



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ  
(национальный исследовательский университет)»**

Институт (Филиал) № 8 «Компьютерные науки и прикладная математика» Кафедра 806  
Группа М8О-408Б-20 Направление подготовки 01.03.02 «Прикладная математика и  
информатика»

## **Профиль Информатика**

## **Квалификация: бакалавр**

# **ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

## **БАКАЛАВРА**

## на тему: Создание системы преобразования рукописного математического текста в LaTeX

Автор ВКРБ: Ядров Артем Леонидович (\_\_\_\_\_)  
Руководитель: Миронов Евгений Сергеевич (\_\_\_\_\_)  
Консультант: — (\_\_\_\_\_  
Консультант: — (\_\_\_\_\_  
Рецензент: — (\_\_\_\_\_)

## **К защите допустить**

Заведующий кафедрой № 806 Крылов Сергей Сергеевич (\_\_\_\_\_  
мая 2024 года)

## СОДЕРЖАНИЕ

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ . . . . .	3
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ . . . . .	4
ВВЕДЕНИЕ . . . . .	5
1 ПОСТАНОВКА ЗАДАЧИ И ТЕОРЕТИЧЕСКИЕ ПРЕДПОСЫЛКИ . . . . .	8
1.1 Пользовательские сценарии . . . . .	8
1.1.1 Авторизация пользователя . . . . .	8
1.1.2 Преобразование фотографии . . . . .	9
1.1.3 Преобразование файла формата .pdf . . . . .	10
1.1.4 Обзор состояния системы . . . . .	11
1.1.5 Контроль обучения . . . . .	11
1.1.6 Замена моделей . . . . .	11
1.2 Метрики . . . . .	12
1.2.1 Метрики нейросетевых моделей . . . . .	13
2 РАЗРАБОТКА ПРОГРАММНОГО ПРОДУКТА . . . . .	14
2.1 Высокоуровневая архитектура нейронного модуля . . . . .	14
2.2 Коррекция перспективы . . . . .	16
2.2.1 Неправильное распознавание . . . . .	20
2.3 Сегментация . . . . .	21
2.3.1 Сегментация на абзацы . . . . .	21
2.4 Выделение формул . . . . .	22
2.4.1 Модель . . . . .	22
2.4.2 Данные для обучения . . . . .	22
2.4.3 Обучение . . . . .	24
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .	29

## **ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ**

В настоящей выпускной квалификационной работе бакалавра применяют следующие термины с соответствующими определениями:

## **ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ**

В настоящей выпускной квалификационной работе бакалавра применяют следующие сокращения и обозначения:

## ВВЕДЕНИЕ

В России на постоянной основе проводятся научные исследования во многих областях. Результаты этих исследований публикуются в виде научных статей, которые являются важным инструментом для распространения новых знаний и научных открытий. Только в электронной версии научной библиотеки опубликовано 52573694 [1] статей, и все они написаны с помощью  $\text{\LaTeX}$  — мощного инструмента для верстки и оформления математических формул и научных текстов, который позволяет создавать качественные и профессионально оформленные статьи. Также с помощью  $\text{\LaTeX}$  можно готовить конспекты к предметам, причем это может делать как преподаватель, так и студент.

Однако, набор даже простых формул в  $\text{LaTeX}$  для неподготовленного человека может оказаться достаточно сложным и трудоемким занятием. Для примера возьмем формулу

$$f(x, y, \alpha, \beta) = \frac{\sum_{n=1}^{\infty} A_n \cos\left(\frac{2n\pi x}{\nu}\right)}{\prod \mathcal{F}_g(x, y)} \quad (1)$$

На рисунке 1 показан листинг  $\text{\LaTeX}$  кода этой формулы:

```
1 f(x,y,\alpha, \beta) = \frac{%
2   \sum \limits_{n=1}^{\infty} A_n \cos%
3     \left( \frac{2 n \pi x}{\nu} \right)%
4   }{%
5     \prod \mathcal{F}_g(x,y)%
6   }%
7 }
```

Рисунок 1 – Листинг формулы 1

Как мы видим, используется много специальных символов (например, символ суммы, произведения, а также дроби, скобки и пр.), которые необходимо знать или тратить время для их поиска на просторах Интернета. В любом случае, требуется каждый раз компилировать pdf-файл для просмотра и проверки результата, что требуется некоторого количества времени.

В настоящее время появляется все больше различных нейросетей

(например, Гигачат, Yandex GPT, ChatGPT, stable diffusion, Midjourney и тд), в том числе преобразующие рукописный текст на изображении в машинный, а также способных генерировать готовый код. Поэтому появляется мотивация для автоматизации процесса преобразования формул из чистового варианта на бумаге в готовый  $\text{\LaTeX}$ -код.

Однако, мир не стоит на месте, и компания *Mathpix* придумала свое решение [2] этой задачи, которое распространяется по платной подписке, что не удовлетворяет требованию доступности ПО.

Целью работы является разработка прототипа платформы, выполняющего распознавание научного текста и генерацию готового  $\text{\LaTeX}$  кода.

Для достижения поставленной цели в работе были решены следующие задачи:

- а) Определить требования к платформе
- б) Спроектировать архитектуру платформы
- в) Разработать:
  - 1) Сервис для распознавания научного текста
  - 2) Сервис для авторизации пользователя и синхронизации папок в облачном хранилище с авторизацией на основе протокола *OAuth 2.0*
  - 3) Приложение для взаимодействия пользователя с сервером
- г) Протестировать прототип платформы на наборе тестовых изображений

Для разработки программного обеспечения необходимо изучить технологии и методы, решающие поставленные задачи. Работа основывается на следующих библиотеках, технологиях, алгоритмах:

- а) *Python* является основным языком программирования, который использовался для решения задач;
- б) *Tensorflow* - библиотека для запуска и обучения моделей
- в) *YOLO* - модель, созданная для классификации объектов на изображении
- г) *openCV* - библиотека для обработки изображения
- д) *tkinter* - *Python* библиотека для работы с *GUI*
- е) *grpc* - фреймворк для удаленного вызова процедур
- ж) Что-то еще

Результатом работы является:

- а) Сервис для коррекции перспективы изображения
- б) Сервис для нахождения формул на изображении
- в) Дообученная модель *YOLO*, позволяющая находить формулы на изображении
- г) Сервис для преобразования найденных формул в  $\text{\LaTeX}$ -код
- д) *GUI* приложение с авторизацией в *Google Drive* и взаимодействием с удаленным сервером с помощью *grpc*

# 1 ПОСТАНОВКА ЗАДАЧИ И ТЕОРЕТИЧЕСКИЕ ПРЕДПОСЫЛКИ

Итак, перед нами стоит задача разработки прототипа платформы, позволяющей преобразовывать изображение с математическим текстом в  $\text{\LaTeX}$ -код.

## 1.1 Пользовательские сценарии

Для того, чтобы определить требования к нашей платформе, были проработаны пользовательские сценарии.

