Б

Проблемы искусственного интеллекта. 2026. N 3 (42). С. 96-106
Problems of Artificial Intelligence. 2026;3(42):96-106.
Системный анализ, управление и обработка информации, статистика
Научная статья
УДК 004.932:004.89 **DOI** 10.24412/2413-7383-2026-3-42-96-106

М. В. Близно
Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, 118 б
bmv.ipai@mail.ru; ORCID 0009-0006-5212-6395

УПРАВЛЯЕМАЯ ДИФФУЗИОННАЯ ГЕНЕРАЦИЯ СИНТЕТИЧЕСКИХ АЭРОФОТОСНИМКОВ БПЛА ДЛЯ СИСТЕМ ПЕРЕКРЕСТНОЙ ГЕОЛОКАЛИЗАЦИИ: АНАЛИТИЧЕСКИЙ ОБЗОР

M. V. Blizno
Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Donetsk, Artema str, 118-b
bmv.ipai@mail.ru; ORCID 0009-0006-5212-6395

CONTROLLABLE DIFFUSION-BASED GENERATION OF SYNTHETIC UAV AERIAL IMAGERY FOR CROSS-VIEW GEO-LOCALIZATION: A SURVEY

Рассмотрены методы формирования синтетических аэрофотоснимков БПЛА для обучения систем перекрестной геолокализации (CVGL). Материал систематизирован по пяти направлениям. Симуляционные данные теряют 10–30 % точности из-за расхождения доменов, а наивная синтетическая аугментация без фильтрации ухудшает результаты. Систематизированы ограничения существующих подходов и построена карта пробелов исследований: среди проанализированных публикаций отсутствуют работы, объединяющие управляемую диффузию, многокомпонентную оценку качества и оценку на целевой задаче CVGL.
Ключевые слова: CVGL; синтетические данные; диффузионные модели; ControlNet; беспилотные летательные аппараты; аэрофотоснимки; дистанционное зондирование; оценка качества изображений.
Финансирование: Работа выполнена в рамках федерального проекта «Развитие человеческого капитала в интересах регионов, отраслей и сектора исследований и разработок» национального проекта «Наука и университеты», работы молодёжной лаборатории «Лаборатория интеллектуальных систем и анализа данных» по теме «Извлечение семантической информации из изображений для автономных систем навигации беспилотных летательных аппаратов», FREN-2024-0002

Methods for generating synthetic UAV aerial imagery for training cross-view geo-localization (CVGL) systems are reviewed. The material is systematized into five directions. Simulation data lose 10–30 % of accuracy due to the domain gap, and naive synthetic augmentation without filtering degrades performance. The limitations of existing approaches are systematized, and a research-gap map is built: among the analyzed publications, none combines controllable diffusion, multi-component quality assessment, and evaluation on the downstream CVGL.
Key words: cross-view geo-localization; synthetic data; diffusion models; ControlNet; unmanned aerial vehicles; aerial imagery; remote sensing; image quality assessment.
Funding: The work was carried out as part of the federal project "Development of Human Capital in the Interests of Regions, Sectors, and Research and Development," within the national project "Science and Universities." The project was conducted by the youth laboratory "Laboratory of Intelligent Systems and Data Analysis" on the topic "Extraction of Semantic Information from Images for Autonomous Navigation Systems of Unmanned Aerial Vehicles," FREN-2024-0002.

Введение

Определение местоположения беспилотного летательного аппарата (БПЛА) только по визуальным данным в отсутствии спутниковой навигации является сложной задачей. Перекрестная геолокализация (cross-view geo-localization, CVGL) решает данную задачу сопоставлением аэрофотоснимка, полученного с камер БПЛА, со спутниковым изображением местности. Положение аппарата определяется тем спутниковым фрагментом, чей вектор признаков ближе всего к вектору признаков аэрофотоснимка. Практическая необходимость в таких системах возникает в задачах автономной навигации, мониторинга территорий и поисково-спасательных работах, где информация о местонахождении устройства критически важна.

Ключевое ограничение при построении систем CVGL лежит в обучающих данных. Обучение требует размеченных пар «аэрофотоснимок – спутниковый фрагмент», охватывающих один и тот же участок с разными сезонами, при разном освещении и погоде, на различных высотах и ракурсах съемки. Ручной сбор таких данных с фиксированной привязкой к координатам связан с существенными трудозатратами и не масштабируется на нужное разнообразие условий наблюдения.

Синтетические наборы на игровых и симуляционных движках масштабируют обучение, но при переносе на реальные данные приводят к снижению точности на 10-30 % из-за разрыва доменов [1], [2]. Управляемая диффузия позволяет варьировать геометрию и семантику через структурные условия [3-5], однако ее адаптация к аэросъемке БПЛА и CVGL остается фрагментарной.

Методика формирования наборов аэрофотоснимков для CVGL [6], генерация синтетических наборов для сегментации снимков БПЛА [7], [8], восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантики [9] так же рассматривается и в русскоязычных работах.

