

doi: 10.17586/2226-1494-2025-25-4-676-683

УДК 004.89

## Генерация шрифтов на основе анализа стиля и структуры символов с использованием диффузионных моделей

Максим Игоревич Маслов<sup>1</sup>, Анна Евгеньевна Авдюшина<sup>2</sup>,  
Мария Александровна Солодкая<sup>3</sup>, Александр Владимирович Кугаевских<sup>4</sup>✉

<sup>1</sup> ООО «Нанософт Разработка», 108811, Москва, Российская Федерация

<sup>1,2,3,4</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

<sup>2</sup> АО «Центр исследований и разработок», Москва, 101000, Российская Федерация

<sup>1</sup> [maxmiriko@gmail.com](mailto:maxmiriko@gmail.com), <https://orcid.org/0009-0005-8198-4703>

<sup>2</sup> [avdiushina@itmo.ru](mailto:avdiushina@itmo.ru), <https://orcid.org/0000-0002-8235-902X>

<sup>3</sup> [msolodkaya@itmo.ru](mailto:msolodkaya@itmo.ru), <https://orcid.org/0009-0000-7047-6862>

<sup>4</sup> [avkugaevskikh@itmo.ru](mailto:avkugaevskikh@itmo.ru)✉, <https://orcid.org/0000-0002-6676-0518>

### Аннотация

**Введение.** Рассмотрена роль генеративных нейросетей в разработке и оптимизации шрифтов, которые играют ключевую роль в создании эстетически привлекательных и функциональных дизайнов. Основное внимание уделено лицензионным ограничениям и недостаточной доступности шрифтов для различных мировых языков, что создает сложности для дизайнеров и типографов в процессе создания текстовых материалов. Новизна подхода заключается в применении диффузионной модели в качестве генеративной нейронной сети для автоматического создания шрифтов, включая недостающие глифы для языков, не поддерживаемых стандартными шрифтами. **Метод.** Разработана диффузионная модель, представляющая собой алгоритм генерации шрифтов на основе анализа закономерностей структуры символов и логики их построения. Модель интегрирована в программное решение, которое автоматизирует процесс создания шрифтовых макетов, позволяя пользователям генерировать новые глифы и шрифты с учетом специфических языковых потребностей. Методика включает предварительную подготовку данных, обучение сети и последующую генерацию символов, имитирующих стиль и композицию исходных шрифтов. **Основные результаты.** В ходе экспериментов диффузионная модель продемонстрировала высокую способность к генерации качественных шрифтовых символов, визуально схожих с оригинальными образцами. В качестве исходных данных использовались шрифтовые наборы с ограниченным набором символов, что позволило оценить возможности модели по созданию недостающих глифов для различных языков. Результаты показали, что разработанная модель успешно воспроизводит стилистические особенности оригинального шрифта, что подтверждает ее потенциал для создания универсальных шрифтовых решений, адаптированных к глобальным потребностям. **Обсуждение.** Предложенная методика генерации шрифтов представляет интерес для специалистов в области дизайна, типографики и разработки текстовых материалов, предназначенных для различных языковых аудиторий. Полученные результаты могут быть полезны при создании шрифтов для использования в многоязычных проектах, особенно если возникает потребность в добавлении недостающих символов.

### Ключевые слова

генерация шрифтов, диффузионные модели, синтез изображений, шрифтовой дизайн, стиль шрифтов

### Благодарности

Работа выполнена в рамках государственного задания (проект FSER-2025-0004).

**Ссылка для цитирования:** Маслов М.И., Авдюшина А.Е., Солодкая М.А., Кугаевских А.В. Генерация шрифтов на основе анализа стиля и структуры символов с использованием диффузионных моделей // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 4. С. 676–683. doi: 10.17586/2226-1494-2025-25-4-676-683

## Font generation based on style and character structure analysis using diffusion models

**Maksim I. Maslov<sup>1</sup>, Anna E. Avdyushina<sup>2</sup>, Maria A. Solodkaya<sup>3</sup>, Alexander V. Kugaevskikh<sup>4✉</sup>**

<sup>1</sup> LLC "Nanosoft Razrabotka", 108811, Moscow, Russian Federation

<sup>1,2,3,4</sup> ITMO University, Saint Petersburg, 197101, Russian Federation

<sup>2</sup> JSC "Research and development center", Moscow, 101000, Russian Federation

