|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ | | | | | | | | | | | | | | | | | | | | | | |
| **ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ** | | | | | | | | | | | | | | | | | | | | | | |
| **УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ** | | | | | | | | | | | | | | | | | | | | | | |
| **«СЕВАСТОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»** | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| **ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ** | | | | | | | | | | | | | | | | | | | | | | | |
| *(название института полностью)* | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| Кафедра/департамент | | | | | «Информационные системы» | | | | | | | | | | | | | | | | | | |
| *( наименование кафедры/департамента полностью)*  09.03.02 | | | | | | | | | | | | | | | | | | | | | | | |
| *(код и наименование направления подготовки/специальности)* | | | | | | | | | | | | | | | | | | | | | | | |
| *(наименование профиля/специальности)* | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| **КУРСОВАЯ РАБОТА / КУРСОВОЙ ПРОЕКТ** | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| по дисциплине | | | | | | | | | | | | | | | | | | | | | | |
| **Практикум по цифровому проектированию** | | | | | | | | | | | | | | | | | | | | | | |
| *(наименование дисциплины)* | | | | | | | | | | | | | | | | | | | | | | |
| на тему | | | Распознавание автомобильных номеров с помощью | | | | | | | | | | | | | | | | | | | |
| глубоких нейронных сетей | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| Выполнили: обучающиеся | | | | | | | | | | | | | | | | | | | | | |
| группы: | | | | | | | | | | | | | ИС/б-21-2-о, ПИ/б-21-1-о | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | |
| Мовенко К. М.  Ларин А. М. | | | | | | | | | | | | | | | | | | | | | |
| *(инициалы, фамилия)* | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | |
| « | | | | | | | | | |  | » | |  | | 20 | | 24 | | г. | | |
|  | | | | | | | | | | | | | | | | | | | | | |
| Научный руководитель: | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | |
| *(инициалы, фамилия)* | | | | | | | | | | | | | | | | | | | | | |
| « | | | | | | | | | |  | | » | |  | | 20 | | 24 | | г. | | |
|  | | | | | | | | | | | | | | | | | | | | | |
| Оценка | | |  | | | | | | |  | | | | | | | | | | | |
|  | | | | | | | | | |  | | | | | | | | | | | |
| « |  | » | |  | | 20 | 24 | г. |
|  | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| Севастополь | | | | | | | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | 24 |  | | | | | | | | | |

# АННОТАЦИЯ

В данном проекте рассматривается проблема распознавания автомобильных номеров по фотографии. Автомобильные номера состоят из ограниченного набора символов, расположенных в случайном порядке, и обладают рядом характеристик: страна, регион, формат, стиль. Это делает задачу их распознавания сложной и интересной.

Для решения поставленной задачи будут применены технологии компьютерного зрения и машинного обучения, а именно свёрточные нейронные сети.

# ВВЕДЕНИЕ

Данный проект направлен на применение глубоких нейронных сетей для распознавания автомобильных номеров. Автомобильные номера представляют собой уникальные идентификаторы, имеющие различные форматы и стили в зависимости от страны или региона. Их распознавание и классификация с помощью искусственного интеллекта являются важной вехой в области обеспечения транспортной безопасности и анализа трафика в режиме реального времени.

На рисунке 1.1 представлен пример автомобильного номера (регистрационного знака) с подробным описанием размерностей, используемых обозначений и деталей. В зависимости региона знак может иметь немного иной формат, однако общая форма у всех знаков едина.



Рисунок 1.1 – Пример регистрационного знака (формат РФ)

Возможность точно распознать номер автомобиля по фото может быть полезна для нескольких сценариев применения. Например, при реализации автоматизированной системы управления дорожным движением, или реализации системы безопасности на охраняемом объекте. Применений технологии много, что свидетельствует об её актуальности.

# 1 АНАЛИЗ ЛИТЕРАТУРЫ И ПОСТАНОВКА ЗАДАЧИ

Распознавание автомобильных номеров последние 10 лет является одним из самых продаваемых решений на основе компьютерного зрения. Опыт последних лет говорит, что для решения данной (и множества других) задач наиболее надёжные и гибкие решения позволяют реализовать алгоритмы CNN (свёрточные нейронные сети).

