МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

ДИФФУЗИОННАЯ МОДЕЛЬ ДЛЯ УСЛОВНОЙ ГЕНЕРАЦИИ РУКОПИСНЫХ СИМВОЛОВ С УЧЁТОМ СТИЛЕВЫХ И СЕМАНТИЧЕСКИХ АТРИБУТОВ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы
направления 02.03.02 — Фундаментальная информатика и информационные технологии
факультета компьютерных наук и информационных технологий
Ноздрова Егора Александровича

Научный руководитель:
доцент, к. ф.-м. н., доцент _____ А. С. Иванов
подпись, дата

Зав. кафедрой:
к. ф.-м. н., доцент _____ С. В. Миронов
подпись, дата

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы. Автоматическая генерация рукописного текста является одной из актуальных задач современного машинного обучения и компьютерного зрения. Интерес к данному направлению обусловлен широким спектром практических приложений: аугментация обучающих данных для систем оптического распознавания символов (OCR), синтез обучающих материалов, верификация личности по динамике почерка, а также создание персонализированных шрифтов для цифровой типографики. Несмотря на значительный прогресс в области синтеза изображений, задача управляемой генерации рукописных символов с одновременным и независимым контролем их семантики и стилевых характеристик остается открытой.

Цель бакалаврской работы – разработать и экспериментально оценить диффузионную модель для условной генерации рукописных символов кириллического алфавита с независимым управлением семантическим классом и стилем написания.

Поставленная цель определила **следующие задачи**:

1. Анализ существующих методов генеративного моделирования и обоснование выбора диффузионного подхода.
2. Постановка задачи условной генерации и формализация математической модели.
3. Проектирование архитектуры DiT с механизмами внедрения стилевых и семантических условий.
4. Реализация полного цикла обучения, включая косинусное расписание шума, EMA-усреднение весов и метод DDIM-сэмплинга.
5. Количественная оценка качества генерации по метрикам FID и Classification Accuracy.

Методологические основы генеративного моделирования представлены в работах зарубежных исследователей. Основы диффузионных вероятностных моделей разработаны в трудах Джонатана Хо, Якоба Соль-Дикстейна, Алекса Никола, Прафуллы Дхариваля и Цзяминя Сонга, заложивших математический аппарат прямого и обратного диффузионных процессов и методы ускоренного сэмплинга. Архитектурные принципы трансформерных моделей и их применение к задачам компьютерного зрения отражены в работах Ашиша Васвани, Алексея Досовицкого, Уильяма Пиблза и Сэйнин Се. Вопросы условной генерации и управления генеративными моделями исследованы в работах Иэна Гудфеллоу, Дидерика Кингмы и Джонатана Хо. Методы оценки качества генеративных моделей по метрике FID разработаны Мартином Хойзелем с соавторами..

Теоретическая значимость бакалаврской работы заключается в систематизации и сравнительном анализе подходов к независимому добавлению условий в архитектуру диффузионных трансформерных моделей. Детально изложен математический аппарат прямого и обратного диффузионных процессов, а также механизм адаптивной нормализации слоёв (adaLN), обеспечивающий раздельное внедрение семантических и стилевых атрибутов в единую архитектуру денойзера.

Практическая значимость бакалаврской работы состоит в разработке готовой модели, способной генерировать неограниченное количество уникальных рукописных символов кириллического алфавита с контролируемым стилем написания, включая возможность смешения стилей. Разработанное решение применимо для аугментации обучающих данных систем оптического распознавания рукописного текста (OCR), синтеза персонализированных шрифтов и создания обучающих материалов.

Структура и объём работы. Бакалаврская работа состоит из введения, 2 разделов, заключения, списка использованных источников и 5 приложений.

Общий объем работы – 95 страниц, из них 63 страницы – основное содержание, включая 8 рисунков и 3 таблицы, цифровой носитель в качестве приложения, список использованных источников информации – 23 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретические основы диффузионных моделей» посвящён формализации задачи условной генерации рукописных символов и анализу теоретического аппарата диффузионных вероятностных моделей.

В разделе вводится формальная постановка задачи: обучающая выборка представляется множеством троек «изображение — метка класса — метка стиля», а цель состоит в обучении генеративной модели, обеспечивающей ортогональное управление этими двумя атрибутами. Проведён сравнительный анализ четырёх основных семейств генеративных моделей — вариационных автоэнкодеров, нормализующих потоков, генеративно-состязательных сетей и диффузионных моделей, — на основании которого обоснован выбор диффузионного подхода. Подробно рассмотрен математический аппарат прямого марковского процесса зашумления, выведена замкнутая формула репараметризации, позволяющая получать зашумлённое состояние напрямую из исходного изображения. Проанализированы линейное и косинусное расписания шума, показано преимущество последнего за счёт более плавного разрушения сигнала на ранних шагах. Рассмотрен обратный генеративный процесс и вывод упрощённой функции потерь на основе предсказания шума. Отдельное внимание уделено методу ускоренного сэмплинга DDIM, позволяющему сократить число шагов генерации с тысячи до нескольких десятков за счёт перехода к немарковской формулировке процесса с сохранением маргинальных распределений. В завершение раздела рассмотрена

архитектура Vision Transformer и обоснована целесообразность перехода от свёрточных архитектур к трансформерным денойзерам ввиду их глобального рецептивного поля и предсказуемой масштабируемости.