<sup>1</sup> maxmirko@gmail.com, <https://orcid.org/0009-0005-8198-4703>

<sup>2</sup> avdyushina@itmo.ru, <https://orcid.org/0000-0002-8235-902X>

<sup>3</sup> mslodkaya@itmo.ru, <https://orcid.org/0009-0000-7047-6862>

<sup>4</sup> avkugaevskikh@itmo.ru✉, <https://orcid.org/0000-0002-6676-0518>

### Abstract

The article discusses the role of generative neural networks in the development and optimization of fonts which play a key role in creating aesthetically attractive and functional designs. The main attention is paid to licensing restrictions and insufficient availability of fonts for various world languages, which creates difficulties for designers and typographers in the process of creating text materials. The novelty of the approach lies in the use of the diffusion model as a generative neural network for automatic font creation, including missing glyphs for languages not supported by standard fonts. To solve the tasks set, a diffusion model has been developed which is an algorithm for generating fonts based on the analysis of patterns in the structure of symbols and the logic of their construction. The model is integrated into an application that automates the process of creating font layouts, allowing users to generate new glyphs and fonts tailored to specific language needs. This technique includes preliminary data preparation, network training, and subsequent character generation that mimic the style and composition of the original fonts. During the experiments, the diffusion model demonstrated a high ability to generate high-quality font characters visually similar to the original samples. Font sets with a limited set of characters were used as source data, which allowed us to evaluate the capabilities of the model to create missing glyphs for various languages. The results showed that the developed model successfully reproduces the stylistic features of the original font, which confirms its potential for application in the development of font solutions for global use. The proposed method of font generation is of interest to specialists working in the field of design, typography, and the creation of text materials for various language audiences. The results obtained can be useful when creating fonts intended for use in multilingual projects that require the presence of missing characters.

### Keywords

font generation, diffusion models, image synthesis, font design, font style

### Acknowledgements

The work was carried out within the framework of the State Assignment (project No. FSER-2025-0004).

**For citation:** Maslov M.I., Avdyushina A.E., Solodkaya M.A., Kugaevskikh A.V. Font generation based on style and character structure analysis using diffusion models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 4, pp. 676–683 (in Russian). doi: 10.17586/2226-1494-2025-25-4-676-683

### Введение

В современном мире компьютерных технологий разработка и оптимизация шрифтов играют важную роль в формировании эстетически привлекательных и функциональных дизайнов [1]. Однако использование шрифтов ограничено различными факторами. В первую очередь, шрифты являются интеллектуальной собственностью компаний, и отдаются в аренду на определенный срок, а также тип соглашения контролирует форму использования шрифта. Так, например, если лицензия позволяет использовать его только на веб-сайте, то заказчик не имеет права его использовать на печати или модифицировать шрифт для применения в других форматах. Определенной проблемой безусловно также является и факт, что большинство шрифтов, кроме лицензионных условий, имеют ограничение в доступности символов разных шрифтов под различные языки. Подавляющее большинство цифровых шрифтов, распространяемых по сети Интернет, рассчитаны на английский язык, и помимо самых популярных и известных, не предоставляют возможность написания текста с использованием символов из расширенной кириллицы, латиницы и других типов буквенных или символьных письменностей. Это может затруднить процесс создания и визуального оформления текстов для публичных материалов.

Современные генеративные нейросетевые модели представляют собой универсальный инструмент прототипирования для дизайнеров и типографов. Это позволяет значительно сократить время, затрачиваемое на разработку шрифтовых макетов, а также предоставляет существенную вариативность, что добавляет многим вариантам уникальности. Генеративные сети способны выявлять закономерности в структуре символов и композиции шрифтов, а также аппроксимировать логику построения глифов для создания недостающих букв для разных мировых языков.

В настоящей работе предложена диффузионная модель, использующаяся в качестве генеративной нейронной сети для автоматической генерации шрифтов. На базе данной модели разработано программное решение, которое реализует автоматизацию процесса создания недостающих глифов и улучшение стилистической консистентности шрифтовых наборов. Анализ основных принципов и характеристик модели демонстрирует ее потенциал для решения задач генерации шрифтов, а также перспективы применения в создании качественных шрифтовых решений для многоязычных текстов.

Цель работы — автоматизация процесса создания шрифтов.

## Обзор тематических научных работ

Современные исследования показывают значительный прогресс в применении генеративных нейронных сетей для создания шрифтов. Одним из наиболее устоявшихся подходов является использование архитектуры U-Net [1]. В исследованиях [2, 3] предложен подход DeepVecFont, в котором применяется двойная модальность для учета текстуры и стиля, что обеспечивает улучшение стилистической согласованности векторных шрифтов. Данный метод получил высокие результаты в задачах генерации высокоточных шрифтов, сохраняя стилистическое соответствие оригиналам.

Особый интерес представляет работа [4], в которой разработана модель FontDiffuser, использующая диффузионные процессы для создания глифов с высокой детализацией. Для художественной генерации шрифтов в исследовании [5] предложен метод GenText, основанный на разделении управления текстурой и стилем шрифтов.

В [6] представлен Generative Adversarial Networks (GAN)-ориентированный подход StrokeGAN, направленный на генерацию китайских иероглифов. Использование штрихового кодирования позволило избежать повторения однородных образцов и повысить разнообразие создаваемых символов.

