УДК 378:004
DOI: 10.25686/978-5-8158-2474-4-2025-659-669

Распознавание рукописных графических элементов и мультиязычного текста: обзор современных подходов, базовый метод и перспективы развития

Н. Ю. Скопин, М. С. Мосалев, Д. А. Королев, А. М. Пискунова
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Аннотация. В данной статье представлен обзор современных методов решения задачи распознавания рукописного контента, включающего мультиязычный текст (русский и английский) и графические элементы, такие как блок-схемы из прямоугольников и стрелок, с акцентом на распознавание изображений маркерных досок. Для мультиязычного рукописного текста рассмотрены архитектуры, способные обрабатывать несколько языков, а для блок-схем — методы распознавания фигур и структуры. Кроме того, в работе предложена базовая модель, в которой для распознавания текста используются решения EasyOCR и TrOCR, а для анализа графических элементов — алгоритмы на базе OpenCV, выделяющие контуры фигур. Цель этой модели — оценить сложность решения поставленной задачи с использованием общедоступных инструментов при минимальной подготовке. Базовая модель ориентирована на специфическую задачу распознавания контента на маркерной доске, а результаты формируются в формате XML для последующей обработки, например визуализации. Система протестирована на специально собранном датасете изображений маркерных досок. В статье также обсуждаются основные сложности задачи распознавания и обозначаются перспективные направления для повышения точности и качества решения.

***Ключевые слова*:** распознавание рукописного контента, мультиязычный текст, блок-схемы, выделение контуров фигур, распознавание диаграмм, обнаружение объектов, распознавание изображений, маркерная доска

Recognition of handwritten graphic elements and multilingual text: a review of modern approaches, a basic method and development prospects

N. Y. Skopin, M. S. Mosalev, D. A. Korolev, A. M. Piskunova
National Research University Higher School of Economics, Moscow, Russia

Abstract. This article presents an overview of modern methods for solving the problem of recognizing handwritten content, including multilingual text (Russian and English) and graphical elements such as block diagrams consisting of rectangles and arrows, with a focus on the recognition of whiteboard images. For multilingual handwritten text, architectures capable of processing multiple languages are considered, and for block diagrams — methods for recognizing shapes and structure. In addition, the paper proposes a baseline model in which EasyOCR and TrOCR are used for text recognition, and OpenCV-based algorithms that extract shape contours are used for analyzing graphical elements. The purpose of this model is to assess the complexity of solving the given problem using publicly available tools with minimal preparation. The baseline model is focused on the specific task of recognizing content on a whiteboard, and the results are generated in XML format for further processing (e.g., visualization). The system was tested on a specially collected dataset of whiteboard images. The article also discusses the main challenges of the recognition task and outlines promising directions for improving the accuracy and quality of the solution.

***Keywords*:** handwriting recognition, multilingual text, flowcharts, shape outline detection, diagram recognition, object detection, image recognition, whiteboard

Введение

Маркерные доски являются неотъемлемой частью современных образовательных и профессиональных сред, выступая важным инструментом для обсуждений, обучения и совместной работы. Контент на таких досках часто представляет собой сочетание рукописного текста на нескольких языках, таких как русский и английский, и графических элементов, таких как блок-схемы, диаграммы и стрелки.

В образовательной сфере автоматизация извлечения графических схем способствует оцифровке лекционных материалов и упрощает поиск информации в архивах. В научных и инженерных лабораториях такие инструменты позволяют структурированно документировать результаты проектных и экспериментальных работ. В бизнес-контексте распознавание блок-схем и графиков ускоряет фиксацию результатов мозговых штурмов и совещаний. Для достижения полной автоматизации обработки содержимого доски необходимо также преобразовывать полученные данные в машиночитаемые форматы.

1. Обзор существующих подходов

В работе Schäfer B. и др. [1] задача распознавания рукописных диаграмм разделяется на два этапа: локальное распознавание, локализация символов и восстановление глобальной структуры. Авторы рассматривают диаграммы как совокупность блоков, стрелок и поясняющего текста — подход, близкий к нашему.

Предлагается модификация ускоренной сверточной нейронной сети на основе регионов (Faster Region-based Convolutional Neural Network, Faster R-CNN) с добавлением предсказания ключевых точек стрелок и пост-обработки для воссоздания общей структуры. Это позволяет интегрировать детекцию объектов и анализ топологии диаграммы. В архитектуру включены пирамида признаков (Feature Pyramid Network) и модуль регрессии ключевых точек стрелки, что улучшает распознавание стрелок и текста. Ключевые точки стрелок используются для определения связей между элементами. Особое внимание уделено аугментации: применялся геометрический пайплайн и специальные преобразования для снижения путаницы между стрелками и текстом. В рамках анализа статьи выделены следующие преимущества описанной модели:

- одновременное распознавание символов и структуры,
- эффективность на малых и крупных датасетах благодаря целевой аугментации,
- высокая скорость — менее 100 мс на диаграмму.

Кроме того, есть недостатки и перспективы:

- проблемы с пересекающимися ограничивающими прямоугольниками (bounding box) стрелок;
- возможность внедрения графовых мер схожести для оценки.

