

Булгаков Денис Геннадьевич

*студент 4 курса бакалавриата, Национальный исследовательский ядерный
университет «МИФИ», Россия, г. Москва*

Заика Михаил Алексеевич

*студент 4 курса бакалавриата, Национальный исследовательский ядерный
университет «МИФИ», Россия, г. Москва*

ГЕНЕРАЦИЯ СИНТЕТИЧЕСКИХ ИЗОБРАЖЕНИЙ РУКОПИСНЫХ ПОДПИСЕЙ НА ОСНОВЕ ВАРИАЦИОННОГО АВТОЭНКODЕРА С СОСТАВНОЙ ФУНКЦИЕЙ ПОТЕРЬ

Аннотация. В статье рассматривается задача синтеза изображений оффлайн рукописных подписей средствами глубокого генеративного моделирования. Актуальность работы определяется дефицитом размеченных данных при обучении систем биометрической верификации и ограничениями на распространение реальных баз подписей, обусловленными требованиями конфиденциальности. Цель исследования — разработка вариационного автоэнкодера, обеспечивающего управляемый синтез визуально правдоподобных вариаций подписи в окрестности заданного эталонного образца, и экспериментальная оценка вклада отдельных компонент функции потерь в качество реконструкции. Основным препятствием на этом пути является характерная для классических вариационных автоэнкодеров размытость выходных изображений, критичная для подписей, где значимы тонкие штрихи и топологическая связность элементов. Предложена составная функция потерь из восьми компонент, контролирующих пиксельную точность, структурное и перцептивное сходство, чёткость границ, геометрическое расположение и связность штрихов, бинарность выхода и регулярность латентного пространства, а также метод локального сэмплирования с параметром отклонения. Экспериментальная проверка выполнена на корпусе из 30 157 изображений подписей.

Ключевые слова: вариационный автоэнкодер, синтез изображений, рукописная подпись, латентное пространство, составная функция потерь, локальное сэмплирование, LPIPS.

Annotation. This article addresses the synthesis of offline handwritten signature images by means of deep generative modelling. The relevance of the work is determined by the shortage of labelled data for training biometric verification systems and by the confidentiality restrictions that limit the distribution of real signature databases. The purpose of the study is to develop a variational autoencoder that provides controlled synthesis of visually plausible signature variations in the neighbourhood of a given reference sample, and to experimentally assess the contribution of individual loss components to reconstruction quality. The main obstacle here is the blurriness of the output images typical of classical variational autoencoders, which is critical for signatures, where thin strokes and topological connectivity of elements matter. A composite loss function of eight components is proposed, controlling pixel accuracy, structural and perceptual similarity, edge sharpness, geometric placement and connectivity of strokes, output binarity and latent space regularity, together with a local sampling method governed by a deviation parameter. The experimental evaluation was carried out on a corpus of 30,157 signature images.

Keywords: variational autoencoder, image synthesis, handwritten signature, latent space, composite loss function, local sampling, LPIPS.

Синтез изображений рукописных подписей востребован в двух прикладных контекстах. Первый — аугментация обучающих выборок систем автоматической верификации, где дефицит размеченных данных остаётся основным ограничивающим фактором качества [15]. Второй — исследования в области почерковедения, где требуется моделирование естественной вариативности почерка одного человека [5]. При этом доступ к реальным базам подписей ограничен требованиями конфиденциальности персональных данных, что превращает разработку методов правдоподобного синтеза в

самостоятельную научно-техническую задачу. Классические подходы опираются на нейромоторное моделирование процесса письма и явное параметрическое описание траектории пера [6]; современные методы обучаются непосредственно на растровых изображениях без реконструкции динамики письма [10].