Для начала определим основные роли целевых пользователей:

- а) Пользователь приложения;
- б) Сторонний разработчик - программист, использующий *API* приложения;
- в) Разработчик - программист, занимающийся разработкой конкретных частей приложения;
- г) Архитектор - работник, занимающийся планированием и разработкой высокоуровневой архитектуры приложения;
- д) Администратор системы - специалист, отвечающий за настройку, управление и контроль работы системы;
- е) Инженер по машинному обучению - специалист, занимающийся разработкой моделей машинного обучения;
- ж) Аналитик данных - специалист, занимающийся оценкой качества обучения модели, выбором подходящих метрик, а также оптимизацией процесса обучения.

Под пользователем приложения подразумевается пользователь одного из типов приложения:

- *WEB* приложение;
- *Desktop* приложение;
- мобильное приложение.

### 1.1.1 Авторизация пользователя

Участники: пользователь любого из типов приложения

Предусловие: пользователь запускает приложение впервые или хочет зайти с другого аккаунта

Постусловие: пользователь авторизован через *Google* и предоставлен

доступ к *Google Drive*

Сценарий: Для авторизации с помощью *Google* необходимо реализовать следующий сценарий, изображенный на рисунке 2:

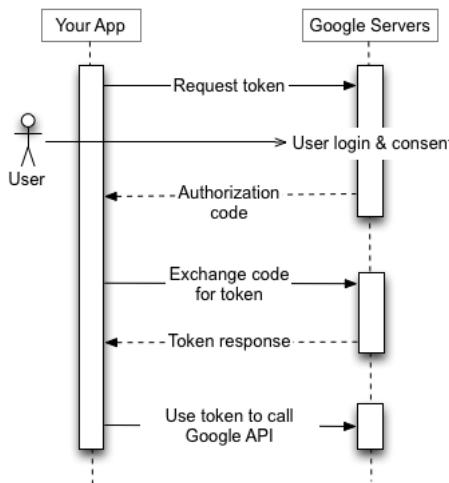


Рисунок 2 – Сценарий сетевого взаимодействия при авторизации пользователя [3]

По сути *Google* делает всё сам на этапе создания сервиса внутри кода. Если приложение не видит токена для авторизации, то происходит автоматическое перенаправление на сайт авторизации. Дальнейший токен сохраняется локально.

Для замены аккаунта достаточно удалить локальный токен пользователя и перезапустить сервис *Google*. Стоит отметить, что лучше сохранять токен для дальнейшего быстрого входа в сервис. Но так как мы разрабатываем прототип, то и этого решения будет вполне достаточно.

### 1.1.2 Преобразование фотографии

Участники: пользователь приложения

Предусловие: пользователь сделал фото научного текста и открыл его в приложении

Постусловие: пользователь получает готовый *LAT<sub>E</sub>X* код текста, полученного на изображении в виде архива из исходного кода и *pdf* файла

Сценарий:

- а) Производится автоматическая коррекция перспективы фотографии
- б) Производится автоматическое сегментация текста на абзацы
- в) С приложения на сервер отправляется изображение, а также таблица

- с координатами начала и конца абзацев
- г) С сервера на приложение отправляется таблица с найденными формулами
  - д) Пользователь проверяет правильность распознавания формул и вносит корректиды
  - е) Приложение отправляет на сервер таблицу с финальными формулами
  - ж) Сервер загружает архив с L<sup>A</sup>T<sub>E</sub>X кодом и pdf файлом в *Google Drive*, а также отсылает его пользователю
  - и) *Desktop* приложение обновляет папку *Google Drive* и загружает в локальное хранилище последний архив

При необходимости пользователь может корректировать точки перспективы, а также точки абзацев на изображении.

### 1.1.3 Преобразование файла формата .pdf

Участники: пользователь приложения

Предусловие: пользователь загрузил файл научного текста в формате .pdf в приложение

Постусловие: пользователь получает готовый L<sup>A</sup>T<sub>E</sub>X код текста, полученного на изображении в виде архива из исходного кода и pdf файла

Сценарий: В случае с .pdf файлом мы предполагаем, что нет необходимости корректировать перспективу.

- а) Производится автоматическое сегментация текста на абзацы
- б) С приложения на сервер отправляется изображение, а также таблица с координатами начала и конца абзацев
- в) С сервера на приложение отправляется таблица с найденными формулами
- г) Пользователь проверяет правильность распознавания формул и вносит корректиды
- д) Приложение отправляет на сервер таблицу с финальными формулами
- е) Сервер загружает архив с L<sup>A</sup>T<sub>E</sub>X кодом и pdf файлом в *Google Drive*, а также отсылает его пользователю
- ж) *Desktop* приложение обновляет папку *Google Drive* и загружает в локальное хранилище последний архив

#### **1.1.4 Обзор состояния системы**

Участники: администратор системы

Предусловие: существует рабочая система.

Постусловие: выводится статистика системы в зависимости от расположения сервиса (*Google* или *Yandex*).

Сценарий:

- a) Пользователь авторизуется в облачном сервисе (*Google* или *Yandex*)
- б) ... (когда-нибудь будет, если понадобится)

#### **1.1.5 Контроль обучения**

Так как самая важная часть системы (распознавание текста) реализуется на основе нейросетевых алгоритмов (о чем будет подробно рассказано позже), важно контролировать процесс обучения. Для этого необходимо иметь доступ к метрикам модели, а также обновлять обучающие наборы данных, проводить тестирование модели на новых данных, проводить валидацию в процессе обучения.

Важно быть уверенным в том, что модель не переобучена и способна обобщать данные.

Для этих целей очень хорошо подходит платформа *Weights & Biases* [4]

Участники: аналитик данных

Предусловие: существует рабочая система, обученная модель

Постусловие: выводится информация об обучении модели

Сценарий:

- a) Пользователь заходит на веб-сайт *wandb* [4];
- б) Пользователь выбирает интересующую его версию модели;
- в) Платформа выводит всю информацию о модели.

#### **1.1.6 Замена моделей**

Необходимо предусмотреть возможность замены одной модели на другую. Это является важным требованием к системе, поскольку позволяет улучшать эффективность и точность работы всей системы:

- a) Это дает возможность адаптировать систему под изменяющиеся условия. Если модель устаревает или не удовлетворяет требованиям, необходимо иметь возможность быстро ее заменить;

- б) Это позволяет снизить затраты на разработку и поддержку системы. Вместо разработки или поддержки своей модели, можно заменить текущую модель на готовое открытое решение;
- в) Это повышает гибкость системы. При изменении требований к системе, текущая модель может не удовлетворять новым требованиям.

Участники: инженер по машинному обучению, разработчик системы, аналитик данных

Предусловие: существует система

Постусловие: существует рабочая система, удовлетворяющая текущим требованиям

Сценарий:

- а) Инженер по машинному обучению разрабатывает новую модель или находит открытое решение;
- б) Разработчик системы встраивает новую модель в систему;
- в) Аналитик проверяет работоспособность модели, а также удовлетворение модели текущим требованиям;
- г) По результатам проверки:
  - 1) в случае удовлетворения требованиям старая модель заменяется на новую;
  - 2) в случае неудовлетворения требованиям цикл повторяется.

## 1.2 Метрики

Итак, мы определили ожидаемое поведение системы. Теперь мы установим критерии для оценки систем, за которыми ведется наблюдение. Оптимальный метод оценки отдельных компонентов системы заключается в использовании метрик, так как они позволяют сравнивать разные состояния системы и выбирать наилучшее из них.