В статье Bluche T. и др. [2] предлагается архитектура со сверточным кодировщиком для извлечения признаков из изображений и особой двунаправленной рекуррентной сетью с долгой краткосрочной памятью (Long short-term memory) в качестве декодировщика для генерации текстовой последовательности. Основная идея — создание универсального кодировщика, обученного однажды на разноязычных данных, который можно использовать без дообучения для распознавания новых языков.

Несмотря на прогресс в задаче распознавания рукописного текста (Handwritten Text Recognition, HTR), например, с использованием многомерной рекуррентной нейронной сети (multidimensional Recurrent Neural Network), ключевые ограничения — построчная обработка и отсутствие поддержки мультиязычности. Авторы решают это, обучая модель на семи языках (включая русский и английский), чтобы кодировщик научился извлекать обобщенные признаки, инвариантные к языку. Декодировщик при этом адаптируется под конкретную задачу.

Модель обучалась на 133 тыс. изображений на английском, французском, немецком, испанском, итальянском, португальском и русском языках. Результаты показали высокую эффективность даже без тонкой настройки (fine-tuning): модель способна распознавать текст неизвестного заранее языка, что открывает путь к обработке мультиязычных и смешанных текстов.

В работе Omasa T. и др. [3] описано, как сделать так, чтобы языковая модель с возможностью обработки визуальных данных (Visual Language Model) могла понять для себя структуру диаграммы, то есть увидеть, где стрелки, как они соединяют блоки и другие детали, и дальше смогла бы ответить на вопросы по ней. Упоминаются модели, учитывающие стрелки, и конкретная архитектура, обзор по статье с данной архитектурой есть в предыдущем пункте.

Вопрос, который ставят себе авторы: как можно совместить детектор, обращающий свое внимание на стрелки (flowchart-aware detector), и языковую модель через составление с помощью первой модели запроса, который содержит геометрическую информацию для второго?

Основной подход, описанный в статье, состоит из семиэтапного конвейера, который авторы подробно описали в своей работе.

2. *Основные проблемы распознавания*

Распознавание контента с маркерных досок сопряжено с рядом технических трудностей, обусловленных особенностями среды и способа записи информации [4, 5]. Системы распознавания сталкиваются со следующими основными проблемами:

- неравномерное освещение и блики: источники света создают блики на глянцевой поверхности доски и тени, что приводит к неравномерной яркости изображения и затрудняет определение границ символов;
- качество маркеров: маркеры различаются по толщине линий и насыщенности цвета, что влияет на контрастность и читаемость надписей;
- вариативность почерка: индивидуальные различия в стилях письма требуют от системы способности адаптироваться к разнообразным почеркам и способам отрисовки графических элементов;
- использование нескольких цветов: многоцветные надписи могут усложнять распознавание, особенно если цвета плохо контрастируют с фоном;
- частичные стирания и переписывания: следы предыдущих записей создают шум, который может быть ошибочно принят за актуальный контент;
- качество изображения и искажение перспективы: фотографии, сделанные под углом, приводят к геометрическим искажениям, требующим коррекции. Качественное распознавание зависит от параметров сделанного изображения доски (разрешение, фокус, экспозиция);
- фоновые помехи: посторонние объекты в кадре необходимо исключить при обработке изображения;
- перекрытие текста и графики: текст и графические элементы могут пересекаться или быть близкими друг к другу, что затрудняет их разделение и распознавание, также необходимо учитывать случай вложенности одного графического элемента в другой, например, прямоугольник в прямоугольнике.

Эти проблемы с качеством изображения напрямую влияют на точность распознавания, так как затрудняют корректную сегментацию текста и графических элементов, а также классификацию символов. Для преодоления этих вызовов необходимо применять специализированные методы предобработки изображений и использовать продвинутые алгоритмы машинного обучения, способные справляться с шумами и искажениями.

Распознавание смешанных русско-английских рукописных текстов предъявляет особые требования к алгоритмам распознавания рукописного текста. Ниже выделены проблемы, возникающие при разработке мультиязычных решений [6–9]. Они препятствуют созданию универсальных моделей, способных работать над смешанными текстами с одинаковой точностью. Список проблем в данном случае следующий:

1. Идентификация языка. При использовании гибридного подхода (раздельное применение одноязычных моделей) возникает задача надежной классификации языка фрагмента на уровне строки или даже слова (Q-сети, D-триггер). Ошибки на этом этапе приводят к каскадному ухудшению качества распознавания.
2. Детектирование строк и слов. Также при таком гибридном проходе возникает задача детектирования слов и строк. Смешанный текст может менять направление или стиль штриха. Это приводит к ошибкам выделения строк и отдельных слов: символы разных алфавитов могут сливаться в одном связном компоненте или, наоборот, разделяться на несколько.
3. Вариативность почерка и артефакты. Индивидуальные особенности письма (разная ширина штриха, наклон, связность букв) в сочетании с шумом и пятнами на доске затрудняют извлечение стабильных признаков.
4. Мультиязычная модель. Разработка единой модели для распознавания сразу двух алфавитов требует особых архитектурных решений: мультиязычная токенизация, расширенный словарь токенов, схема объединения выходов нейронной сети для работы с последовательностями переменной длины (Connectionist Temporal Classification) и языковой модели. Подобные решения обычно сложнее в обучении и требуют мультиязычных датасетов.
5. Различия в наборе символов. Кириллица и латиница имеют перекрывающиеся визуальные символы (A/A, E/E, B/B), что порождает неоднозначности при классификации без учета лингвистического контекста. Для их разрешения необходима интеграция языковых моделей и контекстных алгоритмов постобработки.