Также в работах [7, 8] исследована методика Few-Shot генерации шрифтов, включая VQ-Font. Эти модели демонстрируют способность создавать шрифты на основе ограниченного количества примеров, что особенно актуально для редких языков и новых стилей.

В дополнение к этим подходам в [9] описан метод интеграции диффузионных процессов с кодированием краев, что позволило добиться улучшения визуальной привлекательности шрифтов. Аналогично, работа [10] показывает, что многомасштабная агрегация контента

в диффузионных моделях способствует повышению точности и детализации создаваемых символов.

Таким образом, учитывая крайнюю необходимость в высокой детализации сгенерированных изображений, их визуальную привлекательность и стилевую согласованность для всех символов шрифта, модели на основе автоэнкодеров менее применимы в рассматриваемых задачах. Исходя из этого, разработка универсального решения на основе диффузионной модели, способного генерировать недостающие символы с сохранением стилистической согласованности, является актуальной задачей.

## Архитектура решения

Разработанная архитектура нейронной сети представляет собой универсальное решение для генерации шрифтов с учетом специфики недостающих символов. Основной акцент сделан на интеграции современных подходов, таких как диффузионные процессы, сверточные нейронные сети и механизмы внимания, что обеспечивает высокую детализацию и стилистическую согласованность глифов. Общая архитектура модели представлена на рис. 1.

**Основной генеративный блок.** В основе модели лежит модифицированная диффузионная архитектура U-Net, которая адаптирована для задач генерации шрифтов. Для повышения эффективности работы модели были добавлены следующие ключевые компоненты:

— слои внимания (Self-Attention) обеспечивают выявление закономерностей между различными позициями входных данных, что позволяет улучшить стилистическую согласованность символов. Подход основан на работе [11], где рассмотрено моделирование сложных зависимостей;

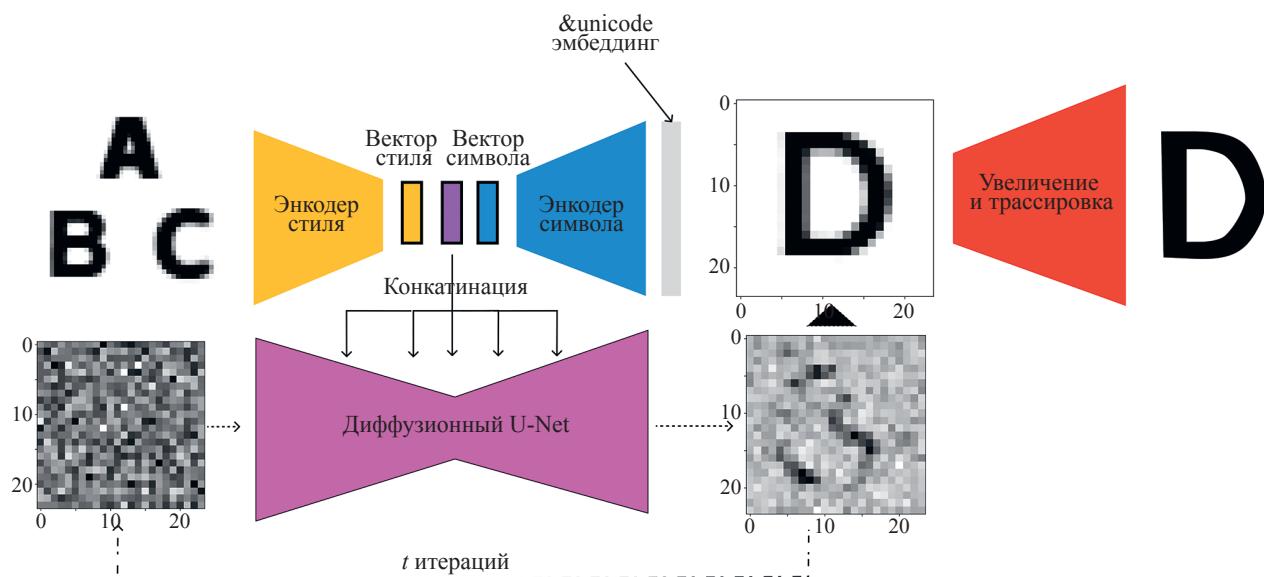


Рис. 1. Общая архитектура сети.

A, B, C и D — примеры символов шрифта для показа стилевого кодирования (примеры входных и выходных данных)

Fig. 1. Neural network architecture.

