

УДК 004.932.75¹, 004.89

 10.25209/2079-3316-2026-17-2-191-262

Roof-DeGAN: гибридная GAN с межмасштабным вниманием для восстановления областей крыш на аэрофотоснимках

Игорь Викторович **Винокуров**^{1✉}, Георгий Михайлович **Лапаньков**²,
Георгий Дмитриевич **Умаров**³

¹⁻³ Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

[✉]igvvinokurov@fa.ru

Аннотация. В работе предложена гибридная генеративно-сопоставительная модель Roof-DeGAN для восстановления повреждённых и скрытых участков изображений крыш на аэрофотоснимках. Архитектура сочетает Vision Transformer с плотными связями в генераторе и многоуровневый дискриминатор с межмасштабным вниманием. Модель объединяет преимущества GAN, элементов диффузионного моделирования и трансформерных механизмов. Эксперименты на данных ППК «Роскадастр» показали превосходство над современными методами: PSNR = 33,7 дБ, SSIM = 0,971, LPIPS = 0,048, FID = 17,8 при времени инференса 0,15 с на изображение. Разработанный подход обладает высокой практической ценностью для задач кадастрового учёта и обновления картографических материалов.

Ключевые слова и фразы: генеративно-сопоставительные сети, Roof-DeGAN, восстановление изображений, аэрофотоснимки, дистанционное зондирование, реконструкция крыш

Для цитирования: Винокуров И. В., Лапаньков Г. М., Умаров Г. Д. *Roof-DeGAN: гибридная GAN с межмасштабным вниманием для восстановления областей крыш на аэрофотоснимках* // Программные системы: теория и приложения. 2026. Т. 17. № 2(71). С. 191–262. https://psta.psiras.ru/read/psta2026_2_191-262.pdf

Введение

Восстановление изображений и контуров зданий на аэрофотоснимках для создания актуальных карт местности и кадастровой оценки объектов капитального строительства является одной из основных задач, решаемых в ППК «Роскадастр». Эффективное решение этой задачи представляет значительный интерес для картографии, мониторинга городской инфраструктуры и градостроительного планирования [1, 2]. Однако на практике аэрофотоснимки часто содержат дефекты, вызванные атмосферными факторами (облачность, туман), временными объектами (строительная техника, транспорт, кроны деревьев) или техническими ограничениями съёмки (шумы сенсоров, низкое разрешение). Качество восстановления напрямую влияет на точность последующего автоматизированного анализа, включая семантическую и инстанс-сегментацию, построение трёхмерных моделей и оценку состояния объектов капитального строительства.

Традиционные методы восстановления изображений, такие как диффузия по Навье–Стоксу [3] и патч-матчинг на основе поиска подобия [4] демонстрируют ограниченную эффективность при работе со сложными структурами крыш. Эти подходы опираются преимущественно на низкоуровневые признаки (интенсивность пикселей, градиенты цветов и структур) и не учитывают семантику объектов, что приводит к размыванию границ, искажению геометрических форм (скаты, коньки, ендовы) и появлению неестественных текстурных артефактов. При реконструкции протяжённых повреждённых областей такие методы не способны генерировать принципиально новое содержание, что критично для аэрофотоснимков с крупными дефектами.

Современные методы глубокого обучения предлагают три основных направления решения задачи: генеративно-сопоставительные сети (GAN), диффузионные вероятностные модели и их гибридные комбинации.

Генеративно-сопоставительные сети остаются востребованными благодаря высокой скорости инференса и устойчивости к ограниченным объёмам обучающих данных. В работе [5] были представлены основы сопоставительного обучения, а условные GAN (*conditional GAN*, cGAN) для трансляции «изображение-в-изображение» стали базовым подходом к восстановлению. Однако классические архитектуры, такие как Pix2Pix [6], ESRGAN [7], демонстрируют падение качества при обработке крупных масок из-за ограниченной ёмкости дискриминатора и недостаточного штрафа за высокочастотные искажения, что особенно заметно на текстурах кровельных материалов.

Диффузионные вероятностные модели демонстрируют высокое качество генерации. Базовая архитектура DDPM [8] и её развитие для задач восстановления [9] позволяют достигать высокого реализма, однако требуют тысяч итераций на инференсе и больших объёмов данных. В 2025 году появились специализированные адаптации для дистанционного зондирования: SatDiff [10] на базе Stable Diffusion обеспечивает высококачественное восстановление спутниковых снимков, метод КАО [11] вводит адаптивную к ядру оптимизацию, превосходящую предыдущие подходы в структурных задачах, метод Image Characteristic-Guided [12] учитывает низкоранговые свойства изображений. Несмотря на превосходные метрики, эти модели сохраняют высокую вычислительную сложность и низкую скорость инференса.

Гибридные методы объединяют преимущества GAN и диффузии. Современные гибридные реализации GAN, такие как DeGAN [13], BD-VITGAN [14] и TAMGAN [15], интегрируют трансформеры, плотные связи и многоуровневое внимание, значительно повышая структурную согласованность и перцептивное качество в задачах удалённого зондирования. Другие подходы, такие как DSEPGAN [16], позволяют достичь баланса между скоростью, детерминированностью и качеством восстановления, что особенно важно для узкоспециализированных задач аэрофотосъёмки крыш.

Современное состояние области характеризуется фрагментарностью: большинство методов разрабатываются для общих датасетов, диффузионные модели требуют избыточных ресурсов, а гибридные подходы недостаточно адаптированы к специфике аэрофотоснимков крыш. Отсутствует специализированный метод, сочетающий устойчивость к малым выборкам, детерминированность инференса, высокое перцептивное качество и умеренную вычислительную сложность. Настоящая работа направлена на устранение этих недостатков и является продолжением исследований, описанных автором в работах [17, 18]. Предложен гибридный подход к восстановлению повреждённых областей крыш на аэрофотоснимках, основанный на условной генеративно-сопоставительной сети DeGAN с трансформерными блоками в генераторе и многоуровневом дискриминаторе с вниманием между уровнями детализации. Составная функция потерь объединяет сопоставительную компоненту, попиксельное сравнение, перцептивные признаки и диффузионное усиление, что позволяет достичь баланса между точностью геометрической реконструкции и естественностью текстур кровельных материалов.

В разделе 1 приводится формальная постановка задачи восстановления скрытых областей крыш. Раздел 2 описывает архитектуру Roof-DeGAN, включая строение генератора с трансформерными блоками (оснащёнными механизмом динамического разрежения внимания) и многоуровневого дискриминатора, а также составную функцию потерь. В разделе 3 представлены используемые датасеты (ZRG для предобучения и целевой датасет с синтетическими повреждениями), методы предобработки данных, а также детали двухэтапного обучения модели (предобучение на сегментацию и дообучение на восстановление) и перечень метрик количественной оценки (PSNR, SSIM, LPIPS, FID). В разделе 4 анализируется динамика обучения и сходимость модели. Раздел 5 представляет количественное сравнение предложенного подхода с современными методами. В разделе 6 проводится абляционное исследование для оценки вклада каждой компоненты функции потерь, а также влияния предобучения на ZRG. Раздел 7 анализирует качество восстановления в зависимости от площади повреждения. В разделе 8 обсуждаются ограничения и перспективные направления дальнейших исследований. Завершают статью раздел 9 и список литературы.

1. Постановка задачи

Задача восстановления повреждённых или скрытых областей крыш зданий на аэрофотоснимках может быть сформулирована как задача условной генерации изображения. Пусть $I_{gt} \in \mathbb{R}^{H \times W \times 3}$ — исходное неповреждённое RGB-изображение крыши здания размером $H \times W$ пикселей. Бинарная маска повреждённой области $M \in \{0, 1\}^{H \times W}$ определяет пиксели, подлежащие восстановлению: $M_{ij} = 1$ соответствует повреждённому пикселю, $M_{ij} = 0$ — сохранным областям. Повреждённое изображение формируется как $I_{in} = I_{gt} \odot (1 - M)$, где $\odot$ — поэлементное умножение.

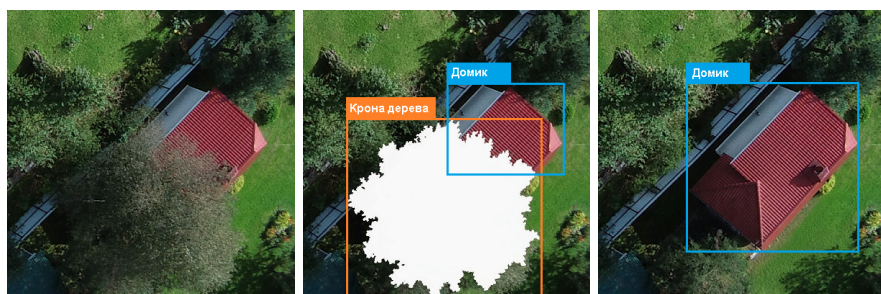
Целью является настройка параметров θ параметрической функции $G_\theta(\cdot)$ такой, что для пары (I_{in}, M) она восстанавливает полное изображение $I_{rec} = G_\theta(I_{in}, M)$, минимизируя функцию потерь генератора $\mathcal{L}_G$. Ключевой особенностью задачи является необходимость генерации нового содержания при строгом сохранении геометрической целостности архитектурных элементов крыши (скаты, коньки, ендовы) и семантической согласованности текстур кровельных материалов (черепица, металлочерепица, шифер) с окружающим контекстом, что отличает её от общих задач восстановления изображений (*inpainting*) и требует специализированных гибридных подходов (например, GAN с диффузией, учитывающей геометрические и семантические признаки).

На рисунке 1 представлена последовательность из трёх изображений, иллюстрирующая процесс восстановления скрытой области крыши – исходное повреждённое изображение, маска повреждения (крона дерева) и результат восстановления.

Необходимость решения данной задачи продиктована практическими потребностями ППК «Роскадастр». При массовой обработке аэрофотоснимков для ведения кадастрового учёта и обновления картографических материалов ключевым этапом является выделение контуров зданий и формирование цифровых карт. Однако наличие на снимках повреждённых или скрытых областей (например, закрытых кронами деревьев, тенями от облаков или временными объектами) приводит к тому, что стандартные алгоритмы сегментации и построения контуров не могут корректно определить границы крыши. Невосстановленное, зашумлённое изображение не позволяет получить достоверную бинарную маску объекта, что непосредственно сказывается на:

- точности контурных карт – на месте закрытого фрагмента крыши контур либо разрывается, либо ошибочно проводится по границе перекрывающего объекта (например, кроны дерева);
- формировании цифровых моделей рельефа и местности – искажается высотная привязка углов здания;
- кадастровом учёте – неправильно вычисляется площадь объекта, что ведёт к юридическим и экономическим последствиям.

