Студенческий научный электронный журнал
StudArctic Forum

ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Т. 11, № 2. С. 24–32 **78-я Международная научная конференция обучающихся и молодых учёных** **2026**

Научная статья
УДК 004.93

ПАВЛОВ
Максим Павлович

магистратура, Петрозаводский государственный университет (Петрозаводск, Россия),
maksim_pavlov_2003@list.ru

КОНТРОЛИРУЕМЫЕ ТРАНСФОРМАЦИИ ИЗОБРАЖЕНИЙ ДЛЯ РАСШИРЕНИЯ ОБУЧАЮЩИХ ВЫБОРОК МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Научный руководитель:
Рогов Александр Александрович
Рецензент:
Корзун Дмитрий Жоржевич
Статья поступила: 17.05.2026;
Принята к публикации: 27.06.2026;
Размещена в сети: 27.06.2026.

Аннотация. Ручная подготовка изображений для задач компьютерного зрения трудоёмка и дорога, что ограничивает обучающую выборку модели ИИ и ведёт к её переобучению, неустойчивости к смене условий съёмки. В работе систематизированы методы контролируемых трансформаций изображений, дополняющих обучающую выборку, и их влияние на качество и безопасность моделей. Рассмотренный подход повышает точность, обобщающую способность и устойчивость к состязательным атакам. Рассмотрен переход к генеративному синтезу сложных помех, позволяющий разрабатывать надёжные системы, работающие при редких искажениях или явлениях.

Ключевые слова: аугментация изображений, контролируемые трансформации, обучающая выборка, устойчивость моделей, состязательная защита, генеративные модели

Для цитирования: Павлов М. П. Контролируемые трансформации изображений для расширения обучающих выборок моделей искусственного интеллекта в задачах компьютерного зрения // StudArctic Forum. 2026. Т. 11, № 2. С. 24–32.

Современные системы компьютерного зрения, основанные на глубоких нейронных сетях, достигли высоких результатов в решении задач классификации, сегментации и обнаружения объектов. Однако это стало возможным во многом благодаря наличию больших, тщательно размеченных наборов данных. Ручная разметка изображений для задач компьютерного зрения – трудоёмкий и дорогостоящий процесс, из-за чего во многих областях обучающие выборки остаются ограниченными. Следствием этого становятся переобучение моделей искусственного интеллекта и их неустойчивость к смене условий съёмки. Контролируемое расширение данных путём аугментации (применения контролируемых трансформаций над изображением, сохраняющих семантику) признано эффективным способом борьбы с указанными проблемами [Shorten].

В современной литературе предложен широкий спектр аугментаций: от геометрических и цветовых трансформаций до регуляризационных методов, таких как методы Cutout [DeVries], MixUp [Zhang], CutMix [Yun] и автоматизированных политик (например, политика AutoAugment) [Cubuk, 2019]. Программная реализация многих из них сосредоточена в библиотеке Albumentations [Buslaev]. Отдельное внимание при контролируемых трансформациях уделяется и вопросам безопасности. Доказано, что целенаправленные состязательные возмущения на изображениях (например, изменение яркости в определенных участках или пары точечных пикселей) способны обманывать нейронные сети [Goodfellow], а состязательное обучение с использованием аугментированных примеров повышает устойчивость моделей [Madry].

Цель данной работы – систематизировать методы контролируемых трансформаций изображений, оценить их влияние на качество и безопасность моделей искусственного интеллекта и обосновать переход к генеративному синтезу сложных помех. Для достижения цели решаются следующие задачи:

1) классифицировать основные типы аугментаций, что необходимо для формирования единой таксономии и понимания природы вносимых искажений;
2) проанализировать их влияние на точность, обобщающую способность и робастность, чтобы

выявить оптимальные методы и алгоритмы расширения данных для прикладных задач;

3) рассмотреть аспекты безопасности, включая защиту от состязательных атак, поскольку это критически важно для внедрения моделей в реальные системы на практике;
4) выявить возможности генеративных моделей для создания реалистичных, трудноформализуемых искажений, что позволит расширять обучающие выборки в условиях задач с "голодом" данных и редко возможных на практике.

Методологическую основу исследования составляет обзорно-аналитический подход, опирающийся на сравнительное изучение опубликованных экспериментальных исследований за период 2017–2024 годов. Отбор источников осуществлялся по критериям релевантности теме, наличия эмпирической валидации и публикации в рецензируемых изданиях.

