УДК 004.89
DOI: 10.26102/2310-6018/2025.50.3.013

Искусственная нейронная сеть подавления артефактов наложения изображений для изменения атрибутов лица на основе дифференциальной активации

Гу Чунюй✉, М.Л. Громов

Национальный исследовательский Томский государственный университет, Томск, Российская Федерация

Резюме. В работе предлагается новый метод подавления артефактов, возникающих при наложении изображений друг на друга. Метод основан на дифференциальной активации. Задача наложения изображений возникает во многих приложениях, однако в данной работе она рассматривается с точки зрения редактирования атрибутов лица. Существующие подходы подавления артефактов имеют существенные ограничения. Они используют дифференциальную активацию для локализации областей редактирования с последующим слиянием признаков, что приводит к потере характерных деталей (например, украшения, прически) и нарушению целостности фона. Передовой метод подавления артефактов основан на энкодер-декодерной архитектуре и иерархической агрегации карт признаков генератора StyleGAN2 с декодером, что приводит к искажению текстур, чрезмерной резкости и эффекту алиасинга. Мы предлагаем метод, объединяющий традиционный алгоритм обработки изображений с методом глубокого обучения. В нем объединены блендинг Пуассона и нейронная сеть MAResU-Net. Блендинг Пуассона используется для создания слитых изображений без артефактов, а сеть MAResU-Net учится сопоставлять изображения, загрязненные артефактами, с чистыми версиями. В результате формируется конвейер преобразования изображений с артефактами наложения в чистые изображения без артефактов. На первых 1000 изображениях базы данных CelebA-HQ разработанный метод демонстрирует превосходство по сравнению с известным методом по пяти метрикам: PSNR: +17,11 % (от 22,24 до 26,06), SSIM: +40,74 % (от 0,618 до 0,870), MAE: −34,09 % (от 0,0511 до 0,0338), LPIPS: −67,16 % (от 0,3268 до 0,1078), FID: −48,14 % (от 27,53 до 14,69) при 26,3 млн параметров (в 6,6 раз меньше, чем 174,2 млн у аналога) и ускорении обработки на 22 %. Метод сохраняет детали аксессуаров, фоновые элементы и текстуру кожи, которые обычно теряются в существующих методах, что подтверждает его практическую ценность для реальных приложений редактирования лиц.

Ключевые слова: глубокое обучение, изменение атрибутов лица, сеть подавления артефактов наложения, преобразование изображений, дифференциальная активация, MAResU-Net, генеративно-состязательная сеть (GAN).

Благодарности: данная работа была поддержана грантом Китайского совета по стипендиям (CSC) № 201908090255.

Для цитирования: Гу Чунюй, Громов М.Л. Сеть подавления артефактов наложения для изменения атрибутов лица на основе дифференциальной активации. *Моделирование, оптимизация и информационные технологии.* 2025;13(3). URL: https://moitvivt.ru/ru/journal/pdf?id=1971 DOI: 10.26102/2310-6018/2025.50.3.013

Artificial neural network for image blending artifact suppression in differential activation-based face attribute editing

Gu Chongyu✉, M.L. Gromov

National Research Tomsk State University, Tomsk, the Russian Federation

Abstract. The paper proposes a new method for suppressing artifacts generated during image blending. The method is based on differential activation. The task of image blending arises in many applications; however, this work specifically addresses it from the perspective of face attribute editing. Existing artifact suppression approaches have significant limitations: they employ differential activation to localize editing regions followed by feature merging, which leads to loss of distinctive details (e.g., accessories, hairstyles) and degradation of background integrity. The state-of-the-art artifact suppression method utilizes an encoder-decoder architecture with hierarchical aggregation of StyleGAN2 generator feature maps and a decoder, resulting in texture distortion, excessive sharpening, and aliasing effects. We propose a method that combines traditional image processing algorithms with deep learning techniques. It integrates Poisson blending and the MAResU-Net neural network. Poisson blending is employed to create artifact-free fused images, while the MAResU-Net network learns to map artifact-contaminated images to clean versions. This forms a processing pipeline that converts images with blending artifacts into clean artifact-free outputs. On the first 1000 images of the CelebA-HQ database, the proposed method demonstrates superiority over existing approach across five metrics: PSNR: +17.11 % (from 22.24 to 26.06), SSIM: +40.74 % (from 0.618 to 0.870), MAE: −34.09 % (from 0.0511 to 0.0338), LPIPS: −67.16 % (from 0.3268 to 0.1078), and FID: −48.14 % (from 27.53 to 14.69). The method achieves these results with 26.3 million parameters (6.6× fewer than the 174.2 million parameters of comparable method) and 22 % faster processing speed. Crucially, it preserves accessory details, background elements, and skin textures that are typically lost in existing methods, confirming its practical value for real-world facial editing applications.

