

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 004.932.2

Латушкин
Никита Сергеевич

Программные средства для автоматического анализа видеоряда

АВТОРЕФЕРАТ

диссертации на соискание степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель
Сиротко С.И.
к.физ-мат.н., доцент

Минск 2020

Работа выполнена на кафедре информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **Сиротко Сергей Иванович**,
кандидат физико-математических наук, доцент
кафедры информатики учреждения образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Рецензент: **ШЕШОЛКО Владимир Константинович**,
кандидат физико-математических наук, доцент
кафедры управления информационными
ресурсами учреждения образования «Академия
управления при Президенте РБ»

Защита диссертации состоится «23» июня 2020 г. года в 10⁰⁰ часов на заседании Государственной экзаменационной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул. Гикало, 9, копр. 4, ауд. 112, тел. 293-85-91, e-mail: inform@bsuir.by

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

ВВЕДЕНИЕ

В последние годы человечество достигло значительного прогресса в области информационных технологий. Значительное увеличение мощности вычислительной техники в том числе позволило добиться прорыва в области глубинного обучения, так как глубокие нейронные сети требуют огромных вычислительных ресурсов, что в свою очередь привело к значительному прогрессу в решении задач из области компьютерного зрения.

Технический прогресс позволяет собирать и накапливать большое количество информации, которое необходимо обрабатывать и анализировать. Всё это требует большого количества времени, ресурсов и работы человека. Поэтому задачей большого количества программного обеспечения является автоматизация обработки информации, а также упрощение работы с ней. Большое количество таких задач находятся в области компьютерного зрения.

Так, например, даже в такой сложной области, как медицина, где знаний обычного человека недостаточно, современные нейросетевые алгоритмы показывают неплохие результаты и позволяют значительно упростить и ускорить работу врачей.

Ещё одной сферой применения компьютерного зрения является создание автоматических систем безопасности и идентификации. Такие системы основаны на распознавании и идентификации лиц, и ими пользуется огромное количество людей в повседневной жизни. В качестве примера можно привести систему идентификации при разблокировке мобильного телефона, созданную компанией Apple; анализ и группировка изображений, созданная Google, для упрощения управления и систематизации фотографий.

Подобные решения также могут автоматизировать рутинный человеческий труд. Так, автоматизируется конвейерная сборка с использованием роботов: им передаётся информация о положении объекта, которая получена с помощью методов компьютерного зрения, путём анализа положения объекта относительно камеры.

Методы компьютерного зрения также находят активное применение в частичной или полной автоматизации управления автомобилем. В свою очередь частичная автоматизация позволяет водителю уменьшить количество аварийных ситуаций. Задача полной автоматизации управления автомобилем на сегодняшний день является приоритетной для многих крупных компаний, в первую очередь это актуально для пассажироперевозок.

Таким образом, нейросетевые методы и компьютерное зрение в частности находят широкое применение в различных сферах жизни человека, как

повседневных, так и узкоспециализированных. На данный момент технологический прогресс позволяет создавать продукты, направленные на автоматизацию рутинной работы для уменьшения количества ресурсов и времени, которое необходимо затратить для её выполнения.

Библиотека БГУИР

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цели и задачи исследования

Целью данной работы является разработка системы для автоматического отслеживания лиц на видео. Для функционирования такой системы, она должна содержать несколько основных компонент:

- обнаружение лиц;
- распознавание лиц;
- трекер объектов;
- модуль анализа информации;

Первый модуль необходим для обнаружения всех лиц на кадрах, а также их последующего сопоставления с людьми. В качестве решения данной задачи был выбран подход на основе методов глубинного обучения, а именно был использована нейросетевая модель обнаружения. В качестве входа этот модуль получает по очереди каждый кадр из видео. В результате работы этого модуля для каждого кадра получаем координаты рамок обрамляющих лица, найденных на текущем кадре. В качестве метрики, оценивающей качество этого модуля используется mean average precision.

Распознавание лиц используется для агрегации рамок, полученных в результате работы модуля обнаружения, в коллекции, внутри которых содержатся рамки одного и того же человека. Для этих целей так же было выбрано решение на основе нейронной сети. В качестве входных данных используем изображения, которые вырезаны по координатам рамок, полученных из модуля обнаружения. Эти изображения приводятся к одному размеру и выравниваются. Результатом работы этого модуля является набор векторов представлений, соответствующих каждой из поданных на вход рамок. Данные вектора обладают следующим свойством: если посчитать метрику косинусной схожести для двух векторов, то у лиц, принадлежащих одному человеку, она будет высокой, а у разных – низкой.