* * * * *

Существующие методы аугментации изображений удобно рассматривать, объединяя их в четыре большие группы (рисунок 1): геометрические преобразования, цветовые и яркостные искажения, внесение шума и структурно-регуляризационные модификации. Такое деление отражает характер изменений, вносимых в исходные данные, и позволяет целенаправленно конструировать конвейеры расширения выборки под конкретные задачи. Однако в последнее время с развитием технологий появилось ещё одно, пятое направление контролируемых трансформаций изображений – генеративное.

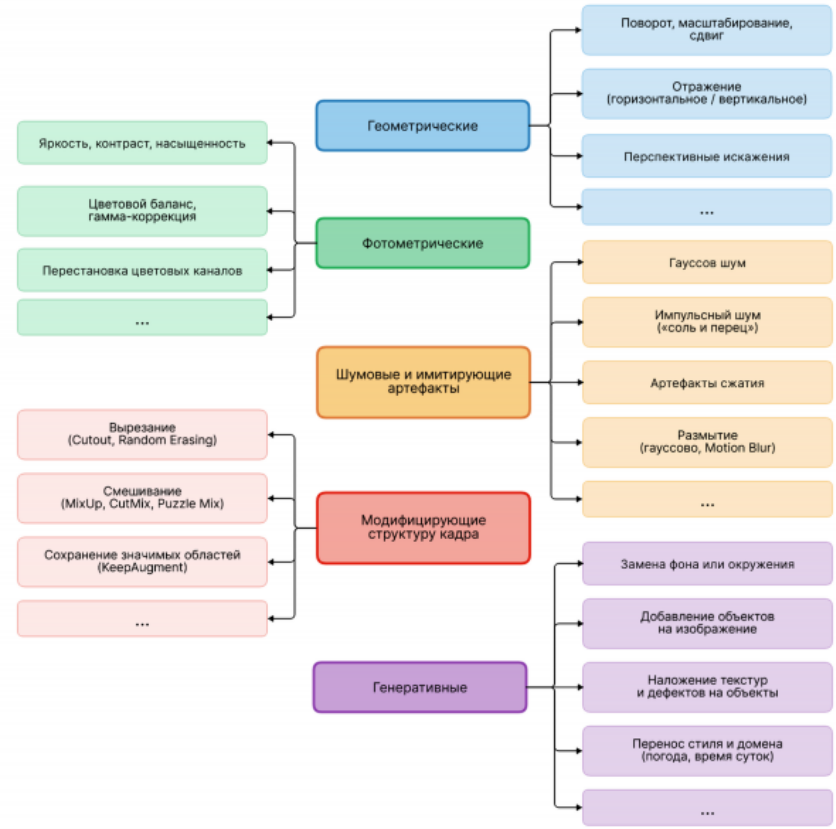

Рис. 1. Таксономия методов контролируемой аугментации изображений

Геометрические преобразования имитируют изменение положения, ракурса и масштаба объектов в кадре. К ним относятся поворот на случайный угол, горизонтальное и вертикальное отражение, масштабирование, сдвиг, аффинные и перспективные искажения, а также случайная обрезка с последующим масштабированием. Эти трансформации «дёшевы» с вычислительной точки зрения и широко применяются практически во всех задачах компьютерного зрения. Не меняя семантики сцены, они вынуждают модель вырабатывать инвариантность к пространственному положению и ориентации объектов [Shorten: 8].

Цветовые и яркостные искажения моделируют вариации освещённости и характеристик регистрирующей аппаратуры. Типичные операции включают регулировку яркости, контраста, насыщенности, сдвиг цветового баланса и гамма-коррекцию. Они позволяют нейронной сети привыкнуть к тому, что одни и те же объекты могут выглядеть по-разному в зависимости от условий съёмки, не разрушая при этом их идентичности. Однако важно соблюдать умеренность, так как

чрезмерное изменение цвета способно исказить семантически значимые свойства (например, сделать покраснения на коже синими) [Shorten: 9].

Внесение шума традиционно служит целям повышения устойчивости моделей к помехам, неизбежно возникающим при передаче, сжатии и съёмке изображений. Наиболее распространены аддитивный гауссов шум и импульсный шум типа «соль и перец». Модели, обученные с добавлением подобных искажений, демонстрируют большую устойчивость к низкокачественным входным данным, что особенно значимо в условиях плохой освещённости или при использовании недорогих камер [Shorten: 11].