Цель обзора – систематизировать методы на пересечении управляемой диффузии, аэросъемки БПЛА и CVGL. Рассмотрены синтетические наборы [1], [2], архитектуры CVGL [10], [11], диффузионные модели [3-5], структурный контроль [12], [13] и оценка качества [14], [15]. Научный значимость состоит в сопоставлении работ по пяти осям и формализации пробела: существующие метрики не различают структурные, семантические и геометрические дефекты отдельного снимка.

Практическая значимость заключается в том, что сравнительная матрица и карта покрытия применимы как инструмент выбора компонентов при проектировании систем генерации обучающих данных для CVGL, а сформулированные требования как основа технического задания на разработку такой системы.

1 Методология отбора источников

Отбор публикаций охватывает ведущие конференции и профильные журналы. Критерий отбора заключается в значимости пересечения направлений областей управляемой диффузионной генерации, аэросъемки и CVGL. Из выборки исключены общие VLM-методы и синтез изображений без связи с CVGL.

Настоящая систематизация отличается от опубликованных обзоров [16], [17] фокусом на пересечении областей. Отобранные работы сопоставлены по пяти направлениям. Первое, это тип генерации [1-5]. Второе различает многокомпонентнось оценки качества [12-15] Третье фиксирует наличие итеративной оптимизации [14]. Четвертое отражает оценку на целевых задачах по метрикам CVGL, реализованных в эталонных наборах [2], [10], [11]. Пятое отражает использование VLM- описаний при формировании условий. На этих пяти осях и строится сравнительная матрица, служащая базой для дальнейшей систематизации и выявления пробелов области.

2 Обзор методов

2.1 Синтетические наборы данных для аэросъемки БПЛА

В работе [1] представлен SynDrone – мультимодальный набор из 72 000 аэрофотоснимков городских сцен, привязанных к трем высотам полета: 20, 50 и 80 м., созданного при помощи симулятора CARLA. Каждый кадр сопровождается RGB-изображением, картой глубины и семантической маской. Полнота пиксельных аннотаций делает такой набор хорошим источником структурных карт, однако в его охвате варьируются лишь высота и ракурс, тогда как сезон, время суток и погода остаются фиксированными, а сцены ограничены городской застройкой, в соответствии с рисунком 1 а.

В работе Game4Loc [2] собран набор GTA-UAV на движке GTA V, содержащий 33 763 снимка, покрывающих территорию 81,3 км² с высот от 80 до 650 м., включая городские, горные, прибрежные и лесные ландшафты, в соответствии с рисунком 1 б, однако вариативность внешних условий по-прежнему отсутствует. Для сопоставления кадров дрона и спутника при частичном перекрытии предложено взвешенное контрастивное обучение, учитывающее долю общей области пары.



а) б)

Рисунок 1 – Пример из наборов данных: а) SynDrone [1]; б) GTA-UAV [2]

Ни один из наборов не использует диффузионные модели и не варьирует условия наблюдения для одной и той же локации, именно эта комбинация остается незакрытой.

2.2 Методы перекрестной геолокализации

Прогресс последних лет задачи CVGL опирается на контрастивное обучение с продуманным отбором отрицательных примеров. Sample4Geo [11] сочетает энкодер ConvNeXt [18] с функцией InfoNCE [19] и стратегией отбора сложных негативных примеров, где похожие спутниковые фрагменты отбираются по географической близости и визуальному сходству. Такой подход дал высокий Recall@1 (доля запросов, для которых верное соответствие оказывается первым в ранжированном списке) на CVUSA [20], CVACT [21] и University-1652 [22]. Ограничение проявляется при смене источника данных – точность заметно снижается, если модель обучена только на синтетических данных, что указывает на чувствительность метода к разрыву доменов между обучающим набором и целевыми данными.

Иной подход представляет GLQINet – трансформерная архитектура (ViT) [10] с квадрантным разбиением карт признаков и согласованием глобальных и локальных дескрипторов (IGLAM). Механизм явно кодирует пространственную структуру сцены и достигает SOTA-показателей на University-1652 [22] и SUES-200 [23]. Изменение внешних условий, таких как сезон, время суток, погоды, при этом не учитывается. Модель настроена на стабильную геометрию, а не на вариативность освещения и атмосферы.

Влияние условий съемки на CVGL количественно оценили в [24]. Авторы показали, что туман, дождь и снег существенно снижают точность сопоставления, и предложили диффузионное восстановление изображений как этап предобработки перед геолокализацией. Диффузионные модели в данном случае применяется для очистки деградированных снимков, а не для формирования обучающих данных с контролируемой вариацией погоды.

Разрыв между чувствительностью моделей к условиям съемки и отсутствием механизмов создания таких условий обусловливает применения синтетического подхода, заключающегося в управляемой генерации аэрофотоснимков с целенаправленной вариацией сезона, времени и погоды.

2.3 Диффузионные модели для дистанционного зондирования

Применение диффузионных моделей в области дистанционного зондирования началось со спутникового домена, где накоплены крупные размеченные архивы. В работе [3] представили DiffusionSat – первую фундаментальную генеративную модель для спутниковых снимков, построенную как латентная диффузия на базе Stable Diffusion 2.1 и дообученную на публичных наборах для дистанционного зондирования. Условием генерации служат числовые геопространственные метаданные – координаты и временная метка, кодируемые синусоидальными векторными представлениями. Такая параметризация избавляет от текстовых описаний и охватывает несколько задач: временную генерацию, сверхразрешение и восстановление содержимого изображения. Структурный контроль через ControlNet [12], при этом, отсутствует, как и связь с прикладными задачами.