Три доминирующие парадигмы глубокого генеративного моделирования различаются по применимости к данной задаче. Генеративно-состязательные сети [7] обеспечивают высокое визуальное качество, однако их обучение сопряжено с нестабильностью и коллапсом мод, что критично при ограниченном объеме данных и затрудняет контроль над результатом синтеза. Диффузионные вероятностные модели [8, 14] превосходят прочие подходы по качеству, но требуют значительных вычислительных ресурсов и большого числа обучающих примеров, а итеративная процедура сэмплирования делает инференс дорогим. Вариационный автоэнкодер [11, 13] обладает явным вероятностным латентным пространством, что позволяет не только порождать образцы, но и управлять их характеристиками. Именно это свойство определило выбор архитектуры: рассматриваемая задача требует не безусловной генерации, а управляемого синтеза вариаций в окрестности заданного эталона. Известным недостатком классических вариационных автоэнкодеров является размытость реконструкций [1, 3] — дефект, критичный для подписей, где значимы тонкие штрихи шириной один-три пикселя и связность элементов. Настоящая работа посвящена преодолению этого недостатка средствами составной функции потерь.

Обучающий корпус составлен из изображений подписей без какой-либо разметки: ни идентификаторы авторов, ни метки подлинности не используются, поскольку модель оптимизирует исключительно реконструкционные и регуляризационные цели. Из 30 286 собранных изображений полный конвейер предобработки прошли 30 157; остальные отбракованы как нечитаемые или превысившие таймаут обработки. Конвейер включает шесть этапов: перевод в оттенки серого; нормализацию полярности

по средней яркости относительно порога 127; морфологическое удаление линейных артефактов бланков операцией открытия с горизонтальным ядром 40×1 и вертикальным 1×40 , выполняемое только при фактическом обнаружении линий; обрезку полей по крайним строкам и столбцам, содержащим пиксели темнее 240, с отступом 8 пикселей; бинаризацию методом Оцу; масштабирование к разрешению 256×256 и нормализацию к диапазону $[-1, 1]$. В режиме обучения применяется случайная аффинная аугментация с намеренно узкими диапазонами — поворот в пределах $\pm 5^\circ$, сдвиг до 5 % по каждой оси, масштабирование в пределах $[0,95; 1,05]$; горизонтальное отражение применяется с вероятностью лишь 0,1, поскольку для большинства подписей оно нарушает естественную направленность штрихов.

Энкодер отображает одноканальное изображение в параметры факторизованного нормального апостериорного распределения. Свёрточная подсеть состоит из пяти блоков с ядром 4×4 и шагом 2, каждый из которых вдвое уменьшает пространственное разрешение; за свёрткой следуют пакетная нормализация и функция активации LeakyReLU с коэффициентом наклона 0,2, причём в первом блоке нормализация не применяется, так как вход уже нормализован. Итоговая карта признаков разворачивается в вектор и проецируется общим полносвязным слоем размерности 1024, из которого две независимые линейные головы предсказывают вектор средних и логарифм вектора дисперсий. Разделение голов является принципиальным архитектурным решением: оно позволяет обучать эти два набора параметров независимо, не создавая конфликта между стремлением к точной реконструкции, снижающим дисперсию, и регуляризацией, удерживающей её вблизи единицы. Логарифм дисперсии ограничивается диапазоном $[-4, 4]$, что предотвращает численное переполнение при последующем экспоненцировании.

Декодер симметричен энкодеру: латентный вектор проходит два полносвязных слоя, разворачивается в начальную карту признаков и

последовательно апсэмплируется пятью блоками, каждый из которых выполняет интерполяцию с масштабом 2, свёртку 3×3 и нормализацию по образцу (Instance Normalization); функцией активации промежуточных блоков служит ReLU, финального — гиперболический тангенс, что гарантирует принадлежность выходных значений диапазону $(-1, 1)$. Апсэмплинг интерполяцией с последующей свёрткой, в отличие от транспонированных свёрток, не порождает артефактов «шахматной доски». Метод интерполяции выбирается в зависимости от разрешения: на первых двух блоках применяется интерполяция ближайшего соседа, сохраняющая дискретные границы, поскольку при малом разрешении билинейное усреднение смешало бы штрих и фон до того, как свёрточный слой сможет их разделить; на последующих блоках, где пространственный шаг пикселя мал относительно толщины штриха, применяется билинейная интерполяция.