Ручная обработка тысяч снимков с такими дефектами крайне трудоёмка и занимает значительное время. Разработка эффективной модели



(а) повреждённое изображение I_{in}

(б) маска повреждения (I_{in}, M)

(в) восстановленное изображение I_{rec}

Рисунок 1. Пример восстановления повреждённой области крыши

восстановления изображений позволяет автоматически заполнять скрытые области семантически согласованным содержанием с сохранением геометрии крыши. В результате алгоритмы картографирования получают на вход целостное изображение, что обеспечивает корректное выделение контуров, вычисление площади и формирование достоверных картографических продуктов. При этом время обработки сокращается с нескольких минут на одно изображение до долей секунды, что критически важно при работе с региональными и федеральными архивами аэрофотосъёмки.

2. Архитектура Roof-DeGAN

Предлагаемая модель представляет собой гибридную генеративно-состязательную сеть, в которой генератор восстанавливает скрытые области изображения, а дискриминатор оценивает правдоподобность полученных фрагментов. Для повышения качества восстановления применяются современные элементы: трансформерные блоки для учёта глобального контекста, плотные связи для лучшего распространения признаков и компонент диффузионного усиления для повышения стабильности и естественности текстур.

Генератор построен на основе архитектуры, сочетающей трансформерные блоки и плотные свёрточные соединения, специально адаптированной для задач восстановления спутниковых и аэрофотоснимков. Он состоит из трёх основных частей: сжимающей ветви (энкодер), промежуточного слоя (бутылочное горлышко, *bottleneck*) и расширяющей ветви (декодер). Между соответствующими уровнями сжимающей и расширяющей ветвей установлены *skip*-коннекторы, передающие низкоуровневые признаки.

Сжимающая ветвь состоит из последовательности блоков, каждый из которых включает механизм самосогласованного внимания, полносвязный слой для учёта дальних зависимостей в изображении и плотный свёрточный блок, выполняющий преобразование:

$$y = \text{LeakyReLU}(\text{BN}(\text{Conv}_{3 \times 3}(x))) + x,$$

где

$\text{Conv}_{3 \times 3}(\cdot)$ – двумерная свёртка с ядром размером 3×3 ,

BN – пакетная нормализация,

$\text{LeakyReLU}(\cdot)$ – функция активации с отрицательным наклоном.

Количество каналов признаков увеличивается от 64 до 1024 (в *bottleneck*), а пространственное разрешение уменьшается в два раза на каждом уровне за счёт шага свёртки равного 2.

Расширяющая ветвь симметрична сжимающей и использует транспонированные свёртки для увеличения разрешения. На каждом уровне выполняется следующее преобразование:

$$y = \text{TransposeConv}_{4 \times 4, s=2}(\text{Concat}(x_{dec}, x_{enc})) + \text{ResidualBlock}(x),$$

где

$\text{TransposeConv}_{4 \times 4, s=2}(\cdot)$ – транспонированная свёртка с ядром 4×4 и шагом (stride) 2,

$\text{Concat}(x_{dec}, x_{enc})$ – конкатенация (объединение по каналам) признаков текущего уровня расширяющей ветви и соответствующего уровня сжимающей ветви через skip-коннектор,

$\text{ResidualBlock}(x)$ – дополнительный блок с остаточным соединением для стабилизации обучения.

Skip-коннекторы передают детализированные низкоуровневые признаки (границы, локальные текстуры), что предотвращает размытие восстановленных областей.

Выходной слой генератора состоит из свёртки размером 1×1 с функцией активации $\tanh(\cdot)$, которая приводит значения пикселей к диапазону от -1 до 1 :

$$\hat{I} = \tanh(\text{Conv}_{1 \times 1}(z)),$$

где $\hat{I}$ – восстановленное изображение, z – выход последнего блока расширяющей ветви.

Дискриминатор выполнен в виде многоуровневой сети с механизмом межмасштабного внимания. Он оценивает правдоподобность не всего изображения целиком, а одновременно на нескольких уровнях детализации (от мелких фрагментов до крупных областей), связывая информацию между разными масштабами признаков для более точной оценки согласованности восстановленных фрагментов. На вход дискриминатора подаётся конкатенация повреждённого изображения и маски повреждения. На выходе формируется набор карт вероятности разного разрешения. Каждый уровень включает свёрточную основу с функцией активации $\text{LeakyReLU}(\cdot)$ и нормализацией спектра, а также модуль межмасштабного внимания, который связывает информацию между уровнями детализации. Такой подход обеспечивает точную оценку высокочастотных деталей (текстур кровельных покрытий) и геометрических элементов крыш, а также повышает устойчивость обучения за счёт разнообразия градиентов.

Стандартные трансформерные блоки в генераторе имеют квадратичную вычислительную сложность $O(n^2)$ и не учитывают специфику аэрофотоснимков крыш, которые содержат большие однородные области (плоские скаты) наряду с мелкими деталями (черепица, стыки). Для решения этой проблемы в Roof-DeGAN предложен механизм динамического разрежения внимания (*Dynamic Sparse Attention*, DSA). На каждом уровне энкодера для каждого патча вычисляется локальная мера вариативности:

$$v_i = \text{Var}(p_i) + \text{Var}(\nabla p_i),$$

где p_i – значения пикселей в i -м патче, ∇p_i – градиент. Для патчей с низкой вариативностью attention вычисляется на объединении:

- предсказуемых соседних патчей (например, в окне 3×3);
- 30% случайно выбранных патчей по всему изображению.

Патчи с высокой вариативностью ($v_i \geq \tau$) обрабатываются полностью. Порог τ динамически подстраивается как перцентиль распределения $\{v_i\}$ на текущем уровне (фиксируется $\tau = P_{50}$ (50-й перцентиль, медиана распределения $\{v_i\}$)). DSA снижает вычислительную сложность с $O(n^2)$ до $O(n \cdot k)$, где $k \ll n$ – эффективный размер внимания, и позволяет обрабатывать изображения большего разрешения без потери качества на текстурированных участках.

Организация Roof-DeGAN приведена на рисунке 2.

Общая функция потерь модели $\mathcal{L}_G$ складывается из нескольких компонентов:

$$(1) \quad \mathcal{L}_G = \lambda_{pix} \mathcal{L}_{pix} + \lambda_{adv} \mathcal{L}_{adv} + \lambda_{perc} \mathcal{L}_{perc} + \lambda_{diff} \mathcal{L}_{diff} + \lambda_{color} \mathcal{L}_{color}.$$

Пиксельная компонента ($\mathcal{L}_{pix}$) отвечает за точное совпадение яркости каждого пикселя в сохранных областях и обеспечивает геометрическую целостность изображения. Эта компонента (функция потерь $\mathcal{L}_1$) гарантирует, что модель не будет произвольно изменять яркость и форму объектов в неповреждённых частях изображения [6].

Состязательная компонента ($\mathcal{L}_{adv}$) заставляет генератор создавать настолько реалистичные фрагменты, чтобы дискриминатор не мог отличить их от настоящих. Она отвечает за реалистичность текстур и резкость границ [5].

Перцептивная компонента ($\mathcal{L}_{perc}$) сравнивает изображения не по-пиксельно, а в пространстве высокоуровневых признаков, извлечённых

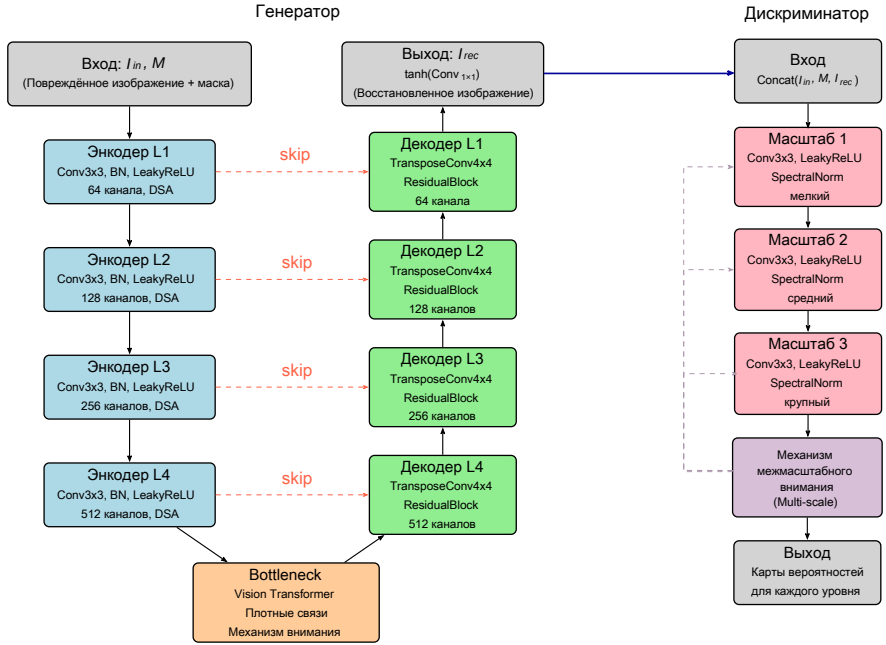


РИСУНОК 2. Основные компоненты модели Roof-DeGAN

предобученной свёрточной сетью, например, VGG-16 [19]. Эта компонента позволяет модели улавливать структуру объекта и текстуру материала (черепица, шифер, металл), игнорируя незначительные сдвиги пикселей, которые не влияют на визуальное восприятие человеком.

Диффузионная компонента ($\mathcal{L}_{diff}$) использует предобученную модель шумоподавления DDPM [8] в качестве эксперта по реалистичности. Она заставляет восстановленное изображение вести себя так же, как эталон, в процессе диффузионного зашумления и восстановления. Подход основан на идеях, изложенных в работах [9, 12], и позволяет эффективно подавлять характерные для GAN высокочастотные артефакты, а также стабилизировать обучение.

Компонента цветовой согласованности ($\mathcal{L}_{color}$) штрафует модель за появление неестественных оттенков, сравнивая распределения цветов в восстановленном и эталонном изображениях. Реализация основана на дифференцируемой гистограммной потере [20]. Эта компонента сохраняет общую цветовую гамму, свойственную конкретному типу кровельного

покрытия (например, терракотовый цвет черепицы или серый оттенок шифера), предотвращая появление «кислотных» или выцветших тонов [12].

