

УДК 004.89

DOI: [10.26102/2310-6018/2025.49.2.046](https://doi.org/10.26102/2310-6018/2025.49.2.046)

Разработка улучшенного модуля дифференциальной активации с использованием Grad-CAM++ и семантической сегментации для изменения атрибутов лица

Гу Чуньюй✉, М.Л. Громов

*Национальный исследовательский Томский государственный университет, Томск,
Российская Федерация*

Резюме. Современные методы изменения атрибутов лица страдают от двух системных проблем: нежелательная модификация второстепенных признаков и потеря контекстных деталей (аксессуаров, фона, текстуры волос и т. д.), что приводит к артефактам и ограничивает их применение в задачах, требующих фотографической точности. Для решения этих проблем мы предлагаем улучшенный модуль дифференциальной активации, предназначенный для точного редактирования с сохранением контекстной информации. В отличие от существующего решения (EOGI), предложенное решение включает: использование градиентной информации второго и третьего порядка для точной локализации редактируемых областей, применение увеличения тестового времени (ТТА) и метода главных компонент (РСА) для центрирования карты активации классов (CAM) вокруг объектов и удаления большого количества шума, интеграцию данных семантической сегментации для повышения пространственной точности. Экспериментальное оценивание на первых 1000 изображениях CelebA-HQ (разрешение 1024×1024 пикселей) демонстрирует значительное превосходство над современным методом EOGI: снижение среднего значения FID на 13,84 % (от 27,68 до 23,85), снижение среднего значения LPIPS на 7,03 % (от 0,327 до 0,304) и снижение среднего значения MAE на 10,57 % (от 0,0511 до 0,0457). Предложенный метод превосходит существующие подходы как в количественной оценке, так и в качественном сравнении. Результаты демонстрируют улучшенное сохранение деталей (например, серьги, фона), что делает метод применимым в задачах, требующих высокой фотореалистичности.

Ключевые слова: глубокое обучение, изменение атрибутов лица, дифференциальная активация, карты активации класса (CAM), семантическая сегментация, генеративно-состязательная сеть (GAN).

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

Для цитирования: Гу Чуньюй, Громов М.Л. Разработка улучшенного модуля дифференциальной активации с использованием Grad-CAM++ и семантической сегментации для изменения атрибутов лица. *Моделирование, оптимизация и информационные технологии*. 2025;13(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1932> DOI: 10.26102/2310-6018/2025.49.2.046

Development of an improved differential activation module using Grad-CAM++ and semantic segmentation for facial attribute editing

Gu Chongyu✉, M.L. Gromov

National Research Tomsk State University, Tomsk, the Russian Federation

Abstract. Modern methods of facial attribute editing suffer from two systemic issues: unintended modification of secondary features and loss of contextual details (such as accessories, background, and hair textures, etc.), which lead to artifacts and restrict their application in scenarios requiring

photographic accuracy. To address these problems, we propose an improved differential activation module designed for precise editing while preserving contextual information. In contrast to the existing solution (EOGI), the proposed solution includes: the use of second- and third-order gradient information for precise localization of editable areas, applying test-time augmentation (TTA) and principal component analysis (PCA) to center the class activation map (CAM) around objects and remove a lot of noise, the integration of semantic segmentation data to enhance spatial accuracy. The evaluation on the first 1,000 images of the CelebA-HQ dataset (resolution 1024×1024) demonstrates significant superiority over the current method EOGI: a 13.84 % reduction in the average FID (from 27.68 to 23.85), a 7.03 % reduction in the average LPIPS (from 0.327 to 0.304), and a 10.57 % reduction in the average MAE (from 0.0511 to 0.0457). The proposed method outperforms existing approaches in both quantitative and qualitative analyses. The results demonstrate improved detail preservation (e.g., earrings and backgrounds), which makes the method applicable in tasks demanding high photographic realism.

Keywords: deep learning, facial attribute editing, differential activation, class activation maps (CAM), semantic segmentation, generative adversarial network (GAN).