Реконструкционный член вариационной нижней оценки в стандартной формулировке для изображений подписей неприемлем: штрихи занимают малую долю площади кадра, и равномерный пиксельный штраф оптимизирует преимущественно точность воспроизведения фона. Для устранения этого дисбаланса каждому пикселю назначается вес, пропорциональный его интенсивности:

$$w_{ij} = 9 (I_{ij} + 1) / 2 + 1, \quad (1)$$

откуда при значении пикселя, соответствующем штриху, вес равен 10, а при значении, соответствующем фону, — единице; ошибки в области штрихов штрафуются в десять раз сильнее. Полная функция потерь строится как взвешенная сумма семи реконструкционных компонент и регуляризатора:

$$\begin{aligned} L = & w_{l1}L_{l1} + w_{ssim}L_{ssim} + w_{perc}L_{perc} + \beta L_{KL} + \\ & + w_{edge}L_{edge} + w_{ch}L_{ch} + w_{bin}L_{bin} + w_{topo}L_{topo}. \end{aligned} \quad (2)$$

Структурная и перцептивная компоненты [9, 16, 17] контролируют соответственно яркостно-контрастно-структурное и признаковое сходство, причём перцептивная вычисляется на фотометрически инвертированных изображениях, поскольку используемая предобученная сеть обучалась на

естественных изображениях с тёмными объектами на светлом фоне. Граничная компонента сопоставляет карты Собеля и Лапласа, компонента формы — мягкие карты расстояний, компонента бинарности — поэлементную энтропию Шеннона, топологическая компонента — карты связности штрихов, формируемые дифференцируемыми операциями эрозии и дилатации.

Коэффициент при регуляризаторе задаёт компромисс между двумя конкурирующими целями. При стремлении его к нулю энкодер коллапсирует апостериорное распределение в точку, превращая модель в детерминированный автоэнкодер: реконструкция высококачественна, однако латентное пространство распадается на изолированные кластеры, не соединённые непрерывными переходами, и любое возмущение выводит вектор в «пустую» область, декодирование из которой даёт артефакты [4, 12]. При значениях порядка единицы структура пространства близка к стандартной нормальной, что обеспечивает непрерывность сэмплируемой области, но избыточная регуляризация подавляет способность энкодера кодировать тонкие детали. Рабочим критерием выбрана средняя KL-дивергенция на размерность в диапазоне от 1,5 до 4,0 нат; ему соответствует значение $\beta = 0,0005$.

Центральным методом генерации является локальное сэмплирование — управляемое стохастическое возмущение латентного вектора, соответствующего эталонной подписи. Эталонное изображение кодируется в параметры апостериорного распределения, после чего формируется возмущённый вектор $z^* = \mu + \lambda \cdot \sigma \odot \varepsilon$, где ε — стандартный гауссовский шум, $\odot$ — поэлементное произведение, а λ — вводимый параметр масштаба. Ожидаемое евклидово расстояние от точки сэмплирования до центра распределения вычисляется аналитически:

$$E\|z^* - \mu\|_2 = \lambda \cdot \|\sigma\|_2. \quad (3)$$

Из соотношения (3) следует, что λ линейно масштабирует ожидаемый радиус сэмплирования относительно нормы вектора апостериорных стандартных отклонений. Существенно, что такая интерпретация инвариантна к

конкретному изображению: значение $\lambda = 1$ отвечает типичному расстоянию между двумя выборками из апостериорного распределения одного и того же образца. Это позволяет выделить три содержательных режима генерации. При λ не более 0,5 синтезированные подписи практически совпадают с детерминированной реконструкцией эталона, различаясь лишь микрогеометрией штрихов. При λ в диапазоне от 0,5 до 1,5 подписи сохраняют общую структуру и авторский стиль, но заметно варьируют пропорции, наклон и форму отдельных элементов — этот диапазон и признан наиболее пригодным для аугментации. При λ свыше 1,5 точка выходит за пределы хорошо структурированной части латентного пространства, и декодирование даёт выраженные артефакты.