Трекинг лиц используется для устранения минусов работы модулей обнаружения и распознавания лиц. Иногда модуль обнаружения лиц может ошибаться и не находить все лица на отдельном кадре, тогда в случае наличия этого человека на других кадрах видео, с помощью трекинга мы имеем возможность найти его на всех кадрах, на которых он присутствует. Для решения этой задачи используется модель, на основе нейронной сети.

Модуль анализа информации используется для улучшения точности работы предыдущих модулей, а также для объединения полученной информации.

Итогом работы системы является файл в формате json. Этот файл содержит полную информацию о каждом из персонажей, встреченных в видео, а именно: уникальный номер персонажа, а также информация о всех его появлениях на кадрах в виде координат рамок его лица на этих кадрах.

Объектом исследования является разработка системы, способной автоматически анализировать видео и выдавать аннотацию к нему.

Предметом исследования являются методы обнаружения и распознавания лиц, трекинга объектов с использованием нейронных сетей, а также алгоритмы анализа полученных данных.

Описанная задача имеет существующие решения. Примером такого решения является FastRedaction. Данное программное обеспечение позволяет находить и отслеживать лица на видео. Это решение имеет ряд минусов. Прежде всего это низкая скорость обработки видео: ее длительность в 4-5 раз превышает продолжительность видео. Также Fastredaction имеет большое количество ложноотрицательных результатов, т.е. данное решение находит не все лица, присутствующие на кадре. Особенно актуальна эта проблема для видео с плохим качеством съёмки, а также решение плохо показывает себя для обнаружения лиц маленького размера. По этой причине данное решение не может надёжно выполнять поставленные в этой работе цели. Помимо описанных проблем, это программное обеспечение является дорогостоящим, создание экземпляра работы приложения стоит 19 долларов, обработка каждой минуты видео добавляет к стоимости ещё 1 доллар. В связи с описанными выше проблемами данного решения, разработка более быстрого и качественного решения является актуальной задачей.

Область исследования

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-40 81 04-2020 специальности 1-40 81 04 – «Обработка больших объемов информации».

Теоретическая и методологическая основа исследования

В основу диссертации легли работы зарубежных ученых в области анализа изображений, а также анализ онлайн-сервисов и технической документации по рассматриваемой тематике.

Информационная база исследования сформирована на основе литературы, открытой информации, технических нормативно-правовых актов, сведений из электронных ресурсов, а также материалов научных конференций и семинаров.

Научная новизна

Научная новизна и значимость полученных результатов работы заключается в разработке алгоритмов, способных получать из видео полную информацию о лицах, встреченных в видео, а именно: уникальный номер лица, а также информация о всех его появлениях на кадрах в виде координат рамок.

Теоретическая значимость работы заключается в детальном анализе существующих алгоритмов обработки изображений с использованием методов глубокого обучения, разработке алгоритма позволяющего компенсировать недостатки распознавания лиц, а также разработке алгоритма позволяющих создать единую систему из отдельных нейронных сетей.

Практическая значимость диссертации заключается в разработке программного средства для автоматического анализа видеоряда.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 56 научной конференции аспирантов, магистрантов и студентов БГУИР.

Работа внедрена в ЗАО “Оксаджайл”.

Опубликованность результатов диссертации

По теме диссертации опубликованы 2 печатные работы в сборнике научной конференции аспирантов, магистрантов и студентов БГУИР.

Структура и объем диссертации

Диссертация состоит из общей характеристики работы, введения, трёх глав, заключения, списка использованных источников.

В первой главе представлен анализ предметной области.

Вторая глава дает краткую характеристику используемых технологий и описывает причины, по которым они были выбраны.

В третьей главе приведены сведения о ходе выполнения работы и детальное описание всех модулей системы.

Общий объем работы составляет 78 страниц, из которых основного текста — 60 страниц, 30 рисунков на 27 страницах, список использованных источников из 39 наименований на 3 страницах и 1 приложение на 19 страницах.

Библиотека БГУИР

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** были рассмотрены актуальные задачи решаемые нейросетевыми подходами. Были приведены конкретные примеры внедрения глубинного обучения в продукты различных компаний. Также описано обоснование актуальности темы.

В **общей характеристике работы** сформулированы цели и задачи диссертации, обозначена область исследований, научная и практическая значимость. Приведено количество опубликованных работ по теме диссертации, а также описана ее структура.

