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

УДК 004.93:004.032.26

Курбанов
Сейитджан Сердарович

Алгоритм распознавания текста на изображениях посредством нейронных сетей

АВТОРЕФЕРАТ
на соискание степени магистра
по специальности 1-40 80 04 – Информатика и технологии программирования

Научный руководитель
Боброва Н.Л.
к.т.н., доцент

Минск 2024

КРАТКОЕ ВВЕДЕНИЕ

В современном обществе наши жизненные сцены полны различной текстовой информации. Текст с конкретной и четкой семантикой является чрезвычайно важным обобщением, описанием и выражением для реальных сцен. Детектирование текста реальной сцены является ключевой технологией для реализации интеллектуального восприятия сцены, имеет важные исследовательское значение. Однако из-за сложного и разнообразного фона, неоднородных текстовых шрифтов, несовместимых размеров и неопределенных направлений текста в реальных сценах текущая обработка этой задачи не достигла идеальных результатов.

Выбор данной темы обусловлен стремлением к разработке и улучшению методов компьютерного зрения и нейронных сетей для эффективного распознавания текста на изображениях. Распознавание текста на изображениях является сложной задачей из-за разнообразия шрифтов, размеров, ориентации и искажений, которые могут присутствовать на изображениях.

Диссертационная работа посвящена разработке и оптимизации алгоритма распознавания текста на изображениях с использованием нейронных сетей. Используя глубокое обучение и подходы, основанные на сверточных нейронных сетях, стремится создать модель, способную автоматически извлекать и распознавать текст из разнообразных визуальных контекстов. Результаты этого исследования будут иметь как научное, так и практическое значение, внося вклад в развитие области компьютерного зрения и обработки текста на изображениях.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цели и задачи исследования

Целью диссертационной работы является разработка и анализ алгоритма распознавания текста на изображениях с использованием нейронных сетей. Основная задача заключается в создании эффективного и точного метода, способного автоматически обнаруживать и распознавать текст на сложных графических сценах.

Для достижения поставленной цели необходимо решить следующие задачи:

1 Провести исследование и изучение различных методов распознавания текста на изображениях, включая как классические подходы, так и современные методы, основанные на нейронных сетях. Это позволит определить преимущества и недостатки существующих решений и выявить наиболее перспективные подходы.

2 Создать архитектуру нейронной сети, которая будет способна локализовать и распознавать текст на изображениях сложных графических сцен. Это может включать комбинацию сверточных и рекуррентных слоев, а также использование алгоритма CTC-loss для обучения сети без необходимости выравнивания между входными и выходными последовательностями.

3 Собрать и предварительно обработать набор данных, содержащий изображения с различными типами текста. Также может потребоваться аугментация данных для обеспечения разнообразия и улучшения обобщающей способности модели.

4 Обучить разработанную архитектуру нейронной сети на подготовленном наборе данных, применяя методы оптимизации и регуляризации для достижения высокой точности распознавания текста. Подобрать оптимальные параметры обучения, такие как размер входных изображений и выбор функции потерь.

Провести эксперименты на тестовых наборах данных и оценить точность распознавания текста, а также другие метрики производительности модели, такие как скорость обработки и устойчивость к шуму и искажениям. Сравнить полученные результаты с другими существующими методами и алгоритмами распознавания текста..

Объектом исследования являются алгоритмы распознавания текста с помощью нейронных сетей..

Предметом исследования является Разработка и оптимизация алгоритмов распознавания текста на изображениях с использованием нейронных сетей.

Основной *гипотезой*, положенной в основу диссертационной работы, является Внедрение глубоких нейронных сетей с архитектурами CNN и RNN в алгоритмы распознавания текста повысит точность и эффективность процесса распознавания по сравнению с традиционными методами. Глубокие нейронные

сети, такие как сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN), проявили себя в различных задачах обработки изображений и обработки последовательностей, включая распознавание текста. Использование сверточных слоев в CNN позволяет модели извлекать локальные признаки из изображений, в то время как рекуррентные слои в RNN могут улавливать контекстуальные зависимости в последовательностях символов.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н. Л. Боброва, заключается в формулировке целей и задач исследования.

Публикации результатов диссертации

По теме диссертации опубликовано 4 печатных работ в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена архитектуре CRNN (Convolutional Recurrent Neural Network). Разработка и реализация модифицированной версии CRNN для задачи распознавания текста на изображениях. Обоснование выбора данного подхода и его преимущества.. В третьей главе обоснование выбора конкретного метода для решения задачи распознавания текста на изображениях. Описание разработанного алгоритма, основанного на выбранном методе. Реализация алгоритма и описание используемых инструментов и технологий. Эксперименты и результаты, полученные с использованием разработанного алгоритма.

Общий объем работы составляет 57 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ

Введение является вступительным разделом работы, в котором осуществляется обоснование выбора темы исследования, описание мотивации исследователя, а также постановка целей и задач работы. Введение создает контекст и подчеркивает значимость выбранной темы, объясняя причины, по которым она является актуальной и важной для исследования. Кроме того, во введении определяются цели исследования, а также формулируются конкретные задачи, направленные на достижение данных целей.