Структурно-регуляризационные методы воздействуют на саму пространственную структуру изображения. Они заставляют модель опираться на глобальный контекст и множественные признаки, а не на отдельные локальные детали. Наиболее известные представители этой группы преобразований:

- *Метод Cutout* – случайное зануление (или замена шумом) прямоугольной области на изображении. Данный приём вынуждает сеть использовать информацию из уцелевших участков и препятствует тому, чтобы решение определялось единственным высокоактивированным фрагментом [DeVries].
- *Метод MixUp* – формирование нового обучающего примера путём линейной комбинации пикселей и соответствующих меток двух случайно выбранных изображений. Этот подход порождает гладкое поведение модели в пространстве между классами и снижает склонность к запоминанию отдельных примеров [Zhang].
- *Метод CutMix* – замена случайного фрагмента одного изображения фрагментом другого с пропорциональным смешиванием меток. В отличие от MixUp, здесь сохраняется пространственная структура, что положительно сказывается на способности модели локализовать объекты [Yun: 6023].

Отдельного упоминания заслуживают автоматические политики аугментации, в которых оптимальные комбинации и параметры преобразований подбираются алгоритмически для конкретного набора данных. Подход политики *AutoAugment* [Cubuk, 2019: 113] использует обучение с подкреплением для поиска наилучших последовательностей операций, а его развитие – стратегия *RandAugment* [Cubuk, 2020: 18613] заменяет трудоёмкий поиск случайным выбором из фиксированного набора преобразований, демонстрируя при этом сопоставимую эффективность и значительно снижая вычислительные затраты.

Применение контролируемых трансформаций приводит к ощутимому росту показателей качества. По обобщённым данным, на малых и средних выборках прирост точности классификации составляет от 5 до 15 процентных пунктов, а при использовании современных регуляризационных методов может достигать 17 % [Shorten: 14]. Механизмы улучшений предсказания моделей ИИ носят комплексный характер:

- геометрические трансформации формируют пространственную инвариантность;
- цветовые и яркостные искажения — фотометрическую инвариантность;
- структурные методы (методы CutMix, MixUp) препятствуют чрезмерной активации нейронов на локальных признаках, вынуждая сеть учитывать глобальный контекст [Yun: 6028].

Совместное действие этих групп преобразований порождает синергетический эффект: модель перестаёт запоминать нерелевантные детали фона и сохраняет точность в условиях, с которыми не сталкивалась при обучении [Shorten: 17]. Количественно это подтверждается результатами на специализированных наборах: например, внесение предметно-ориентированных артефактов в MedMNIST-C значительно повышает устойчивость предсказаний по сравнению с универсальными методами дополнения [Athalye]. Для эффективной реализации конвейеров аугментации на практике используются специализированные библиотеки. Наиболее распространённой является Albumentations — библиотека для Python с открытым исходным кодом, предоставляющая унифицированный интерфейс для построения конвейеров любой сложности [Buslaev: 125]. Её ключевая особенность — синхронное преобразование изображения и связанных целевых структур (масок сегментации, ограничивающих рамок, ключевых точек), что гарантирует сохранение пространственного соответствия. Библиотека интегрирована с фреймворками PyTorch и TensorFlow и оптимизирована под высокопроизводительную пакетную обработку.

Глубокие нейронные сети, несмотря на высокую точность, обладают фундаментальной уязвимостью: они чувствительны к малым, специально сконструированным возмущениям входных данных, которые практически незаметны для человека, но способны полностью изменить предсказание модели [Goodfellow]. Подобные образцы получили название состязательных примеров (adversarial

examples). По данным обзоров, даже модели класса архитектур ResNet и Vision Transformer демонстрируют падение точности на 80–100 % под воздействием атак FGSM, PGD и CW при отсутствии защиты [Ворона: 3–8]. Природа уязвимости связана с тем, что нейронные сети опираются на признаки, статистически значимые, но семантически несущественные. Состязательная атака эксплуатирует эту особенность, добавляя к изображению возмущение, согласованное с градиентом функции потерь [Goodfellow], в результате чего сеть с высокой уверенностью относит образец к неверному классу [Чехонина]. Наиболее действенным методом защиты признано состязательное обучение, при котором модель тренируется не только на чистых, но и на состязательно искажённых примерах [Madry]. Его суть заключается в решении минимаксной задачи: на каждой итерации алгоритм PGD генерирует возмущение, максимизирующее потери, а модель обучается минимизировать ошибку на зашумлённых данных. Такой подход обеспечивает устойчивость к широкому спектру атак, хотя и ценой некоторого снижения точности на чистых данных. Роль аугментации в этом контексте двойственна:

- некорректно спроектированные трансформации могут сместить распределение данных за пределы допустимой вариативности и скомпрометировать семантику;
- точно настроенное дополнение служит мощным регуляризатором, повышающим устойчивость модели, в том числе к намеренным атакам [Li: 929–930].

Метод AROID, автоматически подбирающий политики аугментации в процессе состязательного обучения, показал, что дополнение данных способно превзойти по эффективности специализированные защитные методы и быть совместимым с ними [Li]. Контролируемая аугментация, включая состязательное обучение, представляет собой не только инструмент повышения точности, но и необходимый элемент стратегии обеспечения информационной безопасности систем искусственного интеллекта [Герасимов: 53–60].

Классические параметрические искажения хорошо моделируют простые геометрические и фотометрические вариации. Однако реальные условия эксплуатации порождают помехи значительно более сложной структуры: переменную облачность, туман, локальные блики, артефакты сжатия, текстурные изменения, связанные с износом материалов. Сбор и разметка репрезентативной выборки для каждого из таких случаев трудоёмки, а порой и невозможны. Генеративные модели, прежде всего генеративно-состязательные сети (GAN) и диффузионные модели, позволяют синтезировать подобные искажения непосредственно на имеющихся снимках, не изменяя положения, формы и класса целевых объектов и сохраняя исходную разметку [Athalye].

- Архитектуры для непарного переноса стиля, такие как архитектура CycleGAN, переводят изображения из исходного домена (например, ясная погода) в целевой (туман, сумерки, подводная среда), оставляя неизменными силуэты и взаимное расположение предметов при корректной настройке структурных потерь [Zhu: 2223].
- Диффузионные модели при использовании механизмов дорисовывания по маске (inpainting) способны добавлять локальные артефакты — блики, царапины, пыль — точно в заданные области, не затрагивая семантически значимые детали [Dhariwal].

Практическая реализация такого подхода всё чаще опирается на локальные инструменты, не требующие передачи данных во внешние облачные сервисы, что особенно важно для соблюдения этических норм, запрещающих манипуляции, искажающие интерпретируемость данных. Генеративный синтез сложных помех позволяет имитировать реалистичные, трудноформализуемые условия съёмки без дорогостоящего сбора и разметки новых кадров. Модели, обученные с включением синтетических данных, приобретают устойчивость к широкому спектру редких и экстремальных искажений. Состязательное обучение снижает долю успешных атак с 80–100 % до 10–20 %, хотя и сопровождается снижением точности на чистых данных на 2–5 %. Совмещение с автоматическим подбором стратегий (AROID) позволяет частично компенсировать эти потери [Li].

Таблица 1

Сравнение групп методов контролируемых трансформаций изображений [Shorten]; [Yun]; [Li]

Группа методов	Преимущества	Ограничения	Применимость метода на практике	Вычислительная сложность	Эффект применения

Геометрические	Вычислительная эффективность, инвариантность к ракурсу.	Не моделируют сложные помехи.	Идеальны в качестве базового уровня для любых задач.	Низкая (O(1) для базовых операций)	Прирост точности на 5–15 %
Фотометрические	Устойчивость к изменению освещения.	Риск искажения семантики при избыточной интенсивности.	Применять с ограничением диапазона параметров	Низкая (попарное применение)	Прирост точности на 5–15 %
Шумовые	Устойчивость к артефактам съёмки.	Могут разрушать часть признаков.	Использовать умеренно и валидировать на тестовых данных	Низкая (аддитивный шум)	Повышение устойчивости к низкокачественным данным
Структурно-регуляризационные	Подавление переобучения, улучшение обобщения.	Требуют настройки под предметную область.	Комбинировать с другими группами преобразований.	Средняя (требует модификации обучающих примеров)	Повышение точности до 17 % на малых и средних выборках
Автоматические политики	Адаптивность, снижение ручной настройки.	Вычислительные затраты на поиск стратегии.	Начинать поиск с RandAugment как баланс эффективности и вычислительной стоимости.	Высокая (AutoAugment), Средняя (RandAugment)	Снижение точности на чистых данных на 2–5%, но повышение устойчивости
Генеративные	Моделирование сложных, трудноформализуемых искажений.	Требуют значительных ресурсов и ручной проверки семантики кадра после преобразования.	Для задач с редкими условиями или высокой стоимостью реальной съёмки, но высокой ценой ошибки.	Очень высокая (требует обучения GAN/диффузионных моделей)	Повышение устойчивости к широкому спектру редких и экстремальных искажений.