Системы распознавания автомобильных номеров состоят из двух модулей: детектор (обнаружение номера на фото) и распознаватель символов (Optical Character Recognition, OCR). В рамках данной работы будем считать, что детектор реализован заранее, все фотографии номеров берутся в обрезанном виде.

Также важным допущением является то, что при обучении рассматриваются автомобильные номера лишь китайского. Это связано с выбором набора данных, подробнее в разделе 2.

Таким образом задача работы − самостоятельное обучение модели OCR для автомобильных номеров заданного типа.

# 2 ДАТАСЕТ И ПРЕДОБРАБОТКА

## 2.1 Множество данных

В качестве набора данных был выбран датасет CCPD2019 (Chinese City Parking Dataset). В нём содержится около 300 тысяч изображений китайских номерных знаков при различных условиях.

CCPD – наиболее популярный датасет для начала обучения нейросетей распознаванию автомобильных номеров. Это обусловлено крайне большим числом изображений и выделению из них подмножеств с различными условиями съёмки: погода, размытие, яркость, угол наклона и пр.

Также важное достоинство датасета – наличие изображений с выделенным «регионом интереса». Из фотографий автомобилей обрезано всё, что не является автомобильным номером. Благодаря этому в данном проекте можно пренебречь созданием детектора и сосредоточиться на распознавании текста.

## 2.2 Предобработка

Из исходного датасета были сделаны две выборки. В первой, train, содержатся изображения для обучения нейронной сети. Он большой (199980 файлов), фотографии в нём сделаны при различных условиях. Чёткого размера нет, но все фотографии представляют собой прямоугольники разрешением примерно 170/50.

Во втором наборе, test, хранится набор изображений для тестирования обученной нейросети. Изображений 9999 штук, все они взяты из подмножества CCPD-weather (фотографии в плохих погодных условиях).

## 2.3 Детали набора данных

Формат входных данных следующий: каждое изображение в CCPD имеет один номерной знак китайского типа (рисунок 2.1). Каждый знак состоит из: китайского иероглифа (символ провинции), латинской буквы и набора из пяти букв и цифр.



Рисунок 2.1 – Формат регистрационного знака (Китай)

К каждому номерному знаку в датасете прикреплён набор из 13 цифр. Это – индексы соответствующих символов в номере. Для иероглифов, первой буквы и 5-значного кода индексация берётся по соответствующим массивам (рисунок 3).

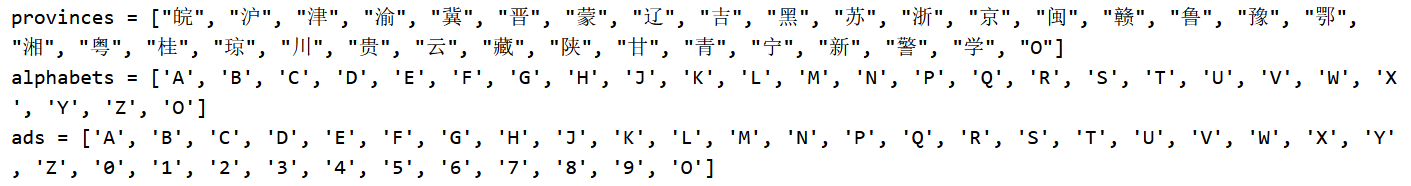


Рисунок 3 – Массивы для индексации символов

Последним символом каждого массива является буква O, которая используется как знак «нет символа», поскольку в китайских номерных знаках нет буквы O.

# 3 МЕТОДЫ

## 3.1 Архитектура модели

Основная проблема заключается в необходимости извлечь текстовую информацию из изображений, которые могут быть искажены, иметь шум или различный угол наклона. Требуется модель, которая способна эффективно и точно распознавать текст в таких условиях.

Для решения задачи была выбрана модель свёрточной рекуррентной нейронной сети (CRNN), состоящей из трёх компонент: свёрточные слои, рекуррентные слои и слой транскрипции (рисунок 3.1). Подобная модель является типичной для указанной проблемы.