Первая глава представляет собой аналитический обзор академических и научно-популярных источников, связанных с темой исследования. Глава начинается с анализа архитектурных решений, используемых при решении задач распознавания текста. Затем рассматривается роль компьютерного зрения в этой области. Далее проводится обзор методов распознавания текста на сложных графических сценах.

Основное внимание уделяется методам детектирования текста. В главе рассматриваются традиционные методы детектирования, а также методы, основанные на глубоком обучении. Кроме того, рассматривается вопрос локализации текста на изображениях. Глава завершается описанием методов распознавания текста.

Целью первой главы является предоставление обзора существующих методов и подходов к детектированию и распознаванию текста на изображениях.

Вторая глава посвящена подходу, основанному на применении CRNN-архитектуры (Convolutional Recurrent Neural Network) для детектирования и распознавания текста на изображениях. В рамках данной главы проводится обзор CRNN-архитектуры и ее компонентов.

CRNN-архитектура объединяет сверточные и рекуррентные слои для обработки изображений с последующим распознаванием текста. В разделе "Обзор CRNN-архитектуры" рассматривается структура и функциональность каждого компонента. Полносвязный слой играет важную роль в CRNN-архитектуре, и его функциональность подробно изучается. Сверточные слои применяются для извлечения признаков из входных изображений, а слои субдискретизации используются для уменьшения размерности признаковых карт после сверточных операций.

Слой нормализации по мини-батчам в CRNN-архитектуре способствует стабилизации обучения и повышению производительности модели. Рекуррентные слои, такие как LSTM (Long Short-Term Memory), используются для моделирования последовательностей и контекста при распознавании текста.

Важным аспектом CRNN-архитектуры является применение CTC (Connectionist Temporal Classification) loss функции. Она позволяет обучать модель на задаче распознавания текста, учитывая переменную длину последовательностей и возможные перестановки символов.

Цель второй главы состоит в том, чтобы предоставить обзор основных компонентов CRNN-архитектуры и объяснить их роль в детектировании и распознавании текста на изображениях. Этот подход обладает преимуществами в области обработки текста на изображениях и находит широкое применение в различных приложениях, таких как оптическое распознавание символов (OCR), автоматическое распознавание номерных знаков и других сценариях, требующих извлечения текста из визуальных данных.

Третья глава посвящена выбору метода и реализации алгоритма для детектирования и распознавания текста на изображениях. В главе рассматриваются различные методы и их реализация с использованием определенных инструментов разработки.

В разделе "Используемые инструменты разработки" приводится обзор инструментов, выбранных для реализации алгоритмов. Рассматриваются используемые библиотеки и фреймворки, такие как TensorFlow и PyTorch, а также другие инструменты, необходимые для разработки и обучения моделей. Затем в разделе "Реализация модели FCN" описывается процесс реализации алгоритма FCN (Fully Convolutional Network) для детектирования текста на изображениях. Рассматривается структура модели и ее особенности.

В разделе "Реализация модели U-net" представлена реализация алгоритма U-net, который также используется для детектирования текста на изображениях. Объясняется архитектура U-net и рассматриваются детали реализации. В разделе "Реализация модели EAST" описывается процесс реализации алгоритма EAST (Efficient and Accurate Scene Text detection), который широко применяется для детектирования текста на сложных графических сценах. Рассматривается структура модели и его реализация с использованием выбранных инструментов.

Далее в разделе "Сравнение результатов распознавания текста" проводится сравнительный анализ результатов распознавания текста с использованием различных реализованных моделей. Оцениваются показатели точности и производительности каждого метода.

В разделе "Реализация финального метода" представлена финальная реализация выбранного метода, основанного на комбинации оптимальных компонентов и алгоритмов, выявленных в предыдущих разделах.

В конце главы, в разделе "Ветка распознавания" рассматриваются дополнительные аспекты распознавания текста, такие как поворот прямоугольника для коррекции наклона текста. Затем описывается реализация и результаты точности алгоритма CRNN+CTC-loss, который используется для распознавания текста.

Цель третьей главы заключается в выборе оптимального метода и реализации алгоритма для детектирования и распознавания текста на изображениях. В главе представлены различные модели, их реализация и результаты сравнения, а также реализация финального метода.

ЗАКЛЮЧЕНИЕ

В заключение, данная работа представляет обзор и анализ существующих методов локализации и распознавания текста на изображениях сложных графических сцен. Был предложен новый подход, основанный на комбинации рекуррентных и сверточных нейронных сетей, а также алгоритма CTC-loss. Разработанный метод детектирования и распознавания надписей на изображениях реальных сцен включает в себя выбор параметров обучения нейронной сети, таких как размер входного изображения, замена базовой сети и выбор функции потерь.