Метрики могут быть различными и зависят от специфики конкретной системы. Например, для оценки производительности может использоваться метрика времени ответа системы на запрос пользователя, а для оценки надежности - частота отказов системы.

Важно отметить, что выбор метрик должен основываться на целях и задачах проекта. Каждая система имеет свои особенности и требования к работе, поэтому необходимо определить наиболее значимые параметры для ее

функционирования.

Кроме того, использование метрик позволяет проводить анализ данных о работе системы и делать выводы об эффективности ее работы. Это помогает оптимизировать процессы и повышать качество продукта.

В данной работе ключевую роль играют метрики нейросетевых моделей.

### 1.2.1 Метрики нейросетевых моделей

Метрики, применяемые к нейросетевым моделям, можно глобально разделить на:

- а) Метрики качества модели - показатели, позволяющие оценивать эффективность работы модели;
- б) Метрики потерь - показатели, позволяющие оценивать степень ошибки или несоответствия между прогнозами модели и реальными данными.

В качестве метрик качества модели мы выбрали следующие:

- а) *Precision* - точность обнаруженных объектов, указывающая, сколько обнаружений было выполнено правильно;
- б) *Recall* - способность модели идентифицировать все экземпляры объектов на изображениях;
- в) *mAP50* - средняя точность, рассчитанная при пересечении, превышает порог объединения (*IoU*), равный 0,50. Это показатель точности модели, учитывающий только "простые" обнаружения;
- г) *mAP50 – 95* - среднее значение средней точности, рассчитанное при различных пороговых значениях долговых расписок, варьируется от 0,50 до 0,95. Это дает полное представление о работе модели при различных уровнях сложности обнаружения.

В качестве метрик потерь мы выступает метрика *clsloss* (*Classification loss*) - потеря классификации, определяющая погрешность в прогнозируемых вероятностях классов для каждого объекта на изображении по сравнению с исходным значением. Меньшее значение этой метрики означает, что модель более точно предсказывает класс объектов

## 2 РАЗРАБОТКА ПРОГРАММНОГО ПРОДУКТА

Конечный продукт является целостной системой, и необходимо рассказать про каждую ее часть. Ключевым элементом системы является нейронный модуль, обеспечивающий обработку данных, предоставленных пользователем, с использованием методов нейронного обучения. Поэтому необходимо начать именно с него.

### 2.1 Высокоуровневая архитектура нейронного модуля

Поскольку задача преобразования изображения в текст, выполняемая модулем, не является тривиальной, было решено разбить ее на несколько этапов. Схема разбиения представлена на рисунке 3

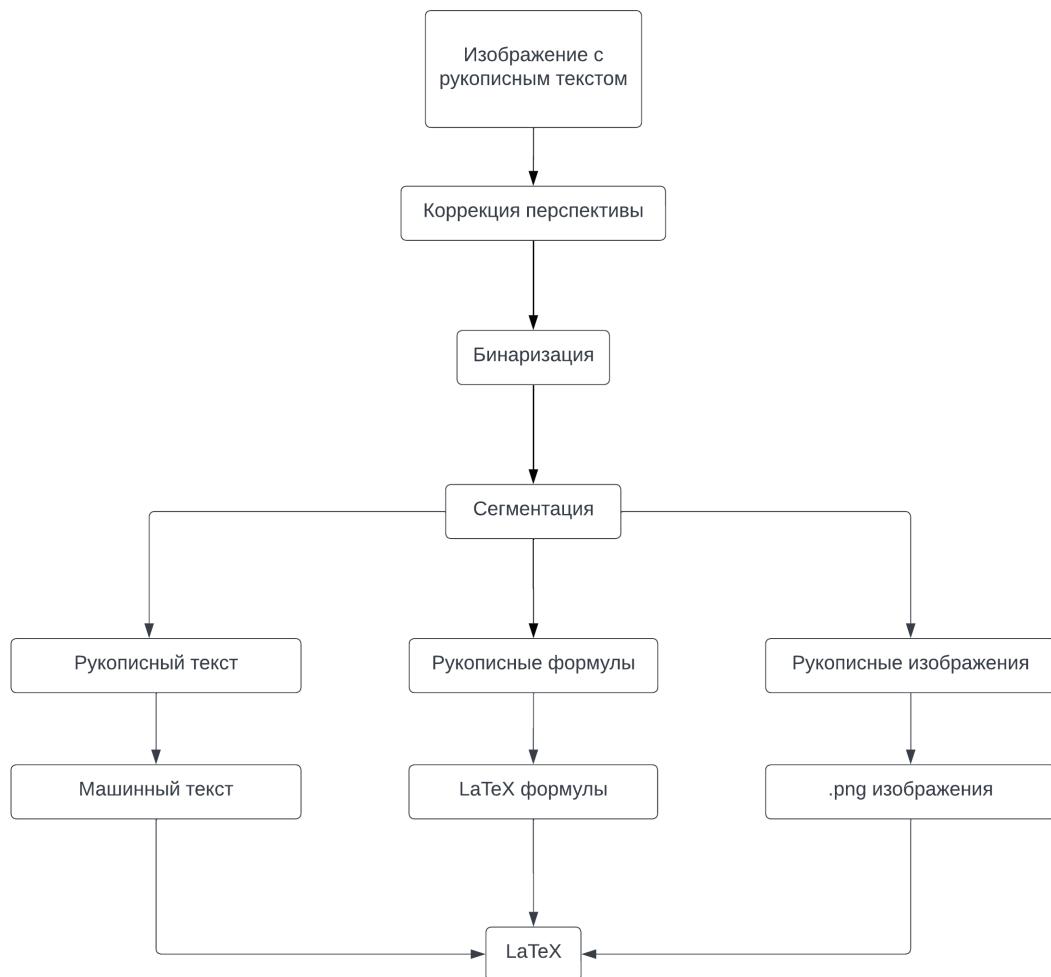


Рисунок 3 – Общая архитектура модели

Каждый из этапов данной схемы можно охарактеризовать входными

данными и результатом работы модуля (выходными данными). Такой подход обладает рядом преимуществ:

- а) Гибкость и масштабируемость: Модульная структура позволяет легко добавлять новые компоненты или модифицировать существующие без необходимости переписывать всю модель.
- б) Ускорение процесса обучения: Благодаря возможности параллельной обработки данных, модульные нейронные сети обучаются быстрее, чем монолитные модели.
- в) Улучшение качества модели: Разделение модели на модули позволяет специалистам сосредоточиться на оптимизации каждого компонента, что в итоге приводит к улучшению общей производительности модели.
- г) Простота внедрения новых технологий: Модульная архитектура облегчает внедрение новых технологий и подходов, таких как трансферное обучение или диффузионные модели.
- д) Улучшение производительности: некоторые модули могут исполняться в препроцессинге на клиентской машине, что ослабляет нагрузку на сервер.

Однако, имеются и недостатки:

- а) Проблемы с совместимостью: Разные модули могут использовать различные архитектуры, форматы данных и методы обучения, что может привести к проблемам совместимости.
- б) Риск ухудшения производительности: Неправильно спроектированные или плохо интегрированные модули могут снизить общую производительность модели.
- в) Необходимость в дополнительных ресурсах: Для обучения и развертывания модульных моделей часто требуются дополнительные ресурсы, такие как GPU или TPU.