6. Доменно-специфичные особенности. Технические термины, аббревиатуры и имена собственные часто выходят за пределы базового словаря. Модели нужно дообучать, если они обучались на определенных корпусах, и расширять лексикон, что усложняет процесс обучения и поддержки системы.

Постановка задачи

Задача данной работы – обзор современных нейросетевых методов распознавания рукописных русско-английских текстов и методов для распознавания фигур, создание базовой модели, сочетающей готовые решения и комбинированный подход для распознавания текста, и оценка работы базовой модели на собранном и размеченном датасете фотографий маркерных досок.

Важно отметить, что задача исследования включает две взаимосвязанные подзадачи: мультиязычное распознавание рукописного текста (на русском и английском языках) и детекцию графических элементов (прямоугольников и стрелок). Основные вызовы заключаются в точной сегментации и идентификации языка каждого текстового фрагмента, а также в отделении графических элементов от фонового шума и текстовых меток и в формировании итогового файла в машиночитаемом формате с результатами распознавания. Распознавание рукописного текста сталкивается с вариативностью индивидуальных стилей письма. Кроме того, наличие нескольких языков усложняет процесс из-за смешения алфавитов и необходимости точной идентификации языка. Одновременно распознавание графических элементов требует применения соответствующих методов для анализа их структуры и значения.

Теория

В данном разделе описана предлагаемая базовая модель для распознавания содержимого маркерной доски (рукописных блок-схем и мультиязычного текста). Модель включает несколько ключевых этапов: предобработку изображения, выделение и распознавание графических элементов через контуры, распознавание мультиязычного текста и генерацию XML-файла для структурированного представления результатов. В данной работе под мультиязычностью понимается поддержка двух языков – русского и английского.

1. *Предобработка изображения*

Первым шагом в обработке изображения является его предобработка, направленная на улучшение качества, включающая корректировку контраста и яркости, и подготовку к последующим этапам анализа. Этот процесс начинается с преобразования исходного цветного изображения в оттенки серого, что упрощает цветовое пространство и позволяет сосредоточиться на интенсивности пикселей, игнорируя цветовые вариации. Далее применяется гауссово размытие, которое сглаживает изображение, уменьшая влияние высокочастотных шумов, сохраняя при этом значимые границы объектов.

Для удаления фоновых теней и слабых пикселей, не относящихся к полезной информации, применяется пороговая операция: из цветного изображения извлекается компонент серого, после чего все пиксели с интенсивностью большей 70 переводятся в чисто белый цвет. Сам порог определен эмпирически по результатам экспериментов. Такая операция точки белого позволяет избавиться от бледных пятен и сгладить неравномерную засветку, обеспечивая более четкий контраст между линиями маркера и фоновым покрытием доски. В результате выполнения всех шагов первого этапа получается изображение с повышенной контрастностью, унифицированным разрешением и минимальным уровнем фонового шума.

Для выделения контуров используется алгоритм детекции краев Кэнни, известный своей способностью находить границы объектов с высокой точностью благодаря двухпороговой фильтрации и отслеживанию связности [10]. После этого применяются морфологические операции: дилатация для соединения разрывов в контурах и эрозия для удаления мелких шумовых артефактов. Эти шаги формируют надежную основу для последующего обнаружения контуров, что критически важно для выделения графических элементов диаграммы.

2. Выделение и распознавание графических элементов

Выделение графических элементов, таких как прямоугольники и стрелки, осуществляется с использованием методов анализа контуров. После предобработки из изображения извлекаются контуры, представляющие замкнутые области. Эти контуры аппроксимируются до полигонов для упрощения их геометрической структуры с сохранением ключевых характеристик. Классификация форм основана на анализе числа сторон полученного полигона: четырехугольники идентифицируются как прямоугольники, тогда как стрелки распознаются по более сложным критериям.

Для обнаружения стрелок используется анализ выпуклой оболочки полигона. Вычисляется количество вершин выпуклой оболочки. Если оно лежит в диапазоне от 4 до 5 (дополнительная сторона в случае, если стрела имеет плоское основание) и если форма стрелки имеет ровно две дополнительные точки, которых нет в выпуклой оболочке (рис. 1), то контур рассматривается как потенциальная стрелка и ищется наконечник и основание [11]. Точка, не принадлежащая выпуклой оболочке, являющаяся самой выпуклой острой вершиной и наиболее выступающая из общего контура, интерпретируются как наконечник стрелки, а наиболее удаленная точка от наконечника стрелки на оболочке определяется как основание. Этот подход позволяет надежно различать направление стрелок, что важно для интерпретации структуры диаграммы.

