

Сравнительный анализ современных методов генерации изображений: VAE, GAN и диффузионные модели

А.В. Катаев, Ю.М. Власова, В.А. Ким, Д.А. Гусынин

Волгоградский государственный технический университет

Аннотация: Статья представляет собой анализ современных методов генерации изображений: вариационных автоэнкодеров (Variational Autoencoder - VAE), генеративных состязательных сетей (Generative Adversarial Networks - GAN) и диффузионных моделей. Основное внимание уделено сравнительному анализу их производительности, качеству генерации и вычислительным требованиям. Для оценки качества изображений используется метрика расстояние Фреше между распределениями признаков (Fréchet Inception Distance - FID). Диффузионные модели показали наилучшие результаты (FID 20.8), превосходя VAE (FID 59.75) и GAN (FID 38.9), но требуют значительных ресурсов. VAE стабильны, но генерируют размытые изображения. GAN обеспечивают высокое качество, но страдают от нестабильности обучения и коллапса моды. Диффузионные модели, благодаря пошаговому декодированию шума, сочетают детализацию и структурированность, что делает их наиболее перспективными. Также рассмотрены методы генерации из изображения в изображение, применяемые для модификации изображений. Результаты исследования полезны для специалистов в области машинного обучения и компьютерного зрения, способствуя улучшению алгоритмов и расширению областей применения генеративных моделей.

Ключевые слова: дипфейк, глубокое обучение, искусственный интеллект, GAN, VAE, диффузионная модель.

Генеративные модели машинного обучения активно развиваются и применяются в задачах создания фотореалистичных изображений, редактирования медиафайлов и автоматической генерации цифрового контента. Современные подходы, такие как вариационные автоэнкодеры (Variational Autoencoder - VAE), генеративные состязательные сети (Generative Adversarial Networks - GAN) и диффузионные модели, обладают различными преимуществами и ограничениями.

VAE обеспечивают устойчивость при обучении и эффективное построение латентных представлений, однако генерируемые изображения часто страдают от размытости и недостатка детализации. GAN способны создавать высококачественные изображения, но их обучение нестабильно, а разнообразие генерируемых данных ограничено из-за проблемы коллапса моды, при которой генератор производит ограниченное множество образцов,

игнорируя полное разнообразие данных. Диффузионные модели демонстрируют высокую детализацию и структурированность генерации, однако требуют значительных вычислительных ресурсов и времени для синтеза изображений.

Одним из ключевых направлений в области генеративных моделей является преобразование существующих изображений с использованием методов генерации из изображения в изображение. Эти методы позволяют модифицировать изображения, сохраняя их структуру, но изменяя стиль, освещение, цветовую палитру или другие атрибуты. Такие технологии применяются для улучшения качества изображений, художественной обработки, удаления шума и восстановления утраченных деталей.

Целью данной статьи является исследование современных методов создания реалистичных изображений, включая анализ их особенностей, преимуществ и недостатков. Особое внимание уделяется диффузионным моделям, которые в последнее время демонстрируют выдающиеся результаты в генерации реалистичных изображений.

Вариационные автоэнкодеры

Вариационные автоэнкодеры (Variational Autoencoder - VAE) представляют собой класс генеративных моделей, способных создавать реалистичные изображения, включая лица. Они основаны на принципах байесовского и вариационного вывода. VAE являются вариантом автоэнкодера, который позволяет не только кодировать изображения в скрытое пространство, но и генерировать новые изображения из этого пространства.

Автоэнкодер состоит из двух основных компонентов: энкодера и декодера. Энкодер преобразует входное изображение в скрытое представление (латентное пространство), а декодер восстанавливает изображение из этого скрытого представления [1, 2].

VAE добавляет вероятностный подход к автоэнкодеру. Он обучает модель, которая генерирует параметры распределения в скрытом пространстве, а не просто кодирует входное изображение в точку в этом пространстве. Это позволяет VAE генерировать новые изображения путем сэмплирования из распределения в скрытом пространстве [1, 2, 3].

Функция потерь VAE (1) состоит из двух частей: потерь восстановления и потери дивергенции Кульбака-Лейблера [3].

$$L_{VAE} = E_{q(z|x)}[\log p(x|z)] - KL(q(z|x) || p(z)), \quad (1)$$

где L_{VAE} - функция потерь VAE, $q(x/z)$ - условное распределение данных, $q(z/x)$ - аппроксимация апостериорного распределения, $p(z)$ - априорное распределение скрытых переменных, $KL(q(z/x)||p(z))$ - дивергенция Кульбака-Лейблера между апостериорным и априорным распределениями.