Второй раздел «Построение диффузионной модели» посвящён практической реализации диффузионной модели на основе архитектуры Diffusion Transformer (DiT) и экспериментальной оценке её качества.

В разделе подробно описан используемый датасет рукописной кириллицы CoMNIST, включающий 15480 изображений 33 букв в 10 стилях написания; ввиду отсутствия в датасете готовой стилевой разметки самостоятельно реализована процедура автоматической группировки образцов почерка методом кластеризации KMeans. Спроектирована и реализована архитектура DiT, включающая операцию патчификации изображения, 12 трансформерных блоков с механизмом self-attention и адаптивной нормализацией слоёв (adaLN-Zero), а также механизм раздельного внедрения векторов условий — метки семантического класса, метки стиля и временного шага — через суммарный вектор обусловливания. Реализован полный цикл обучения модели, включающий косинусное расписание шума, оптимизатор AdamW, смешанную точность вычислений FP16 и экспоненциальное скользящее усреднение весов (EMA) для повышения устойчивости генерации. Проведена количественная оценка качества обученной модели по двум независимым метрикам: Fréchet Inception Distance (FID = 27.0), отражающей качество и разнообразие распределения сгенерированных изображений, и Classification Accuracy (76.5%), отражающей семантическую точность воспроизведения заданного класса символа. Продемонстрирована возможность смешения стилей за счёт интерполяции стилевых эмбеддингов, а также проведён анализ неудачных случаев генерации, связанных с визуально похожими парами символов при ограниченном разрешении изображений.

ЗАКЛЮЧЕНИЕ

В рамках настоящей выпускной квалификационной работы была спроектирована, реализована и экспериментально проверена диффузионная вероятностная модель для условной генерации рукописных символов кириллического алфавита с независимым управлением семантическим классом и стилем написания.

В теоретической части проведён сравнительный анализ основных семейств генеративных моделей - VAE, нормализующих потоков, GAN и диффузионных моделей - и обоснован выбор диффузионного подхода. Детально изложен математический аппарат прямого и обратного диффузионных процессов, проанализированы линейное и косинусное расписания шума, рассмотрены методы classifier-free guidance и ускоренного сэмплинга DDIM.

В практической части реализован полный обучающий конвейер на основе архитектуры Diffusion Transformer (DiT) с механизмом двойного добавления условия через блоки адаптивной нормализации слоев (AdaLN). Разработана процедура раздельного внедрения метки семантического класса и метки стиля, а также применено EMA-усреднение весов для повышения стабильности генерации.

Оценка обученной модели проводилась по двум метрикам. Достигнутое значение Classification Accuracy = 0.765 свидетельствует о том, что в 76.5% случаев модель корректно воспроизводит запрошенный символ, а затруднения связаны преимущественно с визуально схожими парами букв при малом разрешении 32 x 32. Значение FID = 27.0 подтверждает близость статистического распределения сгенерированных изображений к реальным, что является приемлемым результатом для узкоспециализированного малоресурсного датасета.

Полученные результаты подтверждают принципиальную применимость архитектуры DiT с двойным добавлением условия для решения задачи независимого управления семантикой и стилем при синтезе рукописных символов. Перспективными направлениями дальнейшего развития работы являются увеличение разрешения генерируемых изображений, переход к непрерывным стилевым эмбеддингам и исследование методов быстрого дистилляционного сэмплинга для сокращения времени инференса.

Отдельные части бакалаврской работы не были опубликованы и не представлялись на конференциях.

Основные источники информации:

1. Ho, J., Jain, A., Abbeel, P. Denoising Diffusion Probabilistic Models [Электронный ресурс]. — URL: https://arxiv.org/pdf/2006.11239 (Дата обращения 22.04.2026). Загл. с экр. Яз. англ.
2. Song, J., Meng, C., Ermon, S. Denoising Diffusion Implicit Models [Электронный ресурс]. — URL: https://arxiv.org/pdf/2010.02502 (Дата обращения 23.04.2026). Загл. с экр. Яз. англ.
3. Nichol, A., Dhariwal, P. Improved Denoising Diffusion Probabilistic Models [Электронный ресурс]. — URL: https://arxiv.org/pdf/2102.09672 (Дата обращения 23.04.2026). Загл. с экр. Яз. англ.
4. Peebles, W., Xie, S. Scalable Diffusion Models with Transformers [Электронный ресурс]. — URL: https://openaccess.thecvf.com/content/ICCV2023/papers/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.pdf (Дата обращения 27.04.2026). Загл. с экр. Яз. англ.

5. Vaswani, A. et al. Attention Is All You Need [Электронный ресурс]. — URL: https://arxiv.org/pdf/1706.03762 (Дата обращения 25.04.2026). Загл. с экр. Яз. англ.
6. Dosovitskiy, A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Электронный ресурс]. — URL: https://arxiv.org/pdf/2010.11929 (Дата обращения 25.04.2026). Загл. с экр. Яз. англ.
7. Ho, J., Salimans, T. Classifier-Free Diffusion Guidance [Электронный ресурс]. — URL: https://arxiv.org/pdf/2207.12598 (Дата обращения 27.04.2026). Загл. с экр. Яз. англ.
8. Goodfellow, I. et al. Generative Adversarial Nets [Электронный ресурс]. — URL: https://arxiv.org/pdf/1406.2661 (Дата обращения 25.04.2026). Загл. с экр. Яз. англ.