Лучшие значения коэффициентов λ были подобраны экспериментально в ходе исследования функции потерь (см. раздел 6).

3. Детали обучения и оценка точности

3.1. Предобучение на ZRG (этап сегментации)

Для формирования начальных представлений о геометрии и текстурах кровельных покрытий использовался приватный датасет Zeitview Rooftop Geometry (ZRG) [21].¹ Датасет содержит 22 334 аннотированных RGB-изображения жилых домов, полученных с беспилотных летательных аппаратов на высоте 3–5 метров над крышами. Разрешение снимков составляет <1 см/пиксель, что позволяет различать мелкие элементы кровли (отдельные черепицы, стыки, вентиляционные выходы). Каждое изображение сопровождается:

- сегментационной маской крыши (бинарная маска «крыша/фон»);
- 3D-каркасом (полигональная модель граней крыши);
- информацией о наличии нависающей растительности и теней (естественные окклюзии).

Выбор ZRG обусловлен следующими преимуществами по сравнению с классическими датасетами аэрофотоснимков (Inria [22], ISPRS Potsdam [23]). В отличие от Inria и ISPRS Potsdam, которые содержат только бинарную разметку «здание/не здание», ZRG предоставляет детальную сегментационную маску крыши и 3D-каркас. Разрешение ZRG (<1 см/пиксель) значительно превосходит разрешение Inria (30 см/пиксель) и ISPRS Potsdam (5 см/пиксель), что позволяет различать мелкие элементы кровли. Кроме того, ZRG включает естественные окклюзии (тени, нависающие деревья), отсутствующие в классических датасетах, ориентирован на жилую и пригородную застройку (в отличие от городской в Inria) и насчитывает более 22 000 аннотированных объектов против 38 патчей в ISPRS Potsdam.

¹ППК «Роскадастр» использует датасет на законных основаниях

Генератор Roof-DeGAN предварительно обучался на датасете ZRG в режиме бинарной сегментации крыш. Архитектура на этом этапе включала энкодер-декодер с трансформерными блоками и плотными skip-соединениями (как в целевой модели), но без состязательной компоненты. Функция потерь на этапе предобучения:

$$L_{pretrain} = \lambda_{pix} L_{pix} + \lambda_{perc} L_{perc},$$

где

L_{pix} — $\mathcal{L}_1$ -потеря (попиксельное сравнение с маской крыши),
 L_{perc} — перцептивная потеря на признаках VGG-16 [19].

Коэффициенты: $\lambda_{pix} = 1,0$, $\lambda_{perc} = 0,05$.

Обучение продолжалось в течение 65 эпох с использованием оптимизатора Adam ($lr = 10^{-4}$, $\beta_1 = 0,5$, $\beta_2 = 0,999$), размер мини-батча — 16 изображений. Веса дискриминатора на этом этапе не инициализировались.

3.2. Перенос весов и дообучение (этап восстановления)

После завершения предобучения веса энкодера и декодера копировались в генератор целевой модели Roof-DeGAN. Выходной слой генератора заменялся: вместо одного канала (бинарная маска) устанавливалось три канала (RGB), его веса инициализировались случайно. Дискриминатор инициализировался случайным образом.

Дообучение проводилось на целевом датасете (1600 изображений с синтетическими масками повреждений) с использованием полной функции потерь (1). Оптимальные коэффициенты: $\lambda_{pix} = 1,0$, $\lambda_{adv} = 0,1$, $\lambda_{perc} = 0,05$, $\lambda_{diff} = 0,01$, $\lambda_{color} = 0,05$. Обучение продолжалось в течение 30 эпох с ранней остановкой (обучение прекращалось, если функция потерь на валидационной выборке (200 изображений) не уменьшалась в течение 5 последовательных эпох). Согласно динамике обучения (рисунок 4 и 6), наилучшие метрики достигнуты на 25-й эпохе, после чего на 26-й эпохе зафиксировано резкое ухудшение (коллапс состязательного обучения). Поэтому в качестве итоговой модели сохранены веса, соответствующие 25-й эпохе.

Размер мини-батча составил 32 изображения. Благодаря использованию смешанной точности (FP16) и механизма Flash Attention, обучение стабильно выполнялось на GPU NVIDIA A100 80 ГБ. Выбор батча 32 обеспечивает баланс между вычислительной эффективностью и гладкостью градиентов, необходимой для состязательного обучения.

3.3. Целевой датасет и генерация синтетических масок повреждений

Целевой датасет сформирован на основе архивов аэрофотосъёмки ППК «Роскадастр»² и включает 2000 RGB-изображений крыш зданий (1600 – для обучения, 200 – для валидации, 200 – для тестирования). Все изображения получены с беспилотных летательных аппаратов (БПЛА) на высоте 30–50 м над уровнем крыш, что для типовых садово-дачных построек ($\approx 8 \times 8$ м) обеспечивает пространственное разрешение ≈ 3 см/пиксель. Размер каждого изображения после предобработки составляет 256×256 пикселей, что соответствует одному зданию целиком с небольшим окружающим фоном. Датасет охватывает следующие типы кровельных материалов: металлочерепица (около 35%), керамическая черепица (25%), шифер (20%), битумная черепица (10%) и прочие покрытия (рулонные покрытия, плоские крыши, медь) – 10%. По типу застройки снимки распределены следующим образом: городская плотная (10%), пригородная (15%) и садово-дачная (75%). Все изображения проходят ручную верификацию на отсутствие глобальных дефектов (разрывов более 50% площади, сильных атмосферных искажений, например дымки или бликов), что обеспечивает чистоту эталонных данных.

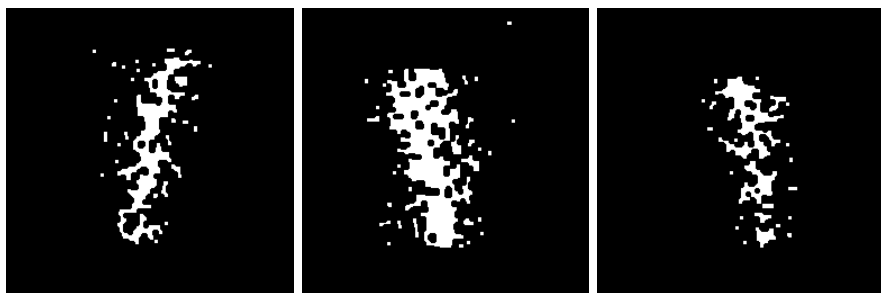
Поскольку реальные аэрофотоснимки с точной маской «скрытая область» (например, под кроной дерева или строительной техникой) получать трудоёмко и дорого, в работе используется синтетическая генерация масок повреждений. В задачах восстановления изображений такой подход является общепринятым и позволяет контролировать форму, размер и положение дефектов. Маска повреждения $M \in \{0, 1\}^{H \times W}$ формируется для каждого обучающего примера по следующему алгоритму. Сначала с заданными вероятностями (40%, 30% и 30% соответственно) выбирается один из трёх типов повреждения: прямоугольное/многоугольное (имитация закрытия строительной техникой или временными сооружениями), реалистичная крона дерева или нерегулярное составное (моделирование теней облаков или спутанных объектов).

Особое внимание уделено генерации масок крон деревьев, которые в отличие от примитивных геометрических фигур (эллипсов или прямоугольников) воспроизводят характерные особенности реальной древесной растительности. В разработанном методе предусмотрены следующие типы крон (примеры приведены на рисунке 3):

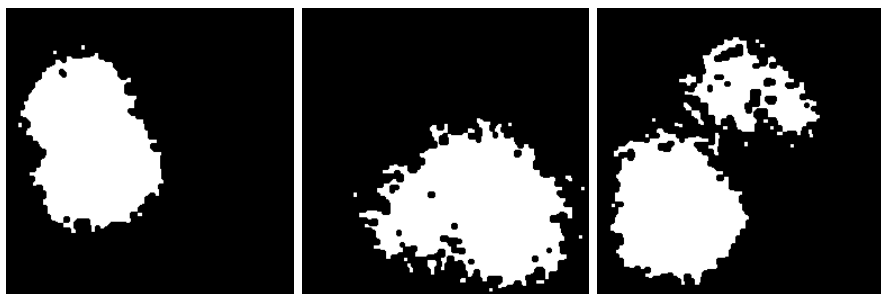
²Собственность ППК «Роскадастр»



(а) Ветвистая крона



(б) Пирамидальная крона



(в) Раскидистая крона

Рисунок 3. Примеры синтетических масок крон деревьев

- Ветвистые (старый дуб, вяз) – кроны с выраженными крупными ветвями, выступающими за пределы основной массы;
- Пирамидальные (ель) – классические конусовидные кроны с характерной формой и «пушистой» текстурой;
- Раскидистые (сосна) – неправильные, асимметричные кроны с неравномерным ростом ветвей.

Для каждого типа кроны задаются следующие параметры: площадь A_m выбирается из равномерного распределения в одном из трёх диапазонов – 10–15%, 20–25% или 30–35% от общей площади изображения, что позволяет оценить устойчивость модели к различной степени перекрытия (см. раздел 7). Для ветвистых форм дополнительно регулируется количество и длина ветвей; для пирамидальных – угол наклона ствола и степень «пушистости»; для раскидистых – степень асимметрии и количество центров роста кроны. Местоположение маски с равной вероятностью выбирается в центре, на краю или в углу крыши, чтобы модель обучалась восстанавливать как изолированные дефекты, так и краевые разрывы.

После генерации формы маска подвергается постобработке: сначала применяется морфологическое открытие с ядром 3×3 для удаления изолированных ложных пикселей на границах маски, затем – гауссово размытие с параметром $\sigma = 1$ и последующая бинаризация по порогу 0,5. Это создаёт плавные, но чёткие границы перехода между повреждённой и сохранной областями, предотвращая появление резких ступенек яркости, которые модель могла бы использовать как «простой ключ» для обхода обучения. Дополнительно для масок крон деревьев с вероятностью 50% генерируются внутренние просветы (имитация промежутков между ветвями), а с вероятностью 40% – торчащие ветки, что повышает реалистичность и усложняет задачу восстановления.