Развитие в сторону многоканального управления предложили в CRS-Diff [4]. Платформа объединяет глобальные условия такие как текст и метаданные, с локальными – картами сегментации, схемами дорог, контурными картами и картами глубины, через ControlNet [12] поверх Stable Diffusion 1.5. Обучение проходит в два этапа: сначала базовая модель текст-изображение, затем управляющая сеть с многоканальными условиями. Управляемость возрастает, но оценка ограничена FID [3], а итеративная оптимизация с обратной связью не предусмотрена. Как и DiffusionSat, работа остается в пределах спутникового домена.

Наиболее близким к рассматриваемой постановке задачи является AeroGen [5]. Авторы применили управляемую структурой расположения объектов условную диффузию для аугментации данных детекции объектов на аэрофотоснимках, впервые реализовав поддержку повернутых ограничивающих рамок как условия генерации. Одним из основных результатов работы является роль фильтрации. Наивная синтетическая аугментация без отбора образцов ухудшает детекцию, тогда как каскадная фильтрация (CLIP-оценка совместно с классификатором ResNet101) дает прирост mAP на 12,6–17,8% для редких классов. Оценка качества сгенерированных данных предстает не дополнительной опцией, а необходимым условием их пригодности.

Сопоставление трех работ выявляет устойчивую границу. DiffusionSat и CRSDiff адаптируют диффузию к спутниковому домену, без структурного контроля геометрии в первом случае и без выхода за пределы FID-оценки во втором. AeroGen обращается к аэрофото, но решает задачу детекции, а не сопоставления ракурсов. Ни один из подходов не адаптирует управляемую диффузионную генерацию к аэросъемке БПЛА и задаче CVGL, оставляя это пересечение незакрытым.

2.4 Структурный контроль диффузионной генерации

Ключевым методом пространственного управления предобученными диффузионными моделями стал ControlNet [12]. Он добавляет управляющие условия, такие как карты границ Canny, глубины, семантической сегментации, не затрагивая веса базовой модели. Однако, практическое применение для каждого типа условий требует отдельного ControlNet, что усложняет многоканальное управление, встроенная оценка соответствия результата заданному условию при этом отсутствует, а геометрический контроль над контурами и глубиной не гарантирует семантической корректности сцены, поскольку сохранение структуры совместимо с искажением содержания.

Часть этих проблем решает ControlNet++ [13], который вводит циклическую функцию потерь согласованности. Из сгенерированного изображения дискриминативная модель восстанавливает управляющий сигнал, который сравнивается с исходным условием, а рассогласование штрафуется при обучении. Такой подход повышает точность следования условию без дополнительных затрат на вывод модели. Однако коррекция действует исключительно на этапе обучения, опирается на единственную метрику согласованности и не проверялась на снимках дистанционного зондирования.

Структурный контроль подробно исследован на уровне архитектуры, при этом ни одно из решений не предлагает диагностики на этапе вывода модели и адресной коррекции конкретных дефектов сгенерированного снимка, рассогласование фиксируется лишь косвенно, через агрегированную функцию потерь при дообучении.

2.5 Итеративная оптимизация и оценка качества генерации

Замыкание генеративного конвейера обратной связью отделяет однопроходный синтез от управляемого. Здесь сформировалось направление, объединяющее обучаемые метрики качества и итеративное дообучение по сигналу вознаграждения. В [15] обучили ImageReward – модель соответствия сгенерированных изображений предпочтениям человека на базе BLIP по массиву из 137 000 попарных сравнений и предложили процедуру ReFL для дообучения диффузии. В IterComp [14] кандидаты оцениваются набором моделей-экспертов (цвет, форма, пространственные отношения, счетность), агрегированная скалярная оценка качества служит сигналом для RL-дообучения, а сама генерация улучшается от итерации к итерации.

Однако, скалярное вознаграждение улавливает факт снижения качества, но не тип дефекта, из-за чего коррекция ведется без привязки к причине дефекта, а сходимость у IterComp требует четырех-пяти итераций. Оба метода привязаны к дообучению модели, тогда как на этапе использования модели распознавание типа дефекта не предусмотрено. Отсутствует и специфика CVGL, ImageReward, калиброванный под общее восприятие изображения, не определяет геометрические деформации и семантические галлюцинации аэроснимков и не коррелирует с эффективностью геолокализации на целевой задаче.

Схожее ограничение имеется у FID [3], использующийся как метрика качества у диффузионных моделей дистанционного зондирования [3], [4]. Он сравнивает распределения признаков, а не качество отдельного снимка, поэтому не обнаруживает геометрические деформации и семантические искажения, критичным именно для CVGL.