Программный комплекс реализован на Python 3.10 с использованием PyTorch 2.x; смешанная точность FP16 с автоматическим масштабированием градиентов удерживает потребление видеопамати ниже 12 ГБ при размере батча 32. Оптимизатор Adam настроен с пониженным значением первого момента 0,5 для уменьшения инерции при нестабильных потерях; скорость обучения снижается планировщиком при стагнации метрик. Обработка изображений распределена между Pillow, scikit-image (бинаризация методом Оцу и морфологические операции) и torchvision (аугментации внутри вычислительного графа). Перцептивная и структурная компоненты потерь подключают библиотеки lpips и pytorch-msssim, тогда как топологическая компонента реализована на примитивах PyTorch, что обеспечивает её совместимость с режимом масштабирования градиентов. Комплекс из восьми подсистем поддерживает шесть режимов генерации: безусловное сэмплирование из априорного распределения, детерминированную реконструкцию, линейную интерполяцию между двумя образцами, двумерное сканирование по выбранным осям, векторную арифметику над латентными кодами и управляемые вариации эталона. Многоуровневое тестирование, охватившее 31 сценарий, верифицировало все программные модули.

Вклад отдельных компонент функции потерь оценивался аблационным экспериментом. Исследовались шесть конфигураций: базовая, включающая только пиксельную, перцептивную и регуляризационную компоненты, четыре варианта с поочерёдным исключением одной из структурных компонент и полная модель. Все конфигурации обучались 50 эпох при одинаковой инициализации весов, размерности латентного пространства 128, скорости обучения $2 \cdot 10^{-5}$ и размере батча 32. Метрики вычислялись на всей обучающей выборке в режиме детерминированной реконструкции; дополнительно фиксировалось число главных компонент, объясняющих 95 % дисперсии распределения апостериорных средних. Результаты приведены в таблице 1.

Таблица 1 – Результаты аблационного исследования компонент функции потерь

Конфигурация	SSIM	LIPS	Компонент до 95 %
Базовая	0,6919	0,1784	105 / 128
Без бинарности	0,7180	0,1522	115 / 128
Без границ	0,7079	0,1803	114 / 128
Без топологии	0,7259	0,1766	112 / 128
Без формы	0,7167	0,1554	117 / 128
Полная	0,7397	0,1419	118 / 128

Полная конфигурация превосходит базовую по обоим метрикам одновременно, однако наиболее содержателен качественный анализ: каждой отключённой компоненте соответствует характерный тип деградации. Без граничной компоненты декодер не получает прямого сигнала о качестве краёв и воспроизводит переход от штриха к фону плавным полутоновым градиентом — реконструкции приобретают «акварельный» вид, при котором линии угадываются, но теряют графическую определённость. Без компоненты бинарности декодер генерирует полутоновые пиксели там, где не уверен в отнесении к штриху или фону, что проявляется как размытая дымка вокруг линий. Обе компоненты действуют синергетически: граничная обеспечивает резкость переходов, а бинарности — чистоту областей штрихов. Без

топологической компоненты возникают два специфических дефекта — слияние близко расположенных штрихов в единую полосу и разрыв длинных линий в местах пересечений. Показательно, что перцептивная метрика этот тип ошибки практически не улавливает: без топологической компоненты она составляет 0,1766 против 0,1784 у базовой конфигурации, то есть отличается в пределах погрешности, тогда как воспринимаемая корректность подписи страдает заметно. Без компоненты формы декодер воспроизводит правильную форму штриха, но со смещением на несколько пикселей, что при попарном сравнении читается как дрожание контуров. Наконец, полная функция потерь повышает число главных компонент, объясняющих 95 % дисперсии, со 105 до 118 из 128: усложнение реконструкционного критерия вынуждает энкодер равномернее распределять информацию по латентным осям.