Keywords: deep learning, facial attribute editing, blending artifact suppression network, image-to-image translation, differential activation, MAResU-Net, generative adversarial network (GAN).

Acknowledgements: This work was supported by China Scholarship Council (CSC) Grant No. 201908090255.

For citation: Gu Chongyu, Gromov M.L. Artificial neural network for image blending artifact suppression in differential activation-based face attribute editing. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=1971 DOI: 10.26102/2310-6018/2025.50.3.013

Введение

Изменение атрибутов лица – это процесс изменения определенных характеристик лица на изображениях с использованием методов компьютерного зрения и машинного обучения. Эта технология значительно продвинулась вперед с развитием глубокого обучения и генеративно-состязательных сетей (GAN) [1].

Существующие методы в основном основаны на генераторе StyleGAN [2, 3] для создания изображений лиц с высоким разрешением. Например, в работе [4] используется пирамидальный энкодер признаков для генерации многомасштабных векторов стиля для StyleGAN, в то время как [5] вводит векторы стохастического шума во время вывода для подавления латентных пространственных неоднозначностей. Хотя эти методы обеспечивают хорошее редактирование, они приводят к существенной потере информации из исходных изображений.

Однако современные подходы сталкиваются с фундаментальными ограничениями в сохранении исходных деталей. Alaluf et al. [6] предлагают итеративный механизм восстановления текстур, который лишь частично сохраняет исходные характеристики, создавая структурные несоответствия. Wang et al. [7] комбинируют низкобитный латентный код для изменения атрибутов и высокобитную латентную карту для восстановления деталей через многоуровневое консультативное слияние признаков. Несмотря на использование адаптивного выравнивания искажений (ADA) для согласования исходных и измененных областей, основной недостаток метода

заключается в размытии восстановленных деталей из-за неоптимального объединения признаков в генераторе.

Самым эффективным методом использует парадигму композиция-декомпозиция [8]: на этапе композиции дифференциальная активация генерирует маску для грубого совмещения исходного и обработанного изображений, а на этапе декомпозиции итоговый результат подавления артефактов формируется иерархическим слиянием признаков из декодера и промежуточных слоев генератора StyleGAN2. Основным ограничением данного подхода является неконтролируемое объединение многоуровневых признаков, что приводит не только к неполному восстановлению исходных деталей, но и вызывает появление синтетических артефактов, включая чрезмерную резкость и алиасинг. Эти артефакты особенно выражены в сложных областях изображения, таких как прически и аксессуары, где механизм слияния не учитывает тонкие структурные особенности, заменяя их усредненными шаблонами из обучающего набора StyleGAN. В результате вместо подавления исходных артефактов метод вносит дополнительные искажения, ухудшающие реалистичность редактирования.

Для решения указанных проблем восстановления уникальных деталей изображения после изменения атрибутов лица мы разработали гибридную архитектуру, интегрирующую блендинг Пуассона для градиентного слияния с базовой сетью MAResU-Net [9]. Поскольку для слитых изображений отсутствуют эталонные данные, мы применяем технологию бесшовного клонирования из OpenCV для генерации реалистичных обучающих выборок без артефактов. В качестве базовой архитектуры используется модифицированная MAResU-Net, принимающая на вход объединенный тензор из исходного и обработанного изображений, что позволяет сети адаптивно выбирать и интегрировать наиболее информативные признаки для восстановления. Предложенный метод демонстрирует значительное превосходство над современным аналогом как по количественным метрикам (улучшение PSNR на 17,11 %, SSIM на 40,74 %, MAE на 34,09 %, LPIPS на 67,16 %, FID на 48,14 %), так и по визуальному качеству результатов – отредактированные изображения сохраняют естественные текстуры, элементы украшений и освещение, максимально приближаясь к исходным данным при полном подавлении артефактов наложения. Экспериментальные результаты подтверждают эффективность предложенного подхода в задачах высококачественного редактирования изображений лиц.

Материалы и методы

Сеть подавления артефактов наложения. Современный подход к изменению атрибутов лица в основном основан на механизме дифференциальной активации. Процесс начинается с создания пространственной маски для локализации атрибутов лица, после чего исходное и обработанное изображения сливаются с использованием этой маски в качестве весовой матрицы. Конечный результат генерируется с помощью сети подавления артефактов, которая снижает эффект «призраков». Однако метод вставки извлеченных признаков из генератора StyleGAN в декодер не позволяет восстановить потерянную информацию в перекрывающихся областях, создает новые артефакты и приводит к чрезмерной резкости и алиасингу. Для подавления артефактов наложения мы предлагаем сеть подавления артефактов, которая основана на сети отображения изображений. Общая схема предложенного подхода представлена на Рисунке 1.

