

MODERN TEXT RECOGNITION MODELS AND METHODS: TECHNOLOGIES, CHALLENGES, AND PERSPECTIVES**Nabiyeva G.***Candidate of Technical Sciences, Associate Professor
Asfendiyarov Kazakh National Medical University***Nurzhanov G.***Master student**Almaty University of Power Engineering and Telecommunications named after G.Daukeev, Almaty, Kazakhstan***СОВРЕМЕННЫЕ МОДЕЛИ И МЕТОДЫ РАСПОЗНАВАНИЯ ТЕКСТА: ТЕХНОЛОГИИ, ВЫЗОВЫ И ПЕРСПЕКТИВЫ****Набиева Г.С.***К.т.н., ассоциированный профессор, Казахский Национальный Медицинский Университет имени
С.Д. Асфендиярова, Алматы, Казахстан***Нуржанов Г.Е.***Магистрант, Алматинский Университет Энергетики и Связи имени Г.Даукеева, Алматы, Казахстан*<https://doi.org/10.5281/zenodo.14499290>**Abstract**

The article focuses on modern methods and models of Optical Character Recognition (OCR). It reviews key approaches, including classical algorithms, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), their combinations (CRNN), and transformers. A comparative analysis of their advantages and limitations is presented, and an optimal approach is proposed for processing textual data in real-world conditions. The practical section demonstrates the efficiency of the CRNN method in text recognition tasks considering factors such as low image quality and complex backgrounds.

Аннотация

Статья посвящена современным методам и моделям оптического распознавания текста (OCR). Рассматриваются основные подходы, включая классические алгоритмы, сверточные нейронные сети (CNN), рекуррентные сети (RNN), их комбинации (CRNN), а также трансформеры. Выполнен сравнительный анализ их преимуществ и недостатков, предложен оптимальный подход для обработки текстовых данных в реальных условиях. Практическая часть демонстрирует эффективность метода CRNN для задач распознавания текста с учетом таких факторов, как низкое качество изображения и сложные фоны.

Keywords: OCR, text recognition, convolutional neural networks, recurrent networks, transformers, CRNN, image processing.

Ключевые слова: OCR, распознавание текста, сверточные нейронные сети, рекуррентные сети, трансформеры, CRNN, обработка изображений.

Введение. С развитием цифровых технологий и автоматизацией обработки данных распознавание текста Optical Character Recognition (OCR) стало важной частью информационных систем, которое часто применяется в бизнесе, медицине, архивном деле и других областях. Данные технологии позволяют конвертировать текст с изображений в машиночитаемый формат, обеспечивая быстрый доступ и высокую точность обработки данных, что особенно востребовано в задачах массового ввода документов, чеков и анкет.

С внедрением методов машинного обучения OCR вышел на качественно новый уровень - модели на основе нейронных сетей, включая сверточные нейронные сети (CNN) и рекуррентные сети (RNN), значительно улучшили точность распознавания, особенно при наличии разнообразия шрифтов, сложных фонов и искажений изображений. Но несмотря на прогресс, OCR сталкивается с рядом ограничений, таких как точное распознавание рукописного текста и слабоконтрастных изображений, а

также необходимость оптимизации моделей для устройств с ограниченными вычислительными возможностями.

Исследование направлено на анализ современных методов и моделей OCR, оценку их эффективности и определение оптимального подхода к распознаванию текста в различных условиях.

Теоретическая часть. Оптическое распознавание текста (OCR) — это технология, которая основывается на методах компьютерного зрения и машинного обучения и преобразует текстовую информацию с изображений в цифровой формат, обеспечивая возможность автоматической обработки и анализа данных [1][2]. Основу работы OCR-систем составляет последовательность процессов, включая предобработку изображения, выделение признаков и классификацию символов. Современные подходы к OCR можно разделить на традиционные методы, опирающиеся на правила и алгоритмы, и методы глубокого обучения, использующие нейронные сети [4].

Классические методы OCR базируются на детерминированных алгоритмах, таких как пороговая обработка, сегментация и извлечение признаков [2][3]. Эти методы успешно применяются для простых задач, например, для распознавания текста, напечатанного стандартными шрифтами на высококачественных изображениях. На этапе сегментации изображение делится на элементы — строки, слова и символы, а затем проводится их анализ. Несмотря на эффективность, классические подходы ограничены в случаях, когда текст включает искажения, сложные фоны или низкий контраст.