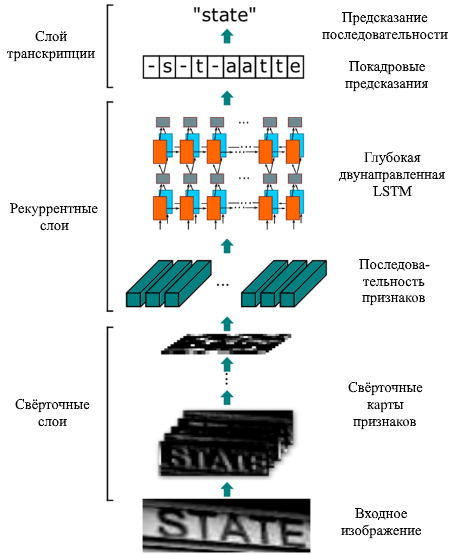


Рисунок 3.1 – Архитектура CRNN

Свёрточные слои обеспечивают надежное извлечение признаков. Рекуррентные слои учитывают контекст последовательности. CTC (метод для слоя транскрипции) эффективно выравнивает и декодирует последовательности, обеспечивая высокую точность распознавания.

## 3.2 Свёрточные слои

Свёрточные слои в указанной модели представляют собой экстрактор фичей. Они предназначены для автоматического выделения признаков из входных изображений. С помощью свёрточных операций они преобразуют изображения в набор признаков, которые затем используются для дальнейшей обработки.

Эти признаки представляют собой активированные выходные карты, которые выявляют различные особенности изображения, такие как: края, текстуры, узоры.

Для извлечения признаков из изображений используется последовательность из нескольких свёрточных блоков (CNNBlock). Эти блоки включают два свёрточных слоя, батч-нормализацию и активацию ReLU.

Свёртка производится через операцию Conv2D:

где — входное изображение, — ядро свёртки, — координаты пикселя.

Таким образом, экстрактор фичей позволяет выделить из изображений набор признаков в виде свёрточных карт, которые затем могут быть использованы для дальнейшей обработки в других нейросетевых архитектурах.

## 3.3 Рекуррентные слои

На рекуррентной части модели карты признаков от свёрточных слоев преобразуются в последовательный формат, подходящий для обработки рекуррентными нейронными сетями (RNN). Пространственная информация эффективно преобразуется во временную последовательность.

Не все данные, с которыми работают нейронные сети, являются независимыми. В номерном знаке при считывании элемента важно учитывать его порядок. Поэтому в качестве рекуррентного слоя использована двунаправленная LSTM (BI-LSTM).

LSTM (Long Short-Term Memory) представляет собой тип рекуррентной нейронной сети, разработанный для преодоления проблемы исчезающего градиента в обычных RNN. Он позволяет модели учиться на последовательных данных, учитывая их контекст, и сохранять информацию на долгосрочной основе.

Двунаправленные LSTM − это вариация стандартной LSTM, которая работает с данными как в прямом, так и в обратном направлении. Это позволяет модели учитывать контекст как слева направо, так и справа налево. При обучении модели оба прохода осуществляются параллельно, и результаты объединяются, что помогает модели лучше понимать контекст и зависимости в последовательностях символов на номерах.

После обработки данных LSTM генерирует скрытое состояние, которое содержит в себе информацию о всей последовательности. Это состояние используется для получения итогового результата – распознанные символы номерного знака.

## 3.4 Слой транскрипции

Слой транскрипции является важным элементом для решения задачи распознавания текста по фото. В контексте CRNN слои транскрипции выполняют функцию перевода выходных данных модели в текстовую последовательность.

Выход двунаправленной LSTM − это последовательность предсказаний, где для каждого символа модель выдает вероятность того, что он находится на данном кадре. Слой транскрипции сжимает эти признаки в пространство классов, соответствующих символам алфавита. Это позволяет модели давать предсказания о том, какие символы содержатся на изображении.

Обучение происходит с помощью CTC-LOSS (Connectionist Temporal Classification Loss) − методу обучения без учителя, который используется для задач распознавания последовательностей. В контексте CRNN, CTC-функция потерь помогает модели сравнивать символы в распознаваемом тексте с соответствующими фрагментами изображения, а также учитывать возможные пропуски.

Во время обучения модели с использованием CTC эта функция потерь сравнивает предсказанные последовательности с истинными метками, что позволяет модели корректировать свои параметры для улучшения точности распознавания.

Перед тем, как подавать предсказания в CTC, необходимо применить к ним активацию softmax и затем взять логарифм:

Где ​ − выходной сигнал модели для символа , а - общее количество символов в алфавите.