Повреждённое изображение формируется как $I_{in} = I_{gt} \odot (1 - M)$, то есть пиксели, соответствующие маске, обнуляются, а сохранные области остаются неизменными. Генерация масок выполняется на лету (*on-the-fly*) в процессе загрузки данных (размер батча составляет 32 изображения). Для валидационной и тестовой выборок маски и повреждённые изображения фиксируются и сохраняются один раз, что обеспечивает воспроизводимость сравнения между разными моделями и конфигурациями.

Предложенный подход к генерации масок отражает наиболее частые на практике случаи перекрытия крыш: древесная растительность (реалистичные кроны деревьев различных типов площадью 15–35%), техника и временные объекты (прямоугольные маски 10–20%), облака и тени (крупные нерегулярные маски 20–35%). К ограничениям метода следует отнести отсутствие имитации полутонových и градиентных перекрытий (например, ажурная тень от листвы), а также отсутствие масок с разрывом более 50% площади (такие случаи требуют уже не восстановления, а генерации «с нуля» и выходят за рамки данной задачи). Тем не менее, как показано в разделе 7 (рисунок 9), уже при 35% перекрытия качество восстановления может становиться неприемлемым, что аргументирует выбор указанных диапазонов площадей повреждений.

3.4. Метрики оценки качества

Для количественной оценки качества восстановления использовались следующие метрики: пиковое отношение сигнала к шуму (PSNR) [24], индекс структурного сходства (SSIM) [25] и метрика перцептивного сходства (LPIPS) [26] на основе признаков предобученной сети VGG-16. Все метрики рассчитывались исключительно по пикселям, заданным маской повреждения M , что обеспечивает корректное сравнение методов в задачах восстановления. Выбор LPIPS обусловлен высокой корреляцией глубоких признаков VGG-16 с человеческим восприятием текстур и форм кровельных материалов [27].

Дополнительно для оценки качества генерации текстур применялась метрика дистанции Фреше между распределениями inception-векторов (FID) [28], вычисляемая между распределениями восстановленных и эталонных изображений на тестовой выборке. Низкое значение FID свидетельствует о том, что сгенерированные фрагменты не только похожи на оригиналы попиксельно, но и статистически неотличимы от реальных изображений крыш.

Для оценки точности восстановления геометрических контуров крыш (скаты, коньки, ендовы) используется метрика Boundary F1 [1]. Границы выделяются детектором Канни (пороги 50, 150), после чего вычисляется F1-мера с допуском 3 пикселя. Метрика рассчитывается только в области повреждения M и характеризует пригодность восстановленного изображения для автоматического контурирования в кадастровых системах.

Для оценки надёжности метрик качества использован метод бутстрэпа (1000 итераций ресэмплинга с возвращением на тестовой выборке из 200 изображений). Результаты представлены как среднее $\pm$ стандартное отклонение; для метрики FID дополнительно приведены 95% доверительные интервалы.

Для визуальной валидации результатов использовался также метод экспертных оценок: три специалиста в области кадастрового учёта проводили сравнительный анализ восстановленных изображений по шкале от 1 до 5, оценивая геометрическую точность и реалистичность текстур. Усреднённая экспертная оценка (*Mean Opinion Score*, MOS) позволила подтвердить корреляцию объективных метрик с субъективным восприятием качества. Все метрики вычислялись как на всей тестовой выборке (200 изображений), так и отдельно для трёх групп повреждений (10–15%, 20–25%, 30–35%), что позволило оценить эффективность работы модели для различных степеней перекрытия крыш.

Эксперименты проводились на вычислительном кластере со следующей конфигурацией: GPU NVIDIA Tesla A100 SXM (80 ГБ памяти), CPU Intel Xeon Gold 6248R (24 ядра), оперативная память 128 ГБ DDR4. Используемое программное обеспечение: Python 3.12, PyTorch 2.5 с поддержкой CUDA 12.x. Для воспроизведения базовой функциональности и экспериментов с небольшими датасетами (предобучение на Ingria) также подготовлена упрощённая версия модели в виде интерактивного блокнота Jupyter в облачной среде Google Colab.

4. Динамика обучения и сходимость модели

На рисунке 4 представлены кривые изменения основных компонентов функции потерь на валидационной выборке в процессе обучения модели.

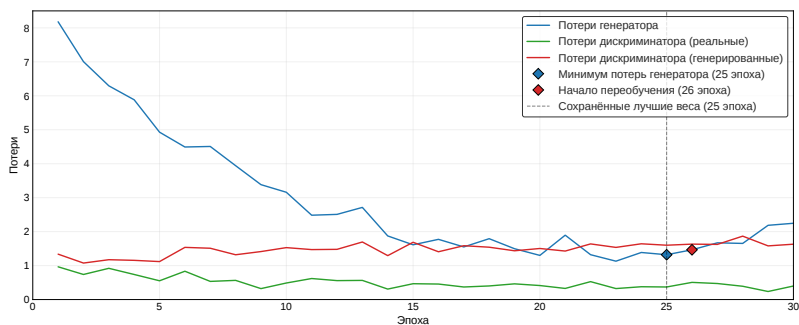


Рисунок 4. Динамика изменения функций потерь генератора и дискриминатора на валидационной выборке при обучении Roof-DeGAN

В первые 10 эпох наблюдается быстрое снижение потерь генератора (с 7,82 до 3,94), что соответствует начальной фазе обучения, когда модель осваивает грубую структуру изображений и общие геометрические особенности крыш. Потери дискриминатора на реальных изображениях монотонно уменьшаются (с 0,89 до 0,58), что указывает на улучшение способности различать эталонные образы. Одновременно потери дискриминатора на сгенерированных изображениях возрастают (с 1,12 до 1,45), отражая типичное для состязательного обучения усиление градиентов против генератора.

В интервале 10–20 эпох темп снижения потерь генератора замедляется, модель начинает уточнять текстуры кровельных материалов и границы архитектурных элементов. Потери дискриминатора на реальных изображениях продолжают плавно снижаться (до 0,41 к 20-й эпохе), а на

сгенерированных – растут (до 1,62), что свидетельствует о сохранении баланса между генератором и дискриминатором.

В интервале 20–25 эпох наблюдается дальнейшее улучшение качества: потери генератора снижаются до минимального значения 1,32 к 25-й эпохе, потери дискриминатора на реальных изображениях стабилизируются на уровне 0,37, на сгенерированных – 1,60. Метрики качества достигают пиковых значений: PSNR – 34,5 дБ, SSIM – 0,972, LPIPS – 0,052, FID – 17,6.

На 26-й эпохе происходит резкое ухудшение: потери генератора возрастают до 2,38, а метрики качества демонстрируют спад (PSNR снижается до 33,42 дБ, LPIPS возрастает до 0,058, FID – до 18,7). Дальнейшее обучение до 30 эпох не приводит к восстановлению качества: потери генератора колеблются в диапазоне 2,35–2,41, метрики остаются на худших значениях. Это послужило основанием для применения ранней остановки (*early stopping*) на 26-й эпохе с сохранением лучшей модели, полученной на 25-й эпохе. В работе не применялись распространённые техники стабилизации GAN (R1-регуляризация и спектральная нормализация). Стабильность достигалась преимущественно архитектурными решениями (DSA, межмасштабный дискриминатор) и композитной функцией потерь. Применение указанных методов – перспективное направление дальнейшей работы.

На рисунке 6 показана динамика изменения метрик качества PSNR, SSIM, LPIPS и FID на валидационной выборке в процессе обучения модели. Анализ этого рисунка подтверждает выводы, полученные на основе кривых потерь. Метрики PSNR и SSIM активно растут до 20-й эпохи, после чего темп роста замедляется. Максимальные значения достигаются на 25-й эпохе: PSNR = 34,5 дБ, SSIM = 0,972. Метрика LPIPS снижается до 0,052, а FID – до 17,6, что указывает на значительное улучшение перцептивного качества и естественности текстур кровельных покрытий. На 26-й эпохе фиксируется ухудшение всех метрик: PSNR снижается до 33,42 дБ, SSIM – до 0,970, LPIPS возрастает до 0,058, FID – до 18,7.

В интервале 27–30 эпох наблюдается плато на ухудшенных значениях, что свидетельствует о начале переобучения – модель начинает чрезмерно подстраиваться под текстуры обучающей выборки в ущерб обобщающей способности на новых изображениях.

Полученные результаты подтверждают эффективность выбранных техник стабилизации (ранняя остановка, сглаживание меток) и демонстрируют, что предложенный гибридный подход Roof-DeGAN достигает баланса между скоростью сходимости, перцептивным качеством и устойчивостью к ограниченному объёму обучающих данных.

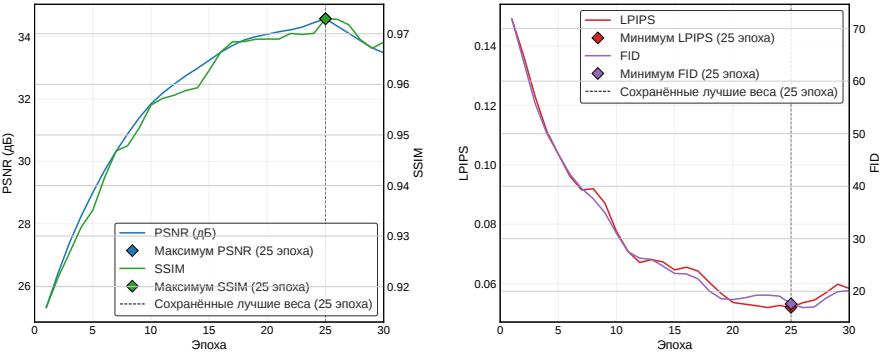


Рисунок 5. Динамика изменения геометрических (PSNR, SSIM) и перцептивных (LPIPS, FID) метрик на валидационной выборке в процессе обучения Roof-DeGAN

5. Количественное сравнение с современными методами

Для объективной оценки эффективности предложенного подхода проведено сравнение с рядом современных методов восстановления изображений, включая традиционные алгоритмы, чистые диффузионные модели и гибридные GAN-архитектуры. Тестирование выполнялось на идентичной тестовой выборке из 200 изображений с одинаковыми синтетическими и естественными масками повреждений площадью 15–25%. Все методы исследовались в стандартных конфигурациях на одном оборудовании (GPU NVIDIA A100 SXM и CPU Intel Xeon Gold 6248R). Результаты представлены в таблице 1. Символ ↑ в заголовке таблицы обозначает, что большее значение лучше, символ ↓ – меньшее значение лучше. Время инференса указано для изображений размером 256 × 256 пикселей.