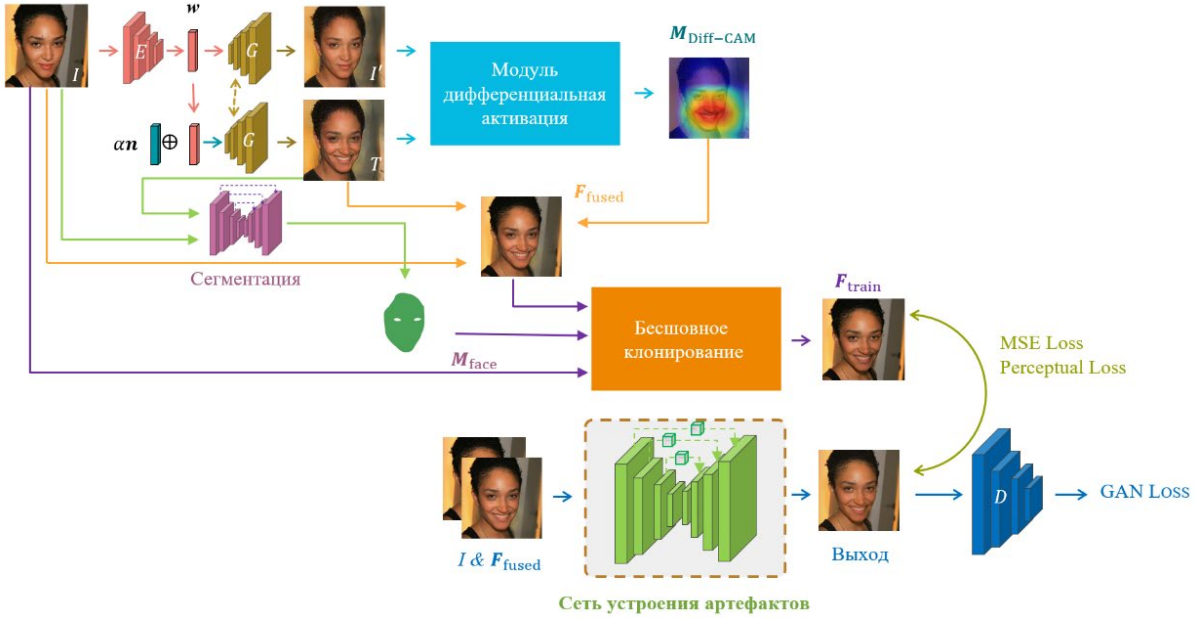

Рисунок 1 – Общая схема предложенного подхода
Figure 1 – Overall pipeline of suggested approach

Процесс инверсии GAN и редактирование. Входное изображение I сначала внедряется в латентное пространство StyleGAN с помощью предобученного энкодера pSp $E_{\text{pSp}}(\cdot)$, выдавая латентный код $w = E_{\text{pSp}}(I)$. Затем изображение I' генерируется с помощью генератора StyleGAN $G(\cdot)$ из этого латентного кода w : $I' = G(w)$. Одновременно выполняется изменение атрибутов путем манипулирования латентным кодом в определенном направлении n с коэффициентом масштабирования α. Обработанное изображение T генерируется следующим образом: $T = G(w + \alpha n)$. Значение коэффициента α фиксировано ($\alpha = 20$) в соответствии с параметрами базового метода EOGI [8], что обеспечивает согласованность сравнения артефактов наложения при одинаковой интенсивности модификации атрибутов.

Расчеты дифференциальной активации. Дифференциальные признаки Δ получаются путем подачи парных изображений $\{I', T\}$ в обучаемый энкодер $\text{E}_{\text{trainable}}$ и вычисления разницы между ними: $\Delta = E_{\text{trainable}}(I') - E_{\text{trainable}}(T)$. Затем дифференциальные признаки классифицируются с помощью классификатора, чтобы предсказать целевые атрибуты лица.

Для вычисления карт активации определяется вес β_c^k, соответствующий k-му каналу карты признаков H и c-му атрибуту, определяется следующим образом:

$$\beta_c^k = \overbrace{\frac{1}{Z}\sum_i \sum_j}^{\text{global average pooling}} \frac{\partial s_c}{\partial H_{ij}^k}, \quad (1)$$

где H представляет собой карту признаков, сгенерированную последним сверточным слоем классификатора, а i и j обозначают высоту и ширину карты признаков соответственно.