VAE обучаются на максимизации нижней оценки правдоподобия L_{VAE} , что делает их более стабильными в обучении, но часто приводит к менее резким и детализированным изображениям [2].

Главной особенностью VAE является его способность генерировать новые изображения путем сэмплирования из латентного пространства. Это позволяет модели генерировать разнообразные варианты изображений, однако генерируемые изображения часто выглядят размытыми. Это происходит из-за использования функции потерь, основанной на расстоянии Кульбака-Лейблера, которая стремится минимизировать различия между распределением данных и априорным распределением в латентном пространстве. Таким образом, модель стремится к усредненным решениям, которые снижают детализацию и четкость генерируемых изображений. VAE широко используются для задач генерации лиц, интерполяции и кастомизации атрибутов. Например, β -VAE [4] позволяет получать более интерпретируемые латентные представления, что полезно для задач редактирования изображений. Пример генерации лица с помощью VAE можно наблюдать на рис. 1 [1].



Рис. 1. – Пример генерации лиц с помощью VAE [1]

Генеративные состязательные сети

Одним из популярных методов генерации лиц являются генеративные состязательные сети (Generative Adversarial Networks - GAN). Они основаны на идее соревнования между двумя нейронными сетями: генератором и дискриминатором. Генератор создаёт изображения, пытаясь "обмануть" дискриминатор, который, в свою очередь, учится отличать реальные изображения от сгенерированных. Этот процесс можно рассматривать как минимаксную игру, где генератор минимизирует вероятность обнаружения, а дискриминатор максимизирует её [5].

В отличие от VAE, у которых качество генерации ограничено использованием среднеквадратичной ошибки в качестве функции потерь, GAN решают проблему размытости изображений за счет минимаксной игры. Этот процесс позволяет GAN одновременно обучаться генерировать правдоподобные изображения и создавать собственную метрику их качества, что устраняет необходимость в явной функции потерь, такой, как среднеквадратичная ошибка.

Генератор принимает на вход случайные векторы - шум, и генерирует изображения, которые старается сделать максимально похожими на настоящие лица. Генератор обучается на основе минимизации функции потерь (2), чтобы создавать максимально реалистичные изображения [6].

$$L_{GAN}(G, D) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (2)$$

где $L_{GAN}(G,D)$ - функция потерь генератора, G - генератор, D - дискриминатор, x - реальное изображение, $p_{data}(x)$ - распределение реальных данных, z - случайный шум, $p_z(z)$ - распределение шума, $D(x)$ - выход дискриминатора для реального изображения, $D(G(z))$ - выход дискриминатора для сгенерированного изображения.

Дискриминатор в GAN анализирует изображения и определяет, являются ли они реальными или сгенерированными. Он обучается на основе различения реальных и сгенерированных изображений (3) и оптимизируется для максимально точной классификации [6].

$$L_{GAN}(D) = -E_{x \sim p_{data}(x)}[\log D(x)] - E_{z \sim p_z(z)}[\log(1 - D(G(z)))], \quad (3)$$

где $L_{GAN}(D)$ - функция потерь дискриминатора, G - генератор, D - дискриминатор, x - реальное изображение, $p_{data}(x)$ - распределение реальных данных, z - случайный шум, $p_z(z)$ - распределение шума, $D(x)$ - выход дискриминатора для реального изображения, $D(G(z))$ - выход дискриминатора для сгенерированного изображения.

Главной особенностью GAN является их способность генерировать реалистичные изображения без явного определения правил генерации. Они обучаются на основе данных и сами находят оптимальные параметры для создания изображений.

Ключевыми преимуществами GAN являются их способность создавать изображения высокого качества, контролировать атрибуты генерируемых изображений через манипуляции в латентном пространстве, что открывает широкие возможности для кастомизации [6, 7].

Однако GAN имеют и ряд ограничений. Во-первых, их обучение часто бывает нестабильным из-за дисбаланса между генератором и дискриминатором. Если дискриминатор становится слишком сильным, генератор перестает обучаться, что приводит к проблеме, известной как коллапс моды, когда генератор производит ограниченное разнообразие изображений. Во-вторых, GAN требуют значительных вычислительных

ресурсов, особенно при работе с большими наборами данных или создании изображений высокого разрешения [6].