A, B, C and D are examples of font characters to show the style coding (examples of input and output data)

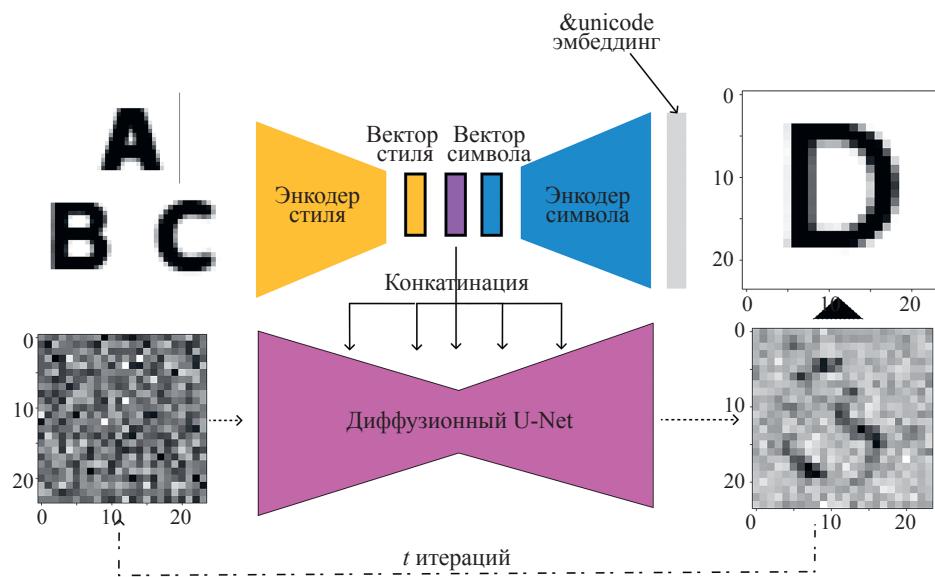


Рис. 2. Архитектура генеративного блока  
Fig. 2. Architecture of the generative block

- синусоидальные позиционные эмбеддинги кодируют временные параметры и повышают устойчивость модели к шумам [12];
- групповая нормализация стабилизирует процесс обучения, что особенно важно для задач генерации на основе диффузионных процессов [13].

Модель принимает на вход трехканальное изображение размером  $32 \times 32$  пикселей и генерирует глифы, соответствующие заданным стилем и классовым условиям. Уникальной особенностью является возможность использования стилевых векторов для сохранения оригинального дизайна шрифта. Входные данные подаются через два энкодера: стилевой и классовый (рис. 2).

**Архитектура энкодера для стиля.** Стилевой энкодер разработан на базе U-Net, представленного в работе [1], и позволяет выделять ключевые визуальные признаки, необходимые для стилистического соответствия генерируемых глифов. Основные компоненты: сверточные слои извлекают важные визуальные характеристики из входных данных [14]; MaxPooling уменьшает размерность изображения для упрощения обработки и ускорения обучения.

Выходной стилевой вектор размером  $4 \times 4 \times 256$  пикселей используется для условной генерации. Такая структура позволяет модели учитывать особенности оригинального шрифта и адаптировать их для недостающих символов. Полная архитектура энкодера стиля изображена на рис. 3.

**Архитектура энкодера для класса.** Классовый энкодер (рис. 4) отвечает за кодирование информации о типе символа, что критически важно для генерации точных и стилистически согласованных глифов. Он состоит из последовательности сверточных слоев, которые выделяют ключевые признаки, необходимые для определения класса символа. Данный подход основан на работах [2, 3], где использовалась двойная модальность для учета текстуры и стиля.