Исследование размерности латентного пространства при прочих равных условиях показало наличие оптимума. При размерности 32 модель демонстрирует наихудшие показатели ($SSIM = 0,6366$, $LIPS = 0,3536$) при практически полной утилизации пространства — 30 активных измерений из 32, что свидетельствует об упоре в потолок ёмкости. При размерности 64 перцептивное расстояние снижается до 0,2845, однако разрыв с более широкими конфигурациями остаётся существенным. Размерность 128 обеспечивает значительный прирост структурного сходства до 0,7097 при высокой утилизации: 118 из 128 компонент содержательно задействованы. Увеличение до 256 формально улучшает метрики реконструкции ($SSIM = 0,7322$, $LIPS = 0,1559$), но для объяснения 95 % дисперсии достаточно лишь 120 компонент из 256, то есть более половины пространства остаётся неструктурированным. Для задачи генерации это неприемлемо: локальное сэмплирование затрагивало бы «мёртвые» оси, не несущие информации об изображении, что снижает предсказуемость синтезируемых вариаций. Размерность 128 принята как финальная.

Отдельно исследовалось влияние ширины свёрточных блоков энкодера и декодера при фиксированных прочих параметрах (таблица 2).

Таблица 2 – Влияние конфигурации свёрточных блоков на качество и стоимость обучения

Конфигурация каналов	Параметров	SSIM	LPIPS	Эпоха, с
Узкая: 8–16–32–64–128	17 млн	0,6921	0,1848	16,6
Сбалансированная: 16–32–64–128–256	35 млн	0,7097	0,1799	20,9
Широкая: 32–64–128–256–512	72 млн	0,6778	0,2115	32,1

Узкая конфигурация при вдвое меньшем числе параметров и заметно меньшем времени эпохи уступает сбалансированной по структурному сходству: свёрточные блоки с ограниченным числом каналов не накапливают достаточного количества признаков и фактически не различают тонкие детали штрихов — реконструкция сохраняет общую форму подписи, но теряет характерные мелкие элементы. Широкая конфигурация, напротив, вдвое увеличивает число параметров и время эпохи, но демонстрирует наихудшие метрики из трёх вариантов. Причина в том, что при фиксированных 50 эпохах обучения оптимизатор не успевает согласовать избыточное число параметров: модель генерирует несвязные полутоновые области вместо чётких штрихов. Сбалансированная конфигурация обеспечивает наилучшее соотношение качества и вычислительных затрат и принята в качестве финальной.

Итоговая модель обучалась 150 эпох в финальной конфигурации. Оценка выполнялась в детерминированном режиме, при котором декодер получает вектор средних без добавления шума, что изолирует качество представления от вариабельности, вносимой методом репараметризации. Значения полного набора метрик приведены в таблице 3.

Таблица 3 – Метрики качества детерминированной реконструкции финальной модели

Метрика	Значение
Взвешенная поэлементная ошибка	0,1058
SSIM	0,7752
LPIPS	0,1259
Граничная ошибка	0,1059

Метрика	Значение
Ошибка формы	0,1462
Бинарная ошибка	0,0271
Топологическая ошибка	0,0165

Низкие значения бинарной и топологической ошибок (0,0271 и 0,0165 соответственно) свидетельствуют о том, что выход модели действительно близок к двухуровневому и сохраняет связность компонент — то есть обе специализированные компоненты функции потерь достигают своей цели. Наибольшими среди метрик остаются ошибка формы и граничная ошибка, что согласуется с качественным анализом: именно точное позиционирование и резкость краёв тонких штрихов составляют основную трудность для декодера. Характеристики обученного латентного пространства приведены в таблице 4.

Таблица 4 – Характеристики обученного латентного пространства
(размерность 128)

Показатель	Значение
Активных измерений	128 / 128
Измерений с KL-вкладом свыше 0,1	128 / 128
Минимальное стандартное отклонение средних	0,8579
Медианное стандартное отклонение средних	0,8938
Максимальное стандартное отклонение средних	1,0575
Компонент до 80 % дисперсии	91
Компонент до 90 % дисперсии	108
Компонент до 95 % дисперсии	118

Приведённые данные свидетельствуют о полном и равномерном использовании латентного пространства. Все измерения активны, а стандартные отклонения апостериорных средних сосредоточены в узком диапазоне вблизи единицы, что соответствует целевому априорному распределению и указывает на эффективную регуляризацию без коллапса. Анализ главных компонент показывает умеренную концентрацию дисперсии: явной главной оси, несущей основную часть вариации, не существует. Двумерная проекция апостериорных средних образует компактное связанное

облако без разрывов, что подтверждает непрерывность пространства. Именно эти свойства являются необходимым условием корректности локального сэмплирования: без них возмущение вблизи эталона порождало бы не вариации почерка, а случайные артефакты.