Из сопоставления обнаруживается устойчивый разрыв – ни агрегированная скалярная оценка качества [14], [15], ни FID [3], [4] не дают для каждого изображения многокомпонентной оценки, которая различала бы структурные, семантические и

геометрические дефекты аэрофотоснимков. Итеративные же процедуры работают на этапе дообучения и лишены оценки созданных изображений, указывающей на причину расхождения. Покомпонентная диагностика сгенерированных изображений и коррекция параметров в процессе генерации, создают открытое направление для методов замкнутой оптимизации генерации. Практическая проверка нейросетевых детекторов и сегментаторов с аугментацией данных выполнена в [25], [26]; инструментарий аннотирования обучающих данных для компьютерного зрения предложен в [27].

3 Сравнительный анализ и научный пробел

Результаты рассмотренных публикаций сведены в две проекции: сводная характеристика рассмотренных работ, в соответствии с таблицей 1; карта покрытия пяти осей, в соответствии с таблицей 2. Совмещение двух проекций показывает, где именно проходят границы освоенного и где остается незакрытое пространство.

Таблица 1 – Сводная характеристика рассмотренных работ

Подход	**Преимущества**	**Ограничения**	**Роль для задачи CVGL**
Симуляция CARLA [1]	Мультимодальность, масштаб, пиксельная разметка	Доменный разрыв, узкий диапазон высот, нет вариативности погоды и сезонов	Источник структурных карт (сегментация, глубина)
GTA-UAV [2]	Масштаб, диапазон высот, частичное сопоставление пар	Доменный разрыв 10–30%, без диффузии, без оценки качества данных	Целевой набор данных для CVGL
ViT [10]	Высокая точность на стандартных наборах данных	Нет работы с условиями наблюдения и синтетикой	Представитель целевых архитектур CVGL
ConvNeXt, InfoNCE [11]	Высокий Recall@1, воспроизводимость	Требует много размеченных пар, деградация на чистой синтетике	Базовая модель для оценки влияния синтетики
Диффузия для восстановления погоды [24]	Явный учет погодных искажений	Диффузия для восстановления, не для генерации выборок	Обеспечение погодной вариативности
DiffusionSat [3]	Первая RS фундаментальная модель, числовое кондиционирование	Только спутники, нет структурного ControlNet, нет целевой связи	Образец адаптации диффузии к RS-домену
CRS-Diff [4]	Многоканальный контроль, специализация на RS	Только спутники, оценка через FID, без оптимизации	Предшественник многоканального контроля
AeroGen [5]	Прирост mAP 12,6–17,8%, доказана нужность фильтрации	Детекция, а не CVGL; две метрики фильтрации	Обоснование многокомпонентной оценки
ControlNet [12]	Надежный контроль геометрии, эффективное дообучение	Нет встроенной оценки и оптимизации, семантика не гарантирована	Базовый метод структурного контроля
ControlNet++ [13]	Улучшение следования условию без затрат на вывод модели	Только на этапе обучения, одна метрика, не проверялось на RS	Принцип согласованности условия
IterComp [14]	Документированное улучшение при итерациях	Скалярная оценка качества без диагностики, 4–5 итераций, нет RS-специфики	Обоснование адресной диагностики
ImageReward [15]	Корреляция с человеческой оценкой, открытые веса модели	Нет RS-специфики, скалярная оценка, нет целевой корреляции	Компонента оценочной системы

Таблица 2 – Карта покрытия ключевых требований

Работа	**Диффузионная генерация**	**Многокомпонентная оценка каждого изображения**	**Итеративная оптимизация**	**Оценка на целевой задаче CVGL**	**VLM-генерация описания**
Game4Loc [2]	Нет, игровой движок	нет	нет	да (Recall@1)	нет
CRS-Diff [4]	Да, Мульти-условный ControlNet	Нет, только FID	нет	нет	нет
DiffusionSat [3]	Да, LDM + мета-данные	Частично – FID, SSIM	нет	нет	нет
AeroGen [5]	Да, Управляемая диффузия	Частично – FID, CAS	нет	Частично – mAP детекции	нет
ControlNet++ [13]	Да, Циклическая оценка качества	Частично, по условию	Частично, в моменте обучения	нет	нет
IterComp [14]	Да, Скалярный оценка качества	Частично, 1 скаляр	Да, скалярная	нет	нет

В соответствии с таблицей 2, ни одна из шести работ не покрывает все пять осей. Диффузионные методы дистанционного зондирования ориентированы на спутниковые снимки [3], [4], AeroGen [5] решает задачу детекции, а симуляционные наборы [1], [2] не устраняют доменный разрыв и не моделируют сезонно-погодную вариативность одной локации.

Не сформирована и многокомпонентная метрика отдельного изображения, связанная с R@1 или mAP: FID [3], [4] нечувствителен к локальным геометрическим и семантическим дефектам, а метрики предпочтений человека не учитывают специфику дистанционного зондирования [15]. Ни одна работа не реализует замкнутый контур «генерация – оценка – оптимизация – проверка на CVGL» с диагностикой причины дефекта. Следовательно, открытым направлением остается управляемый геосемантический синтез аэрофотоснимков БПЛА с контролем качества каждого изображения и адресной обратной связью.