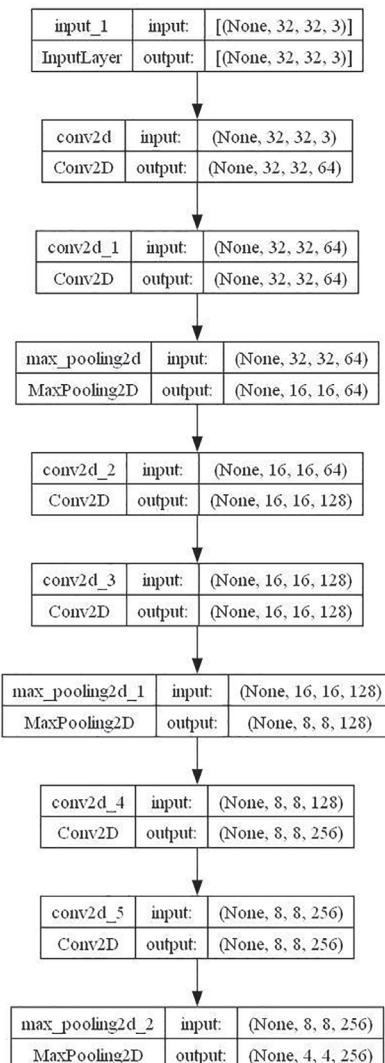


Рис. 3. Архитектура стилевого энкодера  
Fig. 3. Style Encoder Architecture

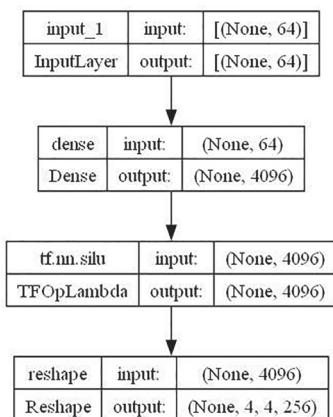


Рис. 4. Архитектура классового энкодера  
Fig. 4. Class Encoder Architecture

**Блок увеличения разрешения.** Для улучшения визуального качества генерируемых глифов применяется модель увеличения разрешения, основанная на Efficient Sub-Pixel Convolutional Neural Network, как описано в [15]. Эта модель позволяет увеличивать изображение с  $32 \times 32$  до  $64 \times 64$  пикселей, обеспечивая высокую детализацию. Основные элементы блока: сверточные слои извлекают признаки на каждом этапе обработки; Depth-to-Space слой объединяет карты признаков для формирования высококачественного изображения.

Архитектура блока показана на рис. 5.

**Интеграция диффузионных процессов и функции потерь.** Диффузионные процессы, реализованные по представленным методикам в работах [16, 17], позволяют модели эффективно устранять шум, добавленный на этапе генерации. Для обучения модели была выбрана комбинация функций потерь L1 Loss и Perceptual Loss, что позволяет достичь более высокой визуальной согласованности [18].

Предложенная архитектура представляет собой универсальное решение для задач генерации шрифтов. Она успешно сочетает диффузионные процессы, стилевые и

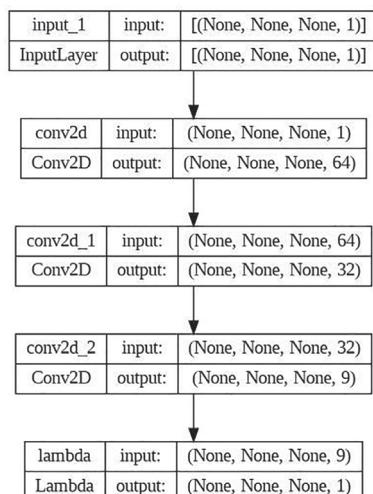


Рис. 5. Архитектура увеличивающего блока  
Fig. 5. Upscale block Architecture

классовые энкодеры, а также блок увеличения разрешения для создания недостающих глифов с высокой степенью стилистической согласованности. Применение синусоидальных эмбеддингов и функций потерь с учетом восприятия позволяет улучшить качество итоговых изображений, открывая новые возможности для использования в многоязычных проектах и дизайне.

### Экспериментальная проверка

**Данные.** Для обучения модели использовались растровые изображения размером  $32 \times 32$  пикселей, представляющие 11 756 шрифтов для латинского алфавита. Обучающая выборка составила 80 % данных, валидационная — 10 %, тестовая — 10 %. Модель обучалась в течение 12 ч на NVIDIA GeForce RTX 3060 Laptop GPU с использованием PyTorch 1.13.0. Размер батча был установлен на 32, скорость обучения —  $2 \cdot 10^{-4}$ , количество эпох — 250. Функции потерь включали L1 Loss и Perceptual Loss, а в качестве оптимизатора использовался Adam. В модели применено 8 групп в GroupNormalization слое, два резидентных блока, размеры каналов [64, 128, 256, 512] на различных уровнях нейронной сети, основное внимание уделено на уровнях с каналами 256 и 512.

**Примеры генерируемых символов.** Для оценки работы модели были сгенерированы символы для различных шрифтов и стилей. На рис. 6 показан полный набор символов, созданный моделью, включая недостающие глифы. Для сравнения приведены оригинальные символы и результаты генерации.

**Текстовые примеры.** Для демонстрации сгенерированного шрифта моделью в реальных условиях были созданы примеры, включающие гарнитуру, отдельные слова и абзацы текста.

На рис. 7 показаны отдельные символы, демонстрирующие консистентность дизайна.

**Слова.** Пример текста «парашютист» приведен на рис. 8. Сочетание символов демонстрирует, как модель сохраняет стиль.

**Оценка модели.** Для оценки качества сгенерированных шрифтов проведена субъективная оценка группой



Рис. 6. Сравнение оригинальных (a) и сгенерированных (b) символов  
Fig. 6. Comparison of the original (a) and generated (b) characters

**A B C D E F G H I J  
K L M N O P Q R S T  
U V W X Y Z**

Рис. 7. Отдельные сгенерированные символы  
Fig. 7. Separate generated characters

# ПАРАШЮТИСТ

Рис. 8. Пример слова, набранного с использованием сгенерированного шрифта

Fig. 8. An example of a word typed using a generated font

профессиональных шрифтовых дизайнеров. Основное внимание уделялось следующим критериям:

- стилистическая консистентность (насколько символы в шрифте соответствуют единому стилю);
- кернинг и трекинг (насколько гармонично символы сочетаются друг с другом в парах и словах);
- плотность текста (визуальная равномерность текста, набранного с использованием сгенерированного шрифта).