Качественный анализ неудачных реконструкций позволяет выделить два устойчивых типа ошибок. Первый — утрата мелких деталей: тонкие штрихи воспроизводятся неуверенно, поскольку декодер усредняет вероятность отнесения к штриху по небольшой окрестности и не восстанавливает границу шириной в один-два пикселя. Второй — сглаживание изолированных элементов: короткие росчерки и особенно буквенные фрагменты воспроизводятся хуже, чем протяжённые связные линии. При этом общая форма подписи, её наклон и характерный ритм штрихов сохраняются во всех случаях, то есть перечисленные ошибки затрагивают локальные детали, не нарушая целостного перцептивного облика подписи.

Проведённое исследование показывает, что основной недостаток вариационных автоэнкодеров применительно к синтезу рукописных подписей — размытость выходных изображений — устраняется не изменением архитектуры, а специализацией реконструкционного критерия. Каждая из четырёх структурных компонент отвечает за отдельный аспект правдоподобия, и ни одна не дублирует другие: граничная контролирует резкость, компонента бинарности — чистоту областей, топологическая — связность, компонента формы — пространственное расположение. Существенно, что топологическая компонента практически не отражается в агрегированных перцептивных метриках, но её отключение приводит к разрывам и слияниям штрихов, то есть к дефектам, прямо нарушающим биометрическую правдоподобность. Это указывает на ограниченность оценки генеративных моделей для подписей исключительно метриками общего назначения и обосновывает необходимость специализированных структурных показателей.

Второй результат состоит в том, что выбор размерности латентного пространства для задачи управляемой генерации определяется не метриками реконструкции, а степенью утилизации пространства: конфигурация с размерностью 256 формально превосходит принятую, но её пространство наполовину неструктурировано, что лишает локальное сэмплирование предсказуемости. Аналогичное расхождение между формальным качеством и пригодностью наблюдается в конфигурациях свёрточных блоков, где наибольшая по числу параметров модель оказывается наихудшей при ограниченном бюджете эпох обучения. Практически значимо и то, что параметр λ измеряется в единицах типичного разброса апостериорного распределения, а не в абсолютных величинах, что позволяет применять единую настройку ко всему корпусу.

Основным ограничением разработанной модели остаётся воспроизведение мелких изолированных элементов подписи. Перспективными направлениями дальнейшей работы, вытекающими из полученных результатов, являются переход к обучению с постепенным повышением разрешения, позволяющий декодеру формировать тонкие штрихи на поздних этапах, и введение условной генерации по идентификатору автора, которая дала бы возможность синтезировать не только вариации конкретного образца, но и новые подписи в заданном стиле.

Список литературы

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2018. 652 с.
2. Марлей В. Е., Терехов А. Н., Гатчин Ю. А., Милушков В. И., Лиманский Н. Н. Сравнительное исследование эффективности автоэнкодеров в задачах обнаружения аномалий // Нейрокомпьютеры: разработка, применение. 2024. Т. 26, № 5. С. 96–106.
3. Николенко С. И., Кадури А. А., Архангельская Е. О. Глубокое обучение. СПб.: Питер, 2018. 480 с.