Анализ этой таблицы позволяет сделать следующие выводы.

Превосходство над традиционными методами. Roof-DeGAN значительно опережает классические алгоритмы Navier-Stokes и PatchMatch по всем метрикам: выигрыш в PSNR составляет 8,0–9,7 дБ, а FID снижается в 8–10 раз. Это визуально соответствует переходу от артефактных и размытых реконструкций к детализированным изображениям с сохранением геометрии и текстур кровельных покрытий.

Таблица 1. Сравнение Roof-DeGAN с современными методами на тестовой выборке ($N = 200$)

Методы	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$	Π^1	t^2
<i>Традиционные</i>						
Navier–Stokes [3]	24,18	0,742	0,312	187,3	–	0,8
PatchMatch [4]	25,92	0,801	0,245	142,8	–	2,3
<i>Диффузионные</i>						
DDPM [8]	28,45	0,873	0,118	78,6	112	28,5
SatDiff [10]	30,45	0,941	0,065	38,2	189	41,3
КАО [11]	31,12	0,952	0,058	32,7	205	38,9
<i>GAN-методы</i>						
Pix2Pix [6]	28,92	0,885	0,142	92,5	54	0,12
ESRGAN [7]	29,34	0,902	0,098	67,3	67	0,18
DeGAN [13]	31,85	0,958	0,082	29,4	62	0,14
<i>Новый метод</i>						
Roof-DeGAN	33,7 $\pm 1,2$	0,971 $\pm 0,008$	0,048 $\pm 0,011$	17,8 $\pm 2,4$	48	0,15

¹ Π – количество параметров модели в миллионах;

² t – время инференса на одно изображение в секундах.

Сравнение с диффузионными моделями. Чистые диффузионные модели (DDPM, SatDiff, KAO) демонстрируют высокое качество, однако требуют десятков секунд на инференс и значительного объёма обучающих данных. Предложенный гибридный метод превосходит их по PSNR на 2,8–5,5 дБ и достигает существенно лучшего FID (17,8 против 32,7–78,6) при времени инференса в 250–270 раз меньшем (0,15 с. против 38–41 с.). Это делает подход практически применимым для обработки больших массивов аэрофотоснимков в реальном времени.

Сравнение с GAN-архитектурами. По сравнению с базовыми GAN-моделями (Pix2Pix, ESRGAN) и другими гибридными подходами, такими как DeGAN, предложенный метод показывает прирост PSNR до 5,0 дБ и значительное улучшение FID (в 1,7–5,2 раза). Значение LPIPS = 0,048 подтверждает высокое перцептивное качество восстановленных текстур, превосходящее показатели сравниваемых GAN-архитектур (в нашем случае это Pix2Pix = 0,142; ESRGAN = 0,098; DeGAN = 0,082) – улучшение на 41–66% относительно ближайшего конкурента. При этом модель имеет меньшее количество параметров (48 М против 54–205 М), что обеспечивает экономию вычислительных ресурсов. Превосходство Roof-DeGAN над ближайшими конкурентами является статистически значимым: 95% доверительные интервалы метрик не перекрываются. Например, разница в PSNR между Roof-DeGAN ($33,7 \pm 1,2$ дБ) и KAO ($31,1 \pm 1,8$ дБ) составляет 2,6 дБ при непересекающихся интервалах, что подтверждает устойчивость преимущества предложенного метода.

Полученные результаты демонстрируют, что интеграция трансформерных блоков, многоуровневого внимания и диффузионного усиления позволяет предложенному методу достигать наилучшего баланса между качеством восстановления и вычислительной эффективностью среди рассмотренных подходов.

6. Исследование влияния компонент функции потерь

Для количественной оценки вклада каждой компоненты композитной функции потерь проведён сравнительный анализ пяти конфигураций модели, различающихся набором оптимизируемых критериев. Пять конфигураций модели обучались с различными комбинациями компонентов до достижения оптимальной эпохи по критерию ранней остановки. Результаты представлены в таблице 2, лучшие показатели выделены красным.

ТАБЛИЦА 2. Абляционное исследование основных компонентов функции потерь на валидационной выборке

Конфигурация (компоненты из (1))	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$	Визуальные особенности
Без предобучения (все компоненты из (1))	28,34	0,912	0,112	54,2	Заметные артефакты, текстуры сглажены
B ($\lambda_{pix}\mathcal{L}_{pix}$)	29,84	0,931	0,098	87,4	Размытые текстуры, сглаженные границы
A ($\lambda_{pix}\mathcal{L}_{pix} + \lambda_{adv}\mathcal{L}_{adv}$)	31,12	0,952	0,082	48,2	Чёткие границы, локальные текстурные артефакты
P ($\lambda_{pix}\mathcal{L}_{pix} + \lambda_{perc}\mathcal{L}_{perc}$)	31,45	0,958	0,068	39,7	Естественные текстуры, избыточное сглаживание
D ($\lambda_{pix}\mathcal{L}_{pix} + \lambda_{adv}\mathcal{L}_{adv} +$ $+ \lambda_{perc}\mathcal{L}_{perc} + \lambda_{diff}\mathcal{L}_{diff}$)	32,18	0,965	0,058	28,9	Хорошая геометрия и текстура, небольшие цветовые сдвиги
C ($\lambda_{pix}\mathcal{L}_{pix} + \lambda_{adv}\mathcal{L}_{adv} +$ $+ \lambda_{perc}\mathcal{L}_{perc} + \lambda_{diff}\mathcal{L}_{diff} +$ $+ \lambda_{color}\mathcal{L}_{color}$)	34,5	0,972	0,052	17,6	Наиболее реалистичные текстуры, геометрия и цветовая согласованность

Анализ таблицы 2 показывает следующее. Базовая попиксельная конфигурация В (Baseline) обеспечивает геометрическую целостность (положение скатов, коньков, ендов), но приводит к характерной размытости текстур кровельных покрытий ($PSNR = 29,84$ дБ, $LPIPS = 0,098$).

Добавление состязательной компоненты А (Adversarial) существенно повышает чёткость границ и добавляет высокочастотные детали, однако без перцептивной и диффузионной составляющих возникают локальные артефакты ($LPIPS = 0,082$, FID остаётся высоким).

Включение перцептивной компоненты Р (Perceptual) без состязательной части позволяет достичь значения $LPIPS = 0,068$, однако визуально изображения выглядят излишне сглаженными из-за отсутствия высокочастотных деталей, которые вносит состязательная компонента.

Наилучшее значение $LPIPS$ (0,058) среди неполных конфигураций достигается при совместном использовании состязательной, перцептивной и диффузионной компонент D (Diffusion). Добавление диффузионного усиления значительно улучшает структурную согласованность и снижает FID (до 28,9), однако наблюдаются небольшие цветовые сдвиги.

Полная конфигурация С (Color), включающая все пять компонентов, объединяет их преимущества и достигает на валидационной выборке лучших значений геометрических и перцептивных метрик точности $PSNR = 34,5$ дБ, $SSIM = 0,972$, $LPIPS = 0,052$, $FID = 17,6$.

Сравнение конфигурации «Без предобучения» и полной конфигурации С показывает, что предложенный двухэтапный подход (предобучение на ZRG с последующим дообучением) обеспечивает прирост $PSNR$ на 6,16 дБ, снижение $LPIPS$ на 0,060 и снижение FID на 36,6 по сравнению с обучением с нуля. Это подтверждает эффективность переноса весов из задачи сегментации крыш в задачу восстановления изображений.

Абляционное исследование (таблица 2) подтвердило необходимость использования всех пяти компонент функции потерь. Для достижения наилучшего баланса между геометрической точностью и перцептивным качеством была проведена настройка весовых коэффициентов λ . Оптимальные значения, полученные в ходе гиперпараметрической оптимизации на валидационной выборке, составили:

- $\lambda_{pix} = 1,0$ (базовое пиксельное соответствие),
- $\lambda_{adv} = 0,1$ (реалистичность),
- $\lambda_{perc} = 0,05$ (коррекция текстур),
- $\lambda_{diff} = 0,01$ (сохранение геометрии),

- $\lambda_{color} = 0,05$ (цветовой баланс).

На рисунке 6 представлена динамика изменения метрик точности на валидационной выборке.

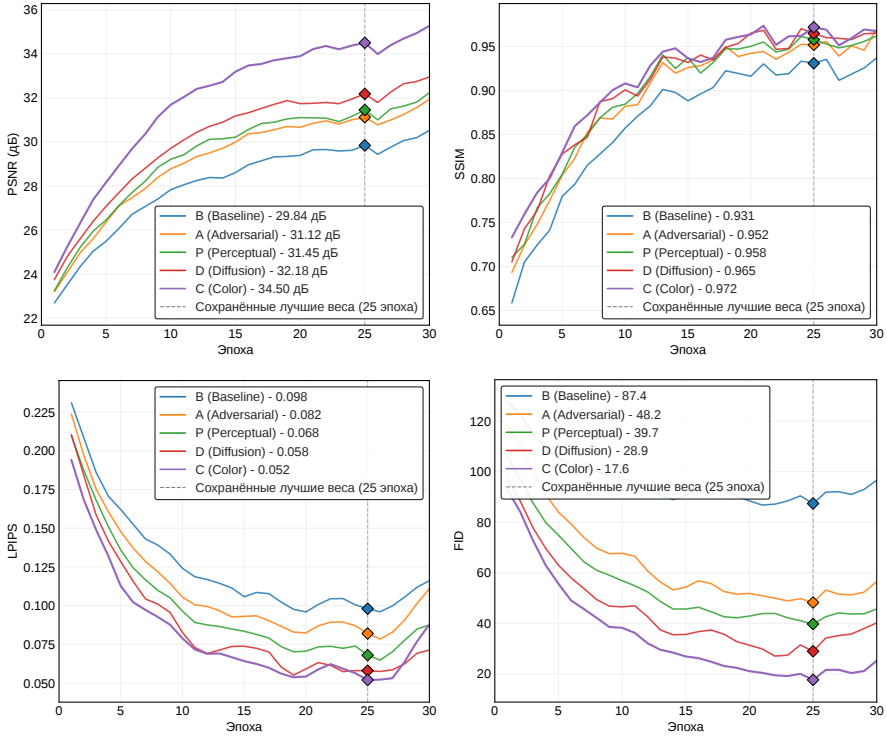


Рисунок 6. Динамика изменения геометрических (PSNR, SSIM) и перцептивных (LPIPS, FID) метрик на валидационной выборке

Пиксельная компонента ($\lambda_{pix} = 1,0$). Данная компонента выполняет роль базового регуляризатора, обеспечивающего сохранение неизменных областей изображения и грубое соответствие восстановленных фрагментов эталону по яркости. Значение $\lambda_{pix} = 1,0$ выбрано в качестве референсного, поскольку $\mathcal{L}_{pix}$ ($\mathcal{L}_1$ -потеря) имеет естественный масштаб, сопоставимый с суммарным вкладом остальных компонент. Уменьшение λ_{pix} до 0,5 приводит к заметному размытию границ архитектурных элементов (PSNR снижается на 1,2 дБ), а увеличение до 2,0 подавляет состязательную и

перцептивную компоненты, делая текстуры излишне сглаженными.