В отличие от приведённых выше результатов из литературы, далее представлен собственный эксперимент автора, направленный на сравнительную оценку влияния различных стратегий аугментации на точность и устойчивость модели классификации изображений на наборе данных CIFAR-10 с использованием архитектуры ResNet-18. Набор данных CIFAR-10 содержит 60000 изображений размером 32×32 пикселя, разделённых на 10 классов. Для обучения использовались 50000 изображений, для тестирования — 10000. Все эксперименты выполнялись в единой среде с фиксированными параметрами обучения и одинаковым разбиением данных. В качестве инструмента реализации конвейеров аугментации использовалась библиотека Albumentations. Обучение выполнялось с использованием оптимизатора AdamW (learning rate=1e−3, batch size=128). Максимальное число эпох составляло 30, применялась ранняя остановка (early stopping). Эксперименты проводились в среде Google Colab с использованием GPU NVIDIA T4. В исследовании сравнивались шесть групп контролируемых трансформаций изображений:

- геометрические;
- фотометрические;
- шумовые;
- структурно-регуляризационные;
- генеративно-подобные;
- комбинированные конвейеры.

Особое внимание уделялось корректности оценки качества. Аугментации применялись только к обучающей выборке, тогда как валидационный и тестовый наборы оставались неизменными. Такой подход исключает искусственное искажение метрик и соответствует общепринятой методологии оценки обобщающей способности моделей. Помимо стандартной точности (accuracy) на исходном тестовом наборе дополнительно проводилась оценка робастности модели. Для этого на этапе тестирования формировались отдельные выборки с контролируемыми искажениями: гауссовым шумом, размытием, туманом и изменением яркости. Итоговая метрика средней робастности вычислялась как среднее значение точности на всех искажённых выборках. Полученные результаты представлены в таблице 2.

Таблица 2

Влияние различных групп контролируемых трансформаций изображений на точность и устойчивость модели к шуму

	Точность (accuracy), %					
Метод аугментации	**Исходные данные**	**Шум**	**Размытие**	**Искусственный туман**	**Изменение яркости**	**Средняя устойчивость**
Базовая модель	85.28	65.92	46.45	37.37	83.72	58.12
Геометрические	90.60	60.61	59.50	56.05	88.88	66.26
Фотометрические	84.38	67.26	43.80	34.30	83.51	57.27
Шумовые	80.19	73.92	68.74	58.58	78.38	69.90
Структурные	84.63	66.90	43.95	36.30	82.83	57.50
Генеративные	85.05	75.24	72.85	68.30	83.33	74.93
Комбинированные	89.42	78.60	58.02	53.07	88.30	69.50

Результаты демонстрируют выраженный компромисс между максимальной точностью на исходных данных и устойчивостью к искажениям. Геометрические преобразования обеспечили наилучшую точность классификации на исходном наборе данных (90.60 %), однако их устойчивость к сложным деградациям изображения оказалась умеренной. Напротив, шумовые и генеративно-подобные аугментации несколько снижали качество на исходных данных, но существенно повышали устойчивость модели к туману, размытию и шумовым помехам. Наиболее сбалансированные результаты показали комбинированные стратегии аугментации, подтвердив наличие синергетического эффекта при совместном использовании нескольких типов преобразований. Вместе с тем генеративные аугментации повысили среднюю устойчивость на 16.81 % по сравнению с базовой моделью, что подтверждает перспективность направления генеративного синтеза сложных искажений.