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

For citation: Gu Chongyu, Gromov M.L. Development of an improved differential activation module using Grad-CAM++ and semantic segmentation for facial attribute editing. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.). URL: <https://moitvvt.ru/ru/journal/pdf?id=1932> DOI: 10.26102/2310-6018/2025.49.2.046

Введение

Изменение атрибутов лица означает изменение черт лица на изображении с помощью методов компьютерного зрения. Этот метод используется во многих приложениях, таких как улучшение распознавания лиц, применение виртуального макияжа и генерация выражений лица [1, 2]. В последние годы было предложено много методов на основе инверсии генеративно-сопоставительной сети (GAN) [3]. GAN-инверсия направлена на обратное преобразование заданного изображения в латентное пространство предварительно обученной модели GAN таким образом, чтобы это изображение могло быть достоверно реконструировано из инверсионного кода с помощью генератора [4]. Сначала эти методы преобразуют исходное изображение в латентное пространство генератора StyleGAN [5, 6], а затем используют измененные векторы стиля для генерации новых изображений. Авторы работы [7] используют сеть пирамидальных признаков в качестве энкодера для создания векторов стиля на разных уровнях изображения для генератора StyleGAN. В отличие от детерминированного энкодера, авторы [8] добавляют небольшие шумовые векторы во время вывода для устранения неоднозначности в процессе инверсии. Хотя эти методы позволяют редактировать изображение, они также теряют значительную часть информации, содержащейся в исходном изображении.

Алалуф и др. [9] используют итеративный механизм уточнения для постепенного восстановления деталей изображения. Хотя восстановленное изображение сохраняет некоторое сходство в общих очертаниях и отдельных текстурах, конкретные детали существенно отличаются от исходного изображения. Ван и др. [10] применяют карты искажений между исходным и обработанным изображениями для высокой точности редактирования, которые содержат много утраченной информации. Метод, при котором карты признаков вставляются в предобученный StyleGAN для восстановления, не обеспечивает достаточного уровня деталей и приводит к появлению новых артефактов. Самым современным методом является использование дифференциальной активации [11]. В этом методе область редактирования активируется с помощью

дифференциального расчета, и эта активация служит маской для изменения соответствующей области изображения. Затем структура энкодер-декодера удаляет артефакты, возникающие при смешивании изображений. Однако современный метод дифференциальной активации охватывает области, выходящие за пределы целевого редактирования. Это приводит к потере информации, например, деталей вокруг носа при редактировании бороды или признаков ушей и фона за контуром лица.

Чтобы решить эти проблемы, мы предлагаем новый модуль дифференциальной активации для точного определения области редактирования. Сначала в качестве базовой модели используется Grad-CAM++ [12]. По сравнению с Grad-CAM [13], Grad-CAM++ добавляет информацию о градиентах второго и третьего порядков, что помогает точнее выделять ключевые области и уменьшает фоновый шум. Далее вводятся два метода сглаживания. Метод увеличения тестового времени применяет набор преобразований к изображению, что позволяет лучше центрировать карту активации на объекте. Метод сглаживания Эйгена [14] вычисляет первую главную компоненту ковариационной матрицы карты активации, что помогает удалить значительную часть шума. Наконец, с помощью сегментации изображения область активации строго ограничивается контуром лица, предотвращая влияние на области вне лица. После этого применяется предобученная сеть подавления артефактов [11] для удаления артефактов, возникающих в процессе слияния изображений. Предложенный метод превосходит существующие подходы как по количественным, так и по качественным анализам. Визуальные результаты демонстрируют лучшее сохранение деталей, таких как серьги и фон.

Улучшенный модуль дифференциальной активации с использованием Grad-CAM++ и семантической сегментации

Существующий метод изменения атрибутов лица, основанный на дифференциальной активации, часто генерирует тепловые карты, выходящие за пределы целевой области. Эти тепловые карты приводят к снижению точности редактирования и потере деталей в нецелевых зонах. Для точного позиционирования редактируемых атрибутов мы предлагаем улучшенный модуль дифференциальной активации. Общая схема метода представлена на Рисунке 1.