С появлением глубокого обучения OCR-системы стали значительно точнее. Сверточные нейронные сети (CNN) [5] широко используются для выделения визуальных признаков символов благодаря своей способности эффективно распознавать пространственные зависимости в изображениях. CNN автоматически выделяют значимые характеристики текста, такие как формы и контуры, что улучшает точность распознавания. Однако сверточные сети менее эффективны при работе с последовательными данными, такими как строки текста.

Для анализа последовательных структур текста применяются рекуррентные нейронные сети (RNN) [8], которые способны учитывать зависимости между символами, что особенно важно при распознавании слов и фраз. Модификация RNN, LSTM (долгая краткосрочная память), способна сохранять контекст на протяжении всей последовательности, что делает её подходящей для задач OCR, где последовательность символов важна для правильного распознавания текста.

Кроме того, современные модели OCR всё чаще включают трансформеры — архитектуры, которые объединяют возможности CNN и RNN для эффективного анализа как пространственных, так и последовательных данных. Например, Vision Transformer (ViT) [7] применяет механизмы внимания для выделения важных элементов изображения, что позволяет решать задачи OCR более эффективно. Трансформеры обеспечивают высокую точность, но требуют значительных вычислительных ресурсов, что ограничивает их использование в некоторых практических приложениях.

Постановка задачи. Целью является анализ существующих методов к распознаванию текста с учетом современных требований к качеству и ско-

рости обработки изображений. Задача состоит в выборе оптимальной системы OCR, способной с высокой точностью и стабильностью распознавать текстовые данные с изображений, независимо от условий их получения, также предполагает учет таких факторов, как разнообразие шрифтов, наличие искажений, низкое качество изображений и неравномерное освещение, что затрудняет на практике традиционное распознавание.

Для решения поставленной задачи необходимо:

- исследовать и проанализировать современные методы и модели OCR, включая подходы на основе сверточных нейронных сетей (CNN), рекуррентных сетей (RNN), а также моделей с использованием трансформеров;

- оценить преимущества и ограничения каждой модели, рассматривая их применимость к различным типам текстовых данных и условиям их обработки;

- протестировать методы на наборе данных, содержащем изображения с текстом и выявить оптимальный.

- Сравнить результаты с существующими OCR-системами, чтобы оценить достигнутые улучшения в точности и скорости распознавания.

Анализ существующих моделей и методов. Рассмотрение моделей и методов, используемых для распознавания текста, выявляет различные подходы, основанные на применении традиционных алгоритмов и современных нейронных сетей [4]. С каждым годом OCR-системы становятся все более совершенными, особенно с интеграцией глубокого обучения. Традиционные методы OCR, такие как пороговая обработка и сегментация, эффективно работают при наличии четкого текста и высококачественных изображений [3]. Однако эти методы ограничены в условиях низкого качества данных или наличия сложных фоновых помех. С другой стороны, глубокие нейронные сети (DNN), включая сверточные (CNN), рекуррентные сети (RNN) и трансформеры, открыли новые возможности для точного распознавания текста в условиях, где традиционные подходы оказались недостаточными [10].

Для более структурного анализа ниже представлена таблица 1 с описанием различных подходов, их преимуществ и недостатков.

Преимущество и недостатки существующих методов и моделей

Метод / Модель	Принцип работы	Преимущества	Недостатки
Классические методы	Основаны на алгоритмах пороговой обработки и сегментации символов.	Простота реализации; подходит для стандартного текста в высоком качестве.	Ограничены в условиях низкого качества, сложных фонов и искажений; требуют ручной настройки.
Сверточные нейронные сети (CNN)	Используют фильтры для выделения визуальных признаков (контуры, формы) текста.	Высокая точность выделения визуальных признаков; устойчивость к искажениям.	Ограничены при работе с последовательными данными; требуют значительных вычислительных ресурсов.
Рекуррентные нейронные сети (RNN)	Анализируют последовательные данные, учитывая контекст текста для распознавания последовательностей.	Способны учитывать контекст символов; подходят для строк текста и рукописного текста.	Ограничены в обработке пространственных зависимостей; могут быть медленными.
CRNN (Сочетание CNN и RNN)	Использует CNN для выделения признаков и RNN для последовательной обработки символов.	Комбинирует преимущества CNN и RNN; подходит для сложных текстовых структур.	Высокие вычислительные требования; может быть ограничен в скорости обработки.
Трансформеры	Используют механизм внимания для определения важных участков изображения и обработки последовательностей.	Способны учитывать пространственные и последовательные зависимости; высокая точность на сложных данных.	Очень высокие вычислительные затраты; требуют значительных ресурсов для обучения и настройки.

