;
- Instance segmentation — определение пикселей, принадлежащих каждому объекту каждого класса по отдельности.

В данной работе мы занимались instance segmentation и создавали попиксельную маску искомых объектов. Для решения поставленной задачи было решено использовать нейронную сеть U-Net.

U-Net [2] — это сверточная нейронная сеть, которая была создана в 2015 году для сегментации биомедицинских изображений в отделении Computer Science Фрайбургского университета. Архитектура сети представляет собой полносвязную сверточную сеть, модифицированную так, чтобы она могла работать с меньшим количеством примеров (обучающих образов) и делала более точную сегментацию.

U-Net [5] считается одной из стандартных архитектур CNN для задач сегментации изображений, когда нужно не только определить класс изображения целиком, но и сегментировать его области по классу, т.е. создать маску, которая будет разделять изображение на несколько классов. Архитектура состоит из стягивающего пути для захвата контекста и симметричного расширяющегося пути, который позволяет осуществить точную локализацию.

Сеть обучается сквозным способом на небольшом количестве изображений и превосходит предыдущий наилучший метод (сверточную сеть со скользящим окном) на соревновании ISBI по сегментации нейронных структур в электронно-микроскопических стеках. Используя ту же сеть, которая была обучена на изображениях световой микроскопии пропускания (фазовый контраст и DIC), U-Net заняла первое место в конкурсе ISBI 2015 года по трекингу клеток в этих категориях с большим отрывом. Кроме того, эта сеть работает быстро. Сегментация изображения 512×512 занимает менее секунды на современном графическом процессоре.

Для U-Net характерно:

- достижение высоких результатов в различных реальных задачах, особенно для биомедицинских приложений;
- использование небольшого количества данных для достижения хороших результатов.

2 Разработка приложения

Для решения поставленной цели в качестве языка программирования был выбран язык Python [3] — высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода. Так же были подключены сторонние библиотеки, список которых можно видеть ниже.

Используемые библиотеки:

- PyTorch 1.1;
- Segmentation models;
- CUDA Toolkit 10.1;
- Albumentations;
- Matplotlib.

PyTorch — библиотека машинного обучения для языка Python с открытым исходным кодом, созданная на базе Torch. Используется для решения различных задач: компьютерное зрение, обработка естественного языка.

Segmentation models [6] — библиотека с готовыми нейронными сетями для сегментирования изображений написанная на основе готовых инструментов PyTorch [7].

Albumentations [8] — быстрая и мощная библиотека для преобразования изображений, использованная для трансформаций над изображениями.

Библиотека matplotlib [9] — это библиотека двумерной графики для языка программирования Python с помощью которой можно создавать высококачественные рисунки различных форматов. Matplotlib представляет собой модуль-пакет для python.

Для того что бы обучение нейронной сети происходило в разумных временных рамках было решено подключить ресурсы видеокарты. Для этого нам понадобилась видеокарта от производителя Nvidia и специальный набор инструментов для нее Cuda Toolkit 10.1 [10]. В ходе текущей практики была использована следующая видеокарта NVIDIA GeForce GTX 1060 6gb, что позволило существенно сократить время обучения нейронной сети.

Обучение нейронной сети происходило в 2 этапа:

- 1) обучение на изображениях с разрешением 256×256 пикселей;
- 2) обучение на изображениях с разрешением 512×512 пикселей.

После чего было получено 2 обученных модели и сравнен результат их работы.

Дальше нам нужно было выбрать и подготовить данные для обучения нейронной сети. В качестве тренировочных и тестовых данных было решено использовать один из известных соревновательных датасетов Carvana Image Masking Challenge [11] представленный сайтом Kaggle.com в 2017 году. Следующим этапом были написаны функции с помощью которых выполнялось разбиение полученных данных для удобной работы с ними.

Далее были объявлены все параметры и константы для обучения нейронной сети, описаны все важные формулировки и определения, хорошо задокументирован разработанный исходный код приложения.

После всех подготовок было произведено обучение нейронных сетей в ходе которого производился контроль над их обучением, поэтому во избежание проблем с переобучением моделей были определены дополнительные метрики с помощью которых производился контроль над обучаемыми нейронными сетями. Обучение всех нейронных сетей происходило в 50 эпох.

Как говорилось ранее, обучение производилось в 2 этапа, поэтому после обучения была отдельно сохранена обученная модель на изображениях с разрешением 256×256 пикселей и внесены соответствующие поправки в коде для обучения модели на изображениях с разрешением 512×512 пикселей. Обучения нейронной сети на изображениях размером 256×256 пикселей составило около 10 часов, обучение на изображениях 512×512 пикселей приблизительно 22 часа. Разница в качестве двух обученных сетей составила около 1.9 раза в лучшую сторону.