4. Bowman S. R., Vilnis L., Vinyals O., Dai A. M., Jozefowicz R., Bengio S. Generating Sentences from a Continuous Space // Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL). 2016. P. 10–21.
5. Diaz M., Ferrer M. A., Impedovo D., Malik M. I., Pirlo G., Plamondon R. A Perspective Analysis of Handwritten Signature Technology // ACM Computing Surveys. 2019. Vol. 51, No. 6. Art. 117. P. 1–39.
6. Ferrer M. A., Diaz-Cabrera M., Morales A. Static Signature Synthesis: A Neuromotor Inspired Approach for Biometrics // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015. Vol. 37, No. 3. P. 667–680.
7. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 2672–2680.
8. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 6840–6851.
9. Johnson J., Alahi A., Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution // Proceedings of the European Conference on Computer Vision (ECCV). 2016. P. 694–711.
10. Kang L., Riba P., Wang Y., Rusiñol M., Fornés A., Villegas M. GANwriting: Content-Conditioned Generation of Styled Handwritten Word Images // Proceedings of the European Conference on Computer Vision (ECCV). 2020. P. 273–289.
11. Kingma D. P., Welling M. An Introduction to Variational Autoencoders // Foundations and Trends in Machine Learning. 2019. Vol. 12, No. 4. P. 307–392.
12. Lucas J., Tucker G., Grosse R., Norouzi M. Don't Blame the ELBO! A Linear VAE Perspective on Posterior Collapse // Advances in Neural Information Processing Systems. 2019. Vol. 32. P. 9403–9413.
13. Rezende D. J., Mohamed S., Wierstra D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models // Proceedings of the 31st International Conference on Machine Learning (ICML). 2014. P. 1278–1286.

14. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 10684–10695.
15. Ruiz V., Linares I., Sanchez A., Velez J. F. Off-line Handwritten Signature Verification Using Compositional Synthetic Generation of Signatures and Siamese Neural Networks // Neurocomputing. 2020. Vol. 374. P. 30–41.
16. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612.
17. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595.

References

1. Goodfellow I., Bengio Y., Courville A. Glubokoe obuchenie [Deep learning]. Moscow, DMK Press Publ., 2018. 652 p. (In Russian).
2. Marley V. E., Terekhov A. N., Gatchin Yu. A., Milushkov V. I., Limanskiy N. N. Comparative study of autoencoder efficiency in anomaly detection problems // Neurocomputers: Development, Application. 2024. Vol. 26, No. 5. P. 96–106. (In Russian).
3. Nikolenko S. I., Kadurin A. A., Arkhangelskaya E. O. Glubokoe obuchenie [Deep learning]. St. Petersburg, Piter Publ., 2018. 480 p. (In Russian).
4. Bowman S. R., Vilnis L., Vinyals O., Dai A. M., Jozefowicz R., Bengio S. Generating Sentences from a Continuous Space // Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL). 2016. P. 10–21.
5. Diaz M., Ferrer M. A., Impedovo D., Malik M. I., Pirlo G., Plamondon R. A Perspective Analysis of Handwritten Signature Technology // ACM Computing Surveys. 2019. Vol. 51, No. 6. Art. 117. P. 1–39.

6. Ferrer M. A., Diaz-Cabrera M., Morales A. Static Signature Synthesis: A Neuromotor Inspired Approach for Biometrics // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015. Vol. 37, No. 3. P. 667–680.
7. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 2672–2680.
8. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 6840–6851.
9. Johnson J., Alahi A., Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution // Proceedings of the European Conference on Computer Vision (ECCV). 2016. P. 694–711.
10. Kang L., Riba P., Wang Y., Rusiñol M., Fornés A., Villegas M. GANwriting: Content-Conditioned Generation of Styled Handwritten Word Images // Proceedings of the European Conference on Computer Vision (ECCV). 2020. P. 273–289.
11. Kingma D. P., Welling M. An Introduction to Variational Autoencoders // Foundations and Trends in Machine Learning. 2019. Vol. 12, No. 4. P. 307–392.
12. Lucas J., Tucker G., Grosse R., Norouzi M. Don't Blame the ELBO! A Linear VAE Perspective on Posterior Collapse // Advances in Neural Information Processing Systems. 2019. Vol. 32. P. 9403–9413.
13. Rezende D. J., Mohamed S., Wierstra D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models // Proceedings of the 31st International Conference on Machine Learning (ICML). 2014. P. 1278–1286.
14. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 10684–10695.
15. Ruiz V., Linares I., Sanchez A., Velez J. F. Off-line Handwritten Signature Verification Using Compositional Synthetic Generation of Signatures and Siamese Neural Networks // Neurocomputing. 2020. Vol. 374. P. 30–41.

16. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. 2004. Vol. 13, No. 4. P. 600–612.

17. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595.