Описание и анализ моделей:

– *классические методы*:

классические подходы OCR основаны на алгоритмах пороговой обработки и сегментации, которые разбивают изображение на отдельные элементы и анализируют их отдельно [2][3]. Эти методы подходят для стандартного печатного текста на изображениях высокого качества и отличаются простотой реализации и меньшими вычислительными затратами. Однако они чувствительны к низкому качеству изображений и требуют ручной настройки, что ограничивает их применение в условиях, требующих гибкости;

– *сверточные нейронные сети (CNN)*:

CNN стали ключевым компонентом современных OCR-систем благодаря способности выделять визуальные признаки, такие как формы и контуры символов, автоматически. Модель CNN обрабатывает изображение через каскад слоев фильтров, которые обучаются выделять значимые характеристики текста, что позволяет системе устойчиво работать в условиях искажений [5][9]. Тем не менее, CNN ограничены при работе с последовательными данными, так как они не учитывают взаимосвязи между символами, что может быть критичным для распознавания целых слов и предложений;

– *рекуррентные нейронные сети (RNN)*:

RNN и их модификации, такие как LSTM и GRU, предназначены для работы с последовательными данными и способны учитывать контекст предыдущих символов, что делает их полезными при распознавании строк текста [6][8]. Они учитывают последовательные связи между символами, что позволяет повысить точность распознавания

целых слов и предложений. Однако RNN требуют значительных вычислительных ресурсов и времени на обработку и не подходят для анализа пространственных зависимостей в изображении;

– *CRNN (Сочетание CNN и RNN)*:

модель CRNN объединяет возможности CNN для выделения признаков и RNN для последовательного анализа текста, что делает её эффективной для задач OCR на сложных данных [10]. Этот подход позволяет одновременно учитывать визуальные и последовательные характеристики текста, что делает модель особенно полезной для распознавания рукописного текста и строк с нечеткими символами. Хотя CRNN обладает высокой точностью, модель требует значительных вычислительных ресурсов, что ограничивает её применение в реальном времени;

– *трансформеры*:

в последние годы трансформеры, включая Vision Transformer, продемонстрировали высокую точность в задачах OCR благодаря механизму внимания, который позволяет модели эффективно анализировать как пространственные, так и последовательные зависимости [4][7]. Эти модели способны обрабатывать сложные данные, такие как тексты с искажениями, рукописный и многоязычный текст. Однако высокие вычислительные требования и потребность в больших объемах данных для обучения делают трансформеры затратными в использовании.

Рассмотрим алгоритм работы каждого из представленных выше методов:

1. Классические методы:

- *предобработка* - бинаризация изображения, удаление шума;
- *сегментация* - разделение изображения на отдельные символы;
- *распознавание шаблонов* - использование алгоритмов вроде сопоставления шаблонов или гистограмм для определения символов;

- *постобработка* - исправление ошибок на основе словаря.

Для наглядности показана схема работы алгоритма классическим методом на рисунке 1 [2].

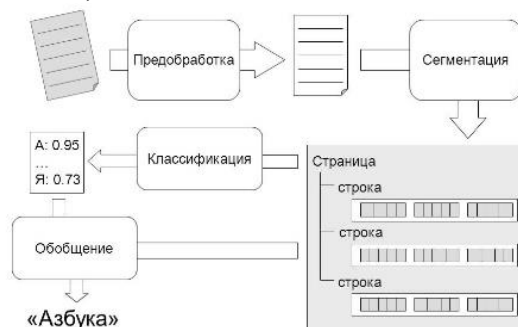


Рисунок 1. Схема алгоритма классического метода

2. Сверточные нейронные сети (CNN):

- *извлечение признаков* - использование сверточных слоев для выделения краев, форм и текстур символов;
- *классификация* - выходной слой классифицирует каждый фрагмент изображения как конкретный символ;

- *распознавание текста* - объединение всех символов в итоговый текст.