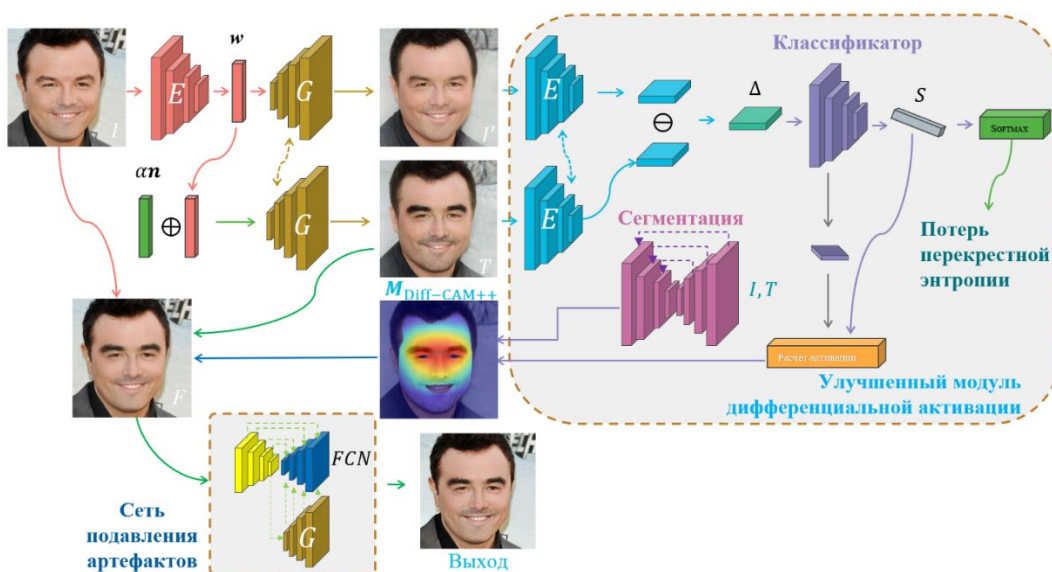


Рисунок 1 – Общая схема предложенного метода
Figure 1 – Overall scheme of suggested method

Инверсия и модификация в латентном пространстве. Сначала входное изображение I преобразуется в латентное пространство генератора StyleGAN2 с помощью предобученного энкодера pSp $E_{pSp}(\cdot)$, получая латентный код:

$$w = E_{pSp}(I). \quad (1)$$

Фиксированный предобученный генератор StyleGAN2 $G(\cdot)$ затем преобразует полученный латентный код в изображение, полученное методом GAN-инверсии:

$$I' = G(w). \quad (2)$$

Семантическая модификация достигается путем смещения w в целевом направлении n с коэффициентом масштабирования α , создавая обработанное изображение:

$$T = G(w + \alpha n). \quad (3)$$

Извлечение дифференциальных признаков. Мы используем первые 17 блоков из обучаемой сети ResNet101 [15] в качестве энкодера $E_{ResNet}(\cdot)$. Эта структура захватывает признаки низкого и среднего уровня, сохраняя определенные пространственные детали (что помогает локализовать измененную область) и предоставляя возможности предварительной семантической абстракции (что помогает различать реальное изменение и естественные изменения). Дифференциальные признаки Δ вычисляются как:

$$\Delta = E_{ResNet}(I') - E_{ResNet}(T). \quad (4)$$

Эти признаки подаются в оставшиеся слои ResNet101 для классификации атрибутов, обучаемой с помощью функции потерь перекрестной энтропии $\mathcal{L}_{ce}$:

$$\mathcal{L}_{ce} = - \sum_{i=1}^M y_i \log \left(\frac{e^{s_i}}{\sum_{j=1}^M e^{s_j}} \right), \quad (5)$$

где $y = \{y_1, y_2, \dots, y_N\}$ – унитарные коды (one-hot vector) атрибутов, $s = \{s_1, s_2, \dots, s_N\}$ – выход классификатора до многопеременной логистической функции (Softmax), M – общее количество атрибутов. Классификатор обучен классифицировать дифференциальные признаки из 7 атрибутов энкодера pSp, а именно: «Борода», «Густые брови», «Очки», «Приоткрытый рот», «Узкие глаза», «Признаки старения» и «Улыбка».