Маска Diff-CAM ($M_{\text{Diff−CAM}}$) может быть представлена как кусочно-линейное преобразование взвешенных дифференциальных признаков:

$$M_{\text{Diff−CAM}} = \text{ReLU}(\textstyle\sum_{\text{k}} \beta_c^k H^k). \quad (2)$$

Затем слитое изображение F_{fused} формируется по формуле средневзвешенного значения:

$$F_{\text{fused}} = T \odot M_{\text{Diff-CAM}} + I \odot (1 - M_{\text{Diff-CAM}}), \quad (3)$$

где символ $\odot$ обозначает произведение Адамара.

Процесс подавления артефактов. Чтобы подавлять артефакты наложения, возникающие при слиянии исходного I и обработанного изображения T, мы предлагаем новую сеть подавления артефактов наложения, основанную на преобразовании изображений.

Наша цель – преобразовывать изображения с артефактами в чистые изображения, сохраняя желаемые изменения атрибутов и восстанавливая утраченные детали. Мы предлагаем гибридную архитектуру, интегрирующую блендинг Пуассона, для градиентного слияния, с базовой сетью MAResU-Net, для генерации более точных целевых отображений при подавлении артефактов. Этот гибридный подход сочетает преимущества классического метода компьютерного зрения с представительной мощностью нейронных сетей.

Поскольку для слитого изображения F_{fused} отсутствуют эталонные данные, мы синтезируем набор изображений с помощью функции бесшовное клонирование (seamless cloning) из OpenCV[1] для обучения сети подавления артефактов наложения. Бесшовное клонирование – это продвинутая технология слияния изображений, предоставляемая библиотекой OpenCV, которая позволяет естественным образом встраивать объект из одного изображения в другое, достигая бесшовного визуального эффекта. Данная технология основана на уравнении Пуассона и использует обработку в градиентной области для сохранения текстурных характеристик исходного объекта, обеспечивая при этом плавный переход с фоном целевого изображения.

Для реализации бесшовного клонирования, помимо исходного изображения с объектом и целевого фонового изображения, необходимо получить точную маску объекта. В представленном методе маска формируется с использованием предобученной сети анализа лица[2], которая последовательно анализирует как слитое изображение, так и оригинальное изображение. Мы выбираем все компоненты, находящиеся внутри контура лица, кроме глаз, что позволяет получить точную маску лица. Окончательная маска M_{face} создается путем вычисления области пересечения масок для обоих изображений с применением произведения Адамара, что обеспечивает сохранение естественных границ лица. Этот подход обеспечивает сохранение деталей глаз из исходного изображения, предоставляя надежную маску сегментации изображения в сложных условиях (например, при частичном перекрытии лица микрофоном). Обучающее изображение, обозначаемое как F_{train}, определяется следующим образом:

$$F_{\text{train}} = seamlessClone(F_{\text{fused}}, I, M_{\text{face}}). \quad (4)$$

Мы используем MAResU-Net [9] в качестве основы сети подавления артефактов. MAResU-Net – это архитектура сети, разработанная специально для эффективной семантической сегментации высококачественных снимков дистанционного зондирования. В традиционной архитектуре ResU-Net [10] используются простые сквозные связи (skip connection), которые «механически» объединяют признаки энкодера и декодера без учета их семантической значимости, что ограничивает эффективность использования многоуровневых признаков. MAResU-Net преодолевает это ограничение за счет внедрения механизма линейного внимания (LAM), который адаптивно

[1] Seamless Cloning. OpenCV. URL: https://docs.opencv.org/4.x/df/da0/group__photo__clone.html (дата обращения: 01.05.2025).
[2] zllrunning. face-parsing.PyTorch: Using modified BiSeNet for face parsing in PyTorch. GitHub. URL: https://github.com/zllrunning/face-parsing.PyTorch (дата обращения: 20.04.2025).

перераспределяет веса признаков на разных уровнях иерархии. Такой подход позволяет интеллектуально комбинировать низкоуровневые текстурные признаки с высокоуровневыми семантическими характеристиками, автоматически выделяя наиболее информативные элементы для задачи сегментации. Благодаря линейной вычислительной сложности LAM, архитектура сохраняет высокую эффективность даже при обработке изображений сверхвысокого разрешения.