Состязательная компонента ($\lambda_{adv} = 0,1$). GAN-обучение склонно к нестабильности при высоких значениях состязательной компоненты. Экспериментально установлено, что при $\lambda_{adv} > 0,2$ дискриминатор слишком быстро сходится, порождая ослабевающие градиенты для генератора и приводя к коллапсу мод уже после 10–12 эпох. При $\lambda_{adv} < 0,05$ влияние состязательной составляющей становится незначительным: модель генерирует геометрически корректные, но излишне гладкие текстуры (LPIPS $> 0,07$). Значение $\lambda_{adv} = 0,1$ обеспечивает устойчивое состязательное равновесие, при котором дискриминатор остаётся достаточно «сильным», чтобы предоставлять информативный градиент, но не подавляет генератор. Следует отметить, что увеличение λ_{adv} с 0,05 до 0,1 повышает Boundary F1 с 0,89 до 0,93, подтверждая важность состязательной компоненты для чёткости геометрических границ.

Перцептивная компонента ($\lambda_{perc} = 0,05$). Перцептивная потеря, вычисляемая на признаках VGG-16, имеет существенно больший масштаб по сравнению с $\mathcal{L}_1$ -потерей. Прямое использование $\lambda_{perc} = 1,0$ приводит к доминированию этой компоненты и появлению характерных артефактов – чрезмерной текстурированности и «галлюцинаций» мелких деталей, отсутствующих в эталоне. Понижение коэффициента до 0,05 позволяет сохранить положительный эффект перцептивного обучения (естественность текстур черепицы и шифера) без перекоса общей функции потерь. Абляционное исследование (таблица 2) подтверждает, что исключение $\mathcal{L}_{perc}$ увеличивает LPIPS на 0,030, а удвоение λ_{perc} до 0,1 не даёт значимого улучшения, но замедляет сходимость.

Диффузионная компонента ($\lambda_{diff} = 0,01$). Диффузионное усиление основано на предобученной модели DDPM, выступающей в роли «эксперта по реалистичности». Масштаб диффузионной потери существенно варьируется в зависимости от уровня шума и текущего состояния генератора. Значение $\lambda_{diff} = 0,01$ подобрано таким образом, чтобы компонента оказывала стабилизирующее влияние (подавление высокочастотных GAN-артефактов, снижение FID) без доминирования над пиксельной и состязательной составляющими. При $\lambda_{diff} > 0,05$ модель начинает копировать текстурные особенности предобученной диффузионной модели, что приводит к избыточному сглаживанию мелких деталей кровли. При $\lambda_{diff} < 0,005$ положительный эффект диффузионного усиления становится статистически незначимым.

Компонента цветовой согласованности ($\lambda_{color} = 0,05$). Цветовая потеря, основанная на дифференцируемой гистограмме, имеет масштаб, чувствительный к размеру изображения и количеству бинов гистограммы. Для изображений 256×256 пикселей коэффициент 0,05 является оптимальным: он эффективно подавляет неестественные цветовые сдвиги (например, появление «кислотных» оттенков при восстановлении терракотовой черепицы), но не приводит к усреднению цветовых кластеров. При $\lambda_{color} = 0,1$ наблюдается лёгкое «выцветание» насыщенных цветов кровельных материалов (снижение цветового разнообразия по метрике Colorfulness Index [29] на 12%).

Итоговый баланс. Таким образом, выбранные коэффициенты обеспечивают сбалансированный вклад каждой компоненты в общую функцию потерь, что подтверждается результатами абляционного исследования: полная конфигурация достигает на валидационной выборке PSNR = 34,5 дБ, SSIM = 0,972, LPIPS = 0,052 и FID = 17,6, превосходя все неполные конфигурации (таблица 2) по совокупности геометрических и перцептивных метрик.

7. Анализ качества восстановления в зависимости от площади повреждения

Для оценки устойчивости модели к различным масштабам повреждений тестовая выборка из 200 изображений была разделена на три группы: малые повреждения (10–15% площади), средние (20–25%) и обширные (30–35%). Малые повреждения составляют порядка 85% от общего количества изображений в тестовой выборке, средние и обширные – 10% и 5% соответственно.

Результаты сравнения с современными методами приведены в таблице 3. Лучшие показатели выделены красным. Как видно из таблицы, Roof-DeGAN стабильно превосходит все сравниваемые подходы во всех диапазонах повреждений. При малых повреждениях (10–15%) выигрыш в PSNR составляет 1,5–9,7 дБ по сравнению с традиционными методами и 1,5–2,1 дБ относительно современных диффузионных моделей (SatDiff, KAO). При этом SSIM улучшается до 0,986 против 0,742–0,958 у конкурентов, а FID снижается до 18,2 против 32,7–187,3.

При средних повреждениях (20–25%) преимущество усиливается: прирост PSNR достигает 2,0–8,3 дБ, SSIM растёт до 0,971 против 0,718–0,952, LPIPS снижается до 0,059 против 0,058–0,328 у конкурентов, а FID уменьшается до 22,5 против 35,8–192,6.

ТАБЛИЦА 3. Качество восстановления в зависимости от площади повреждения

Методы	10–15%				20–25%				30–35%			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
<i>Традиционные методы</i>												
Navier-Stokes [3]	24,18	0,742	0,312	187,3	23,45	0,718	0,328	192,6	22,10	0,682	0,356	205,4
PatchMatch [4]	25,92	0,801	0,245	142,8	25,12	0,778	0,262	148,5	23,85	0,741	0,289	162,1
<i>Диффузионные модели</i>												
DDPM [8]	28,45	0,873	0,118	78,6	27,80	0,852	0,132	82,4	26,15	0,814	0,158	89,7
SatDiff [10]	31,78	0,949	0,068	38,2	30,45	0,941	0,065	41,5	28,92	0,912	0,092	52,3
КАО [11]	32,41	0,958	0,061	32,7	31,12	0,952	0,058	35,8	29,67	0,923	0,085	44,1
<i>Гибридные и GAN-методы</i>												
Pix2Pix [6]	28,92	0,885	0,142	92,5	28,15	0,865	0,156	98,2	26,78	0,828	0,182	110,6
ESRGAN [7]	29,34	0,902	0,098	67,3	28,65	0,882	0,112	72,1	27,40	0,845	0,138	84,9
DeGAN базовый [13]	31,85	0,958	0,082	29,4	31,02	0,948	0,088	33,7	29,78	0,922	0,102	41,2
<i>Новый метод</i>												
Roof-DeGAN	33,87	0,986	0,045	18,2	33,42	0,971	0,059	22,5	31,25	0,948	0,074	29,8

Особенно показательно поведение при обширных повреждениях (30–35%): предложенный метод сохраняет высокое качество (PSNR = 31,25 дБ, SSIM = 0,948, LPIPS = 0,074, FID = 29,8), тогда как традиционные методы падают до 22,10–23,85 дБ (FID до 205,4), а чистые диффузионные модели – до 28,92–29,67 дБ (FID до 44,1–89,7). Выигрыш в PSNR по сравнению с лучшими конкурентами (KAO, DeGAN) составляет 1,6–2,3 дБ, SSIM улучшается на 0,025–0,026, LPIPS сопоставим (0,074 против 0,058–0,085 у конкурентов), а FID уменьшается на 11,3–14,4 единицы. Результаты сравнения геометрических и перцептивных метрик модели Roof-DeGAN с лучшими конкурентами приведены на рисунке 7.

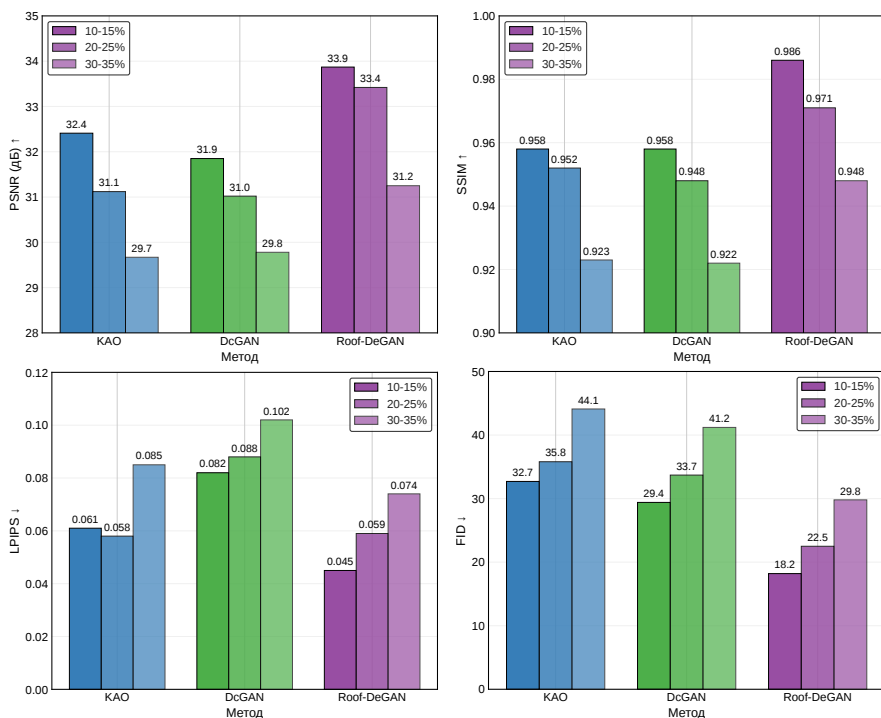


Рисунок 7. Сравнение геометрических (PSNR, SSIM) и перцептивных метрик (LPIPS, FID) Roof-DeGAN с лучшими конкурентами

В таблице 4 приведено сравнение методов по метрике качества границ Boundary F1. Красным выделен лучший результат.