Одним из самых известных примеров является StyleGAN [7], который позволяет генерировать фотореалистичные лица с контролируемыми атрибутами (например, возраст, причёска, выражение лица). Пример генерации лица с помощью GAN можно наблюдать на рис. 2 [5].



Рис. 2. – Пример генерации лиц с помощью StyleGAN [5]

Диффузионные модели

Диффузионные модели - это класс генеративных моделей, которые создают изображения через процесс постепенного удаления шума из исходного распределения данных. В отличие от GAN и VAE, диффузионные модели работают в два этапа: прямой процесс, где данные постепенно зашумляются, и обратный процесс, где шум преобразуется в реалистичное изображение. Этот подход позволяет создавать изображения с высокой детализацией и разнообразием, что делает диффузионные модели одним из самых перспективных направлений в генеративном моделировании [8].

Прямой процесс в диффузионных моделях описывается как постепенное добавление гауссовского шума к данным (4) [8].

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I), \quad (4)$$

где $q(x_t/x_{t-1})$ - прямой процесс диффузионной модели, x_t - изображение на текущем шаге диффузии, x_{t-1} - изображение на предыдущем шаге диффузии, β_t - распределение шума, $N(\mu, \Sigma)$ - нормальное распределение

шума с математическим ожиданием μ и ковариацией Σ , I - единичная матрица.

Обратный процесс, который восстанавливает изображение из шума, основан на предсказании шума на каждом шаге (5) [8].

$$p_{\theta}(x_{t-1} | x_t) = N(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)), \quad (5)$$

где $p_{\theta}(x_{t-1}|x_t)$ - обратный процесс диффузионной модели, x_t - изображение на текущем шаге диффузии, x_{t-1} - изображение на предыдущем шаге диффузии, $N(\mu_{\theta}, \Sigma_{\theta})$ - нормальное распределение шума с математическим ожиданием μ_{θ} и ковариацией Σ_{θ} .

Функция потерь для обновления изображения на каждом шаге диффузии определена в формуле 6 [8].

$$L_{diffusion}(x_t, x_{t+1}) = \|x_{t+1} - x_t\|_2^2, \quad (6)$$

где $L_{diffusion}(x_t|x_{t+1})$ - функция потерь для диффузионной модели, x_t - изображение на текущем шаге диффузии, x_{t+1} - изображение на следующем шаге диффузии.

Stable Diffusion [8] является одной из диффузионных моделей и включает методику управления диффузией с использованием специальных алгоритмов, таких как метод низкоранговой адаптации [9] и ControlNET [10]. Эти методы направлены на улучшение стабильности процесса диффузии и повышение качества генерируемых изображений.

Функция потерь может быть дополнена другими компонентами, такими как регуляризация или перенос стиля (7) [11].

$$L_{total}(x_t, x_{t+1}) = L_{diffusion}(x_t, x_{t+1}) + \lambda \cdot L_{reg}(x_{t+1}) + \mu \cdot L_{style}(x_{t+1}), \quad (7)$$

где $L_{total}(x_t|x_{t+1})$ - общая функция потерь, x_t - изображение на текущем шаге диффузии, x_{t+1} - изображение на следующем шаге диффузии, $L_{diffusion}(x_t|x_{t+1})$ - функция потерь для диффузионной модели, λ, μ - коэффициенты для регуляризации и переноса стиля соответственно, $L_{reg}(x_{t+1})$ - функция регуляризации, $L_{style}(x_{t+1})$ - функция потерь переноса стиля.

Диффузионные модели способны создавать изображения высокого качества, часто превосходя по этому параметру такие методы, как GAN и VAE. Это подтверждается метриками, такими как расстояние Фреше между распределениями признаков (Fréchet Inception Distance - FID), где диффузионные модели демонстрируют лучшие результаты [8, 12, 13].

Также, процесс обучения диффузионных моделей более стабилен по сравнению с GAN, поскольку они не страдают от проблем, связанных с дисбалансом между генератором и дискриминатором. Это делает их более предсказуемыми и менее требовательными к тонкой настройке гиперпараметров.

Кроме того, диффузионные модели отличаются высокой гибкостью: они могут быть адаптированы для широкого спектра задач, включая генерацию изображений, интерполяцию между изображениями и редактирование существующих изображений. Например, модели вроде DALL-E 2 и Stable Diffusion успешно применяются для создания изображений на основе текстовых описаний, что открывает новые возможности для творчества и автоматизации [9, 10, 11].