Генерация маски дифференциальной активации. Для вычисления тепловой карты используется метод Grad-CAM++, который учитывает градиенты второго и третьего порядка, обеспечивая более точную локализацию значимых признаков. В рамках этого подхода для каждой позиции (i, j) в карте активации A^k последнего сверточного слоя вычисляется весовой коэффициент α_{ij}^{kc} , отражающий вклад соответствующей активации в выход модели для класса c :

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_{a,b} A_{ab}^k \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}}, \quad (6)$$

где y^c – выход модели для класса c , A_{ij}^k – значение в позиции (i, j) карты активации A^k . A_{ab}^k – значение в позиции (a, b) , используемое в суммировании для расчета третьего порядка производной.

На основе этих коэффициентов рассчитывается важность канала w_k^c как взвешенная сумма градиентов с применением функции активации ReLU:

$$w_k^c = \sum_{i,j} \alpha_{ij}^{kc} \cdot \text{ReLU}\left(\frac{\partial y^c}{\partial A_{ij}^k}\right). \quad (7)$$

Итоговая тепловая карта активации формируется как линейная комбинация активационных карт каналов с учетом их важности и последующим применением функции ReLU:

$$M_{\text{Diff-CAM++}} = \text{ReLU}\left(\sum_k w_k^c \cdot A^k\right). \quad (8)$$

Для улучшения тепловой карты применяем:

1. Увеличение тестового времени (ТТА): изображение трансформируется с комбинацией отражений по горизонтали и умножения изображения на [1, 0, 1, 1, 0, 9]. Полученные карты активации всех преобразований объединяются для создания более устойчивой тепловой карты. Это позволяет лучше центрировать CAM вокруг объектов.

2. Сглаживание Эйгена: применяем метод главных компонент (PCA) к 2D-картам активации для подавления шума при сохранении значимых областей. Сглаженная CAM вычисляется как:

$$M_{\text{Diff-CAM++}} = \text{Reshape}(p_1) \in \mathbb{R}^{H \times W}, \quad (9)$$

где p_1 – первый главный компонент, полученный методом сингулярного разложения (SVD) матрицы активации.

Чтобы редактирование происходило только внутри контура лица, используется семантическая сегментация. Предобученная сеть анализа лица¹ генерирует карты сегментации для исходного (S_{org}) и модифицированного (S_{modif}) изображений. Сеть анализа лица способна выделить 19 компонентов лица из набора данных CelebAMask-HQ [16]. Мы используем различные компоненты лица для формирования масок сегментации изображения по различным атрибутам. Для исходного изображения выбор компонентов осуществляется следующим образом: Густые брови (Bushy_Eyebrow): {1, 2, 3, 10}; Борода (Beard): {1, 2, 3, 10, 14}; Признаки старения (Old): {1, 2, 3, 10, 11, 12, 13, 14}; Остальные атрибуты: {1, 2, 3, 10, 11, 12, 13}. Для модифицированного изображения выбор компонентов осуществляется таким образом: Борода (Beard) и признаки старения (Old): {1, 2, 3, 4, 5, 6, 10, 11, 12, 13, 14}; Остальные атрибуты: {1, 2, 3, 4, 5, 6, 10, 11, 12, 13}. Уточненная маска дифференциальной активации вычисляется как:

$$M_{\text{Diff-CAM++}} = S_{\text{org}} \odot S_{\text{modif}} \odot M_{\text{Diff-CAM++}}, \quad (10)$$

где $\odot$ – это произведение Адамара.

Смешивание и подавление артефактов. Для предварительного слияния изображений F обработанное изображение T объединяется с исходным изображением I с помощью маски $M_{\text{Diff-CAM++}}$:

$$F = T \odot M_{\text{Diff-CAM++}} + I \odot (1 - M_{\text{Diff-CAM++}}). \quad (11)$$

Чтобы подавить артефакты, возникающие после слияния изображений, используется предобученная сеть с архитектурой «энкодер-декодер», которая извлекает признаки на стадии декодирования и внедряет их в генератор StyleGAN.