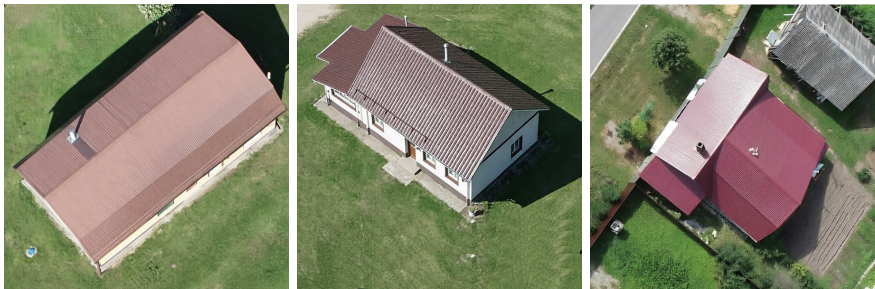
ТАБЛИЦА 4. Сравнение методов по метрике качества границ Boundary F1 (усреднение по всем типам повреждений)

Методы		Boundary F1↑
Традиционные методы	Navier-Stokes [3]	0,38
	PatchMatch [4]	0,46
Диффузионные модели	DDPM [8]	0,62
	SatDiff [10]	0,79
	KAO [11]	0,82
Гибридные и GAN-методы	Pix2Pix [6]	0,54
	ESRGAN [7]	0,58
	DeGAN базовый [13]	0,81
Новый метод	Roof-DeGAN	0,91

Результаты успешного восстановления крыш на тестовом датасете ППК «Роскадастр» показаны на рисунке 8. На нём для каждого из



(а) Исходные изображения

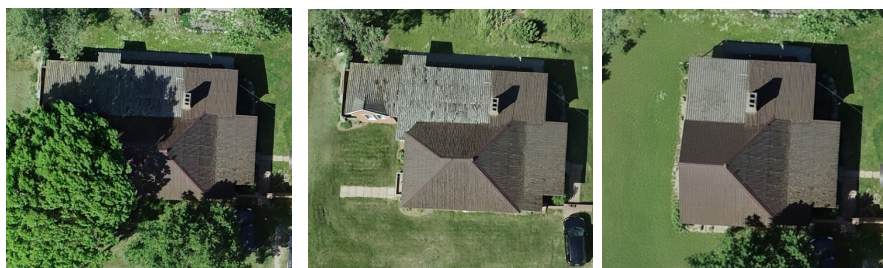


(б) Результаты успешного восстановления

РИСУНОК 8. Крыши дачных домиков до и после их успешного восстановления с использованием предлагаемого метода

трёх примеров приведены исходное изображение (частично закрытое кронами) и результат работы предлагаемой модели. Степень сложности восстановления возрастает слева направо: 10% перекрытия (многоскатная крыша) – минимальное восстановление; 20% перекрытия (двускатная крыша) – восстановление текстуры и геометрии скатов; 35% перекрытия (многоскатная крыша) – значительная реконструкция углов и текстуры.

Для малых и средних повреждений (10–20%) качество восстановления является более чем приемлемым: геометрия скатов сохраняется, текстура кровельного материала воспроизводится достоверно. При обширных повреждениях (35% и более) на восстановленном изображении могут наблюдаться значительные искажения текстур кровельного покрытия и формы крыши. При этом доля полностью неудачных результатов не превышает 3–5% от общего числа обработанных снимков; в этих случаях качество остаётся неприемлемым для последующей автоматизированной кадастровой обработки, рисунок 9.



(a) Исходное изображение

(б) Результаты неуспешного восстановления

Рисунок 9. Пример неуспешного восстановления изображения крыши с использованием предлагаемого метода

8. Ограничения метода и перспективы дальнейших исследований

Несмотря на высокие количественные и качественные результаты, модель Roof-DeGAN обладает рядом ограничений, открывающих направления для дальнейшего развития.

Зависимость от качества маски. Модель требует бинарной маски повреждённой области M . В реальных условиях автоматическая сегментация окклюзий (крон деревьев, теней, техники) редко достигает идеального

качества. Качественный анализ показывает, что при незначительных ошибках сегментации ($\text{IoU} \in [0,85; 0,95]$) модель сохраняет высокую устойчивость — артефакты локализуются преимущественно на границах маски и не затрагивают внутреннюю геометрию скатов благодаря skip-соединениям и многоуровневому дискриминатору. При существенных искажениях маски ($\text{IoU} < 0,8$) наблюдается деградация качества восстановления: появляются эффекты «двойных контуров», локальные искажения текстур и нарушение геометрии коньков, поскольку генератор либо пытается восстановить уже видимые участки, либо оставляет часть окклюзии нетронутой.

Разрешение изображений. Для обработки аэрофотоснимков высокого разрешения, значительно превышающих размер 256×256 пикселей, перспективным направлением является интеграция предложенной архитектуры Roof-DeGAN с фреймворками фрагментарного вывода, такими как SAHI (Slicing Aided Hyper Inference). Данный подход предполагает разбиение исходного изображения и соответствующей маски повреждения M на перекрывающиеся патчи фиксированного разрешения, независимую обработку каждого патча обученной моделью и последующую агрегацию результатов с применением взвешенного усреднения в зонах перекрытия для минимизации граничных артефактов. Такая стратегия позволяет сохранить вычислительную эффективность инференса при работе с изображениями размером 1024×1024 и более, обеспечить непрерывность текстур кровельных материалов и геометрическую согласованность архитектурных элементов крыши. В качестве альтернативного пути масштабирования может быть рассмотрен переход на более эффективные трансформерные блоки с линейной сложностью внимания (Swin Transformer v2 [30], EfficientViT [31] и др.).

Представленность редких классов и сценариев. Датасет охватывает основные типы кровельных покрытий (черепица, металлочерепица, шифер, битумная черепица), однако недостаточно представляет редкие материалы (солома, медь, сланец, зелёная кровля, мембранные покрытия) и сложные погодные условия (снег, дождь, тени от соседних зданий). Расширение датасета за счёт синтетических изображений и применение методов доменной адаптации позволит существенно повысить обобщающую способность модели.

Вычислительная сложность обучения. На этапе инференса модель демонстрирует высокую эффективность (0,15 с на изображение 256×256 на GPU NVIDIA Tesla A100 SXM). Однако обучение требует около 6 часов

машинного времени на том же оборудовании. Значительная ресурсоёмкость обучения ограничивает масштабирование на большие датасеты и изображения высокого разрешения. Перспективными направлениями оптимизации являются дистилляция знаний, квантование весов и замена базовой архитектуры на более лёгкие свёрточные сети [28].

Ограничения диффузионного усиления. Диффузионная компонента повышает стабильность генерации и реалистичность текстур, однако увеличивает вычислительную нагрузку и в отдельных случаях приводит к лёгкому сглаживанию мелких деталей при очень плотных масках. Дальнейшая оптимизация (работа в латентном пространстве или сокращение числа шагов диффузии) позволит устранить данный недостаток.

Чувствительность к локализации повреждений. Хотя модель демонстрирует высокую среднюю точность восстановления границ по метрике Boundary F1 (таблица 4), этот результат достигается в условиях, когда маски повреждений преимущественно расположены в центральных областях скатов (что соответствует 85% тестовой выборки). При обширных повреждениях (30–35%), затрагивающих границы крыш, точность падает до 0,68–0,74 (рисунок 9). Данное ограничение связано с тем, что skip-соединения и межмасштабное внимание не могут передать геометрию, если повреждена вся граница. Перспективным решением является интеграция 3D-каркасов крыш из датасета ZRG в процесс обучения.

Ограниченная применимость. Экспериментальная валидация модели проводилась исключительно на данных ППК «Роскадастр». Полученные результаты могут не обобщаться на аэрофотоснимки, полученные в иных условиях. Для расширения области применения модели требуется дополнительная валидация на других датасетах и, при необходимости, дообучение модели.

В текущей версии исследования целевой датасет ограничен 2000 изображениями. Для дальнейшего повышения обобщающей способности модели планируется расширение выборки до 5000+ изображений с включением редких типов кровельных покрытий и естественных окклюзий (облачность, сезонные изменения растительности). Перспективами развития также являются использование временных рядов аэрофотоснимков (восстановление по нескольким датам) и переход к трёхмерному восстановлению геометрии крыш.

9. Заключение

В ходе выполнения работы разработана гибридная генеративная модель Roof-DeGAN для восстановления скрытых областей крыш зданий на аэрофотоснимках. Предложенная архитектура сочетает трансформерные блоки для учёта глобального контекста, плотные свёрточные связи для улучшения распространения признаков и механизм межмасштабного внимания в многоуровневом дискриминаторе для повышения стабильности обучения.

Основные результаты работы:

- Разработана архитектура генератора типа энкодер–декодер с трансформерными блоками с динамическим разрежением внимания, снижающим сложность с $O(n^2)$ до $O(n \cdot k)$ за счёт адаптивного пропуска однородных областей снимка.
- Создан многоуровневый дискриминатор, оценивающий правдоподобность восстановленных фрагментов на разных масштабах, что повышает устойчивость обучения и качество текстур.
- Предложен и экспериментально обоснован двухэтапный метод обучения: предобучение на датасете ZRG в режиме сегментации крыш с последующим переносом весов в задачу восстановления. Показано, что такой подход обеспечивает прирост PSNR на 6,16 дБ по сравнению с обучением с нуля (таблица 2).
- Экспериментально подтверждена высокая эффективность предложенного подхода: на тестовой выборке достигнуты PSNR = 33,7 дБ, SSIM = 0,971, LPIPS = 0,048 и FID = 17,8, что превосходит современные методы на датасете ППК «Роскадастр» (таблица 1). На тестовой выборке средняя метрика Boundary F1 составила 0,91. При этом для повреждений площадью 10–15% значение достигает 0,96, для 20–25% – 0,88, а для 30–35% снижается до 0,74 (таблица 4). Это подтверждает, что модель уверенно восстанавливает геометрию крыш при умеренных повреждениях, однако при обширных окклюзиях точность границ ожидаемо падает.

Полученные результаты могут быть использованы в системах автоматизированной обработки данных дистанционного зондирования Земли, при обновлении картографических материалов, в задачах мониторинга городской застройки и состояния кровель зданий, а также в смежных областях, требующих восстановления скрытых фрагментов изображений.