Рисунок 1. Контур стрелки

Результаты распознавания контуров сохраняются в виде двух списков для дальнейшей обработки. Для прямоугольников каждый элемент — это список из четырёх пар координат (x,y), соответствующих вершинам прямоугольника в порядке обхода. Для стрелок — это кортеж из двух элементов, где первый элемент — координаты наконечника стрелки, а второй элемент — координаты основания.

Чтобы обеспечить устойчивость метода к вариациям качества изображения и стилей рисования, применяется итеративный подход к настройке параметров аппроксимации контуров. Модель тестирует различные пороговые значения, выбирая конфигурацию, которая максимизирует количество распознанных элементов. Такой адаптивный механизм повышает работоспособность системы, как было выявлено на этапе тестирования, и делает её применимой к более широкому спектру входных данных.

Результат такой предобработки изображения и выделения контуров изображен на рисунке 2. Красным цветом визуализированы распознанные прямоугольники, зеленым – стрелки.

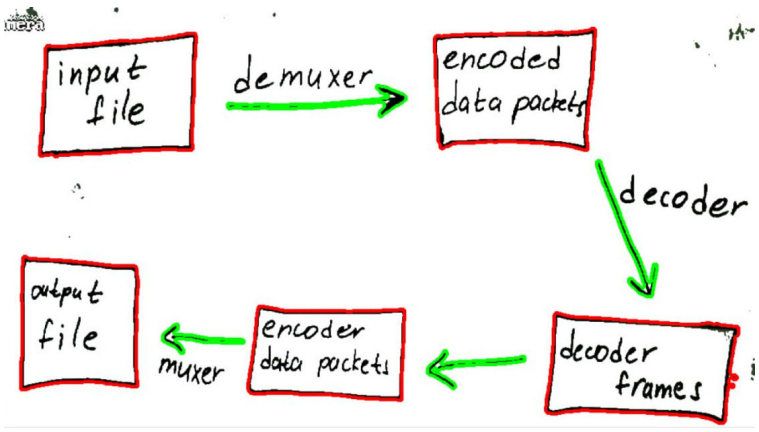

Рисунок 2. Результат распознавания контуров

3. Методы распознавания мультиязычного текста

Распознавание текста в базовой модели реализовано с использованием гибридного подхода, сочетающего детекцию слов, классификацию языка и последующее распознавание, а также готовых инструментов. Особое внимание было уделено точности распознавания, скорости работы и способности адаптироваться к смешанному тексту. Мы протестировали такие инструменты, как EasyOCR [12], Shiftlab [13], Tesseract [14] и TrOCR [15, 16], для распознавания рукописного текста. В итоговую реализацию базовой модели взяли модели на основе TrOCR для решения задачи распознавания текста и EasyOCR для решения задачи детекции слов. Ниже перечислены гибридные методы, которые были протестированы для базовой модели распознавания мультиязычного текста (русский и английский языки).

Детекция слов и формирование строк. Первым этапом в обработке текстового контента стало выделение отдельных слов и формирование строк, так как модель TrOCR принимает на вход только строки рукописного текста или отдельные слова. Данная модель не приспособлена для распознавания многострочного контента. Для детекции слов использовался инструмент EasyOCR, который предоставляет координаты ограничивающего прямоугольника (bounding box) для каждого распознанного слова. На основе этих данных был реализован алгоритм группировки слов по строкам: объекты объединялись в группы по вертикальной координате Y правого нижнего угла ограничивающих их прямоугольников с допустимой погрешностью d . Таким образом, каждая группа представляла собой потенциальную строку текста. После группировки производилось формирование единого ограничивающего прямоугольника для каждой строки путем объединения отдельных ограничивающих прямоугольников слов, входящих в эту группу. Полученные области использовались далее для последовательного применения моделей HTR к целым строкам или отдельным словам.

На основе анализа существующих технологий и результатов предварительных экспериментов были протестированы три подхода для базовой модели к распознаванию смешанного русско-английского текста.

Подход 1: гибридный метод с классификацией языка. Данный подход основан на комбинации EasyOCR, двух моделей TrOCR, обученных для русского и английского языков соответственно, и классификатора языка. После детекции слов с помощью EasyOCR для каждого слова осуществляется предсказание языка с использованием дообученной модели ResNet [12]. Затем применяется соответствующая модель TrOCR для распознавания текста. Недостатком этого подхода является необходимость запуска нескольких моделей, что увеличивает общее время обработки и снижает производительность, а также накопление ошибки при неправильной детекции слова или классификации языка, что может отразиться на итоговом распознавании. Важно отметить, что классификатор на основе ResNet срабатывал не идеально, тем не менее было принято решение протестировать данный подход более подробно.

