

Московский Государственный Университет  
имени М. В. Ломоносова  
Филиал в г. Ташкенте

Факультет Прикладной Математики и  
Информатики

---

**Тема:** Решение задачи семантического разделения  
совмещённых изображений с использованием искусственных  
нейронных сетей.

Студент: Шабанов Данил Сергеевич

Научный руководитель: к.ф.-м.н., с.н.с. Половников Владимир Сер-  
геевич

Работа просмотрена и рекомендована к защите: зав. кафедрой ПМИИ,  
к.ф.-м.н., профессор Гасанов Эльяр Эльдарович

Ташкент 2025

# Содержание

<b>1</b>	<b>Аннотация</b>	<b>1</b>
<b>2</b>	<b>Введение</b>	<b>1</b>
<b>3</b>	<b>Цели, задачи, возможности</b>	<b>3</b>
3.1	Цели . . . . .	3
3.2	Задачи . . . . .	3
3.3	Возможности . . . . .	4

# 1 Аннотация

В современном инфопространстве остро стоит тема применения возможностей нейронных сетей для решения множества задач. И для удовлетворения этого спроса требуется большой и гибкий ресурс для обучения различных моделей. Данная дипломная работа берет на себя решение задачи *семантического разделения совмещённых изображений*, что позволит «очищать» датасеты с изображениями от различных артефактов, что искажают распределение признаков и снижают качество обучаемых моделей компьютерного зрения. Детектируя и выделяя исходное и наложенное изображение, мы получаем возможность использовать «чистые» изображения для дальнейшего использования «сырых» данных при обучении систем искусственного интеллекта и сам выделенный артефакт для визуальной проверки качества процесса или изучения.

Предлагается двухстадийная архитектура U-Net с механизмом Attention Gate, обученная сначала на реальном наборе LSCWD, а затем дообученная на синтетически сгенерированных парах «фон/наложение». Для генерации таких пар разработан Python-скрипт, применяющий случайные повороты, прозрачность и JPEG-шум к любой базе. Практическая ценность этого решения заключается в возможности автоматически получать «очищенные» изображения и точные маски наложений, что делает его практичным инструментом для предварительной подготовки датасетов из нестерильных баз и их реставрации.

## 2 Введение

Обилие визуальных данных, доступных из открытых источников, сопровождается высокой долей изображений с наложенными графическими элементами, далее артефактами. Подобные артефакты искажают содержимое изображения и могут негативно влиять на процесс обучения нейронных сетей, приводя к деградации качества моделей компьютерного зрения, обученных на «загрязнённых» выборках. Поэтому актуальной задачей является разработка методов автоматического обнаружения и удаления таких наложенных артефактов на изображениях. Так же выявление данных артефактов дает возможность изучить природу данных помех, изучив артефакт, либо визуально выявить успешность процесса.

Решение данной задачи осложняется нехваткой специализированных

размеченных наборов данных и специфичность их составления в виду контекста конкретной задачи. Для обучения сегментационных моделей необходимы изображения с точной пиксельной разметкой областей, содержащих наложенные артефакты (например, маски, выделяющие положение артефакта). Однако ручная аннотация тысяч изображений трудоёмка и затратна. Этот дефицит данных существенно затрудняет применение современных методов глубокого обучения к различным задачам машинного зрения.

Предлагаемый в работе подход не ограничивается узкой задачей удаления артефактов на конкретной картинке. Он адресует более широкую проблему искаженных данных, когда искусственные элементы незаметно влияют на статистику выборки и транслируются далее в генеративные или классификационные модели. Автоматическое отделение наложения от фона позволяет не только очищать датасеты, но и создавать банк извлечённых артефактов — это даёт исследователям возможность анализировать причину зарождения помех и строить детекторы для будущих, ещё неизученных типов наложений. Таким образом, работа вносит вклад одновременно в качество исходных данных и в репертуар инструментов аудита визуальных коллекций.

В качестве инструментов и среды решения проблемы стали язык программирования Python, универсальных для работы всех составных частей проекта, что дает больший и простой размах использования результата и быстрое выполнение задачи.

Для преодоления дефицита данных разработан генератор синтетических пар «фон / наложение», реализованный на Python с использованием Pillow и OpenCV для программного взаимодействия со всеми популярными форматами изображений. Генератор накладывает PNG-объекты на фоновые снимки, управляя прозрачностью, перспективой и шумом, и автоматически формирует точные маски.

Сегментация выполняется моделью U-Net, показавшей успешность в пиксельных задачах. Реализованы две её модификации:

- **Базовая** — классический U-Net, обученный на датасете Large-scale Common Watermark Dataset (LSCWD), где исходная bbox-разметка, для унификации датасета конвертирована в бинарные маски скриптом приводящим yolo-формат к требуемым для обучения объектам;
- **Усовершенствованная** — та же U-Net-архитектура, но дополненная механизмом внимания Attention Gate, который фокусирует мо-

дель на областях наложения, и подходом восстанавливающим фон внутри маски. Модель до обучается на смеси реальных и синтетических изображений, благодаря чему надёжно сегментирует более нестандартные артефакты.

Все модели реализованы во фреймворке PyTorch, что и обучает нейросетевые модели архитектуры U-Net, инициализируя обучение как на CPU, так и на GPU устройствах. В рамках прикладного применения и демонстрации данного решения создан REST-сервис на FastAPI: он принимает изображение, возвращает маску наложения, сам артефакт и, при использовании улучшенной версии, восстановленный фон. Через тот же интерфейс доступен и генератор синтетики.

Результатом работы является воспроизводимый инструмент для автоматического разделения совмещённых изображений, пригодный для очистки больших датасетов и их последующего использования в задачах машинного зрения.

## 3 Цели, задачи, возможности

### 3.1 Цели

1. Разработать воспроизводимый метод семантического разделения совмещённых изображений, позволяющий автоматически получать бинарную маску наложения и восстановленный фон;
2. Продемонстрировать, что синтетически расширенный датасет улучшает качество сегментации по сравнению с обучением только на реальных данных;
3. Реализовать прикладной сервис, обеспечивающий интеграцию в существующие конвейеры обработки изображений.

### 3.2 Задачи

1. Построить генератор синтетических пар «фон / наложение» с автоматическим формированием пиксельных масок;
2. Конвертировать разметку набора LSCWD из YOLO-bbox в бинарные маски и сформировать обучающую и валидационную выборки;
3. Реализовать и обучить две версии U-Net (базовую и Attention Gate) с использованием PyTorch и выполнить эксперименты для оценки вклада синтетики и механизма внимания;
4. Интегрировать лучшую модель в FastAPI-сервис;
5. Оценить решение по mIoU, PSNR, SSIM и сопоставить результаты с существующими методами.

### 3.3 Возможности

- Автоматическая очистка больших изображенийных баз изображений от искажающих артефактов без ручной разметки;
- Формирование собственного синтетического датасета с контролем типа, прозрачности и геометрии наложений;
- Получение не только маски, но и отдельного слоя артефакта и реконструированного фона — удобно для визуальной экспертизы;
- Развёртывание сервиса «из коробки» (Docker), поддержка интерфейса на CPU и GPU, а также масштабирование через REST API;
- Расширяемость: возможность дообучения на новых типах наложений или адаптации под другие задачи сегментации.