Заключение

Обзор публикаций на стыке управляемой диффузии, аэрофотосъемки с помощью БПЛА и CVGL выявляет область, не охваченную ни одной из существующих работ. Управляемая генерация методически проработана: механизм нулевой свертки обеспечивает структурный контроль без разрушения предобученного распределения [12], а перенос согласованности циклов на этап обучения повышает точность следования условию [13]. Для дистанционного зондирования диффузионные модели дополнены числовыми метаданными [3] и мультиусловным управлением генерацией [4]. Несмотря на развитие перечисленных направлений, их интеграция в рамках единого конвейера генерации синтетических аэрофотоснимков для задач CVGL в настоящее время отсутствует.

Систематизация литературы показывает, что ни в одной из рассмотренных работ управляемая диффузионная генерация, адаптированная к ракурсу и специфике аэрофотосъемки с помощью БПЛА, многокомпонентная оценка качества каждого синтетического изображения, калиброванная по результату целевой задачи (R@1/mAP), и адресная обратная связь по типу дефекта не объединены в единый конвейер «генерация — оценка — оптимизация». Управляемый геосемантический синтез аэрофотоснимков с БПЛА с контролем качества каждого изображения и обратной связью, замкнутой на метрики геолокации, остается открытой областью.

Список источников

1. Rizzoli G., Barbato F., Caligiuri M., Zanuttigh P. SynDrone – Multi-modal UAV Dataset for Urban Scenarios // 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Paris: IEEE, 2023. P. 2202–2212. DOI: 10.1109/ICCVW60793.2023.00235
2. Ji Y., He B., Tan Z., Wu L. Game4Loc: A UAV Geo-Localization Benchmark from Game Data // Proceedings of the AAAI Conference on Artificial Intelligence. 2025. Vol. 39, No. 4. P. 3913–3921. DOI: 10.1609/aaai.v39i4.32409
3. Khanna S., Liu P., Zhou L., Meng C., Rombach R., Burke M., Lobell D. B., Ermon S. DiffusionSat: A Generative Foundation Model for Satellite Imagery // The Twelfth International Conference on Learning Representations (ICLR). 2024. URL: https://openreview.net/forum?id=I5webNFDgQ (дата обращения: 06.07.2026).
4. Tang D., Cao X., Hou X., Jiang Z., Liu J., Meng D. CRS-Diff: Controllable Remote Sensing Image Generation with Diffusion Model // IEEE Transactions on Geoscience and Remote Sensing. 2024. Vol. 62. Art. 5638714. DOI: 10.1109/TGRS.2024.3453414
5. Tang D., Cao X., Wu X., Li J., Yao J., Bai X., Jiang D., Li Y., Meng D. AeroGen: Enhancing Remote Sensing Object Detection with Diffusion-Driven Data Generation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2025. P. 3614–3624. DOI: 10.1109/CVPR52734.2025.00342
6. Павленко Б.В., Пикалев Я.С. Методика создания набора аэрофотоснимков для задачи перекрестной геолокализации // Проблемы искусственного интеллекта. 2024. № 4 (35). С. 101–112. DOI: 10.24412/2413-7383-2024-4-101-112
7. Молин А.Е., Блеканов И.С., Митрофанов Е.П., Митрофанова О.А. Методы генерации синтетических данных для обучения нейросетей в задаче сегментации уровня азотного режима растений на снимках беспилотных летательных аппаратов на сельскохозяйственном поле // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2024. Т. 20, № 1. С. 20–33. DOI: 10.21638/11701/spbu10.2024.103
8. Устенко, В. Ю. Методика расширения данных для сегментации ключевых объектов на аэрофотоснимках БПЛА / В. Ю. Устенко, М. В. Близно // Проблемы искусственного интеллекта. 2025. № 4(39). С. 84-97. DOI 10.24412/2413-7383-2025-4-39-84-97
9. Фаворская М.Н., Пахирка А.И. Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей // Информатика и автоматизация. 2024. Т. 23, № 4. С. 1047–1076. DOI: 10.15622/ia.23.4.5
10. Xu J., Yin J., Zhang J., Gao T. Enhancing Cross View Geo-Localization Through Global-Local Quadrant Interaction Network // Scientific Reports. 2025. Vol. 15. Art. 33431. DOI: 10.1038/s41598-025-18935-6
11. Deuser F., Habel K., Oswald N. Sample4Geo: Hard Negative Sampling for Cross-View Geo-Localisation // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. P. 16801–16810. DOI: 10.1109/ICCV51070.2023.01545
12. Zhang L., Rao A., Agrawala M. Adding Conditional Control to Text-to-Image Diffusion Models // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. P. 3836–3847. DOI: 10.1109/ICCV51070.2023.00355
13. Li M., Yang T., Kuang H., Wu J., Wang Z., Xiao X., Chen C. ControlNet++: Improving Conditional Controls with Efficient Consistency Feedback // Computer Vision – ECCV 2024. Lecture Notes in Computer Science, Vol. 15085. Milan: Springer, 2024. P. 129–147. DOI: 10.1007/978-3-031-72667-5_8
14. Zhang X., Yang L., Li G., Cai Y., Xie J., Tang Y., Yang Yu., Wang M., Cui B. IterComp: Iterative Composition-Aware Feedback Learning from Model Gallery for Text-to-Image Generation [Electronic resource] // The Thirteenth International Conference on Learning Representations (ICLR). 2025. P. 8608–8628. URL: https://openreview.net/forum?id=4w99NAikOE (Date of access: 06.07.2026).
15. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation / Xu J., Liu X., Wu Y., Tong Y., Li Q., Ding M., Tang J., Dong Yu. // Advances in Neural Information Processing Systems (NeurIPS). 2023. Vol. 36. P. 15903–15935.
16. Liu Y., Yue J., Xia S., Ghamisi P., Xie W., Fang L. Diffusion Models Meet Remote Sensing: Principles, Methods, and Perspectives // IEEE Transactions on Geoscience and Remote Sensing. 2024. Vol. 62. Art. 4708322. DOI: 10.1109/TGRS.2024.3464685
17. Durgam A., Paheding S., Dhiman V., Devabhaktuni V. Cross-view Geo-localization: A Survey // IEEE Access. 2024. Vol. 12. P. 192028–192050. DOI: 10.48550/arXiv.2406.09722
18. Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., Xie S. A ConvNet for the 2020s // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. P. 11976–11986. DOI: 10.1109/CVPR52688.2022.01167
19. Van den Oord A., Li Y., Vinyals O. Representation Learning with Contrastive Predictive Coding // arXiv preprint. 2018. arXiv:1807.03748. DOI: 10.48550/arXiv.1807.03748
20. Workman S., Souvenir R., Jacobs N. Wide-Area Image Geolocalization with Aerial Reference Imagery // Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. P. 3961–3969. DOI: 10.1109/ICCV.2015.451