Несмотря на недостатки, в современных реалиях важно уметь быстро и без проблем масштабироваться и заменять при необходимости один компонент другим, поэтому было принято решение использовать модульную архитектуру.

Подробнее разберем каждый этап данной схемы.

## 2.2 Коррекция перспективы

Коррекция перспективы необходима для устранения шума на изображении и получения лучшего результата. Она состоит из нескольких этапов, представленных на рисунке 4. Также на рисунке представлены результаты, получаемые на каждом из этапов обработки.

Стоит отметить, что коррекция перспективы осуществляется с использованием только алгоритмов обработки изображения без использования каких-либо нейросетей. Поэтому в целях экономии ресурсов сервера, а в следствии улучшения производительности было принято решение выполнять данный этап на машине клиента. Схема данного алгоритма представлена на рисунке 4.

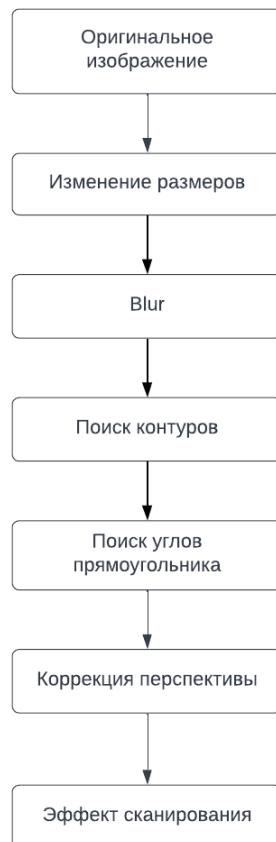


Рисунок 4 – Этапы коррекции перспективы изображения

На начальном этапе мы имеем изображение, показанное на рисунке 5

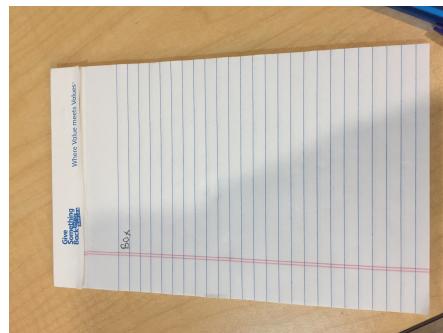


Рисунок 5 – Начальное изображение

Для начала необходимо удалить текст с изображения. Для этого преобразуем изображение в серый цвет и применим к нему размытие Гаусса [5]. На выходе данного этапа имеем изображение, представленное на рисунке 6

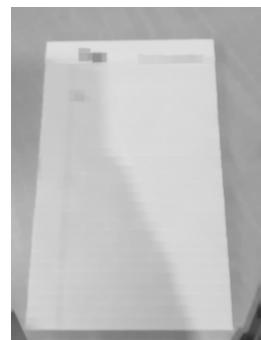


Рисунок 6 – Изображение после размытия Гаусса

Для поиска контуров необходимо выделить ребра. Для этого используется алгоритм Канни [6]. На выходе имеем изображение, представленное на рисунке 7

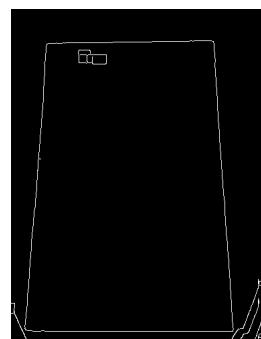


Рисунок 7 – Ребра, найденные на изображении

После нахождения ребер поиск контуров осуществляется двумя способами:

- a) С помощью алгоритма *LineSegmentDetector* [7]
- б) С помощью встроенного в *openCV* алгоритма поиска контуров [8]

Опишем подробнее поиск контуров на основе найденных линий: после прохода алгоритма имеем изображение, представленное на рисунке 8

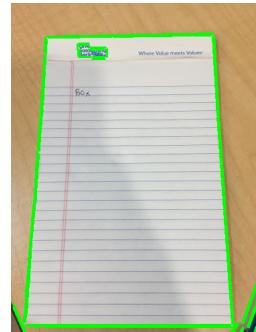


Рисунок 8 – Найденные на изображении линии с помощью алгоритма *LSD*

Контур определяется как пересечение горизонтальных и вертикальных линий. На выходе имеем найденные углы контура, показанные на рисунке 9

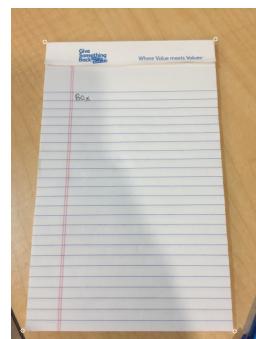


Рисунок 9 – Найденные на изображении контуры на основе линий

С помощью библиотеки *openCV* контуры находятся следующим образом:

- а) Находятся 5 наибольших по площади контуров
- б) Найденные контуры проверяются на количество углов, минимальную площадь контура

Среди всех подходящих контуров выбирается наибольший по площади, как показано на рисунке 10

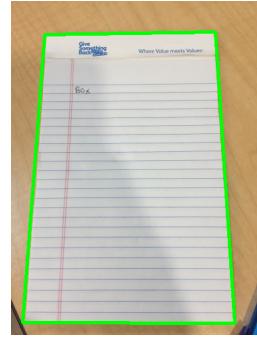


Рисунок 10 – Найденные алгоритмом контуры

На основе найденного контура, представляющего лист бумаги, осуществляем коррекцию перспективы. Для этого находим матрицу коррекции [9] и применим ее к изображению [10]. Получаем результирующее изображение, показанное на рисунке 11

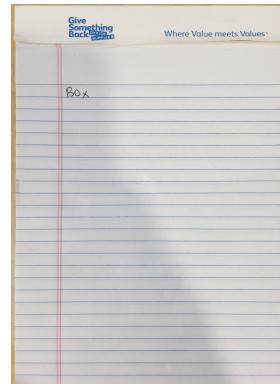


Рисунок 11 – Изображение с коррекцией перспективы

Далее необходимо добавить эффект сканирования. Эффект достигается путем применения к композиции небольшого размытия и серого изображения алгоритма сегментации *AdaptiveThreshold* [11]. В конечном итоге имеем результирующее изображение показанное на рисунке 12

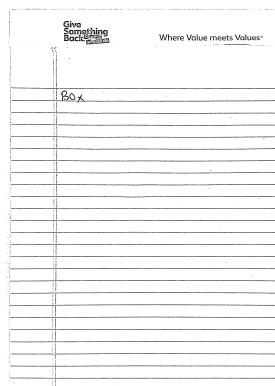


Рисунок 12 – Результирующее изображение

### 2.2.1 Неправильное распознавание

Стоит отметить, что данный алгоритм не является универсальным. Например, если на изображении находится посторонний шум (например, палец на бумаге, экран монитора, большое здание в виде прямоугольника), то вероятность получения неверного результата кратно возрастает. Именно с целью защиты от таких случаев в конечном продукте пользователь должен удостовериться в правильности найденного контура и отредактировать границы контура при необходимости. Пример входного изображения, дающего неверный результат, и найденные на нем контуры приведены на рисунках 13 и 14 соответственно.