Несмотря на свои преимущества, диффузионные модели имеют и ряд ограничений. Одним из главных недостатков является их высокая вычислительная сложность. Процесс генерации изображений требует множества итераций, что делает его медленным по сравнению с GAN и VAE. Это ограничивает их применение в реальном времени или на устройствах с ограниченными вычислительными ресурсами [8].

Кроме того, настройка диффузионных моделей может быть сложной задачей, поскольку требует тщательного выбора параметров, таких как количество шагов и уровень шума. Неправильная настройка может привести к ухудшению качества изображений или увеличению времени генерации.

Диффузионные модели представляют собой одно из современных направлений в области генеративных моделей, и их применение может привести к созданию высококачественных и реалистичных изображений.

Пример генерации лица с помощью диффузионных моделей можно наблюдать на рис. 3 [11].

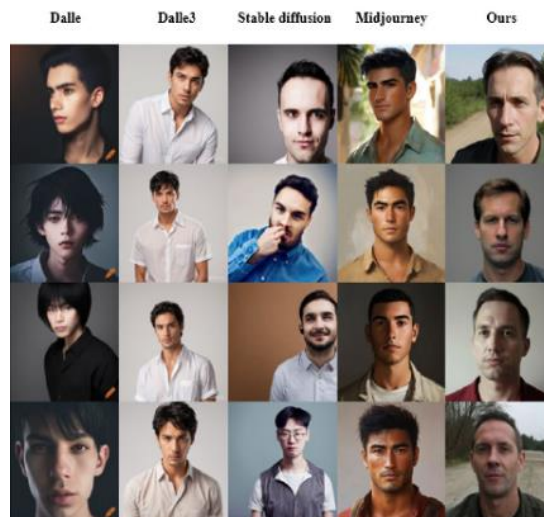


Рис. 3. – Пример генерации лиц с помощью диффузионных моделей [11]

Сравнение методов генерации

Для объективной оценки эффективности генеративных моделей используются различные метрики, такие как FID, индекс Insertion, а также минимальные вычислительные ресурсы графического процессора для обучения модели.

FID - это метрика, которая измеряет качество и реалистичность сгенерированных изображений. Она вычисляет расстояние между распределениями реальных и сгенерированных изображений в пространстве признаков, извлеченных с помощью сети Insertion. Чем ниже значение FID, тем ближе сгенерированные изображения к реальным.

FID вычисляется как расстояние Фреше между двумя многомерными гауссовскими распределениями (8) [14].

$$FID = \left\| \mu_r - \mu_g \right\|^2 + Tr \left(\Sigma_r + \Sigma_g - 2 \cdot \sqrt{\Sigma_r + \Sigma_g} \right), \quad (8)$$

где μ_r, Σ_r - среднее и ковариационная матрица для реальных изображений, μ_g, Σ_g - среднее и ковариационная матрица для сгенерированных изображений.

Индекс Inception - это метрика, оценивающая качество и разнообразие сгенерированных изображений. Она основана на предобученной сети Inception.

Индекс Inception вычисляет расхождение Кульбака-Лейблера между распределением классов, предсказанных InceptionV3, и их средним распределением. Чем выше значение индекса (9), тем выше качество изображения [14].

$$IS = \exp(E_x D_{KL}(p(y|x) \| p(y))), \quad (9)$$

где $p(y/x)$ - распределение классов, предсказанное InceptionV3 для изображения x , $p(y)$ - среднее распределение классов по всем изображениям.

В таблице 1 указано сравнение метрик для каждого из методов генерации изображений [8, 12, 13, 15, 16, 17, 18].

Таблица №1

Сравнение метрик для методов генерации изображений

Название модели	FID	Индекс Inception	Минимальные вычислительные ресурсы графического процессора, гигабайты
VAE	59.753	11.133	4
VQGAN-f16-16384	38.912	13.647	18
Stable Diffusion	20.8	12.5	10

Каждая из представленных моделей обучена на наборе данных “Общие объекты в контексте”. Этот набор является одним из наиболее популярных и широко используемых наборов данных в задачах компьютерного зрения. Он является стандартом для оценки моделей в таких задачах, как обнаружение объектов, сегментация, подписание изображений и генерация изображений. Он содержит более 330 тысяч изображений, на которых размечено около 2,5

миллионов объектов, относящихся к 80 категориям. Каждое изображение сопровождается аннотациями, включающими ограничивающие рамки, маски для сегментации и текстовые описания.

