## Цель и задачи работы

Одной из наиболее сложных форм CAPTCHA являются изображения, содержащие множество объектов с размытыми контурами, шумами и низким разрешением, что затрудняет автоматическое распознавание.

В процессе автоматизированного тестирования web-приложений возникает необходимость обхода подобных CAPTCHA, что требует разработки устойчивых и точных методов распознавания визуального контента.

Целью данной работы является разработка и обучение нейронной сети с поддержкой сегментации, способной автоматически распознавать объекты на изображениях CAPTCHA и выполнять задания, формируемые системой защиты.

Для достижения поставленной цели необходимо решить следующие задачи:

1. проанализировать типы CAPTCHA, применяемых на web-ресурсах;
2. выбрать подходящую архитектуру нейронной сети, обеспечивающую высокую скорость и точность;
3. собрать и разметить датасет реальных CAPTCHA с изображениями объектов;
4. провести предварительную обработку изображений и формирование структуры датасета;
5. обучить выбранную модель на собранных данных;
6. разработать скрипт для автоматизированного прохождения CAPTCHA с использованием обученной модели;
7. протестировать модель в реальных условиях и оценить её эффективность.

## Выбор модели нейронной сети для обучения

CAPTCHA в формате изображений широко используется для защиты ресурсов от автоматизированных ботов и может быть реализована несколькими способами. Как правило, такие CAPTCHA направлены на проверку способности пользователя распознавать и интерпретировать объекты на изображении. Наиболее распространены два варианта реализации (оба варианта реализации проиллюстрированы на слайде:

1. цельное изображение, содержащее несколько объектов, частично размытых или искажённых, при этом изображение разбито на сетку 3×3 или 4×4. Пользователю предлагается выбрать ячейки, содержащие объекты определённого класса (например, автобусы или светофоры);
2. составное изображение, сформированное из 9 или 12 отдельных фрагментов (изображений), каждый из которых представляет собой независимое изображение – зачастую низкого качества, с наложением артефактов или шумов. Задача пользователя – выбрать те изображения, где присутствует нужный объект.

Такие CAPTCHA требуют от системы автоматического анализа способности как к глобальному восприятию изображения, так и к локальной интерпретации его фрагментов. Соответственно, модель, предназначенная для решения данной задачи, должна поддерживать:

1. классификацию объектов на уровне отдельных изображений (для CAPTCHA, основанных на отдельных картинках в сетке);
2. локализацию и сегментацию объектов с высокой точностью, чтобы корректно определить границы объектов в пределах ячеек, особенно в случаях, когда объект может частично заходить за границу между ячейками.

**Переключение слайда**

Для решения этих задач были рассмотрены следующие современные архитектуры нейронных сетей:

1. YOLO;
2. Faster R-CNN;
3. DETR.

Среди этих архитектур было принято решение использовать YOLOv8 по причинам, указанным на слайде:

Кроме того, модель YOLOv8 была успешно протестирована в задачах, близких по структуре к CAPTCHA: детекции дорожных знаков, транспортных средств, пешеходов и других объектов в сложных условиях съёмки, что подтверждает её универсальность и применимость к рассматриваемой задаче.

Таким образом, YOLOv8 является наиболее сбалансированным выбором, обеспечивающим как точную классификацию, так и локализацию объ-

ектов в условиях ограниченных ресурсов и с возможностью адаптации под специфику визуальных CAPTCHA.

## Парсинг CAPTCHA для создания датасета

Для обеспечения высокой точности в задаче автоматического решения CAPTCHA необходимо подготовить собственный набор данных, приближённый к реальным условиям использования. Наиболее эффективным методом является автоматизированный парсинг изображений CAPTCHA, представленных на веб-сайтах, использующих визуальные CAPTCHA-решения, такие как Google reCAPTCHA v2.

Использование реальных CAPTCHA, собранных в автоматическом режиме, имеет ряд преимуществ по сравнению с синтетической генерацией данных:

1. изображения содержат разнообразные сцены, освещение, углы обзора и уровни шума, что положительно влияет на способность модели к обобщению;
2. присутствует большое количество уникальных объектов на фоне, в том числе в частично перекрытых и смазанных вариантах;
3. отсутствует необходимость в ручной генерации изображений и создании дополнительных искажений для повышения реалистичности;
4. возможно извлекать текстовые инструкции к CAPTCHA, что позволяет соотносить каждое изображение с требуемым классом.

Для парсинга CAPTCHA был реализован автоматизированный сценарий взаимодействия с браузером с использованием библиотеки Selenium, блок-сехма которого показана на данном слайде.

## Предобработка изображений датасета

После получения достаточного количества изображений для составления датасета необходимо провести их предварительную обработку и разметку. Это один из самых важных этапов работы, поскольку от качества разметки напрямую зависит точность и эффективность последующей работы модели.

Для создания меток используется инструмент CVAT – многофункциональное веб-приложение с поддержкой аннотации объектов с помощью полигонов, прямоугольников и других форм. CVAT позволяет экспортировать разметку напрямую в формат, совместимый с YOLO.

Поскольку CAPTCHA-изображения часто содержат объекты с нечёткими контурами, наложением и визуальными искажениями, особенно важно использовать ручную точную разметку, а не ограничиваться автоматическими методами. Выделение объектов должно проводиться как можно точнее, с учётом геометрии контуров. На слайде представлен пример изображения с размеченными объектами.

## Обучение модели на датасете

В качестве основной архитектуры была выбрана модель YOLOv8m-seg, поддерживающая сегментацию объектов.

Преимущества YOLOv8m-seg заключаются в следующем:

1. наличие встроенной поддержки сегментации объектов;
2. возможность использования предобученных весов;
3. высокая скорость инференса по сравнению с другими моделями сегментации;
4. встроенные средства аугментации;
5. удобный интерфейс через библиотеку ultralytics, позволяющий быстро запускать обучение, логировать метрики и визуализировать результаты;
6. полная совместимость с аннотациями в формате YOLO, полученными из CVAT.

Обучение проводилось на 35 эпохах при размере изображений 640×640 пикселей и размере батча 8. Использование предобученных весов позволило достичь стабильного снижения функции потерь с первых эпох, а встроенные механизмы аугментации способствовали улучшению обобщающей способности модели.

Результаты обучения отслеживались по ключевым метрикам (IoU, Precision, Recall, Loss), которые визуализировались автоматически. Примеры графиков с результатами обучения приведены на данных слайдах.

## Тестирование модели

После завершения обучения модель была протестирована на реальных CAPTCHA, собранных с помощью автоматического парсера, реализованного на базе библиотеки Selenium.

Тестирование было организовано в виде цикла, позволяющего автоматически проходить CAPTCHA до тех пор, пока не будет достигнут положительный результат. Это позволило зафиксировать частоту ошибок модели и определить случаи, в которых требуются дообучение или оптимизация.

Рабочий процесс тестирования и взаимодействия модели с CAPTCHA представлен на блок-схеме.

## Заключение

В рамках данной работы была реализована система автоматического распознавания и прохождения CAPTCHA с изображениями, основанная на использовании нейросетевой модели YOLOv8 с поддержкой сегментации и были решены все поставленныые задачи.