На рисунке 2 показана в пример последовательность, которая классифицирует рукописные цифры [5].

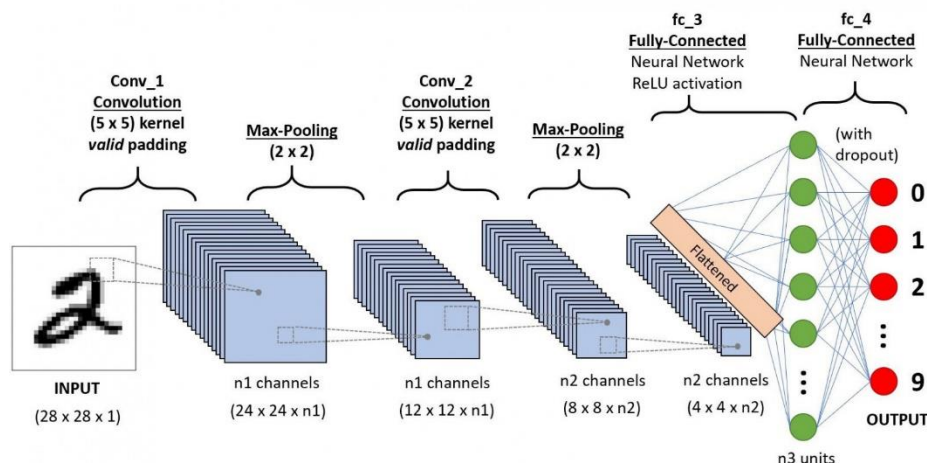


Рисунок 2. Последовательность CNN для классификации рукописных цифр

3. Рекуррентные нейронные сети (RNN):

- *извлечение последовательности* - преобразование изображения текста в последовательность символов или признаков;
- *обработка RNN* - обработка последовательных данных для анализа текста с учетом предыдущих символов;
- *декодирование текста* - генерация итогового текста на основе анализа RNN [6].

4. CRNN (Комбинация CNN и RNN):

- *CNN для выделения признаков* - преобразует изображение в компактную матрицу признаков;
- *RNN для последовательности* - анализирует последовательность признаков, учитывая контекст;
- *CTC-декодер* - использует Connectionist Temporal Classification для получения текста.

На рисунке 3 показана схема последовательности работы алгоритма CRNN [11].

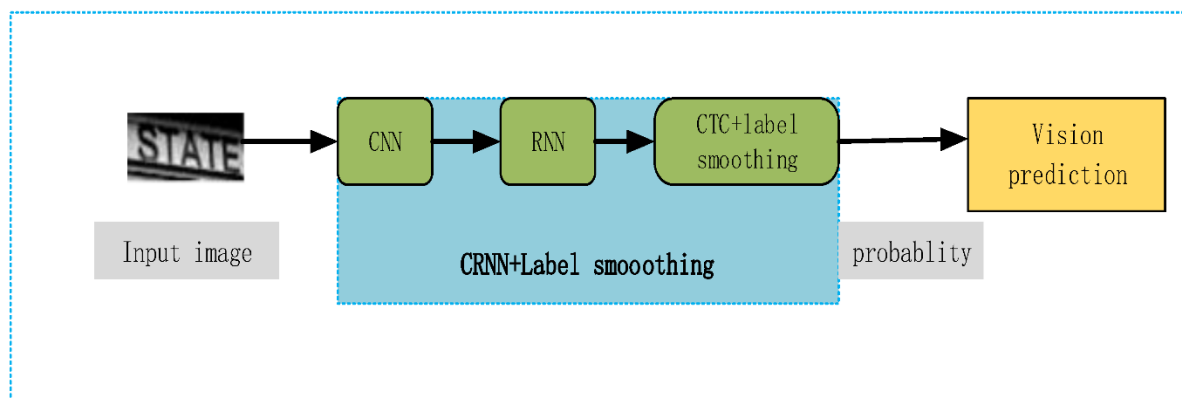


Рисунок 3. Алгоритм CRNN

5. Трансформеры:

- *предобработка* - разделение изображения на патчи или последовательности признаков;
- *механизм внимания (Attention)* - анализирует все элементы изображения одновременно, определяя важные участки;

– *декодирование текста* - формирует текст с учетом глобального контекста изображения.