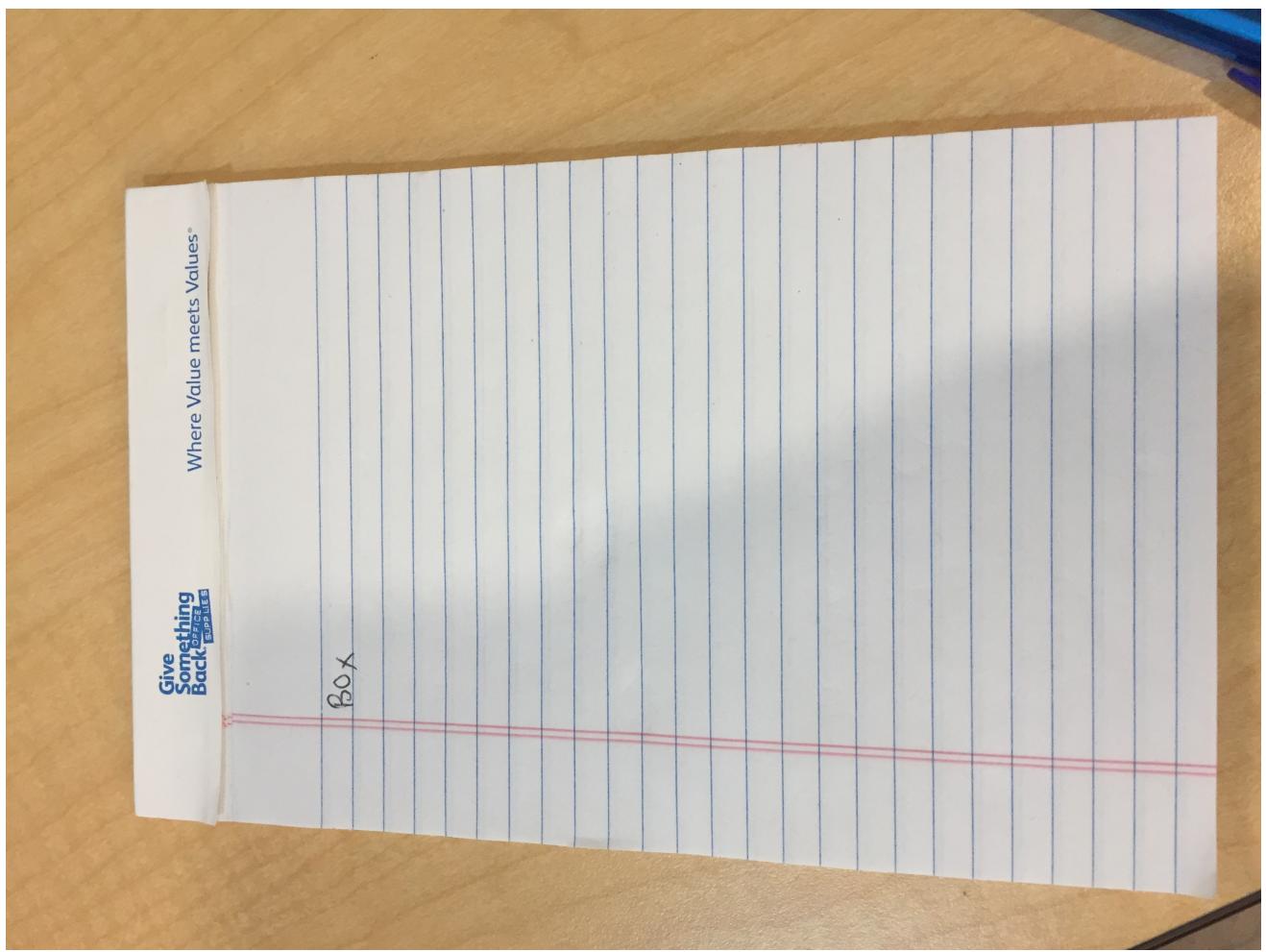


Рисунок 13 – Пример неправильно распознанного изображения:  
входное изображение

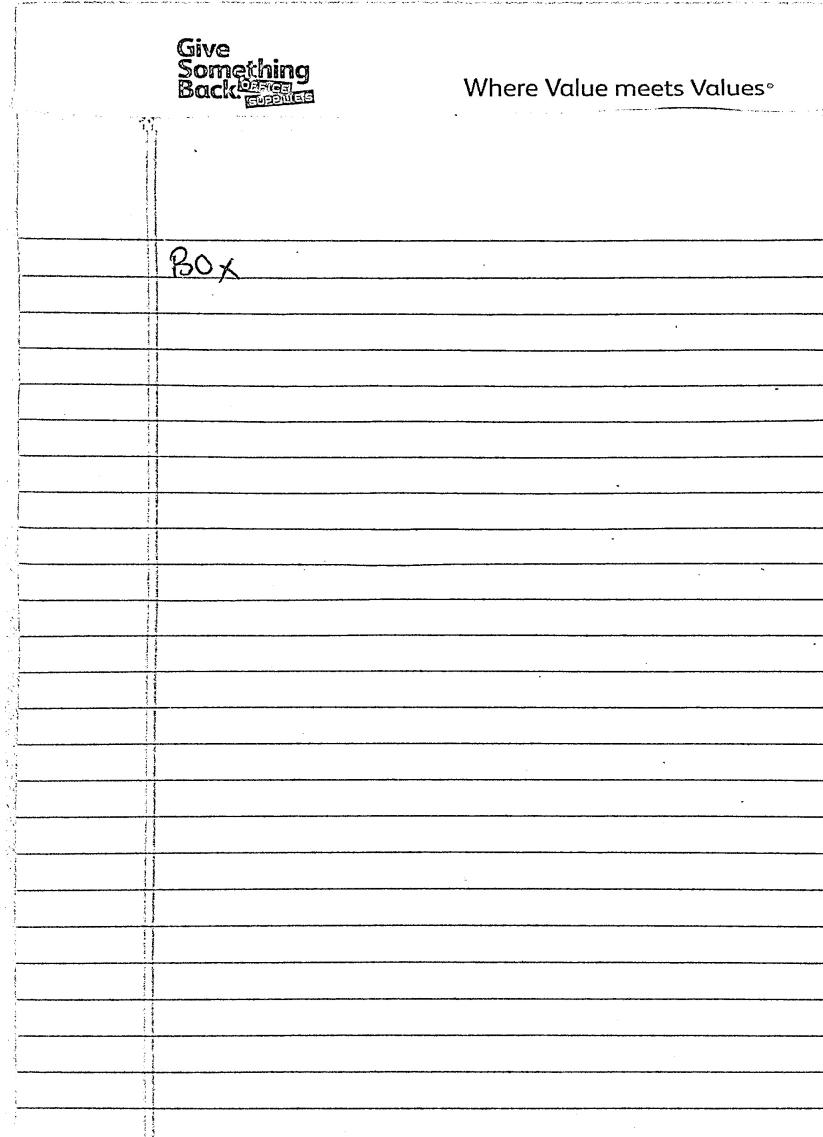


Рисунок 14 – Пример неправильно распознанного изображения:  
найденные контуры

## 2.3 Сегментация

Сегментация состоит из двух отдельных этапов:

- Сегментация на абзацы
- Отделение текста от формул, рисунков и пр.