Результаты

Экспериментальные данные. Предложенный метод был реализован на языке Python с использованием фреймворка PyTorch. Все эксперименты проводились на ПК с видеокартой NVIDIA GeForce RTX 4090, процессором Intel i5-13600KF и 32 ГБ

¹ zllrunning/face-parsing.PyTorch. GitHub. URL: <https://github.com/zllrunning/face-parsing.PyTorch> (дата обращения: 20.04.2025).

оперативной памяти. Для обучения и тестирования использовались два стандартных набора данных: FFHQ [6] и CelebA-HQ [17]. Все лицевые изображения, представленные в иллюстративных материалах данного исследования, получены из набора CelebA-HQ². В качестве оптимизатора применялся Adam с параметрами $\beta_1 = 0,9$ и $\beta_2 = 0,99$. Первоначальный темп обучения $1 \cdot 10^{-4}$ постепенно снижался до $1 \cdot 10^{-7}$ с помощью косинусного отжига. Обучение проводилось в течение 3500 шагов при размере пакета 20, параметр α в уравнении (3) был установлен на значении 20.

Эффективность улучшенного модуля дифференциальной активации. Чтобы доказать эффективность предложенных улучшений, мы меняем SAM-модули и проверяем, как маски отличаются и влияют на редактирование. Как показано на Рисунке 2, предложенный метод генерирует тепловые карты с более сконцентрированными активациями и меньшим уровнем шума по сравнению с базовым подходом. Дифференциальная маска Diff-CAM выходит за пределы целевых областей, особенно вдоль контуров лица, что ухудшает бесшовное объединение обработанных и исходных изображений. В отличие от этого, предложенный модуль дифференциальной активации создает точные маски, которые строго ограничивают редактируемые области. Это обеспечивает эффективное слияние исходного и обработанного содержимого.



Рисунок 2 – Сравнение Diff-CAM с предложенным методом: *a* – исходное изображение; *б* – обработанное изображение; *в* – активация методом Diff-CAM; *г* – редактирование на основе Diff-CAM; *д* – активация предложенным в данной работе методом Diff-CAM++; *е* – редактирование на основе Diff-CAM++

Figure 2 – Comparison Diff-CAM with suggested method: *a* – original image; *b* – manipulated image; *c* – activation map generated by Diff-CAM; *d* – editing based on Diff-CAM; *e* – activation map generated by proposed Diff-CAM++; *f* – editing based on Diff-CAM++

Изменения, обусловленные процессом оптимизации модуля дифференциальной активации, представлены на Рисунке 3.

² tkarras/progressive_growing_of_gans. GitHub. URL: https://github.com/tkarras/progressive_growing_of_gans (дата обращения: 07.05.2025).

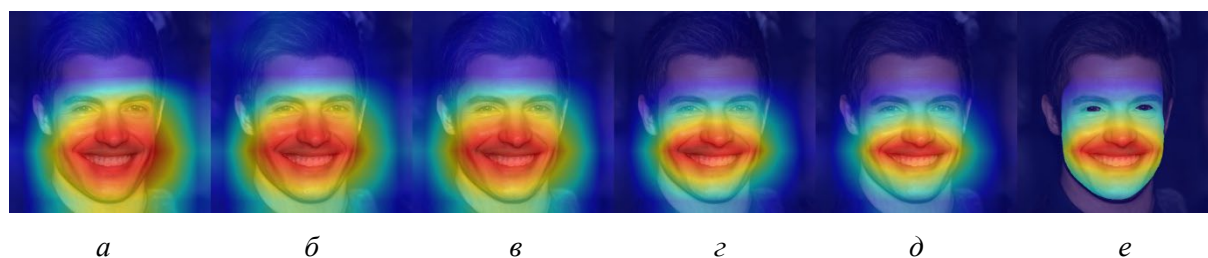


Рисунок 3 – Процесс оптимизации модуля дифференциальной активации: *a* – современный метод Diff-CAM; *б* – Grad-CAM++; *в* – Grad-CAM++ с ТТА (увеличение тестового времени); *г* – Grad-CAM++ с PCA (сглаживание Эйгена); *д* – Grad-CAM++ с ТТА и PCA; *е* – предложенный Diff-CAM++ (Grad-CAM++ с ТТА, PCA и сегментацией)