21. Liu L., Li H. Lending Orientation to Neural Networks for Cross-View Geo-Localization // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. P. 5617–5626. DOI: 10.1109/CVPR.2019.00577
22. Zheng Z., Wei Y., Yang Y. University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization // Proceedings of the 28th ACM International Conference on Multimedia (ACM MM). Seattle: ACM, 2020. P. 1395–1403. DOI: 10.1145/3394171.3413896
23. Zhu R. et al. SUES-200: A multi-height multi-scene cross-view image benchmark across drone and satellite //IEEE Transactions on Circuits and Systems for Video Technology. 2023. Т. 33. №. 9. С. 4825-4839.
24. Feng T., Li Q., Wang X., Wang M., Li G., Zhu W. Multi-weather Cross-view Geo-localization Using Denoising Diffusion Models // Proceedings of the 2nd Workshop on UAVs in Multimedia: Capturing the World from a New Perspective (ACM MM Workshop). New York: ACM, 2024. DOI: 10.1145/3689095.3689103
25. Медведева, Е.В. Исследование нейросетевых алгоритмов распознавания объектов железнодорожной инфраструктуры на видеоизображениях / Е.В. Медведева, А.А. Перевощикова // Компьютерная оптика. 2025. Т. 49, № 3. С. 443–450. DOI: 10.18287/2412-6179-CO-1563
26. Близно, М. В. Сравнение различных методов аугментации на обучение модели нейронной сети / М. В. Близно // Искусственный интеллект: теоретические аспекты и практическое применение : материалы Донецкого Международного круглого стола. Донецк : ФГБНУ «ИПИИ», 2025. 296 с. С. 99-102.
27. Устенко В.Ю., Бондаренко В.И. Разработка программного комплекса аннотирования данных для компьютерного зрения: объектно-ориентированный подход на основе WinForms // Проблемы искусственного интеллекта. 2024. № 4 (35). С. 151–163. DOI: 10.24412/2413-7383-2024-4-151-163