Было сгенерировано 1150 символов, среднее время генерации составило 3,6 секунд на символ. В исследовании приняли участие 15 профессиональных дизайнеров. Для оценки использовалась шкала от 1 до 10, где 1 означает «неудовлетворительно», а 10 — «отлично». Усредненные оценки дизайнеров приведены в таблице.

Проведенные эксперименты подтвердили высокую эффективность предложенной модели для задач генерации шрифтов. На основании полученных результатов можно сделать следующие выводы:

- 1) разработанная модель показала способность сохранять единый стиль символов в рамках шрифтового набора, однако были выявлены случаи, когда отдельные символы, такие как «ш» и «ю» (рис. 9), выбивались из общего стилистического ряда, что указывает на необходимость доработки алгоритма обработки особенностей сложных символов;
- 2) модель демонстрирует хорошую работу с межбуквенными расстояниями. Тем не менее, выявлены недостатки в обработке определенных пар симво-

Таблица. Средние оценки по критериям

Table. Average scores according to criteria

| Критерии                       | Средняя оценка | Стандартное отклонение |
|--------------------------------|----------------|------------------------|
| Стилистическая согласованность | 5,7            | 1,7                    |
| Кернинг/межбуквенные интервалы | 7,3            | 1,3                    |
| Плотность текста               | 5,0            | 1,5                    |

# ПАРАШЮТИСТ

Рис. 9. Пример слова, с не консистентными символами «ш» и «ю»

Fig. 9. Example of a word with inconsistent “ш” and “ю” characters

лов, например, «AV» и «То». Это указывает на необходимость улучшения механизма регулировки кернинга, особенно для пар с высокой частотой использования;

- 3) основные проблемы связаны с визуальным шумом из-за избыточных штрихов и недостаточной равномерностью толщины линий.

Полученные результаты подтверждают, что разработанная архитектура успешно решает задачу генерации шрифтов, включая восполнение недостающих символов и адаптацию для различных стилей. Тем не менее, выявленные замечания дизайнеров указывают на необходимость повышения стилистической однородности символов, включая доработку обработки сложных букв, оптимизацию алгоритма кернинга для улучшения межбуквенных расстояний в парах букв и текстах.

## Заключение

Проведенные исследования и эксперименты показали, что предложенная архитектура нейронной сети, основанная на диффузионных процессах, успешно решает задачу генерации недостающих символов в шрифтовых наборах. Разработанная модель демонстрирует высокую стилистическую консистентность символов, что подтверждается оценкой профессиональных шрифтовых дизайнеров. Результаты экспериментов показали, что использование стилевых и классовых энкодеров в сочетании с диффузионной архитектурой U-Net позволяет создавать глифы, соответствующие исходному стилю шрифта, даже для языков с недостаточной шрифтовой поддержкой.

Особое внимание уделено анализу кернинга и трекинга, где модель продемонстрировала гармоничное сочетание символов в парах букв и словах. Примеры текста, набранного с использованием сгенерированных шрифтов, показали высокую визуальную плотность и читабельность. Тем не менее, выявлены некоторые ограничения, включая трудности с генерацией символов для шрифтов с сильно выраженным декоративными элементами и нестандартными структурами.

Дальнейшая доработка модели может быть направлена на расширение набора данных за счет более разнообразных шрифтов, интеграцию техник Few-Shot Learning для улучшения генерации символов при ограниченном объеме данных, оптимизацию кернинга и трекинга через внедрение дополнительных метрик качества в процесс обучения.

Предложенная архитектура имеет значительный потенциал для практического применения, включая автоматизацию шрифтового дизайна, создание многоязычных текстовых решений и адаптацию шрифтов для специализированных проектов. Результаты исследования подчеркивают важность интеграции современных подходов, таких как диффузионные модели, для решения задач, связанных с генерацией высококачественных шрифтов.