Подход 2: выбор наиболее вероятного результата между двумя языками. В данном случае также используются EasyOCR для детекции и две модели TrOCR, но вместо предварительной классификации языка текст каждого слова распознаётся обеими моделями. Итоговый результат выбирается на основе сравнения нормализованных значений логитов. Однако, как показали эксперименты, значения логитов разных моделей не всегда сравнимы между собой даже после нормализации, что может приводить к ошибкам выбора языка. Это связано с различиями в процессе обучения исходных моделей, которые не были специально адаптированы под такую задачу.

Подход 3: единая мультиязычная модель TrOCR. Третий подход заключается в применении уже обученной модели TrOCR [16], поддерживающей распознавание текста сразу на двух языках — русском и английском. Такой подход позволяет обрабатывать текст за один проход, что положительно влияет на скорость работы. Однако, как показали первоначальные тесты, доступные предобученные модели не обеспечивают достаточного качества распознавания на нашем наборе данных, особенно в случае сложных технических терминов и аббревиатур.

Все три рассмотренных подхода имеют свои преимущества и ограничения. Первый обеспечивает гибкий выбор модели в зависимости от языка, но требует дополнительных вычислений. Второй

предлагает более простую логику выбора, однако сталкивается с проблемой интерпретации логитов. Третий требует дообучения или тонкой настройки модели под специфику данных.

В рамках разработанной базовой модели был выбран третий подход, поскольку он позволил достичь наилучшего баланса между точностью и надежностью распознавания.

Текст внутри и вне прямоугольников блок-схемы распознается отдельно, чтобы затем можно было точно привязать текст к соответствующему прямоугольнику при генерации XML файла. Для этого после распознавания расположения прямоугольников эти области вырезаются и отдельно подаются в модель для распознавания текста. Затем области с прямоугольниками исключаются из изображения, чтобы избежать повторного учета текста при распознавании текста вне прямоугольников на изображении целиком.

4. Формирование XML-файла

Итоговым результатом работы модели является генерация файла в формате XML, который структурирует распознанные элементы в машиночитаемом формате. На вход функции генерации подаются координаты и значения всех распознанных элементов. В XML-файле указываются стили, размер шрифта и толщина линий. При генерации выделяются три основных типа данных.

- Прямоугольники представлены как вершины с указанием их позиций (координаты левого верхнего угла) и размеров (длина и ширина). В этом же объекте сохраняется текст, если он был написан внутри соответствующего прямоугольника.
- Стрелки – координаты, содержащие исходные (sourcePoint) и конечные точки (targetPoint), по которым отрисовывается направление стрелки.
- Текст – самостоятельные блоки с текстом вне прямоугольников. Указывается координата левого верхнего угла, ширина и высота блока с текстом, а также значение самого текста.

Структура XML обеспечивает точное воспроизведение пространственных отношений между элементами и текстом, что позволяет использовать результат для дальнейшей обработки, анализа или визуализации (например, отрисовки в графическом редакторе).

Общая структура предлагаемой базовой модели изображена на рисунке 3.

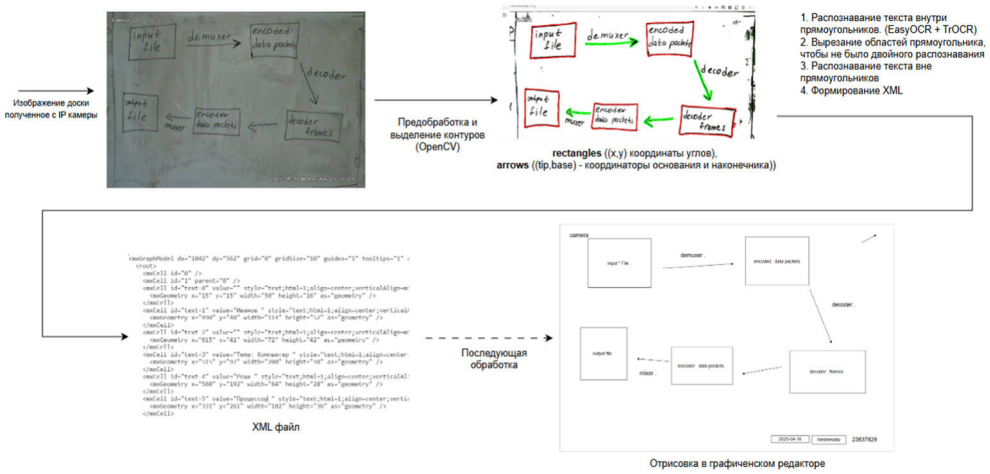

Рисунок 3. Общая структура базовой модели

Результаты экспериментов

1. Датасет для тестирования

Для оценки алгоритмов сформирован датасет изображений маркерной доски в учебных аудиториях. Основная часть изображений снята с помощью IP-камеры, установленной в фиксированном положении и направленной на доску, что обеспечивает неизменный ракурс и перспективу. Для проверки устойчивости к вариациям угла съемки дополнительно включены фотографии, сделанные с мобильного телефона, а также изображения блок-схем и текста на бумажных листах. Изображения содержат рукописный текст на русском и английском языках, а также графические элементы — прямоугольники и стрелки в виде блок-схем. Датасет состоит из 50 изображений с разметкой в виде

правильной детекции прямоугольников, стрелок и текста. Для решения задачи распознавания текста был сформирован отдельный датасет, состоящий из изображений строк и их расшифровок. Количество объектов в нем равно 21.