References

1. Rizzoli G., Barbato F., Caligiuri M., Zanuttigh P. SynDrone – Multi-modal UAV Dataset for Urban Scenarios // 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). – Paris: IEEE, 2023. – P. 2202–2212. – DOI: 10.1109/ICCVW60793.2023.00235
2. Ji Y., He B., Tan Z., Wu L. Game4Loc: A UAV Geo-Localization Benchmark from Game Data // Proceedings of the AAAI Conference on Artificial Intelligence. – 2025. – Vol. 39, No. 4. – P. 3913–3921. – DOI: 10.1609/aaai.v39i4.32409
3. Khanna S., Liu P., Zhou L., Meng C., Rombach R., Burke M., Lobell D.B., Ermon S. DiffusionSat: A Generative Foundation Model for Satellite Imagery // The Twelfth International Conference on Learning Representations (ICLR). – 2024. – URL: https://openreview.net/forum?id=I5webNFDgQ (accessed: 06.07.2026).
4. Tang D., Cao X., Hou X., Jiang Z., Liu J., Meng D. CRS-Diff: Controllable Remote Sensing Image Generation with Diffusion Model // IEEE Transactions on Geoscience and Remote Sensing. – 2024. – Vol. 62. – Art. 5638714. – DOI: 10.1109/TGRS.2024.3453414
5. Tang D., Cao X., Wu X., Li J., Yao J., Bai X., Jiang D., Li Y., Meng D. AeroGen: Enhancing Remote Sensing Object Detection with Diffusion-Driven Data Generation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Seattle: IEEE, 2025. – P. 3614–3624. – DOI: 10.1109/CVPR52734.2025.00342
6. Pavlenko B.V., Pikalev Ya.S. A methodology for creating an aerial-photo dataset for the cross-view geolocalization task // Problems of Artificial Intelligence. – 2024. – No 4 (35). – P. 101–112. – DOI: 10.24412/2413-7383-2024-4-101-112
7. Methods for generating synthetic data to train neural networks for segmenting plant nitrogen-regime levels in UAV imagery of agricultural fields / Molin A.E., Blekanov I.S., Mitrofanov E.P., Mitrofanova O.A. // Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes. – 2024. – Vol. 20, No. 1. – P. 20–33. – DOI: 10.21638/11701/spbu10.2024.103
8. Ustenko V.Yu., Blizno M.V. A data-augmentation methodology for segmenting key objects in UAV aerial imagery // Problems of Artificial Intelligence. – 2025. – No 4 (39). – P. 84–97. – DOI: 10.24412/2413-7383-2025-4-39-84-97
9. Favorskaya M.N., Pakhirka A.I. Semantically-aware restoration of ultra-high-resolution aerial photographs // Informatics and Automation. – 2024. – Vol. 23, No. 4. – P. 1047–1076. – DOI: 10.15622/ia.23.4.5
10. Xu J., Yin J., Zhang J., Gao T. Enhancing Cross-View Geo-Localization Through Global-Local Quadrant Interaction Network // Scientific Reports. 2025. Vol. 15. Art. 33431. DOI: 10.1038/s41598-025-18935-6
11. Deuser F., Habel K., Oswald N. Sample4Geo: Hard Negative Sampling for Cross-View Geo-Localisation // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). - Paris: IEEE, 2023. – P. 16801–16810. – DOI: 10.1109/ICCV51070.2023.01545
12. Zhang L., Rao A., Agrawala M. Adding Conditional Control to Text-to-Image Diffusion Models // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – Paris: IEEE, 2023. - P. 3836–3847. – DOI: 10.1109/ICCV51070.2023.00355
13. ControlNet++: Improving Conditional Controls with Efficient Consistency Feedback / Li M., Yang T., Kuang H., Wu J., Wang Z., Xiao X., Chen C. // Computer Vision – ECCV 2024. Lecture Notes in Computer Science. – Vol. 15085. – Milan: Springer, 2024. – P. 129–147. – DOI: 10.1007/978-3-031-72667-5_8

14. IterComp: Iterative Composition-Aware Feedback Learning from Model Gallery for Text-to-Image Generation / Zhang X., Yang L., Li G., Cai Y., Xie J., Tang Y., Yang Yu., Wang M., Cui B. // The Thirteenth International Conference on Learning Representations (ICLR). – 2025. – P. 8608–8628. – URL: https://openreview.net/forum?id=4w99NAikOE (accessed: 06.07.2026).
15. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation / Xu J., Liu X., Wu Y., Tong Y., Li Q., Ding M., Tang J., Dong Yu. // Advances in Neural Information Processing Systems (NeurIPS). – 2023. – Vol. 36. – P. 15903–15935.
16. Liu Y., Yue J., Xia S., Ghamisi P., Xie W., Fang L. Diffusion Models Meet Remote Sensing: Principles, Methods, and Perspectives // IEEE Transactions on Geoscience and Remote Sensing. – 2024. – Vol. 62. – Art. 4708322. DOI: 10.1109/TGRS.2024.3464685
17. Durgam A., Paheding S., Dhiman V., Devabhaktuni V. Cross-view Geo-localization: A Survey // IEEE Access. – 2024. – Vol. 12. – P. 192028–192050. – DOI: 10.48550/arXiv.2406.09722
18. A ConvNet for the 2020s / Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., Xie S. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – New Orleans: IEEE, 2022. – P. 11976–11986. – DOI: 10.1109/CVPR52688.2022.01167
19. Van den Oord A., Li Y., Vinyals O. Representation Learning with Contrastive Predictive Coding // arXiv preprint. – 2018. – arXiv:1807.03748. – DOI: 10.48550/arXiv.1807.03748
20. Workman S. Wide-Area Image Geolocalization with Aerial Reference Imagery / S. Workman, R. Souvenir, N. Jacobs // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – Santiago: IEEE, 2015. – P. 3961–3969. – DOI: 10.1109/ICCV.2015.451
21. Liu L. Lending Orientation to Neural Networks for Cross-View Geo-Localization / L. Liu, H. Li // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Long Beach: IEEE, 2019. – P. 5617–5626. – DOI: 10.1109/CVPR.2019.00577
22. Zheng Z. University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization / Z. Zheng, Y. Wei, Y. Yang // Proceedings of the 28th ACM International Conference on Multimedia (ACM MM). – Seattle: ACM, 2020. – P. 1395–1403. – DOI: 10.1145/3394171.3413896
23. Zhu R. et al. SUES-200: A Multi-height Multi-scene Cross-view Image Benchmark Across Drone and Satellite // IEEE Transactions on Circuits and Systems for Video Technology. – 2023. – Vol. 33, No. 9. – P. 4825–4839.
24. Multi-weather Cross-view Geo-localization Using Denoising Diffusion Models / Feng T., Li Q., Wang X., Wang M., Li G., Zhu W. // Proceedings of the 2nd Workshop on UAVs in Multimedia: Capturing the World from a New Perspective (ACM MM Workshop). – New York: ACM, 2024. – DOI: 10.1145/3689095.3689103
25. Medvedeva E.V. A study of neural-network algorithms for recognizing railway-infrastructure objects in video imagery / A.A. Perevoshchikova // Computer Optics. – 2025. – Vol. 49, No. 3. – P. 443–450. – DOI: 10.18287/2412-6179-CO-1563
26. Blizno M.V. Comparison of different augmentation methods for neural network model training // Artificial Intelligence: Theoretical Aspects and Practical Application: Proceedings of the Donetsk International Round Table. – Donetsk: FSBSI «IPAI», 2025. – 296 p. – P. 99–102.
27. Ustenko V.Yu., Bondarenko V.I. Development of a data-annotation software package for computer vision: an object-oriented approach based on WinForms // Problems of Artificial Intelligence. – 2024. – No. 4 (35). – P. 151–163. - DOI: 10.24412/2413-7383-2024-4-151-163