Эти характеристики полностью соответствуют требованиям нашей задачи преобразования изображений. В частности, способность MAResU-Net адаптивно комбинировать признаки исходного и обработанного изображения позволяет эффективно подавлять артефакты, одновременно восстанавливая утраченные детали изображения. Механизм линейного внимания (LAM) обеспечивает точное выделение и обработку дефектных областей без значительного увеличения вычислительных затрат. Кроме того, архитектура демонстрирует высокую эффективность при работе с изображениями высокого разрешения (1024 на 1024 пикселей), что критически важно для получения качественных результатов в задачах преобразования изображений. Благодаря этим свойствам MAResU-Net представляет собой идеальную базовую модель для предложенного метода подавления артефактов, сочетая в себе высокую точность обработки с вычислительной эффективностью.

В отличие от существующих подходов, основанных на иерархическом объединении карт признаков генератора с декодером для подавления артефактов и восстановления утраченной информации, предложенный метод использует объединенный входной тензор, содержащий исходное и слитое изображения. Это позволяет MAResU-Net адаптивно выбирать и интегрировать наиболее информативные признаки для восстановления изображения. Для этого мы увеличиваем число входных каналов первого сверточного слоя MAResU-Net с 3 до 6, поскольку на вход подаются 2 RGB-изображения, объединенных по каналам.

Обучающий набор данных был сформирован следующим образом: из всего набора данных FFHQ (70000 изображений) [2] для каждого экземпляра случайно выбирался 1 из 4 возможных атрибутов («Густые брови», «Борода», «Старения», и «Улыбка»), после чего с использованием метода бесшовного клонирования создавалась модифицированная версия изображения.

Общая функция потерь L для оптимизации сети подавления артефактов, определяется следующим образом:

$$L = \lambda_{\text{MSE}} \cdot L_{\text{MSE}} + \lambda_{\text{percept}} \cdot L_{\text{percept}} + \lambda_{\text{adv}} \cdot L_{\text{adv}}, \tag{5}$$

где L_{MSE} – функция средней квадратичной ошибки, L_{percept} – перцептуальная функция потерь, а L_{adv} – состязательная функция потерь. Формальные выражения этих компонент задаются следующим образом:

$$L_{\text{MSE}} = \frac{1}{N} \parallel F_{\text{train}} - G_{\text{MAResU}}(F_{\text{fused}}, I) \parallel^2, \tag{6}$$

$$L_{\text{percept}} = \frac{1}{N} \parallel VGG(F_{\text{train}}) - VGG(G_{\text{MAResU}}(F_{\text{fused}}, I)) \parallel^2, \tag{7}$$

$$L_{\text{adv}} = \mathbb{E}_{F_{\text{train}} \sim p_{\text{data}}(F_{\text{train}})}[log\, D\,(F_{\text{train}})] + \mathbb{E}_{G \sim p_{\text{data}}(G)}\left[log\left(1 - D\big(G_{\text{MAResU}}(F_{\text{fused}}, I)\big)\right)\right], \tag{8}$$

где N – общее число пикселей изображения, а G_{MAResU} – предложенная в работе архитектура, предназначенная для подавления артефактов. В перцептуальной функции потерь используется предобученная сеть VGG-16 [11], из которой извлекаются активации слоев conv2_2, conv3_2, conv4_3 и pool5. Для каждого из этих слоев применяются весовые коэффициенты $\{1/16, 1/8, 1/4, 1\}$. $D(\cdot)$ – дискриминатор StyleGAN2.

Результаты и обсуждение

Экспериментальная установка и гиперпараметры обучения. Эксперименты проводились на стандартных наборах данных FFHQ [2] и CelebA-HQ [12]. Метод был реализован на языке Python 3.7.12 с использованием фреймворка PyTorch 1.10.1 и поддержкой CUDA 11.3. Все вычисления выполнялись на системе с графическим процессором NVIDIA GeForce RTX 4090, процессором Intel Core i5-13600KF и 32 ГБ оперативной памяти. Для оптимизации применялся алгоритм Adam с коэффициентами $\beta_1 = 0{,}9$ и $\beta_2 = 0{,}999$, начальный темп обучения составлял 1×10^{-4} и снижался по косинусному отжигу до 1×10^{-7}. Обучение проводилось в течение 200000 итераций с размером пакета 4. В наших экспериментах гиперпараметры функции потерь устанавливались следующим образом: $\lambda_{\text{MSE}} = 0{,}1$, $\lambda_{\text{perc}} = 1$, $\lambda_{\text{adv}} = 0{,}1$. Все изображения лица, использованные в иллюстративных материалах данного исследования, взяты из набора данных CelebA-HQ[3].

Вычислительная эффективность метода. Предложенная сеть подавления артефактов демонстрирует значительные преимущества по вычислительным параметрам:

– общее количество параметров сети составляет 26285115, что в 6,6 раз меньше по сравнению с существующим методом (174196615 параметров);

– среднее время обработки 1000 изображений сокращено до 1 минуты 45 секунд против 2 минут 15 секунд у аналога (ускорение на 22 %).