Условия эксперимента:

- константные параметры: положение IP-камеры, угол обзора, разрешение изображений, рабочая область доски;
- изменяющиеся параметры: освещение (естественное, искусственное, смешанное), почерк и способ изображения графических элементов (индивидуальные особенности письма), цвет и насыщенность маркера, размеры изображения.

Эти параметры отражают реальные условия использования и позволяют оценить устойчивость алгоритмов к внешним факторам.

2. Метрики для оценки

Для анализа эффективности модели выбраны метрики, обеспечивающие простоту разметки и вычисления.

Распознавание текста:

- уровень ошибок в символах (Character Error Rate, CER): доля ошибок на уровне символов между предсказанным и истинным текстом;
- уровень ошибок в словах (Word Error Rate, WER): доля ошибок на уровне слов.

Эти метрики стандартны для задач распознавания текста и подходят для оценки мультиязычного текста.

Идеальным значением для метрик является 0. Это означает, что распознавание прошло идеально — без ошибок, то есть метрика уровня ошибок в символах показывает, насколько точно система распознает отдельные символы, а метрика уровня ошибок в словах важна для оценки понятности текста, она отражает смысловые ошибки.

Выделение и распознавание графических элементов:

- пересечение по блокам (Intersection over Union, IoU): степень пересечения предсказанных и истинных ограничивающих прямоугольников (bounding box) для оценки точности локализации прямоугольников и стрелок;
- F1-мера: гармоническое среднее точности (precision) и полноты (recall) для оценки качества обнаружения графических элементов;
- точность (accuracy): доля правильно распознанных объектов среди всех объектов

Метрики разделены по задачам, что позволяет детально оценить производительность модели.

Диаграмма считается правильно распознанной, если:

- каждый объект (блок-прямоугольник или текст) правильно классифицирован и локализован с значением метрики пересечения по блокам (IoU) не менее 70 %, то есть пересечение истинного ограничивающего прямоугольника и полученного с помощью нашей базовой модели - 0,7;
- для каждой стрелки правильно идентифицированы узлы основания и наконечника. Это определяется через оценку расстояния между точками основания и наконечника предсказанного и размеченного объекта. Если евклидовы расстояния между этими точками меньше 100, то считается, что стрелка распознана правильно.

3. Результаты базовой модели

В данном разделе представлены результаты количественной оценки производительности базовой модели по четырем ключевым задачам: детекция прямоугольников, локализация текстовых строк, распознавание стрелок и распознавание мультиязычного текста на основе собственного датасета диаграмм и текстов преимущественно на белой маркерной доске. Для оценки использовались стандартные метрики, которые были ранее упомянуты в статье. При оценке локализации строк текста выделялся весь текст на изображение, а не только текст вне блоков, для большего количества примеров.

Оценка детекции прямоугольников, локализации строк текста и детекции стрелок представлена в таблице 1, оценка качества распознавания мультиязычного текста – в таблице 2.

Таблица 1. **Оценка детекции объектов**

Метрика	Оценка детекции прямоугольников	Оценка локализации строк текста	Оценка детекции стрелок
IoU	0.5003	0.6316	-
precision	0.5729	0.6433	0.693
recall	0.5359	0.6040	0.522
accuracy	0.5359	0.6040	0.539
F1 score	0.5538	0.6231	0.595

Таблица 2. **Оценка мультиязычного распознавания текста**

Подход	Подход 1 (с ResNet)	Подход 2 (анализ логитов)	Подход 3 (мультиязычная TrOCR)
CER	0.85	0.51	0.43
WER	1.67	0.89	0.84

Как упоминалось ранее при предобработке изображения для удаления фоновых теней и слабых пикселей, не относящихся к полезной информации, применяется пороговая операция: из цветного изображения извлекается компонент серого, после чего все пиксели с интенсивностью большей 70 переводятся в чисто белый цвет. В таблице 3 представлены результаты распознавания при разных показателях этого порога.

Таблица 3. **Метрики при разном пороге предобработки**

Метрика	Порог (threshold)				
	50	60	70	80	90
IoU rectangle	0.4212	0.4647	0.5003	**0.5043**	0.4910
F1 score rectangle	0.5117	0.5415	0.5538	0.5710	**0.5747**
Accuracy rectangle	0.4524	0.4995	0.5359	**0.5421**	0.5315
IoU text	0.5876	0.6085	**0.6316**	0.6250	0.6215
F1 score text	0.5193	0.5538	**0.6231**	0.6084	0.6062
Accuracy text	0.5028	0.5336	**0.6040**	0.5832	0.5867
F1 score arrow	0.5924	0.5836	**0.5952**	0.5830	0.5410
Accuracy arrow	**0.5543**	0.5472	0.5388	0.5531	0.5200

Обсуждение результатов

Несмотря на полученные результаты и выявленный потенциал базового метода, эксперименты продемонстрировали существенные ограничения текущей реализации, которые обусловливают низкую точность и нестабильность работы в реальных условиях. В связи с этим целесообразно рассмотреть векторы работы дальнейших исследований, направленных на преодоление выявленных проблем и повышение качества распознавания. Ниже представлены ключевые перспективы развития системы, базирующиеся на анализе выявленных недостатков и современных подходах в обработке рукописного контента.