В **первой** главе был описан математический аппарат, необходимый для использования нейронных сетей, так как основной задачей является анализ видеоряда, а при работе с изображениями свёрточные нейронный сети показывают себя в качестве наиболее эффективного подхода для извлечения информации из них. Описаны слои нейронных сетей, функции активаций, а также подходы для предобработки данных и борьбы с переобучением. Также была описана архитектура глубокой нейронной сети, которая используется в различных модулях систем для извлечения информации из кадров видео.

Во **второй** главе были представлены краткие сведения об основных и наиболее важных технологиях используемых при разработке системы. Для разработки системы был выбран язык Python. Numpy является основной библиотекой для хранения данных и операций над ними, производимых на процессоре. OpenCV является библиотекой, с помощью которой производятся все необходимые операции с видео и изображениями. PyTorch основной фреймворк для создания и запуска нейронных сетей, также для удобства само обучение нейронных сетей происходило в Jupyter Notebook. Для быстрых вычислений система использует графические ускорители, и таким образом CUDA и cuDNN являются необходимыми технологиями для ускорения работы нейронных сетей.

В **третьей** главе были описаны все модули, необходимые для разработки системы: обнаружение лиц, распознавание лиц, трекинг объектов, а также алгоритм их общей интеграции в единую систему, способный решать поставленную задачу анализа видеоряда. Созданное программное средство было протестировано на большом объеме тестового материала и было внедрено в реальную эксплуатацию.

ЗАКЛЮЧЕНИЕ

Основные научные результаты

Предложена эффективная архитектура для решения задачи автоматического анализа видеоряда, обеспечивающая качественное объединение и анализ данных модулей системы. Ее особенности позволяют не простаивать ресурсам вычислительной машины. Также, в первую очередь, она разрабатывалась под особенности решения с наличием графического ускорителя. Хотя без наличия графического ускорителя система будет выдавать не менее качественное решение, ее работа значительно замедляется, это вызвано неэффективностью вычислений нейронных сетей на процессоре по сравнению с вычислениями на графическом ускорителе. Именно по этой причине, архитектура была выбрана с ориентацией на наличие графического ускорителя.

В работе было предложено решение обнаружения лиц на кадре с использованием алгоритма RetinaFace. Это решение хорошо справляется со своей задачей и способно найти большинство лиц на кадре даже в трудных условиях, а наличие публичных тестов подтвердило высокое качество этого решения.

Для идентификации лиц был реализован подход называющийся ArcFace. Эффективность его реализации и обучения показывают довольно высокие показатели на публичном тесте. Также решение было доработано для выдачи показателя качества изображения лица, что является необходимым условием для хорошей работы системы.

Предложенный трекер и алгоритм анализа данных из модулей позволили нивелировать ошибки и недочеты работы предыдущих алгоритмов, что значительно повысило качество итогового решения.

Рекомендации по практическому использованию результатов

Итоговое решение способно, получая на вход видео, создавать к нему аннотацию в виде json файла. Аннотация включает в себя отслеживание лиц, которые присутствуют в данном видео. Аннотация сделана таким образом, чтобы была возможность ее легко интегрировать в специальные видеоредакторы, способные исправлять и валидировать автоматическую разметку.

Также система представляет собой интерес с той стороны, что алгоритмы трека объектов, анализа и объединения данных довольно универсальны и способны в дальнейшем быть использованы для отслеживания любых объектов, алгоритмы обнаружения и идентификации которых, схожи с алгоритмами обнаружения и идентификации лиц.

Система показывает себя крайне эффективно с точки зрения времени обработки видео при запуске ее на следующей конфигурации: Ubuntu в качестве операционной системы; графический ускоритель Nvidia RTX 2080; Intel Core i5-8700. Данная аппаратная часть и операционная система позволяет решению обрабатывать видео за время сопоставимое с его продолжительностью.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1-А. Латушкин Н. С. Отслеживание лиц на видео с использованием классических и нейросетевых подходов в компьютерном зрении / Н. С. Латушкин // Компьютерные системы и сети: 56-я научная конференция аспирантов, магистрантов и студентов, Минск / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020 (В печати).

2-А. Латушкин Н. С. Распознавание лиц на изображениях плохого качества с использованием нейросетевых алгоритмов / Н.С. Латушкин, Н.Ю. Сухов, А.В. Козак // Компьютерные системы и сети: 56-я научная конференция аспирантов, магистрантов и студентов, Минск / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2020 (В печати).