Формула для функции потерь CTC определяется следующим образом:

где − выходные предсказания модели, − множество всех допустимых последовательностей длины , − вероятность последовательности π при условии входной последовательности .

Таким образом, слои транскрипции и метод CTC вместе обеспечивают модели CRNN способность эффективно и точно распознавать текст на изображениях, учитывая изменчивость длины текста и различные искажения или шум на изображении.

## 3.5 Альтернативные подходы

Были рассмотрены следующие альтернативные подходы:

Традиционные методы компьютерного зрения: эти методы включают в себя использование классических алгоритмов и признаков, таких как SIFT (масштабно-инвариантная трансформация признаков) или HOG (гистограмма направленных градиентов), в сочетании с традиционными методами машинного обучения, например, SVM. Они позволяют выделить ключевые особенности изображения (углы, градиенты), однако имеют свои ограничения в случае сложных данных, где могут потерять информацию о важных деталях.

Чисто свёрточные сети, CNN: эти сети используются для автоматического извлечения признаков из изображений, чередуя свёрточные слои с полностью связанными слоями для классификации. Однако без рекуррентных сетей чистые CNN не учитывают последовательности данных или долгосрочные зависимости, что может снижать их эффективность при распознавании наборов символов.

# 4 ОЦЕНИВАНИЕ РЕЗУЛЬТАТОВ

## 4.1 Метрики для оценки качества модели

Для оценки качества модели использовались четыре метрики: потери, точность, Character Error Rate (CER) и расстояние Левенштейна.

Потери представляют собой числовое значение, которое указывает на «стоимость» ошибочных предсказаний модели. Чем выше потери, тем хуже модель выполняет свою задачу.

Точность (accuracy) – это доля правильно предсказанных примеров относительно всех предсказанных примеров, т.е. процент правильно распознанных изображений.

Если значение высоко, модель правильно распознает большую часть номеров. Низкое значение точности свидетельствует, что модель часто ошибается и требует более лучшего обучения.

Character Error Rate (CER) – это доля неправильно предсказанных символов к общему количеству символов. Эта метрика показывает, насколько точно модель распознает отдельные символы, из которых состоят номера.

Вставки, замены, удаления – операции над символами строк, с помощью которых можно из предсказанной строки получить эталонную (ожидаемое значение на выходе). Чем выше коэффициент, тем больше ошибок допускает модель и тем хуже её обучение.

Расстояние Левенштейна – это минимальное количество операций (вставка, удаление, замена), необходимых для перехода от предсказанной строки к эталонной. Чем больше расстояние, тем дальше предсказания модели от эталона и тем они хуже.

Также при обучении следует рассмотреть отличие тренировочных метрик от валидационных. Тренировочные метрики оценивают производительность модели на наборе данных, на котором модель обучается, и используются находятся многократно на каждой эпохе для оценки процесса обучения и корректировки параметров модели. Валидационные же в конце обучения оценивают производительность модели на отдельном наборе данных, который не используется для обучения, но представляет собой ту же область задачи.

## 4.2. Результаты обучения

Нейросеть была обучена со следующими параметрами обучения:

* количество эпох = 5;
* количество скрытых нейронов RNN-слоя = 256;
* количество выходных каналов CNN = 512;
* масштабирование изображений до (96, 448).

После последней эпохи были сняты результаты обучения (рисунок 4.1).

Тренировочные метрики таковы: потери от 0.01161 до 0.01208, расстояния Левенштейна от 0.01376 до 0.01546, значения CER от 0.001966 до 0.002209, точность от 0.985 до 0.9865.

Валидационные метрики таковы: потеря 0.01783, Левенштейна 0.03246, CER 0.004637, точность 0.9741.