Таблица 1 – Сравнение методов по метрикам пиксельной точности
Table 1 – Comparison of pixel-level accuracy metrics

	PSNR↑		MAE↓	
	EOGI	Предложенный	EOGI	Предложенный
Борода	22,05±1,32	**26,76±1,43**	0,0514±0,008	**0,0332±0,005**
Брови	22,84±1,19	**27,32±1,30**	0,0468±0,007	**0,0302±0,004**
Старение	21,73±1,25	**23,31±1,44**	0,0563±0,009	**0,0393±0,006**
Улыбка	22,34±1,36	**26,85±1,65**	0,0499±0,008	**0,0323±0,005**

Таблица 2 – Сравнение методов по метрикам перцептивного качества
Table 2 – Comparison of perceptual metrics

	SSIM↑		LPIPS↓		FID↓	
	EOGI	Предложенный	EOGI	Предложенный	EOGI	Предложенный
Борода	0,615±0,076	**0,871±0,060**	0,329±0,048	**0,108±0,022**	27,68	**13,56**
Брови	0,635±0,078	**0,881±0,061**	0,310±0,051	**0,092±0,023**	20,34	**9,2**
Старение	0,592±0,077	**0,851±0,062**	0,350±0,046	**0,132±0,025**	39,2	**23,66**
Улыбка	0,631±0,076	**0,876±0,062**	0,318±0,049	**0,099±0,025**	22,89	**12,33**

Количественная оценка. В рамках сравнения эффективности подавления артефактов и качества генерации между предложенным методом и современным аналогом мы провели оценку по пяти метрикам: Пиковое отношение сигнала к шуму (PSNR), Индекс структурного сходства (SSIM) [13], Learned perceptual image patch similarity (LPIPS) [14], Средняя абсолютная ошибка (MAE) и Fréchet inception distance (FID) [15]. Оценка проводилась на первой 1000 изображений набора данных CelebA-HQ. В качестве ключевых атрибутов для анализа были выбраны: «густые брови», «борода», «старение» и «улыбка». Такой подход позволяет комплексно оценить как пиксельную

[3] tkarras. progressive_growing_of_gans: Progressive Growing of GANs for Improved Quality, Stability, and Variation. GitHub. URL: https://github.com/tkarras/progressive_growing_of_gans (дата обращения: 26.05.2025).

точность (PSNR, MAE), так и перцептивное качество (SSIM, LPIPS, FID) на основе трех критериев: объективных метрик, визуального соответствия и устойчивости к артефактам.

Как видно из данных Таблиц 1 и 2, предложенный нами метод продемонстрировал наилучшие результаты по всем пяти метрикам, значительно превзойдя современный аналог. Это свидетельствует о том, что предложенный метод не только обеспечивает высокую точность на пиксельном уровне (PSNR, MAE), эффективно уменьшая ошибки реконструкции, но и достигает более естественного качества генерации (SSIM, LPIPS, FID) с точки зрения восприятия.