## Литература

1. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // *Lecture Notes in Computer Science*. 2015. V. 9352. P. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
2. Wang Y., Lian Z. DeepVecFont: synthesizing high-quality vector fonts via dual-modality learning // *ACM Transactions on Graphics (TOG)*. 2021. V. 40. N 6. P. 1–15. <https://doi.org/10.1145/3478513.3480488>
3. Wang Y., Wang Y., Yu L., Zhu Y., Lian Z. DeepVecFont-v2: Exploiting Transformers to Synthesize Vector Fonts with Higher Quality // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023. P. 18320–18328. <https://doi.org/10.1109/CVPR52729.2023.01757>
4. Yang Z., Peng D., Kong Y., Zhang Y., Yao C., Jin L. FontDiffuser: One-shot font generation via denoising diffusion with multi-scale content aggregation and style contrastive learning // Proc. of the AAAI Conference on Artificial Intelligence. 2024. V. 38. N 7. P. 6603–6611. <https://doi.org/10.1609/aaai.v38i7.28482>
5. Huang Q., Fu B., Zhang A., Qiao Y. GenText: Unsupervised artistic text generation via decoupled font and texture manipulation // *arXiv*. 2022. [arXiv:2207.09649](https://doi.org/10.48550/arXiv.2207.09649). <https://doi.org/10.48550/arXiv.2207.09649>
6. Zeng J., Chen Q., Liu Y., Wang M., Yao Y. StrokeGAN: Reducing mode collapse in Chinese font generation via stroke encoding // *arXiv*. 2020. [arXiv:2012.08687](https://doi.org/10.48550/arXiv.2012.08687). <https://doi.org/10.48550/arXiv.2012.08687>
7. Park S., Chun S., Cha J., Lee B., Shim H. Few-shot font generation with localized style representations and factorization // Proc. of the AAAI Conference on Artificial Intelligence. 2021. V. 35. N 3. P. 2393–2402. <https://doi.org/10.1609/aaai.v35i3.16340>
8. Yao M., Zhang Y., Lin X., Li X.; Zuo W. VQ-Font: Few-shot font generation with structure-aware enhancement and quantization // Proc. of the AAAI Conference on Artificial Intelligence. 2024. V. 38. N 15. P. 16407–16415. <https://doi.org/10.1609/aaai.v38i15.29577>
9. Ding M. An edge-directed diffusion equation-based image restoration approach for font generation // *IEEE Access*. 2023. V. 11. P. 141435–141444. <https://doi.org/10.1109/ACCESS.2023.3342026>
10. Jeong J., Shin J. Multi-scale diffusion denoised smoothing // Proc. of the 37<sup>th</sup> International Conference on Neural Information Processing Systems. 2023. P. 67374–67397.
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need // *Advances in Neural Information Processing Systems*. 2017. V. 30. P. 1–11.
12. Voronov G., Lighthead R., Davison J., Kretler C.A., Healey D., Butler T. Multi-scale sinusoidal embeddings enable learning on high resolution mass spectrometry data // *arXiv*. 2022. [arXiv:2207.02980](https://doi.org/10.48550/arXiv.2207.02980). <https://doi.org/10.48550/arXiv.2207.02980>
13. Dhariwal P., Nichol A. Diffusion models beat GANs on image synthesis // *Advances in Neural Information Processing Systems*. 2021. V. 34. P. 8780–8794.
14. Convolutional Layer – Building Block of CNNs // *Towards Data Science*. 2024 [Электронный ресурс]. URL: <https://towardsdatascience.com/convolutional-layer-building-block-of-cnns-501b5b643e7b> (дата обращения: 30.01.2024).
15. Xu M., Du X., Wang D. Super-resolution restoration of single vehicle image based on ESPCN-VISR model // *IOP Conference Series: Materials Science and Engineering*. 2020. V. 790. N 1. P. 012107. <https://doi.org/10.1088/1757-899X/790/1/012107>
16. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // Proc. of the 34<sup>th</sup> International Conference on Neural Information Processing Systems. 2020. P. 6840–6851.
17. Nichol A.Q., Dhariwal P. Improved denoising diffusion probabilistic models // Proc. of the 38<sup>th</sup> International Conference on Machine Learning. 2021. V. 139. P. 8162–8171.
18. Lin S., Yang X. Diffusion model with perceptual loss // *arXiv*. 2023. [arXiv:2401.00110](https://doi.org/10.48550/arXiv.2401.00110). <https://doi.org/10.48550/arXiv.2401.00110>

## Авторы

**Маслов Максим Игоревич** — программист, ООО «Нанософт Разработка», 108811, Москва, Российская Федерация; студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0009-0005-8198-4703>, maxmiriko@gmail.com