### 2.3.1 Сегментация на абзацы

Выделение абзацев требуется для дальнейшего сохранения структуры текста в L<sup>A</sup>T<sub>E</sub>X коде. Данный вид сегментации, как и коррекция перспективы, осуществляется при помощи алгоритмов обработки изображения. Поэтому данный этап можно также выполнять на ЭВМ пользователя.

## 2.4 Выделение формул

### 2.4.1 Модель

Для выделения формул недостаточно одних алгоритмов обработки изображений. Существует множество моделей, выполняющих распознавание объектов на изображении. Одной из таких моделей является *YOLO* [12]. Данная модель обладает следующими преимуществами:

- Быстродействие. Нейросеть работает в реальном времени, поэтому её используют для распознавания объектов на фото и видео "здесь и сейчас".
- Точность. Нейросеть *YOLO* умеет распознавать объекты разных размеров в пределах одного кадра.
- Универсальность. Нейросеть *YOLO* способна определять как хорошо знакомые ей объекты, так и те, с которыми она ещё не сталкивалась.
- Простота. Модель *YOLO* можно запросто запускать и дообучать с помощью *Tensorflow* [13].

Однако, данная модель не специализирована на какой-то одной задаче, поэтому будет проигрывать специализированным моделям.

На основании плюсов данной модели, а также на основании самой архитектуры системы, позволяющую при необходимости легко заменить выбранную модель на другую, было принято решение использовать для распознавания формул модель *YOLO* последней версии *v8*.

### 2.4.2 Данные для обучения

В качестве датасета для обучения был взят датасет *ICDAR – 2021* (*International Conference on Document Analysis and Recognition*). Он содержит набор изображений текстовых документов, содержащих формулы, а также для каждого изображения имеется набор координат ограничивающих рамок ("bounding boxes") для каждой формулы. Размер датасета составляет 5171 изображений. На рисунках 15 и 16 показаны примеры входного изображения и изображения с метками формул соответственно.

- Three-Family Condition: To have three families we must require

$$6 = \lambda\eta(\eta - nc_1). \quad (2.9)$$

The second such rule is associated with the anomaly cancellation requirement that

$$[W_Z] = c_2(TZ) - c_2(V_{Z1}) - c_2(V_{Z2}), \quad (2.10)$$

where  $[W_Z]$  is the class associated with non-perturbative five-branes in the bulk space of the Hořava-Witten theory. Vector bundles  $V_{Z1}$  and  $V_{Z2}$  are located on the “observable” and “hidden” orbifold planes respectively. In this talk, for simplicity, we will always take  $V_{Z2}$  to be the trivial bundle. Hence, gauge group  $E_8$  remains unbroken on the “hidden” sector,  $c_2(V_{Z2})$  vanishes and condition (2.10) simplifies accordingly. Using the definition

$$[W_Z] = \frac{1}{2}q_*[W], \quad (2.11)$$

condition (2.10) can be pulled-back onto  $X$  to give

$$[W] = c_2(TX) - c_2(V). \quad (2.12)$$

It follows that

$$[W] = \sigma_*W_B + c(F - N) + dN \quad (2.13)$$

where

$$W_B = 12c_1 - \eta \quad (2.14)$$

and

$$c = c_2 + \left( \frac{1}{24}(n^3 - n) + 11 \right) c_1^2 - \frac{1}{2} \left( \lambda^2 - \frac{1}{4} \right) nn\eta(\eta - nc_1) - \sum_i \kappa_i^2, \quad (2.15)$$

$$d = c_2 + \left( \frac{1}{24}(n^3 - n) - 1 \right) c_1^2 - \frac{1}{2} \left( \lambda^2 - \frac{1}{4} \right) nn\eta(\eta - nc_1) - \sum_i \kappa_i^2 + \sum_i \kappa_i. \quad (2.16)$$

The class  $[W_Z]$  must represent an actual physical holomorphic curve in the Calabi-Yau threefold  $Z$  since physical five-branes are required to wrap around it. Hence,  $[W_Z]$  must be an effective class and, hence, its pull-back  $[W]$  is an effective class in the covering threefold  $X$ . This leads to the following rule.

- Effectiveness Condition: For  $[W]$  to be an effective class, we require

$$W_B \text{ is effective in } B, \quad c \geq 0, \quad d \geq 0. \quad (2.17)$$

## Рисунок 15 – Пример входного изображения

- Three-Family Condition: To have three families we must require

$$\delta = \lambda\eta(\eta - nc_1). \quad (2.9)$$

The second such rule is associated with the anomaly cancellation requirement that

$$[W_Z] = c_2(TZ) - c_2(V_{Z1}) - c_2(V_{Z2}). \quad (2.10)$$

where  $[W_Z]$  is the class associated with non-perturbative five-branes in the bulk space of the Hořava-Witten theory. Vector bundles  $V_{Z1}$  and  $V_{Z2}$  are located on the “observable” and “hidden” orbifold planes respectively. In this talk, for simplicity, we will always take  $V_{Z2}$  to be the trivial bundle. Hence, gauge group  $\mathbb{Z}_2$  remains unbroken on the “hidden” sector.  $c_2(V_{Z2})$  vanishes and condition (2.10) simplifies accordingly. Using the definition

$$[W_Z] = \frac{1}{2}q_*[W]. \quad (2.11)$$

condition (2.10) can be pulled-back onto  $X$  to give

$$[W] = c_2(TX) - c_2(V). \quad (2.12)$$

It follows that

$$[W] = \sigma_*W_B + c(F - N) + dN \quad (2.13)$$

where

$$W_B = 12c_1 - \eta \quad (2.14)$$

and

$$c = c_2 + \left(\frac{1}{24}(n^3 - n) + 11\right)c_1^2 - \frac{1}{2}\left(\lambda^2 - \frac{1}{4}\right)n\eta(\eta - nc_1) - \sum_i \kappa_i^2, \quad (2.15)$$

$$d = c_2 + \left(\frac{1}{24}(n^3 - n) - 1\right)c_1^2 - \frac{1}{2}\left(\lambda^2 - \frac{1}{4}\right)n\eta(\eta - nc_1) - \sum_i \kappa_i^2 + \sum_i \kappa_i. \quad (2.16)$$

The class  $[W_Z]$  must represent an actual physical holomorphic curve in the Calabi-Yau threefold since physical five-branes are required to wrap around it. Hence,  $[W_Z]$  must be an effective class and, hence, its pull-back  $[W]$  is an effective class in the covering threefold  $X$ . This leads to the following rule.

- Effectiveness Condition: For  $[W]$  to be an effective class, we require

$$W_B \text{ is effective in } B, \quad c \geq 0, \quad d \geq 0. \quad (2.17)$$

Рисунок 16 – Пример изображения с метками формул

### 2.4.3 Обучение

Обучение проводилось на платформе *Kaggle* [14] с логированием в системе *wandb* [4]. В качестве графического процессора использовались две видеокарты *Tesla T4*. После обучения модели мы переходим к оценки ее эффективности. Посмотрим на полученные с помощью *wandb* графики описанных выше метрик.