Figure 3 – The process of optimizing the differential activation module: *a* – SOTA method Diff-CAM; *b* – Grad-CAM++; *c* – Grad-CAM++ with TTA (test-time augmentation); *d* – Grad-CAM++ with PCA (Eigen smoothing); *e* – Grad-CAM++ with TTA and PCA; *f* – proposed Diff-CAM++ (Grad-CAM++ with TTA, PCA, and segmentation)

Как показано на Рисунке 3, тепловая карта становится более сфокусированной на области редактирования, при этом уровень шума снижается. Добавление семантической сегментации дополнительно улучшает фокусировку, ограничивая зону активации по контуру лица, точно выделяет нужную область.

Количественная оценка. Мы оценили четыре атрибута («Густые брови», «Борода», «Признаки старения», и «Улыбка») с использованием трех метрик: Средняя абсолютная ошибка (MAE), Learned perceptual image patch similarity (LPIPS) [18] и Fréchet inception distance (FID) [19]. Более низкие значения указывают на лучшее качество.

Таблица 1 – Оценка LPIPS

Table 1 – LPIPS score

	pSp [7]	e4e [8]	Restyle [9]	EOGI [11]	Предложенный
Борода	0,433±0,056	–	–	0,329±0,048	0,308±0,049
Брови	0,407±0,060	–	–	0,310±0,051	0,296±0,051
Старение	0,462±0,054	0,449±0,062	0,437±0,062	0,350±0,046	0,313±0,048
Улыбка	0,416±0,059	0,435±0,061	0,434±0,061	0,318±0,049	0,298±0,049

Таблица 2 – Оценка MAE

Table 2 – MAE score

	pSp [7]	e4e [8]	Restyle [9]	EOGI [11]	Предложенный
Борода	0,083±0,016	–	–	0,0514±0,008	0,0469±0,007
Брови	0,075±0,015	–	–	0,0468±0,007	0,0436±0,007
Старение	0,108±0,018	0,105±0,021	0,104±0,022	0,0563±0,009	0,0473±0,007
Улыбка	0,076±0,015	0,103±0,019	0,108±0,022	0,0499±0,008	0,0449±0,007

Таблица 3 – Оценка FID

Table 3 – FID score

	pSp [7]	e4e [8]	Restyle [9]	EOGI [11]	Предложенный
Борода	57,88	–	–	27,68	23,85
Брови	39,85	–	–	20,34	19,28
Старение	69,59	57,24	47,04	39,20	27,67
Улыбка	42,47	48,23	42,58	22,89	19,78

Как видно из таблиц, предложенный метод демонстрирует наилучшие оценки, что подтверждает сохранение неотредактируемых областей. Отметим, что методы e4e и ReStyle поддерживают лишь 2 из 4 направлений редактирования, выбранных в оценке.

Качественное сравнение. На Рисунке 4 представлены результаты редактирования разными методами. Методы pSp и e4e не восстанавливают исходный фон после редактирования. Restyle использует итеративное уточнение, но создает текстуры, отличающиеся от исходного изображения. Кроме того, e4e и Restyle используют направление редактирования InterFaceGAN [20], которое будет случайно генерировать очки при редактировании атрибута «старения». EOGI сохраняет большинство деталей, но вносит артефакты в лицо и фон, а также теряет информации вокруг контуров лица. В отличие от них, предложенный метод на основе улучшенного модуля дифференциальной активации обеспечивает значительно лучшее качество редактирования. Изображения демонстрируют, что предложенный подход точнее сохраняет детали по сравнению с другими методами.