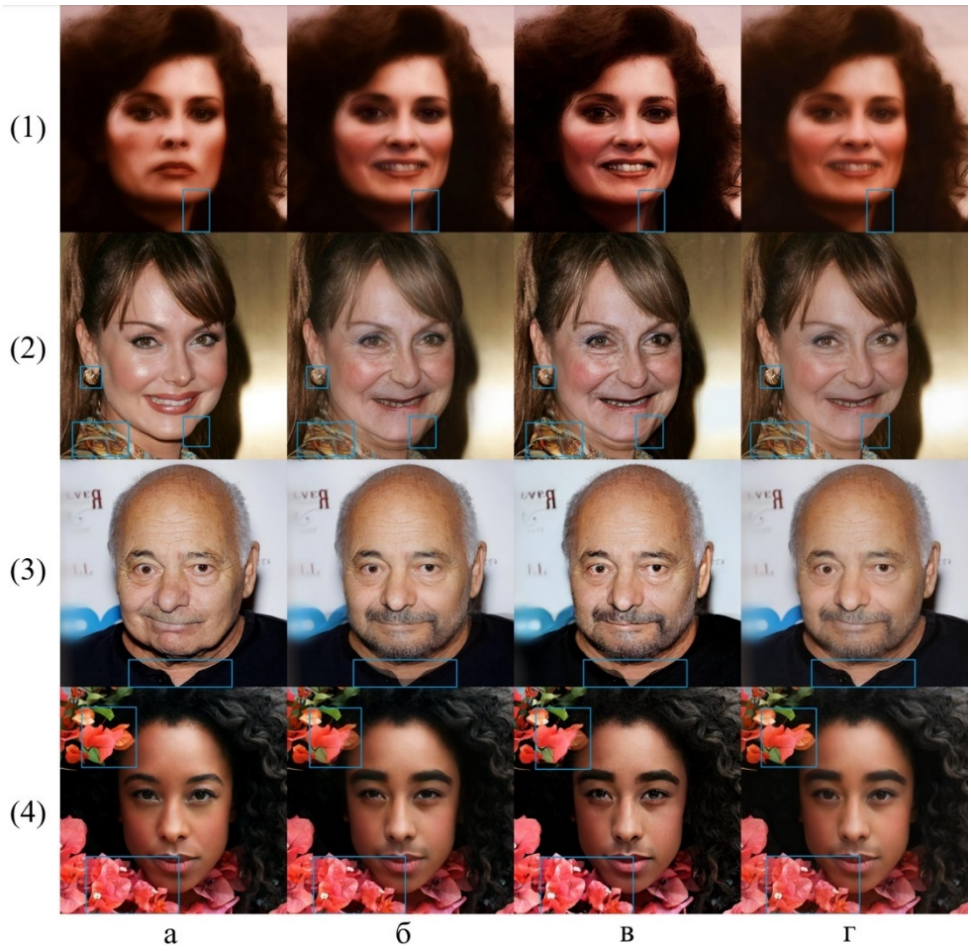

Рисунок 2 – Сравнение методов подавления артефактов: (1) изменение атрибута «улыбка»; (2) изменение атрибута «старение»; (3) изменение атрибута «борода»; (4) изменение атрибута «густые брови»; (а) исходное изображение; (б) слитое изображение с артефактами; (в) подавление артефактов методом EOGI; (г) подавление артефактов предложенным методом в данной работе

Figure 2 – Comparison of artifact suppression methods: (1) editing of the "smiling" attribute; (2) editing of the "aging" attribute; (3) editing of the "beard" attribute; (4) editing of the "bushy eyebrows" attribute; (a) original image; (b) fused image with artifacts; (c) artifact suppression using the EOGI method; (d) artifact suppression using the proposed method in this work

Визуальный анализ результатов. На Рисунке 2 представлено сравнение существующего и предложенного методов при различных изменениях атрибутов на одном и том же слитом изображении с артефактами. Визуальный анализ результатов демонстрирует значительное преимущество предложенного метода по сравнению с современным методом. Предложенный подход обеспечивает более полное подавление

артефактов при сохранении высокого качества изображения, особенно в областях вокруг лица и фона. В отличие от современных аналогов, мы эффективнее восстанавливаем утраченные детали и текстуры. Кроме того, предложенный метод минимизирует размытие и искажения, обеспечивая более естественное и визуально правдоподобное восстановление изображения после редактирования. Эти преимущества подтверждаются как визуальным сравнением, так и количественными метриками, что свидетельствует о превосходстве предложенного подхода.

Ограничения метода. Хотя предложенный метод демонстрирует значительные успехи в сокращении параметров модели, повышении вычислительной эффективности и подавлении артефактов, он имеет определенные ограничения. Основное ограничение связано с обработкой зубов при изменении таких атрибутов, как «улыбка» или «старение». В случаях, когда на исходном изображении присутствуют зубы, модификация этих атрибутов может приводить к появлению артефактов наложения в зубной области. Это объясняется тем, что предложенный метод преимущественно сосредоточен на подавлении артефактов в периферийных областях лица и фона, тогда как проблема корректного отображения зубов требует более специализированного подхода. Данное ограничение определяет важное направление для будущих исследований и усовершенствований метода, в частности, разработку механизмов более точного учета семантики зубного ряда при редактировании изображений. Решение этой задачи позволит расширить применимость метода, сохраняя при этом все его текущие преимущества в скорости работы и качестве обработки изображений.

Заключение

В данной работе предложен новый метод подавления артефактов наложения на лицевых изображениях после изменения атрибутов. Основная идея заключается в комбинации классического метода обработки изображений (бесшовного клонирования) с методом глубокого обучения. Такой гибридный подход позволяет эффективно подавлять артефакты наложения, сохраняя при этом важные детали изображения.

Эксперименты показали, что предложенный метод превосходит существующий аналог по всем основным метрикам качества. В частности, достигается лучшее сохранение текстур и естественного вида лица после редактирования. Важным преимуществом является способность сети восстанавливать утраченные детали в областях вокруг лица.