Рисунок 17 – Графики потерь

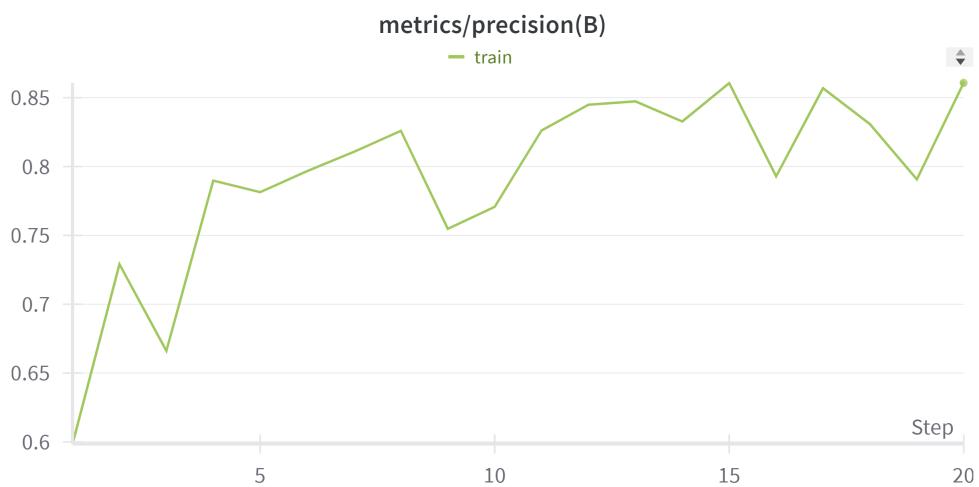


Рисунок 18 – График точности

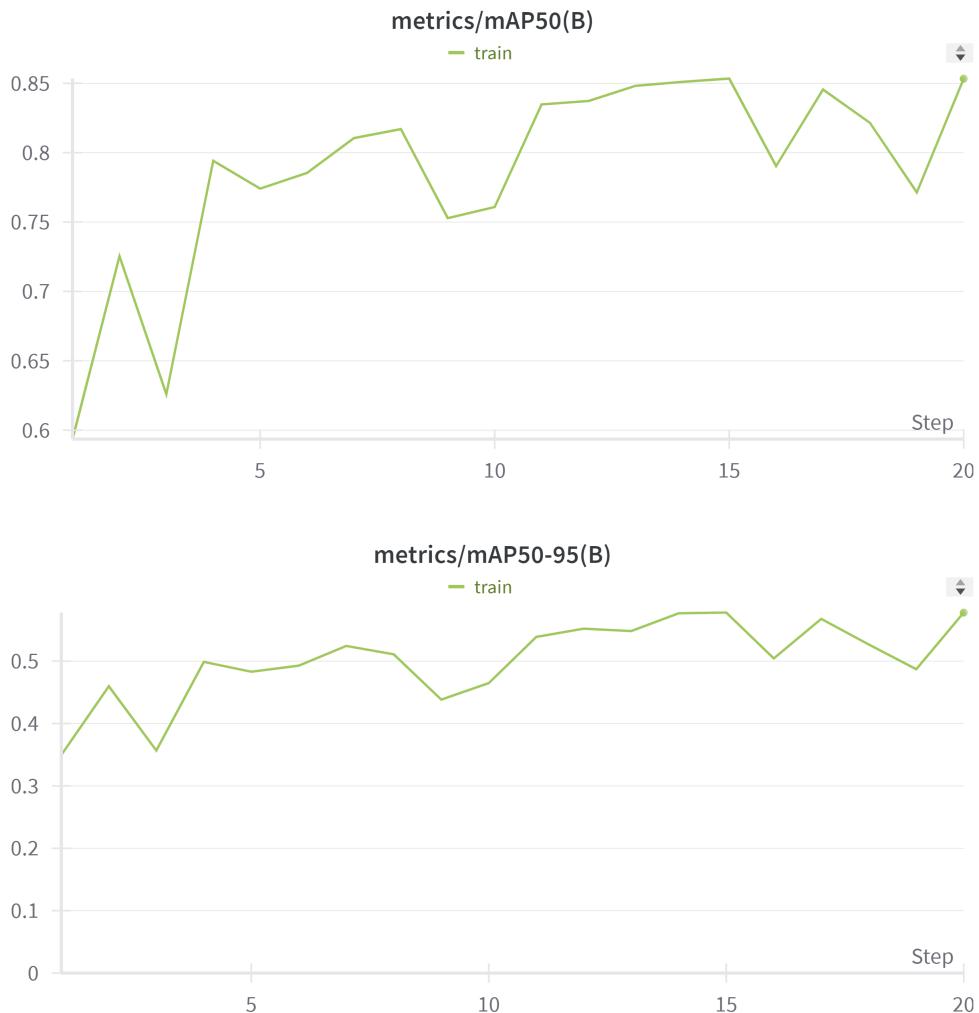


Рисунок 19 – Графики  $mAP$

На каждом из графиков 17 ... 19 по оси  $oX$  отложен номер шага, на котором снимались показатели. Весь этап обучения, состоящий из  $N$  эпох равномерно делится на 20 этапов. На каждом этапе снимаются метрики, попадающие в результирующий график.

После обучения модель была протестирована на тестовых изображениях. На рисунках 20 и 21 показаны результаты тестирования.

**Formula: 0.847** The  $C_{\alpha MN}$  are uniquely determined by the  $\Lambda_{\alpha M}^N$  via **Formula: 0.362**. What have to do then is to classify the possible  $K'$  representation matrices **Formula: 0.850**. In our discussion around (3.18), however, it follows that the possible  $\Lambda_{\alpha M}^N$  correspond with  $m$ -dimensional representations of  $K'$ . Since  $K'$  is compact, any representation of  $K'$  can be **Formula: 0.374** unitary, and any such unitary representation can be **Formula: 0.865** into  $(2m \times 2m)$  matrices **Formula: 0.681** to form a possible set of  $\Lambda_{\alpha M}^N$  or, equivalently, a possible set of  $C_{\alpha MN}$ .

(v)  $C_{\alpha M}$  **Formula: 0.727**

The  $C_{\alpha MN}$  also give rise to the **Formula: 0.850** matrices **Formula: 0.841**. Since **Formula: 0.782**, the  $K = K' \times U(1)$  gauge group is already generate  $K'$ , the  $\Lambda_{\alpha M}^N$  are either zero or they correspond to **Formula: 0.839**. **Formula: 0.918** we already know that the (non-vanishing) matrix  $\Lambda_{0M}^N$  generates this **U(1)**. **Formula: 0.922** **Formula: 0.910** beginning of this subsection assumed only one **U(1)** factor, the  $\Lambda_{\alpha M}^N$  can be at most proportional to  $\Lambda_{0M}^N$ , otherwise there would be another, independent, Abelian factor in the **Formula: 0.834**. For the  $C_{\alpha MN}$  this means that they can be at most (remember that  $C_{0MN} = -(1/2)\delta_{MN}$ )

$$C_{\alpha MN} = d_\alpha \delta_{MN}$$

for some constants  $d_\alpha$ . In that case, the **U(1)** gauge field would be the linear combination

$$A_\mu[U(1)] = \left[ -\frac{1}{2} A_\mu^0 + d_\alpha A_\mu^\alpha \right].$$

**Formula: 0.692**

3.3.2 **Formula: 0.692**  
 $K = U(1)^r$

We now assume when  $K = U(1)^r$  is purely Abelian. For simplicity that all the **U(1)** factors act on the same set of tensor fields. If there were Abelian groups acting on mutually disjoint sets of tensor fields, the cubic polynomial would simply decompose into several subpieces of the type to be described below.