1. Улучшение качества изображения через попиксельные (Pixel-to-Pixel) преобразования

Одним из перспективных направлений является применение генеративных моделей с попиксельными преобразованиями для перевода изображений маркерной доски в более структурированный вид, близкий к цифровым чертежам. Такой подход позволяет устранить неравномерное освещение, блики, фоновые помехи и повысить четкость линий и символов. Это особенно важно при работе с изображениями, сделанными под углом, с пятнами и разводами или при плохом освещении. Предварительная обработка с использованием таких моделей может значительно улучшить качество входных данных для последующих этапов распознавания.

Кроме того, использование моделей с попиксельным преобразованием может способствовать унификации почерка и стиля рисования графических элементов, что положительно скажется на стабильности работы классификаторов форм и моделей распознавания рукописного текста.

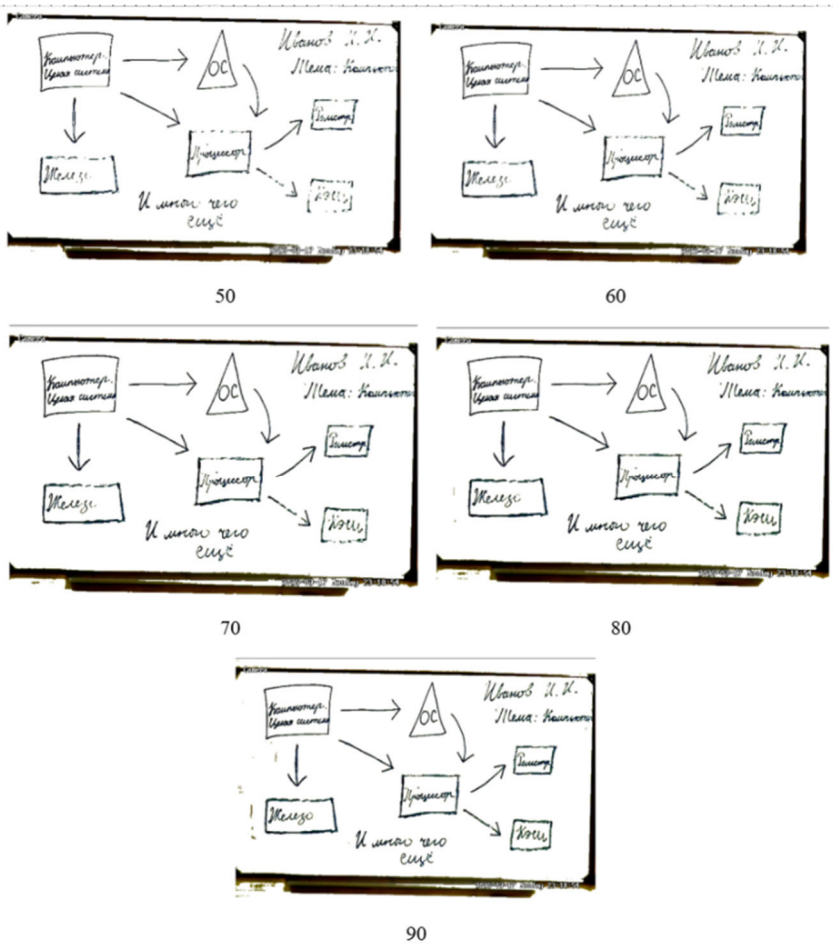

Рисунок 4. Изображения после предобработки с разным порогом белого

2. Совместное распознавание текста и графики: монолитные архитектуры(end-to-end)

Следующим важным направлением является переход от модульной архитектуры к монолитной модели, которая будет одновременно обрабатывать как текстовые, так и графические элементы. Такие модели могут использовать архитектуры типа трансформер или их модификации, учитывающие пространственное расположение объектов.

Интеграция информации о взаимном расположении текста и графики позволит улучшить интерпретируемость результата и повысить точность определения принадлежности текстовых блоков к графическим элементам, например, к стрелкам или прямоугольникам. Также это поможет в автоматическом восстановлении структуры диаграммы.

3. Дообучение мультиязычных моделей распознавания рукописного текста (HTR) на специализированных датасетах

Основным недостатком гибридного подхода остается необходимость запуска нескольких моделей, что увеличивает вычислительную нагрузку. Поэтому перспективным направлением является разработка и дообучение универсальной мультиязычной модели распознавания рукописного текста на датасетах, содержащих смешанные русско-английские рукописные образцы.

Такой подход позволит повысить стабильность распознавания смешанных фраз и аббревиатур.

4. Расширение функциональности: поддержка новых типов графических элементов

На данном этапе модель поддерживает распознавание только прямоугольников и стрелок. Для более широкого применения в образовательной и профессиональной среде целесообразно расширить спектр распознаваемых графических примитивов, включив окружности, ромбы, параллелограммы и другие элементы блок-схем. Это потребует как доработки модуля выделения контуров, так и расширения XML-структуры выходного файла.