На рисунке 4 показан общий принцип работы Visual Transform [7].

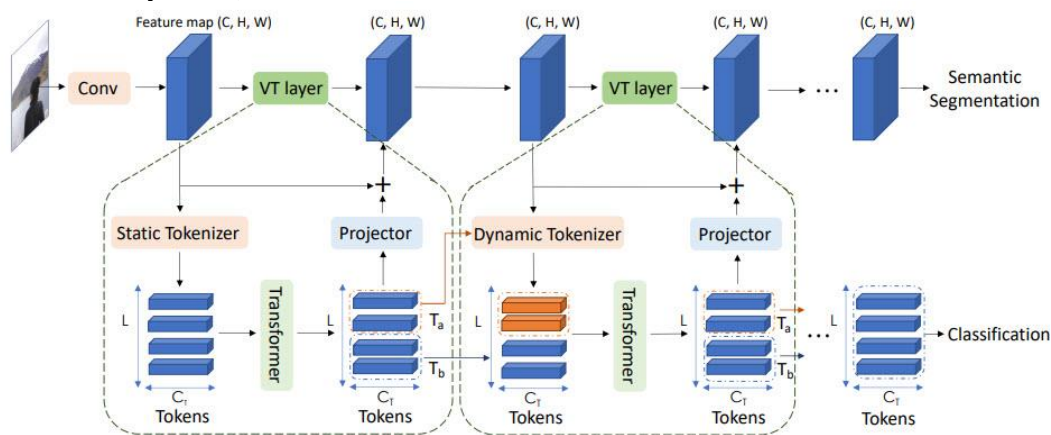


Рисунок 4. Visual Transform

Описание алгоритма из источника [7]:

- вначале используется обычный backbone для получения feature maps
- далее feature map конвертируется в визуальные токены
- токены подаются в трансформеры
- выход трансформера может использоваться для задач классификации

– а если объединить выход трансформера с feature map, то можно получить предсказания для задач сегментации

Практическое применение - решение задачи по выбранному оптимальному методу. Проведем сравнение методов, на вход поступает фотография уличного знака с текстом на русском языке, содержащая небольшие искажения.

В результате получаем следующие результаты, показанные на рисунке 5 и таблице 2.

Таблица 2.

Сравнение результатов методов

Метод	Распознанный текст	Точность, %	Время обработки, мс
Классические методы	"Ули Лена, 2"	70	10
CNN	"Улица Ленина, 2"	85	35
RNN	"Улица Ленниа, 2"	88	55
CRNN	"Улица Ленина, 2"	95	60
Трансформеры	"Улица Ленина, 2"	98	120

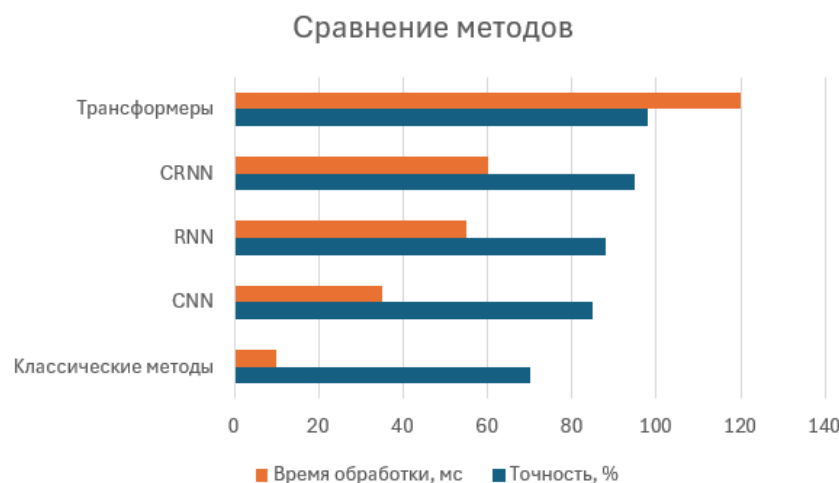


Рисунок 5. Диаграмма сравнения результатов методов и моделей

CRNN и трансформеры показывают наивысшую точность, при этом CRNN имеет лучшую производительность. Классические методы и CNN работают быстрее, но за счет потери точности.