Рисунок 4 – Качественное сравнение с современными методами (верхние два ряда демонстрируют изменение атрибута «Старение», нижние два ряда – изменение атрибута «Улыбка»): *a* – исходное изображение; *b* – результат pSp; *c* – результат e4e; *d* – результат Restyle; *e* – результат EOGI; *f* – результат предложенного метода

Figure 4 – Qualitative comparison with SOTA methods (top two rows demonstrate "Aging" attribute editing, bottom two rows show "Smiling" attribute editing): *a* – source image; *b* – result of pSp; *c* – result of e4e; *d* – result of Restyle; *e* – result of EOGI; *f* – result of the proposed method

Ограничение. Несмотря на достижение высококачественных результатов при изменении атрибутов лица, предложенный метод имеет несколько ограничений. Во-первых, тепловая карта чрезмерно фокусируется на ключевых областях редактирования, что приводит к неполному редактированию таких атрибутов, как возраст, которые

требуют более широкого охвата. Во-вторых, мы использовали предобученную сеть подавления артефактов без тонкой настройки (fine-tuning), что может снизить производительность из-за различий в наборах данных. В-третьих, процесс вставки карты признаков в генератор StyleGAN приводит к новым видимым артефактам. Эти проблемы будут определять нашу будущую работу, включая улучшение распределения тепловой карты, оптимизацию адаптации сети и сокращение нежелательных артефактов.

Заключение

В данной работе мы представили новый модуль дифференциальной активации, предназначенный для точного определения областей, критически важных при изменении атрибутов лица. Предлагаемый метод основан на Grad-CAM++, который учитывает градиенты второго и третьего порядка и обеспечивает более точную оценку значимости пикселей. Дополнительно, использование увеличения тестового времени позволяет точнее центрировать карты активации на целевых объектах, а сглаживание Эйгена способствует подавлению шумов. Включение информации семантической сегментации значительно ограничивает влияние редактирования на нецелевые области вне границ лица. Как качественные, так и количественные анализы подтверждают эффективность и превосходство предложенного метода.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. He Zh., Zuo W., Kan M., Shan Sh., Chen X. AttGAN: Facial Attribute Editing by Only Changing What You Want. *IEEE Transactions on Image Processing*. 2019;28(11):5464–5478. <https://doi.org/10.1109/TIP.2019.2916751>
2. Qiu H., Yu B., Gong D., Li Zh., Liu W., Tao D. SynFace: Face Recognition with Synthetic Data. In: *2021 IEEE/CVF International Conference on Computer Vision (CVPR), 10–17 October 2021, Montreal, QC, Canada*. IEEE; 2021. P. 10860–10870. <https://doi.org/10.1109/ICCV48922.2021.01070>
3. 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].
4. Xia W., Zhang Yu., Yang Yu., Xue J.-H., Zhou B., Yang M.-H. GAN Inversion: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023;45(3):3121–3138. <https://doi.org/10.1109/TPAMI.2022.3181070>
5. 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>
6. 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>
7. 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>
8. 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>
9. 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>
10. 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>
11. 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
12. Chattopadhyay A., Sarkar A., Howlader P., Balasubramanian V.N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 12–15 March 2018, Lake Tahoe, NV, USA. IEEE; 2018. P. 839–847. <https://doi.org/10.1109/WACV.2018.00097>
13. Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, 22–29 October 2017, Venice, Italy. IEEE; 2017. P. 618–626. <https://doi.org/10.1109/ICCV.2017.74>
14. Muhammad M.B., Yeasin M. Eigen-CAM: Class Activation Map Using Principal Components. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, 19–24 July 2020, Glasgow, UK. IEEE; 2020. P. 1–7. <https://doi.org/10.1109/IJCNN48605.2020.9206626>
15. He K., Zhang X., Ren Sh., Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27–30 June 2016, Las Vegas, NV, USA. IEEE; 2016. P. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
16. Lee Ch.-H., Liu Z., Wu L., Luo P. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13–19 June 2020, Seattle, WA, USA. IEEE; 2020. P. 5548–5557. <https://doi.org/10.1109/CVPR42600.2020.00559>
17. 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].
18. 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>
19. 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 19th April 2025].
20. Shen Yu., Yang C., Tang X., Zhou B. InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022;44(4):2004–2018. <https://doi.org/10.1109/TPAMI.2020.3034267>

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](https://orcid.org/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](https://orcid.org/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 29.04.2025; одобрена после рецензирования 03.06.2025; принята к публикации 18.06.2025.

The article was submitted 29.04.2025; approved after reviewing 03.06.2025; accepted for publication 18.06.2025.