Заключение

В данной работе представлен комплексный подход к распознаванию рукописного мультиязычного текста и графических элементов на маркерных досках. Предложена базовая модель, объединяющая методы распознавания текста (EasyOCR и TrOCR) и анализ контуров через OpenCV для выделения фигур. Результаты представлены в структурированном формате, что позволяет использовать их в системах автоматической обработки данных. Основные результаты включают решение ключевых проблем: детекции строк в смешанном тексте, классификации языка, отделения графики от фонового шума. В дальнейшем планируется улучшение качества распознавания через попиксельные преобразования, разработку монолитных моделей для совместного распознавания текста и графики. Полученные результаты открывают возможности для преобразования диаграмм в машиночитаемые форматы, что актуально для образовательных и профессиональных сред.

Ноутбуки для расчета метрик, отрисовки результата и собранный датасет с разметкой (директория Research), а также Телеграмм-бот как интерфейс взаимодействия для распознавания базовой моделью расположены в репозитории GitLab – Режим доступа: https://git.miem.hse.ru/msmosalev1/board-recognizer.

Список литературы

1. Schäfer B., Keuper M., Stuckenschmidt H. Arrow R-CNN for handwritten diagram recognition // International Journal on Document Analysis and Recognition (IJDAR). 2021. Режим доступа: https://link.springer.com/article/10.1007/s10032-020-00361-1
2. Bluche T., Messina R. Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition // Proceedings of the 14th International Conference on Document Analysis and Recognition (ICDAR). 2017. Режим доступа: https://tbluche.com/files/icdar17_gnn.pdf
3. Omasa T., Koshihara R., Morishige M. Arrow-Guided VLM: Enhancing Flowchart Understanding via Arrow Direction Encoding // arXiv. 2025. arXiv:2505.07864. Режим доступа: https://arxiv.org/pdf/2505.07864
4. DrawnNet: Offline Hand-Drawn Diagram Recognition Based on Keypoint Detection // PMC. 2022. Режим доступа: https://pmc.ncbi.nlm.nih.gov/articles/PMC8947756/
5. Sabeghi Saroui B. Recognition of Mathematical Handwriting on Whiteboards // University of Birmingham. 2015. Режим доступа: https://etheses.bham.ac.uk/6251/1/SabeghiSaroui15PhD.pdf
6. Ali S. H., Abdulrazzaq M. B. A Comprehensive Overview of Handwritten Recognition Techniques: A Survey // Journal of Computer Science. 2023. Vol. 19, no. 5. P. 569–587. Режим доступа: https://doi.org/10.3844/jcssp.2023.569.587.
7. Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey / W. AlKendi, F. Gechter, L. Heyberger, C. Guyeux // Journal of Imaging. 2024. Vol. 10, iss. 1. Art. 18. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10817575/
8. Hamad, H. Handwritten Recognition Techniques: A Comprehensive Review / H. Hamad, S. T. Al-Dhaher, M. S. Al-Hakeem // Symmetry. 2024. Vol. 16, Iss. 6. Art. 681. URL: https://www.mdpi.com/2073-8994/16/6/681
9. OCR рукописного текста: почему это так важно и трудно // Яндекс Образование : [сайт]. 2023. 17 ноября. URL: https://education.yandex.ru/journal/kak-raspoznat-rukopisnyi-tekst
10. Canny J. A computational approach to edge detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986. Vol. 8, no. 6. P. 679–698.
11. How to detect different types of arrows in image // Stack Overflow. Режим доступа: https://stackoverflow.com/questions/66718462/how-to-detect-different-types-of-arrows-in-image
12. A Novel Technique for Handwritten Text Recognition Using Easy OCR / B. K. Pattanayak, A. K. Biswal, S. R. Laha [и др.] // 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) : Proceedings. IEEE, 2023. P. 1115–1119. Режим доступа: https://thegrenze.com/pages/servej.php?fn=640_1.pdf&name=Efficient+Text+Extraction+and+Summarization+usingEasyocr+and+GPT-3&id=2886&association=GRENZE&journal=GIJET&year=2024&volume=10&issue=2
13. shiftlab_ocr: A python OCR library to read and generation // GitHub repository. Режим доступа: https://github.com/konverner/shiftlab_ocr (дата обращения: 20.05.2025).
14. Ракшит С., Басу С. Recognition of Handwritten Roman Script Using Tesseract Open source OCR Engine // Proc. National Conference on NAQC. 2008. P. 141-145. Режим доступа: https://arxiv.org/pdf/1003.5891.pdf
15. A Comprehensive Evaluation of TrOCR with Varying Image Effects // The National High School Journal of Science. 2024. Режим доступа: https://nhsjs.com/2024/a-comprehensive-evaluation-of-trocr-with-varying-image-effects/
16. Sergak0. text-recognition // Репозиторий GitHub. Режим доступа: https://github.com/sergak0/text-recognition