Assuming now the above gauge group structure, the  $n$  of **SO(n)** decomposes as follows:

**Formula: 0.096**

$n = \text{singlets}(K) \oplus \text{non-singlets}(K)$

**Formula: 0.739** **0.809**

We denote **0.778** singlets of  $K$  by  $\{\alpha, \beta, \dots = 1, \dots, r\}$  and the non-singlets by  $\{M, N, \dots = 1, \dots, 2m\}$ , i.e., we split

$$i = (\alpha, M).$$

**Formula: 0.865**

The possible non-vanishing  $C_{ijk}$  are now the following.

**Formula: 0.133**; **73**

(i)  $|C_{\alpha\beta\gamma}|$ :

**Formula: 0.654** These coefficients are automatically singlets of  $K$  and can therefore be chosen arbitrarily.

(ii)  $C_{\alpha MN}$ :

Via (3.13), these coefficients are related to the  $K$ -transformation matrices **Formula: 0.393** **Formula: 0.843** **Formula: 0.843**

Рисунок 20 – Предсказание модели на тестовом изображении

this dual situation, the physics is described by the dual of the condensed moose in figure 21. The condensates in this case are

$$\langle \psi_{i,i+1} \chi_{i+1,i+1} \rangle \sim 4\pi f^3 V_{i,i+1} \quad i = 1, \dots, N. \quad (5.1)$$

Again the physics is five dimensional for  $a \ll r \ll R$ , but it is a different fifth dimension, dynamically generated by a different set of interactions and with a different set of gauge bosons.

The transition from (2.3) to (5.1) is theoretically fascinating, but somewhat puzzling, and we will not discuss it in detail here. But it is important to understand the approach to the transition because it bears on the possibly phenomenologically relevant question of how large the five dimensional coupling can be. For example the heaviest of the KK modes has a mass of order  $g(\Lambda_s) f_s$  parametrically lighter than the scale where  $G_s$  gets strong,  $\Lambda_s$ . How similar can we make these scales? Can we increase  $g$  to the region of strong coupling as well?

For simplicity, let us take  $\Lambda$  to infinity so that the physics appears five dimensional at arbitrarily long distances. What happens as we change the ratio of  $\Lambda$  to  $\Lambda_s$ ? For  $\Lambda \ll \Lambda_s$  where the analysis of section 2 applies, the residual gauge interactions at distances large compared to  $a$  are very weak. In the four dimensional language, this is obvious because the gauge coupling  $g_5^2 = g^2 a$ , and its effects at distances of order  $a$  are suppressed by powers of  $g_5^2/a$ . In the four dimensional language, one might worry that there is something wrong with this argument at distances large compared to  $a$ . This worry is groundless. The weak gauge group breaks the condensate (2.3) down to a residual gauge group with coupling of order  $g^2/N$  and the infrared weak  $\Lambda \rightarrow \infty$ .

What happens as we increase  $\Lambda/\Lambda_s$ ? The gauge coupling  $g_5^2 = g^2 a$  but its effects remain tiny at large distances. We know that at some point as  $\Lambda \rightarrow \Lambda_s$  an ecological disaster will occur, dramatically changing the nature of the long distance physics. It is reasonable to suppose that the cataclysm will happen abruptly at some point  $\Lambda \approx \Lambda_s$  where both  $g_5^2$  and  $g^2 a$  get large compared to  $a$ . The only signal at large distances of impending doom is that as  $\Lambda \rightarrow \Lambda_s$ ,  $g_5^2 = g^2 a$  gets large compared to  $a$ . This signals the imminent breakdown of the effective theory because dimensional couplings in an effective theory must not be large compared to the appropriate power of the cut-off. Even though the tree level interactions are still weak at long distances, the theory is losing control of its quantum corrections, a warning that anarchy is about to be loosed upon the world.

## 6 Other completions

The fifth dimension has appeared in the condensed moose because the non-linear sigma model fields allow the gauge field to “hop” from one site to the next. Since we could have obtained this directly as a latticization of the five-dimensional gauge theory, we might ask why we need the original moose model at all. The reason is that latticization in the fifth dimension does not cut-off divergences from large four-momenta: the four-dimensional non-linear sigma model of (2.5) is non-renormalizable, becoming strongly coupled at a scale  $\sim 4\pi f_s$ . That is, this theory requires a UV completion. But this is familiar problem, with familiar solution. The moose model we have constructed provides a UV completion in the same way that QCD

Рисунок 21 – Предсказание модели на тестовом изображении

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. eLibrary. Научная электронная библиотека. — 2000. — URL: <https://www.elibrary.ru> (дата обращения 07.03.2024).
2. Mathpix. PDF to LaTeX. — URL: <https://mathpix.com/pdf-to-latex> (дата обращения 09.03.2024).
3. Google. Using OAuth 2.0 to Access Google APIs. — URL: <https://developers.google.com/static/identity/protocols/oauth2/images/flows/authorization-code.png> (дата обращения 07.03.2024).
4. Weights & Biases. — URL: <https://wandb.ai> (дата обращения 24.03.2023).
5. OpenCV. Smoothing Images. — URL: [https://docs.opencv.org/4.x/d4/d13/tutorial\\_py\\_filtering.html](https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html) (дата обращения 25.04.2024).
6. OpenCV. Canny Edge Detection. — URL: [https://docs.opencv.org/4.x/da/d22/tutorial\\_py\\_canny.html](https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html) (дата обращения 25.04.2024).
7. pylsd. Line Segment Detector. — URL: <https://github.com/primetang/pylsd> (дата обращения 25.04.2024).
8. OpenCV. Contours : Getting Started. — URL: [https://docs.opencv.org/3.4/d4/d73/tutorial\\_py\\_contours\\_begin.html](https://docs.opencv.org/3.4/d4/d73/tutorial_py_contours_begin.html) (дата обращения 25.04.2024).
9. OpenCV. Geometric Image Transformations. — URL: [https://docs.opencv.org/4.x/da/d54/group\\_\\_imgproc\\_\\_transform.html](https://docs.opencv.org/4.x/da/d54/group__imgproc__transform.html) (дата обращения 25.04.2024).
10. OpenCV. Geometric Image Transformations. — URL: [https://docs.opencv.org/4.x/da/d54/group\\_\\_imgproc\\_\\_transform.html](https://docs.opencv.org/4.x/da/d54/group__imgproc__transform.html) (дата обращения 25.04.2024).
11. OpenCV. Image Thresholding. — URL: [https://docs.opencv.org/4.x/d7/d4d/tutorial\\_py\\_thresholding.html](https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html) (дата обращения 25.04.2024).
12. YOLO v.8. — URL: <https://github.com/ultralytics/ultralytics> (дата обращения 25.04.2024).
13. Tensorflow. — URL: <https://www.tensorflow.org> (дата обращения 25.04.2024).
14. Kaggle. — URL: <https://www.kaggle.com/> (дата обращения 24.03.2023).