По результату данного практического сравнения, был выбран метод CRNN (Convolutional Recurrent Neural Network) как оптимальный для реализации, т. к. данный метод сочетает в себе подходы сверточных и рекуррентных нейронных сетей, что обеспечивает высокую точность при работе с последовательными данными, такими как текст.

Заключение. В результате проведенного исследования можно сделать несколько важных выводов относительно методов распознавания текста и их применения в современных задачах. Технологии оптического распознавания текста (OCR) значительно развились за последние десятилетия, что позволило достичь высоких результатов в различных областях, включая цифровизацию документов, автоматическое извлечение данных и анализ изображений с текстом.

Классические методы распознавания текста, хотя и применяются в некоторых нишевых задачах, уступают по точности и универсальности более современным подходам. Эти методы ограничены в обработке искаженных или сложных изображений, поскольку не учитывают контекст текста.

Сверточные нейронные сети (CNN) продвигают распознавание текста на новый уровень, позволяя эффективно извлекать визуальные признаки из изображений. Однако они не учитывают последовательность символов, что ограничивает их возможности при распознавании сложных текстов, таких как рукописные или искаженные.

Рекуррентные нейронные сети (RNN), в свою очередь, идеально подходят для анализа последовательностей, что делает их полезными в задачах, где важен контекст и порядок символов. Но проблемы с долгосрочной зависимостью ограничивают их использование в сложных задачах.

CRNN (Комбинированные CNN и RNN) объединяют достоинства обеих технологий, обеспечивая высокую точность и способность работать с

длинными текстами. Этот подход доказал свою эффективность в ряде применений, таких как распознавание текста на изображениях и видео, а также в ситуациях с искажениями.

Трансформеры представляют собой самый современный подход в OCR, благодаря механизму внимания, который позволяет эффективно учитывать глобальный контекст и долгосрочные зависимости. Несмотря на высокую точность, их вычислительная нагрузка остается значительной, что ограничивает использование в реальном времени.

Выбор метода зависит от конкретных поставленных целей и задач, доступных вычислительных ресурсов и требований к точности. На сегодняшний день CRNN представляет собой оптимальный компромисс между точностью и производительностью, особенно при решении задач распознавания текста в реальных условиях, где важен как контекст, так и пространственные признаки. В то время как трансформеры продолжают оставаться перспективным направлением для обработки сложных текстов, в будущем можно ожидать их все более широкого применения в задачах OCR благодаря дальнейшему улучшению вычислительных мощностей и оптимизации моделей.

Список литературы:

1. Распознавание текста с помощью OCR. URL: <https://habr.com/ru/articles/471542/>
2. Методы распознавания текстов. URL: <https://habr.com/ru/articles/112442/>
3. Ocr — что такое. URL: <https://skyeng.ru/magazine/wiki/it-industriya/chto-takoe-ocr/>
4. Методы распознавания образов: от простых до сложных. URL: <https://www.simbirsoft.com/blog/metody-raspoznaniya-obrazov-ot-prostyx-doslozhnykh/>
5. Наглядно о том, как работает свёрточная нейронная сеть. URL: <https://habr.com/ru/companies/skillfactory/articles/565232/>
6. Кустикова В., к.т.н., ст.преп. каф. МОСТ ИИТММ, ННГУ им. Н.И. Лобачевского, Рекуррентные нейронные сети, Образовательный курс «Введение в глубокое обучение с использованием

Intel® neon™ Framework», Нижний Новгород, 2018 г.

7. Обзор на статью Visual Transformers — новый подход к тренировке моделей компьютерного зрения на основе visual tokens. URL: https://habr.com/ru/companies/ru_mts/articles/512258/

8. Что такое RNN (рекуррентная нейронная сеть)?. URL: <https://aws.amazon.com/ru/what-is/recurrent-neural-network/>

9. П.П. Алексеев, И.Ю. Квятковская, Применение нейронных сетей для распознавания принци-

альных условно-графических электрических обозначений, Вестник Астраханского государственного технического университета, Серия: Управление, вычислительная техника и информатика. 2021, 48-50.

10. A. Shrivastava, J. Amudha, D. Gupta, K. Sharma, Deep Learning Model for Text Recognition in Images, 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019.

11. Снова о распознавании рукописного текста, на этот раз с помощью CRNN. URL: <https://habr.com/ru/articles/720614/>