Заключение

В данной статье исследуются современные методы генерации реалистичных изображений, включая вариационные автоэнкодеры, генеративные состязательные сети и диффузионные модели. Каждый из рассмотренных подходов обладает своими особенностями.

VAE эффективны для построения латентных представлений, однако генерируемые изображения часто страдают от размытости. GAN способны создавать высококачественные изображения, но их обучение сопряжено с проблемами нестабильности и ограниченным разнообразием генерируемых данных. Диффузионные модели выделяются способностью генерировать детализированные изображения, что подтверждается метриками FID и индексом Inception. В частности, Stable Diffusion демонстрирует наилучший результат по FID (20.8), что свидетельствует о высокой реалистичности генерируемых изображений.

Анализ требований к вычислительным ресурсам показал, что диффузионные модели, несмотря на свои преимущества, требуют значительных затрат на обучение и синтез изображений. Однако их способность генерировать высококачественные изображения делает их перспективным инструментом для задач, где важна точность и детализация.

Таким образом, проведенное исследование подтверждает, что диффузионные модели являются наиболее перспективным направлением в области генерации реалистичных изображений, несмотря на их вычислительную сложность. Дальнейшие исследования могут быть направлены на оптимизацию этих моделей для снижения требований к

ресурсам и повышения их эффективности в задачах, требующих высокой точности и детализации.

Литература

1. Lijie H. Comparison of improved variational autoencoder models for human face generation // Journal of Physics Conference Series. 2023. pp. 1-8. URL: iopscience.iop.org/article/10.1088/1742-6596/2634/1/012042.
2. Diederik P. K., Max W. An Introduction to Variational Autoencoders // Foundations and Trends in Machine Learning. 2019. Vol. 12. pp. 307-392. URL: doi.org/10.1561/22000000056.
3. Stephen O. Tutorial: Deriving the Standard Variational Autoencoder (VAE) Loss Function // URL: doi.org/10.48550/arXiv.1907.08956.
4. Christopher P. B., Irina H., Arka P., Loic M., Nick W., Guillaume D., Alexander L. Understanding disentangling in β -VAE // Conference: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2018. pp. 1-11. URL: doi.org/10.48550/arXiv.1804.03599.
5. Захаров Е., Волхонский Д. Генеративно-сопоставительные сети (GAN) // Яндекс Образование URL: [education.yandex.ru/handbook/ml/article/generativno-sostyazatelnye-seti-\(gan\)](https://education.yandex.ru/handbook/ml/article/generativno-sostyazatelnye-seti-(gan)).
6. Mario L., Karol K., Marcin M., Sylvain G., Olivier B. Are GANs Created Equal? A Large-Scale Study // Conference: 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada. 2018. pp. 1-21. URL: doi.org/10.48550/arXiv.1711.10337.
7. Tero K., Samuli L., Timo A. A Style-Based Generator Architecture for Generative Adversarial Networks // Neural and Evolutionary Computing. 2018. pp. 1-12. URL: doi.org/10.48550/arXiv.1812.04948.
8. Robin R., Andreas B., Dominik L., Patrick E., Björn O. High-Resolution Image Synthesis with Latent Diffusion Models // Conference: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

-
2022. pp. 10674-10685. URL: doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042.
9. Yang Y., Wen W., Liang P., Chaotian S., Yao C., Hengjia L., Xiaolong Y., Qinglin L., Deng C., Boxi W., Wei L. LoRA-Composer: Leveraging Low-Rank Adaptation for Multi-Concept Customization in Training-Free Diffusion Models // URL: doi.org/10.48550/arXiv.2403.11627.
10. Lvmin Z., Anyi R., Maneesh A. Adding Conditional Control to Text-to-Image Diffusion Models // URL: doi.org/10.48550/arXiv.2302.05543.
11. Xiang W., Xu S., Lv C., Wang S. A Customizable Face Generation Method Based on Stable Diffusion Model // IEEE Access. 2024. Vol. 12. pp. 195307-195318. URL: ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10810380&isnumber=10380310.
12. Shanchuan L., Xiao Y. Diffusion Model with Perceptual Loss // URL: doi.org/10.48550/arXiv.2401.00110.
13. Surojit S., Sarang J., Ross W. ARD-VAE: A Statistical Formulation to Find the Relevant Latent Dimensions of Variational Autoencoders // Conference: IEEE/CVF Winter Conference. 2025. pp. 1-19. URL: doi.org/10.48550/arXiv.2501.10901.
14. Eric J. N., Pejman K., Shadrokh S. Compound Frechet Inception Distance for Quality Assessment of GAN Created Images // URL: doi.org/10.48550/arXiv.2106.08575.
15. CLIP-Guided-Diffusion // URL: github.com/nerdyrodent/CLIP-Guided-Diffusion.
16. PyTorch-VAE // URL: github.com/AntixK/PyTorch-VAE.
17. stable-diffusion-v1-4 // URL: huggingface.co/CompVis/stable-diffusion-v1-4.
18. tuned-vq-gan // URL: github.com/ai-forever/tuned-vq-gan.
-