После обучения нейронной сети на выход было получено 4 готовых модели с помощью которых можно делать предсказание масок на тестовых данных. Две из них являются моделями обученными на изображениях 256×256 пикселей, где одна из них с наилучшим результатом на этапе валидации и вторая переобученная. Остальные две модели обучены на изображениях 512×512 пикселей соответственно. Далее было продемонстрирована работа данных моделей на тестовых данных и был сделан вывод, что нейронная сеть обученная на изображениях большего разрешения показала лучший результат.

Для того что бы показать проблемы недоученных моделей, было заново запущено обучение нейронных сетей на изображениях 256×256 и 512×512 пикселей, обучение которых было прервано после первой эпохи. На основе полученных данных был сделан вывод, что у модели обученной на разрешении

256 × 256 пикселей окантовка пограничных пикселей предсказана более грубо, нежели чем у нормально обученной модели. Модель обученная на изображениях 512 × 512 пикселей с этим справилась намного лучше, но у нее появился другой дефект — связанный с ошибочным предсказанием области не принадлежащим к искомому объекту.

Ранее было сказано о важности контроля обучения нейронной сети для избежания ее переобучения, поэтому мы изменили загружаемую модель на полученную после последней эпохи обучения и посмотрели как влияет качество сети в предсказаниях на одном и том же изображении. На основе данных полученных по предсказанным маскам на переобученных моделях можно увидеть неверно предсказанные пиксели находящиеся за пределами автомобиля. Это еще раз доказывает нам необходимость следить и контролировать качество обучаемой нейронной сети.

ЗАКЛЮЧЕНИЕ

С развитием технологий и увеличения количества данных возникает необходимость автоматизировать их обработку. Для этих задач активно используются компьютерное зрение и нейронные сети. В данной работе был показан пример решения задачи сегментации изображений для автоматического выделения искомого объекта на изображении.

В настоящей работе были исследованы возможности библиотеки глубокого машинного PyTorch с помощью которой была обучена нейронная сеть для автоматической сегментации изображений и создания попиксельной маски для них. Рассмотрено отличие работы обученной нейронной сети на изображениях разного разрешения.

В ходе работы были решены следующие задачи:

- подготовлен набор данных для последующего обучения нейронной сети;
- обучена модель нейронной сети для автоматической сегментации изображения;
- продемонстрирована работа обученной нейронной сети на тестовом наборе данных;
- продемонстрирована погрешность предсказаний на недоученной и переобученной модели сегментации.

Таким образом, были выполнены все поставленные задачи и, тем самым, достигнута заданная цель.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Ватник, В. Н.* Теория распознавания образов. / В. Н. Ватник, А. Черво-ненкис. — Наука, 1974.
- 2 *Ronneberger, O.* U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // *ArXiv e-prints*. — 2015. — URL: <https://arxiv.org/abs/1809.06839>.
- 3 *Raschka, S.* Python Machine Learning / S. Raschka. — Packt Publishing Ltd, 2015.
- 4 Сегментация изображений [Электронный ресурс]. — URL: http://neerc.ifmo.ru/wiki/index.php?title=%D0%A1%D0%B5%D0%B3%D0%BC%D0%B5%D0%BD%D1%82%D0%B0%D1%86%D0%B8%D1%8F_%D0%B8%D0%B7%D0%BE%D0%B1%D1%80%D0%B0%D0%B6%D0%B5%D0%BD%D0%B8%D0%B9 (Дата обращения 22.04.2019) Загл. с экр. Яз. рус.
- 5 U-Net: нейросеть для сегментации изображений [Электронный ресурс]. — 2018. — URL: <https://neurohive.io/ru/vidy-nejrosetej/u-net-image-segmentation/> (Дата обращения 03.04.2019) Загл. с экр. Яз. рус.
- 6 Segmentation models [Электронный ресурс]. — URL: https://github.com/qubvel/segmentation_models.pytorch (Дата обращения 09.05.2019) Загл. с экр. Яз. англ.
- 7 PyTorch documentation [Электронный ресурс]. — URL: <https://pytorch.org/docs/stable/index.html> (Дата обращения 15.04.2019) Загл. с экр. Яз. англ.
- 8 *Buslaev, A.* Albumentations: fast and flexible image augmentations / A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, A. A. Kalinin // *ArXiv e-prints*. — 2018. — URL: <https://arxiv.org/abs/1809.06839>.
- 9 Matplotlib 3.1.0 documentation [Электронный ресурс]. — URL: https://matplotlib.org/3.1.0/api/_as_gen/matplotlib.pyplot.html (Дата обращения 22.05.2019) Загл. с экр. Яз. англ.
- 10 CUDA Toolkit Documentation v10.1.168 [Электронный ресурс]. — URL: <https://docs.nvidia.com/cuda/> (Дата обращения 15.04.2019) Загл. с экр. Яз. англ.

- 11 Segmentation models (Automatically identify the boundaries of the car in an image) [Электронный ресурс]. — 2017. — URL: <https://www.kaggle.com/c/carvana-image-masking-challenge> (Дата обращения 12.05.2019) Загл. с экр. Яз. англ.