Предложенный метод демонстрирует хороший баланс между качеством обработки и вычислительной эффективностью. В будущей работе планируется улучшить обработку сложных областей, таких как зубы, а также расширить набор поддерживаемых атрибутов. Предложенный подход может быть полезен для создания более совершенных систем редактирования портретных изображений.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Goodfellow I.J., Pouget-Abadie J., Mirza M., et al. Generative Adversarial Networks. arXiv. URL: https://arxiv.org/abs/1406.2661 [Accessed 19th April 2025].
2. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15–20 June 2019, Long Beach, CA, USA*. IEEE; 2019. P. 4401–4410. https://doi.org/10.1109/TPAMI.2020.2970919
3. Karras T., Laine S., Aittala M., Hellsten J., Lehtinen J., Aila T. Analyzing and Improving the Image Quality of StyleGAN. In: *2020 IEEE/CVF Conference on Computer Vision*

and Pattern Recognition (CVPR), 13–19 June 2020, Seattle, WA, USA. IEEE; 2020. P. 8107–8116. https://doi.org/10.1109/CVPR42600.2020.00813
4. Richardson E., Alaluf Yu., Patashnik O., et al. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 20–25 June 2021, Nashville, TN, USA*. IEEE; 2021. P. 2287–2296. https://doi.org/10.1109/CVPR46437.2021.00232
5. Tov O., Alaluf Yu., Nitzan Yo., Patashnik O., Cohen-Or D. Designing an Encoder for Stylegan Image Manipulation. *ACM Transactions on Graphics (TOG)*. 2021;40(4). https://doi.org/10.1145/3450626.3459838
6. Alaluf Yu., Patashnik O., Cohen-Or D. ReStyle: A Residual-Based StyleGAN Encoder via Iterative Refinement. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV), 10–17 October 2021, Montreal, QC, Canada*. IEEE; 2021. P. 6691–6700. https://doi.org/10.1109/ICCV48922.2021.00664
7. Wang T., Zhang Yo., Fan Ya., Wang J., Chen Q. High-Fidelity GAN Inversion for Image Attribute Editing. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18–24 June 2022, New Orleans, LA, USA*. IEEE; 2022. P. 11369–11378. https://doi.org/10.1109/CVPR52688.2022.01109
8. Song H., Du Yo., Xiang T., Dong J., Qin J., He Sh. Editing Out-of-Domain GAN Inversion via Differential Activations. In: *Computer Vision – ECCV 2022: 17th European Conference: Proceedings: Part XVII, 23–27 October 2022, Tel Aviv, Israel*. Cham: Springer; 2022. P. 1–17. https://doi.org/10.1007/978-3-031-19790-1_1
9. Li R., Zheng Sh., Duan Ch., Su J., Zhang C. Multistage Attention ResU-Net for Semantic Segmentation of Fine-Resolution Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*. 2021;19. https://doi.org/10.1109/LGRS.2021.3063381
10. Zhang Zh., Liu Q., Wang Yu. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*. 2018;15(5):749–753. https://doi.org/10.1109/LGRS.2018.2802944
11. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. URL: https://arxiv.org/abs/1409.1556 [Accessed 26th May 2025].
12. Karras T., Aila T., Laine S., Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv. URL: https://arxiv.org/abs/1710.10196 [Accessed 19th April 2025].
13. Wang Zh., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*. 2004;13(4):600–612. https://doi.org/10.1109/TIP.2003.819861
14. Zhang R., Isola Ph., Efros A.A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18–23 June 2018, Salt Lake City, UT, USA*. IEEE; 2018. P. 586–595. https://doi.org/10.1109/CVPR.2018.00068
15. Heusel M., Ramsauer H., Unterthiner Th., Nessler B., Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv. URL: https://arxiv.org/abs/1706.08500 [Accessed 1st April 2025].

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Гу Чунюй, аспирант, Национальный исследовательский Томский государственный университет, Томск, Российская Федерация.
e-mail: chongyugu@gmail.com
ORCID: 0000-0003-4103-2036

Gu Chongyu, Postgraduate, National Research Tomsk State University, Tomsk, the Russian Federation.

Громов Максим Леонидович, кандидат физико-математических наук, доцент, Национальный исследовательский Томский государственный университет, Томск, Российская Федерация.
e-mail: maxim.leo.gromov@gmail.com
ORCID: 0000-0002-2990-8245

Maxim L. Gromov, Candidate of Physical and Mathematical Sciences, Associate Professor, National Research Tomsk State University, Tomsk, the Russian Federation.

Статья поступила в редакцию 26.05.2025; одобрена после рецензирования 26.06.2025; принята к публикации 07.07.2025.

The article was submitted 26.05.2025; approved after reviewing 26.06.2025; accepted for publication 07.07.2025.