Для реализации данной архитектуры нейронных сетей был использован язык программирования Python с применением библиотеки Tensorflow. Проведены эксперименты на двух наборах данных, и полученные результаты были анализированы и представлены в виде графиков изменения функции потерь, расстояния Левенштейна и точности распознавания на тренировочном и двух тестовых наборах данных.

Сравнение результатов данной работы с другими исследованиями показало, что разработанная система достигает результатов, близких к state-of-the-art решениям, при этом требуя меньших вычислительных затрат. Это стало возможным благодаря использованию дополнительных слоев нормализации по мини-батчам.

В целом, представленный подход к распознаванию локализованного текста на изображениях сложных графических сцен является перспективным и может быть применен в различных областях, таких как компьютерное зрение, автоматическое распознавание документов и разработка систем анализа изображений. Дальнейшие исследования могут быть направлены на улучшение точности и эффективности данной системы, а также на ее адаптацию для работы с другими типами данных и задачами распознавания текста.

Была составлена таблица с результатами точности распознавания текста, где было произведено сравнение двух способов распознавания: с использованием дополнительного словаря фиксированного размера и без его использования. На наборе данных SVT была достигнута точность 76.2% без словаря и 95.3% с использованием словаря размером в 50 слов. На наборе данных IIIT5K точность составила 74.0% без словаря, 96.7% с использованием словаря размером в 50 слов и 93.2% со словарем размером в 1000 слов.

Достижения:

1. Высокая точность детектирования: Модель EAST (Efficient and Accurate Scene Text Detection) имеет доказанную эффективность в обнаружении текста на сложных графических сценах. Ее использование в комбинации с CRNN (Convolutional Recurrent Neural Network) и CTC-loss (Connectionist Temporal Classification loss) позволяет достичь высокой точности детектирования текста на изображениях.

2. Хорошая устойчивость к изменениям в условиях окружения: Модель CRNN+CTC-loss+East может успешно справляться с некоторыми вызовами, такими как вариации освещения, фоновый шум и искажения текста, которые могут быть характерны для набора данных III5Tk SVT.

3. Поддержка контекстуального распознавания: CRNN-модель, используемая в этой комбинации, позволяет учитывать контекст и последовательно обрабатывать извлеченные признаки. Это способствует более точному распознаванию текста на изображениях из набора данных III5Tk SVT.

Проблемы:

1. Ограниченное разнообразие данных: Набор данных III5Tk SVT может быть относительно ограничен в терминах разнообразия сцен, шрифтов, языков и других вариаций текста. Это может привести к недостаточной обобщающей способности модели и снижению ее производительности на реальных уличных сценах, отличных от тех, что представлены в наборе данных.

2. Проблемы с шумом и низким качеством изображения: Набор данных III5Tk SVT может содержать изображения с шумом и низким разрешением, что может затруднить распознавание текста. Модель CRNN+CTC-loss+East может испытывать трудности в таких условиях, что может привести к снижению точности и общей производительности.

3. Ограниченная обучающая выборка: Если набор данных III5Tk SVT является относительно небольшим, это может ограничивать способность модели эффективно обучаться и достигать высокой точности. Недостаточное количество размеченных данных может привести к переобучению или недообучению модели.

4. Необходимость оптимизации процесса обучения и инференса: Процесс обучения и инференса модели может быть длительным и требовательным к ресурсам. Оптимизация этих процессов может быть ключевым аспектом для успешного использования модели в реальных приложениях.

Учитывая эти достижения и проблемы, важно применять методы предобработки данных, аугментации и обучения, чтобы улучшить производительность модели CRNN+CTC-loss+East с набором данных III5Tk SVT и справиться с вызовами, связанными с этим набором данных.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Курбанов С. С., Боброва Н.Л., Распознавание текста на изображениях посредством нейронных сетей / С. С. Курбанов, Н.Л. Боброва // Электронные системы и технологии: сборник материалов 60-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 2024 г. / Белорусский государственный университет информатики и радиоэлектроники.

2. Курбанов С. С., Боброва Н.Л., Распознавание и анализ текста с помощью нейронных сетей / С.С. Курбанов, Н.Л. Боброва // Электронные системы и технологии: сборник материалов Международная научно-методическая конференция «Инженерное образование в цифровом обществе», Минск, 14 март 2024 г.: в 2 ч. / Белорусский государственный университет информатики и радиоэлектроники.

3. Курбанов С. С., Боброва Н.Л., Локализация текстовых областей на изображениях с использованием сверточной нейронной сети / С.С. Курбанов, Н.Л. Боброва // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник научный статей X Международной научно-практической конференции, Минск, 13 марта 2024 года. / Белорусский государственный университет информатики и радиоэлектроники

4. Курбанов С. С., Использование нейронных сетей в образовательной сфере / С. С. Курбанов // Электронные системы и технологии: сборник материалов Международная научно-методическая конференция «Инженерное образование в цифровом обществе», Минск, 14 март 2024 г.: в 1 ч. / Белорусский государственный университет информатики и радиоэлектроники.