RESUME

M. V. Blizno

Controllable diffusion-based generation of synthetic UAV aerial imagery for cross-view geo-localization: a survey

Training cross-view geolocalization (CVGL) systems for unmanned aerial vehicles (UAVs) requires labeled "aerial photograph–satellite image" pairs spanning seasonal, weather, and altitude variability for the same location. Manual collection of such pairs is labor-intensive and does not scale, while simulated datasets (CARLA, GTA V) carry over a residual domain gap when transferred to real imagery.

An analytical review of publications at the intersection of controlled diffusion generation, UAV aerial photography, and CVGL has been conducted. The material is organized into five areas: synthetic datasets, CVGL methods, diffusion models for remote sensing, structural control of generation, and iterative optimization and quality assessment, and compared across five axes in a comparative matrix.

The review shows that simulated data loses 10–30% accuracy when transferred to real imagery, naive synthetic augmentation without filtering degrades results, the distribution-based FID metric correlates weakly with Recall@1, and scalar quality scoring in iterative pipelines fails to diagnose defect type and requires four to five iterations to converge.

A research gap map is constructed: none of the reviewed works combines controlled diffusion generation, multi-component per-image quality assessment, and validation against CVGL metrics. The following remain open problems: adapting diffusion generation to the specifics of UAV aerial imagery, a quality metric calibrated against CVGL performance, targeted defect diagnosis during model operation, and closing the "generation–assessment–optimization" loop on CVGL metrics.

РЕЗЮМЕ

М. В. Близно

Управляемая диффузионная генерация синтетических аэрофотоснимков БПЛА для систем перекрестной геолокализации: аналитический обзор

Обучение систем перекрестной геолокализации (CVGL) беспилотных летательных аппаратов требует размеченных пар «аэрофотоснимок – спутниковое изображение», охватывающих сезонную, погодную и высотную изменчивость одной и той же локации. Ручной сбор таких пар трудозатратен и не масштабируется, а симуляционные наборы (CARLA, GTA V) переносят на реальные снимки остаточный разрыв доменов.

Выполнен аналитический обзор публикаций на пересечении управляемой диффузионной генерации, аэросъемки БПЛА и CVGL. Материал систематизирован по пяти направлениям: синтетические наборы данных, методы CVGL, диффузионные модели для дистанционного зондирования, структурный контроль генерации, итеративная оптимизация и оценка качества, и сопоставлен по пяти осям в сравнительной матрице.

Показано, что симуляционные данные теряют 10–30 % точности при переносе на реальные снимки, наивная синтетическая аугментация без фильтрации ухудшает результаты, распределенческая метрика FID слабо коррелирует с Recall@1, а скалярная оценка качества в итеративных схемах не диагностирует тип дефекта и требует четырёх-пяти итераций до сходимости.

Построена карта пробелов исследований: ни одна из рассмотренных работ не объединяет управляемую диффузионную генерацию, многокомпонентную оценку качества каждого изображения и валидацию по метрикам CVGL. Открытыми остаются адаптация диффузионной генерации к специфике аэросъёмки БПЛА, метрика качества, калибруемая по CVGL эффективности, адресная диагностика дефектов на этапе работы модели и замыкание контура «генерация – оценка – оптимизация» на метриках CVGL.

Близно Максим Витальевич – младший научный сотрудник лаборатории интеллектуальных систем и анализа данных, ФГБНУ «Институт проблем искусственного интеллекта», 283048, г. Донецк, ул. Артема, д. 118 б, bmv.ipai@mail.ru. *Область научных интересов:* системы распознавания образов, нейронные сети, генеративные модели. ORCID 0009-0006-5212-6395

Blizno Maxim Vitalievich – Junior Researcher, Laboratory of Intelligent Systems and Data Analysis, Federal State Budgetary Scientific Institution "Institute of Artificial Intelligence Problems," 118b Artema St., Donetsk, 283048, Russia, bmv.ipai@mail.ru. Research interests: pattern recognition systems, neural networks, generative models. ORCID 0009-0006-5212-6395.

Статья поступила в редакцию 13.04.2026;
одобрена после рецензирования 07.05.2026; принята к публикации 02.09.2026
The article was submitted 13.04.2026; approved after reviewing 07.05.2026;
accepted for publication 02.09.2026.