## References

1. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 2015, vol. 9352, pp. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
2. Wang Y., Lian Z. DeepVecFont: synthesizing high-quality vector fonts via dual-modality learning. *ACM Transactions on Graphics (TOG)*, 2021, vol. 40, no. 6, pp. 1–15. <https://doi.org/10.1145/3478513.3480488>
3. Wang Y., Wang Y., Yu L., Zhu Y., Lian Z. DeepVecFont-v2: Exploiting Transformers to Synthesize Vector Fonts with Higher Quality. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023. P. 18320–18328. <https://doi.org/10.1109/CVPR52729.2023.01757>
4. Yang Z., Peng D., Kong Y., Zhang Y., Yao C., Jin L. FontDiffuser: One-shot font generation via denoising diffusion with multi-scale content aggregation and style contrastive learning. *Proc. of the AAAI Conference on Artificial Intelligence*. 2024. V. 38. N 7. P. 6603–6611. <https://doi.org/10.1609/aaai.v38i7.28482>
5. Huang Q., Fu B., Zhang A., Qiao Y. GenText: Unsupervised artistic text generation via decoupled font and texture manipulation. *arXiv*, 2022, [arXiv:2207.09649](https://doi.org/10.48550/arXiv.2207.09649). <https://doi.org/10.48550/arXiv.2207.09649>
6. Zeng J., Chen Q., Liu Y., Wang M., Yao Y. StrokeGAN: Reducing mode collapse in Chinese font generation via stroke encoding. *arXiv*, 2020, [arXiv:2012.08687](https://doi.org/10.48550/arXiv.2012.08687). <https://doi.org/10.48550/arXiv.2012.08687>
7. Park S., Chun S., Cha J., Lee B., Shim H. Few-shot font generation with localized style representations and factorization. *Proc. of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, no. 3, pp. 2393–2402. <https://doi.org/10.1609/aaai.v35i3.16340>
8. Yao M., Zhang Y., Lin X., Li X.; Zuo W. VQ-Font: Few-shot font generation with structure-aware enhancement and quantization. *Proc. of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 15, pp. 16407–16415. <https://doi.org/10.1609/aaai.v38i15.29577>
9. Ding M. An edge-directed diffusion equation-based image restoration approach for font generation. *IEEE Access*, 2023, vol. 11, pp. 141435–141444. <https://doi.org/10.1109/ACCESS.2023.3342026>
10. Jeong J., Shin J. Multi-scale diffusion denoised smoothing. *Proc. of the 37<sup>th</sup> International Conference on Neural Information Processing Systems*, 2023, pp. 67374–67397.
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 1–11.
12. Voronov G., Lighthead R., Davison J., Kretler C.A., Healey D., Butler T. Multi-scale sinusoidal embeddings enable learning on high resolution mass spectrometry data. *arXiv*, 2022, [arXiv:2207.02980](https://doi.org/10.48550/arXiv.2207.02980). <https://doi.org/10.48550/arXiv.2207.02980>
13. Dhariwal P., Nichol A. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 8780–8794.
14. Convolutional Layer – Building Block of CNNs. *Towards Data Science*, 2024. Available at: <https://towardsdatascience.com/convolutional-layer-building-block-of-cnns-501b5b643e7b> (accessed: 30.01.2024).
15. Xu M., Du X., Wang D. Super-resolution restoration of single vehicle image based on ESPCN-VISR model. *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 790, no. 1, pp. 012107. <https://doi.org/10.1088/1757-899X/790/1/012107>
16. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. *Proc. of the 34<sup>th</sup> International Conference on Neural Information Processing Systems*, 2020, pp. 6840–6851.
17. Nichol A.Q., Dhariwal P. Improved denoising diffusion probabilistic models. *Proc. of the 38<sup>th</sup> International Conference on Machine Learning*, 2021, vol. 139, pp. 8162–8171.
18. Lin S., Yang X. Diffusion model with perceptual loss. *arXiv*, 2023, [arXiv:2401.00110](https://doi.org/10.48550/arXiv.2401.00110). <https://doi.org/10.48550/arXiv.2401.00110>

## Authors

**Maksim I. Maslov** — Software Developer, LLC “Nanosoft Razrabotka”, 108811, Moscow, Russian Federation; Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0009-0005-8198-4703>, maxmiriko@gmail.com

**Авдюшина Анна Евгеньевна** — аналитик, АО «Центр исследований и разработок», Москва, 101000, Российская Федерация; ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57221719751](https://orcid.org/0000-0002-8235-902X), <https://orcid.org/0000-0002-8235-902X>, [avdiushina@itmo.ru](mailto:avdiushina@itmo.ru)

**Солодкая Мария Александровна** — ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0009-0000-7047-6862>, [msolodkaya@itmo.ru](mailto:msolodkaya@itmo.ru)

**Кугаевских Александр Владимирович** — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56442745400](https://orcid.org/0000-0002-6676-0518), <https://orcid.org/0000-0002-6676-0518>, [avkugaevskikh@itmo.ru](mailto:avkugaevskikh@itmo.ru)

**Anna E. Avdyushina** — Analyst, JSC “Research and development center”, Moscow, 101000, Russian Federation; Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57221719751](https://orcid.org/0000-0002-8235-902X), <https://orcid.org/0000-0002-8235-902X>, [avdiushina@itmo.ru](mailto:avdiushina@itmo.ru)

**Maria A. Solodkaya** — Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0009-0000-7047-6862>, [msolodkaya@itmo.ru](mailto:msolodkaya@itmo.ru)

**Alexander V. Kugaevskikh** — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56442745400](https://orcid.org/0000-0002-6676-0518), <https://orcid.org/0000-0002-6676-0518>, [avkugaevskikh@itmo.ru](mailto:avkugaevskikh@itmo.ru)

*Статья поступила в редакцию 20.01.2025  
Одобрена после рецензирования 25.05.2025  
Принята к печати 21.07.2025*

*Received 20.01.2025  
Approved after reviewing 25.05.2025  
Accepted 21.07.2025*



Работа доступна по лицензии  
Creative Commons  
«Attribution-NonCommercial»