Дальнейшие исследования будут направлены на адаптацию разработанной модели для учёта временной динамики изменения растительности, интеграцию с данными других спектральных диапазонов, а также на применение предложенного подхода для решения смежных задач: устранения теней, восстановления повреждённых архивных снимков и улучшения качества изображений, полученных в неблагоприятных метеоусловиях.

Список использованных источников

- [1] S. May, Y. Wang, L. Zhang *Building damage assessment with deep learning* // ISPRS Archives.– 2022.– Vol. **XLIII-B3-2022**.– Pp. 1133–1138.  ↑228, 241
- [2] L. Dong, J. Shan *A comprehensive review of earthquake-induced building damage detection with remote sensing techniques* // ISPRS Journal of Photogrammetry and Remote Sensing.– 2013.– Vol. **84**.– Pp. 85–99.  ↑228
- [3] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester *Image inpainting* // *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000* (New Orleans, LA, USA, July 23–28, 2000).– ACM.– 2000.– ISBN 1-58113-208-5.– Pp. 417–424.   ↑228, 245, 252, 254
- [4] C. Barnes, E. Shechtman, A. Finkelstein, D. B. Goldman *PatchMatch: a randomized correspondence algorithm for structural image editing* // *ACM Transactions on Graphics*.– 2009.– Vol. **28**.– No. 3.– id. 24.– 11 pp.  ↑228, 245, 252, 254
- [5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio *Generative adversarial networks* // *Communications of the ACM*.– 2020.– Vol. **63**.– No. 11.– Pp. 139–144.  ↑228, 234
- [6] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros *Image-to-image translation with conditional adversarial networks* // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Honolulu, HI, USA, July 21–26, 2017).– IEEE.– 2017.– ISBN 978-1-5386-0457-1.– Pp. 5967–5976.  ↑228, 234, 245, 252, 254
- [7] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. C. Loy *ESRGAN: Enhanced super-resolution generative adversarial networks* // *Computer Vision – ECCV 2018 Workshops, Proceedings*.– V. V (Münich, Germany, September 8–14, 2018), *Lecture Notes in Computer Science*.– vol. **11133**.– Springer.– 2019.– ISBN 978-3-030-11020-8.– Pp. 63–79.  ↑228, 245, 252, 254
- [8] J. Ho, A. Jain, P. Abbeel *Denosing diffusion probabilistic models* // *Advances in Neural Information Processing Systems 33, 34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (virtual, December 6–12, 2020).– 2020.– ISBN 9781713829546.– Pp. 6840–6851.  arXiv  2006.11239 ↑229, 235, 245, 252, 254

- [9] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, M. Norouzi *Image super-resolution via iterative refinement* // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2023.– Vol. **45**.– No. 4.– Pp. 4713–4726.  ↑229, 235
- [10] T. Panboonyuen, C. Charoenphon, C. Satirapod *SatDiff: A stable diffusion framework for inpainting very high-resolution satellite imagery* // IEEE Access.– 2025.– Vol. **13**.– Pp. 51617–51631.  ↑229, 245, 252, 254
- [11] T. Panboonyuen *KAO: Kernel-adaptive optimization in diffusion for satellite image* // IEEE Transactions on Geoscience and Remote Sensing.– 2025.– Vol. **63**.– id. 5531217.– 17 pp.  ↑229, 245, 252, 254
- [12] Y. Zhou, X. Gao, X. Wu, F. Wang, W. Jing, X. Hu *Image characteristic-guided learning method for remote-sensing image inpainting* // Remote Sensing.– 2025.– Vol. **17**.– No. 13.– id. 2132.– 22 pp.  ↑229, 235, 236
- [13] R. Li, L. Wen, S. Shao, M. Yu, L. Mohaisen *A novel generative adversarial network framework for super-resolution reconstruction of remote sensing* // Frontiers in Earth Science.– 2025.– Vol. **13**.– id. 578321.– 17 pp.  ↑229, 245, 252, 254
- [14] Z. Zhang, W. Feng, M. Zhong, M. Yang *BD-VITGAN: A blind dense VITGAN for satellite remote sensing images super-resolution reconstruction* // Geo-spatial Information Science.– 2025.– Pp. 1–23.  ↑229
- [15] Y. Wang, W. Wu, Z. Zhang, Z. Li, F. Zhang, X. Li *A temporal attention-based multi-scale generative adversarial network to fill gaps in time series of MODIS data for land surface phenology extraction* // Remote Sensing of Environment.– 2025.– Vol. **318**.– id. 114507.  ↑229
- [16] D. Zhou, L. Xu, K. Wu, H. Liu, M. Jiang *DSEPGAN: A dual-stream enhanced pyramid based on generative adversarial network for spatiotemporal image fusion* // Remote Sensing.– 2025.– Vol. **17**.– No. 24.– id. 4050.– 25 pp.  ↑229
- [17] И. В. Винокуров *Повышение точности сегментирования объектов с использованием генеративно-состязательной сети* // Программные системы: теория и приложения.– 2025.– Т. **16**.– № 2.– С. 111–152 (Англ., Рус.).  ↑229
- [18] И. В. Винокуров *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках* // Программные системы: теория и приложения.– 2025.– Т. **16**.– № 1.– С. 3–44 (Англ., Рус.).  ↑229
- [19] J. Johnson, A. Alahi, L. Fei-Fei *Perceptual losses for real-time style transfer and super-resolution* // *Computer Vision - ECCV 2016*, Proceedings.– V. II, 14th European Conference (Amsterdam, The Netherlands, October 11–14, 2016), Lecture Notes in Computer Science.– vol. **9906**.– Springer.– 2016.– ISBN 978-3-319-46474-9.– Pp. 694–711.  ↑235, 237
- [20] J. Zhang, Y. Xiao, G. Chen, Q. Sun, F. Xu, C.-S. Leung *Histogram-guided semantic-aware colorization* // *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022* (Virtual and Singapore, May 23–27, 2022).– IEEE.– 2022.– ISBN 978-1-6654-0541-6.– Pp. 2549–2553.  ↑235

- [21] I. Corley, J. Lwowski, P. Najafirad *ZRG: A dataset for multimodal 3D residential rooftop understanding* // *2024 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2024* (Waikoloa, HI, USA, January 03–08, 2024).– IEEE.– 2024.– ISBN 979-8-3503-1893-7.– Pp. 4623–4631.  [arXiv](#)  2304.13219 %  [↑236](#)
- [22] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez *Can semantic labeling methods generalize to any city? The INRIA aerial image labeling benchmark* // *2017 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2017* (Fort Worth, TX, USA, July 23–28, 2017).– IEEE.– 2017.– ISBN 978-1-5090-4951-6.– Pp. 3226–3229.  [↑236](#)
- [23] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Bailard, S. Benitez, U. Breitkopf *The ISPRS benchmark on urban object classification and 3D building reconstruction* // *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.*– 2012.– T. I.– № 3.– C. 293–298.  [URL](#) [↑236](#)
- [24] Q. Huynh-Thu, M. Ghanbari *Scope of validity of PSNR in image/video quality assessment* // *Electronics Letters.*– 2008.– Vol. 44.– No. 13.– Pp. 800–801.  [↑241](#)
- [25] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *Image quality assessment: from error visibility to structural similarity* // *IEEE Transactions on Image Processing.*– 2004.– Vol. 13.– No. 4.– Pp. 600–612.  [↑241](#)
- [26] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang *The unreasonable effectiveness of deep features as a perceptual metric* // *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018* (Salt Lake City, UT, USA, June 18–22, 2018).– IEEE.– 2018.– ISBN 978-1-5386-6421-6.– Pp. 586–595.  [↑241](#)
- [27] A. Sekrecka, K. Karwowska *Classical vs. machine learning-based inpainting for enhanced classification of remote sensing image* // *Remote Sensing.*– 2025.– Vol. 17.– No. 7.– id. 1305.– 36 pp.  [↑241](#)
- [28] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter *GANs trained by a two time-scale update rule converge to a local Nash equilibrium* // *Advances in Neural Information Processing Systems 30*, 31st Annual Conference on Neural Information Processing Systems 2017 (Long Beach, CA, USA, December 4–9, 2017).– 2017.– ISBN 9781510860964.– Pp. 6626–6637.   [URL](#) [↑241, 257](#)
- [29] D. Hasler, S. E. Süssstrunk *Measuring colourfulness in natural images*, SPIE/IS&T Human Vision and Electronic Imaging (Santa Clara, CA, United States, 20 January 2003), *Proceedings of SPIE.*– vol. 5007.– 2003.– Pp. 87–95.  [↑251](#)
- [30] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo *Swin Transformer V2: Scaling up capacity and resolution* // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022* (New Orleans, LA, USA, June 18–24, 2022).– IEEE.– 2022.– ISBN 978-1-6654-6946-3.– Pp. 11999–12009.  [↑256](#)

- [31] H. Cai, J. Li, M. Hu, C. Gan, S. Han *EfficientViT: Multi-scale linear attention for high-resolution dense prediction*. – 2024. – 12 pp. arXiv:2205.14756   2205.14756  ↑256

Поступила в редакцию 30.04.2026;
одобрена после рецензирования 01.06.2026;
принята к публикации 10.06.2026;
опубликована онлайн 20.06.2026.

Рекомендовал к публикации

к.т.н. В. П. Фраленко

Информация об авторах:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных



0000-0001-8697-1032

e-mail: igvvinokurov@fa.ru



Георгий Михайлович Лапаныков

Выпускник (бакалавр) Финансового Университета при Правительстве Российской Федерации. Область научных интересов: информационные системы, разработка мобильных приложений, анализ данных



0009-0007-0511-628X

e-mail: goshmen2004@gmail.com



Георгий Дмитриевич Умаров

Выпускник (бакалавр) Финансового Университета при Правительстве Российской Федерации. Область научных интересов: информационные технологии, веб-разработка, анализ данных



0009-0007-0364-8477

e-mail: goshaumarov0609@mail.ru

Вклад авторов: *И. В. Винокуров* – 70% (разработка модели и методики проведения экспериментов, реализация обучения и исследования модели, интеграция результатов в информационные системы ППК «Роскадастр»); *Г. М. Лапаныков* – 15% (реализация предобучения на ZRG); *Г. Д. Умаров* – 15% (формирование синтетических масок, визуализация результатов обучения модели).

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*