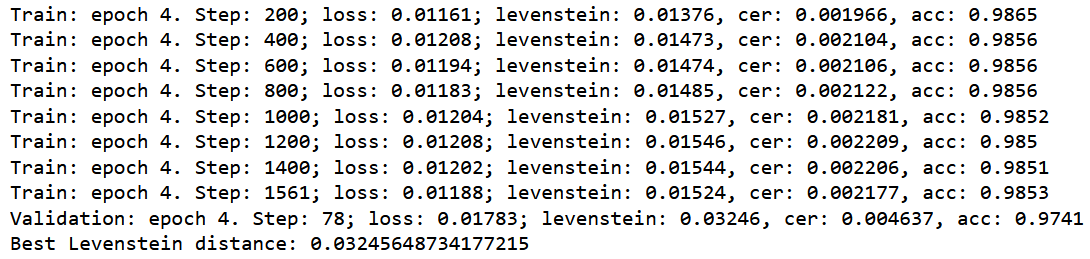


Рисунок 4.1 – Метрики на последней эпохе обучения

Модель показывает стабильно низкие значения потерь на тренировке, что указывает на хорошую сходимость. Тем не менее наблюдается разрыв: валидационные метрики несколько хуже наилучших значений, полученных в процессе тренировки. Это нормально, однако может свидетельствовать о небольшом переобучении.

Валидационная точность − 97%, потери − 1%. Это весьма хороший результат, но не идеальный. Потенциально модель может быть улучшена при увеличении числа эпох и параметров обучения.

## 4.3. Анализ ошибок

Для анализа ошибок и их причин была произведена фильтрация результатов работы модели на тестовом наборе изображений с целью выделить те случаи, в которых модель совершила ошибку.

Результаты показали, что, в процессе распознавания номеров по изображениям, модель совершила 247 ошибок. Были визуализированы ошибочные распознавания (рисунок 4.2).



Рисунок 4.2 – Примеры ошибочно распознанных номеров

Для статистики были выведены 20 символов, распознаваемых ошибочно наиболее часто (рисунок 4.3).

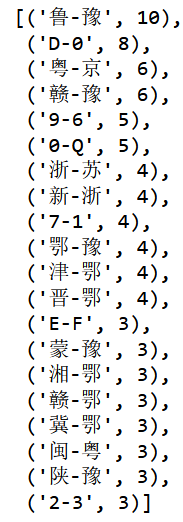


Рисунок 4.3 – Символы с наихудшим распознаванием

Из результатов видно, что наибольшее число ошибок модель допускает при распознавании сложных китайских иероглифов.

Также можно заметить ошибки в распознавании символов: D, 0, 9, 6, Q, E, F, 3, 7, 1. Есть смысл предположить, что модель периодически путает пары символов D-0-Q, E-F-3, однако таких ошибок немного. Возможные решения проблемы − увеличение размеров входных изображений, построение более точной модели свёрточной сети.

# ЗАКЛЮЧЕНИЕ

В рамках проекта удалось успешно обучить модель OCR для распознавания автомобильных номеров. Для этого была использована архитектура CRNN, включающая свёрточные слои, рекуррентные слои (BI-LSTM) и слой транскрипции с CTC-LOSS методом. Обучение проводилось на датасете CCPD2019.

Модель показала хорошие результаты. Валидационная точность составила 97%, потери – 1,7%. Анализ ошибок на тестовом датасете выявил 247 ошибок распознавания, связанных в основном с китайскими иероглифами и парами символов D-0-Q, E-F-3.

Для улучшения модели можно увеличить число эпох и параметры обучения, расширить датасет, включив номера из других регионов и условий съёмки, и экспериментировать с архитектурами моделей.

Также в перспективы развития проекта входит разработка детектора номерных знаков для того, чтобы определять местоположение автомобильного номера на более общих фотографиях. Потенциально модель может быть применена при интеграции с системами реального времени (система дорожной безопасности, охрана объектов и др.)

По результатам работы был сделан вывод, что свёрточные нейронные сети являются отличным средством решения задач визуального распознавания данных.

# СПИСОК ЛИТЕРАТУРЫ

1. Анирад К., Сиддха Г., Мехер К. Искусственный интеллект и компьютерное зрение. Реальные проекты на Python, Keras и TensorFlow. — СПб.: Питер, 2023 — 624 с.
2. «Нейронные сети в детектировании номеров». Habr, 29.05.2017, <https://habr.com/ru/companies/recognitor/articles/329636/> [Электронный ресурс];
3. Shi B., Bai X., Yao C. (2015). An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. arXiv: 1507.05717;
4. Клетте Р. Компьютерное зрение. Теория и алгоритмы / пер. с англ. А. А. Слинкин. – М.: ДМК Пресс, 2019. – 506 с.
5. Liu Y., Wang Y. (2023). A Convolutional Recurrent Neural-Network-Based Machine Learning for Scene Text Recognition Application. Symmetry, 15(4), 849.