References

1. Lijie H. Comparison of improved variational autoencoder models for human face generation Journal of Physics Conference Series. 2023. pp. 1-8. URL: iopscience.iop.org/article/10.1088/1742-6596/2634/1/012042.
 2. Diederik P. K., Max W. An Introduction to Variational Autoencoders Foundations and Trends in Machine Learning. 2019. Vol. 12. pp. 307-392. URL: doi.org/10.1561/22000000056.
 3. Stephen O. Tutorial: Deriving the Standard Variational Autoencoder (VAE) Loss Function URL: doi.org/10.48550/arXiv.1907.08956.
 4. Christopher P. B., Irina H., Arka P., Loic M., Nick W., Guillaume D., Alexander L. Understanding disentangling in β -VAE Conference: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2018. pp. 1-11. URL: doi.org/10.48550/arXiv.1804.03599.
 5. Zakharov E., Volkonskiy D. Generativno-sostyazatel'nye seti (GAN) [Generative Adversarial Networks (GAN)] Yandex Obrazovanie [Yandex Education] URL: [education.yandex.ru/handbook/ml/article/generativno-sostyazatelnye-seti-\(gan\)](https://education.yandex.ru/handbook/ml/article/generativno-sostyazatelnye-seti-(gan)).
 6. Mario L., Karol K., Marcin M., Sylvain G., Olivier B. Are GANs Created Equal? A Large-Scale Study Conference: 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada. 2018. pp. 1-21. URL: doi.org/10.48550/arXiv.1711.10337.
 7. Tero K., Samuli L., Timo A. A Style-Based Generator Architecture for Generative Adversarial Networks Neural and Evolutionary Computing. 2018. pp. 1-12. URL: doi.org/10.48550/arXiv.1812.04948.
 8. Robin R., Andreas B., Dominik L., Patrick E., Björn O. High-Resolution Image Synthesis with Latent Diffusion Models Conference: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. pp. 10674-10685. URL: doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042.
-

9. Yang Y., Wen W., Liang P., Chaotian S., Yao C., Hengjia L., Xiaolong Y., Qinglin L., Deng C., Boxi W., Wei L. LoRA-Composer: Leveraging Low-Rank Adaptation for Multi-Concept Customization in Training-Free Diffusion Models URL: doi.org/10.48550/arXiv.2403.11627.
10. Lvmin Z., Anyi R., Maneesh A. Adding Conditional Control to Text-to-Image Diffusion Models URL: doi.org/10.48550/arXiv.2302.05543.
11. Xiang W., Xu S., Lv C., Wang S. A Customizable Face Generation Method Based on Stable Diffusion Model IEEE Access. 2024. Vol. 12. pp. 195307-195318. URL: ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10810380&isnumber=103803
- 10.
12. Shanchuan L., Xiao Y. Diffusion Model with Perceptual Loss. URL: doi.org/10.48550/arXiv.2401.00110.
13. Surojit S., Sarang J., Ross W. ARD-VAE: A Statistical Formulation to Find the Relevant Latent Dimensions of Variational Autoencoders Conference: IEEE/CVF Winter Conference. 2025. pp. 1-19. URL: doi.org/10.48550/arXiv.2501.10901.
14. Eric J. N., Pejman K., Shadrokh S. Compound Frechet Inception Distance for Quality Assessment of GAN Created Images. URL: doi.org/10.48550/arXiv.2106.08575.
15. CLIP-Guided-Diffusion URL: github.com/nerdyrodent/CLIP-Guided-Diffusion.
16. PyTorch-VAE URL: github.com/AntixK/PyTorch-VAE.
17. stable-diffusion-v1-4 URL: huggingface.co/CompVis/stable-diffusion-v1-4.
18. tuned-vq-gan URL: github.com/ai-forever/tuned-vq-gan.

Дата поступления: 24.02.2025 Дата публикации: 25.04 2025