Полученные результаты согласуются с выводами современных исследований о том, что классические геометрические и фотометрические преобразования преимущественно улучшают обобщающую способность, тогда как более сложные шумовые и генеративные методы повышают устойчивость моделей к реальным деградациям среды наблюдения. Для наглядного сопоставления точности и робастности моделей была построена диаграмма распределения результатов (рисунок 2).

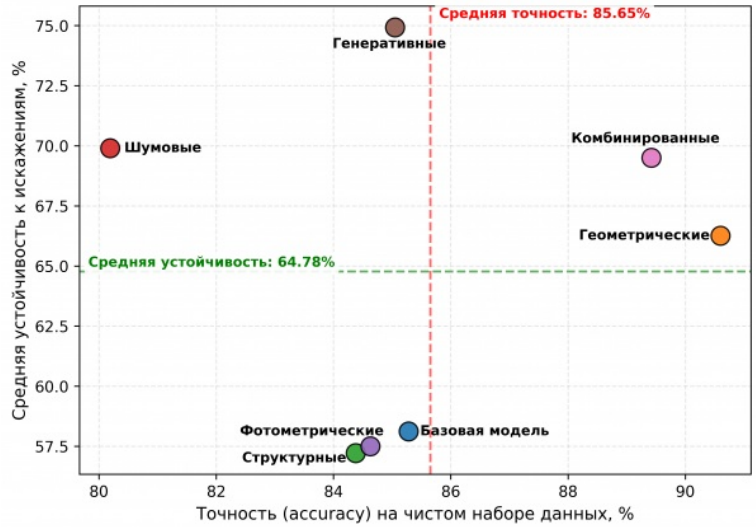

Рис. 2. Соотношение точности на исходных данных и средней устойчивости модели при различных стратегиях аугментации

На графике отчётливо наблюдается разделение методов на две группы: методы, ориентированные преимущественно на повышение точности на исходных данных (геометрические и комбинированные аугментации), и методы, направленные на повышение устойчивости к искажениям (шумовые и генеративно-подобные преобразования). Это подтверждает необходимость подбора стратегий

аугментации с учётом особенностей прикладной задачи и допустимого компромисса между точностью и надёжностью модели. Полный код экспериментов, конфигурации обучения и сценарии воспроизведения результатов опубликованы в открытом репозитории [1] для обеспечения воспроизводимости исследования. К ограничениям проведённого исследования следует отнести использование относительно компактного набора данных CIFAR-10 и одной архитектуры ResNet-18. Полученные результаты демонстрируют общие тенденции влияния аугментации, однако количественные показатели могут отличаться для более сложных наборов данных, задач детекции и сегментации, а также современных архитектур компьютерного зрения. Кроме того, в работе не рассматривались вычислительные затраты различных стратегий аугментации и их влияние на время обучения моделей.

* * * * *

В ходе исследования были последовательно решены поставленные задачи:

1) Классифицированы основные типы аугментаций. Сформирована единая таксономия методов контролируемых трансформаций, объединяющая геометрические, фотометрические, шумовые, структурно-регуляризационные и генеративные подходы.
2) Проанализировано влияние трансформаций на точность, обобщающую способность и устойчивость моделей. Установлено, что аугментация повышает точность классификации на ограниченных выборках на 5–17 %. Наибольший вклад в улучшение обобщения вносят структурно-регуляризационные приёмы (методы Cutout, MixUp, CutMix) и автоматические политики, тогда как шумовые и генеративные методы обеспечивают максимальный прирост устойчивости к реальным искажениям среды наблюдения.
3) Рассмотрены аспекты информационной безопасности, включая защиту от состязательных атак. Выявлена двойственная роль аугментации: некорректно спроектированный конвейер трансформаций может создавать новые уязвимости, тогда как состязательное обучение снижает успешность атак с 80–100 % до 10–20 %. Интеграция состязательного обучения с автоматическим подбором стратегий (метод AROID) позволяет компенсировать сопутствующее снижение точности на чистых данных на 2–5 %.
4) Выявлены возможности генеративных моделей для создания реалистичных, трудноформализуемых искажений. Показано, что в отличие от классических параметрических методов, архитектуры CycleGAN и диффузионные модели способны синтезировать сложные помехи (туман, блики) без нарушения семантической разметки и изменения геометрии целевых объектов.

Решение поставленных задач в совокупности позволило достичь главной цели исследования: провести систематизацию методов контролируемых трансформаций, оценить их влияние на качество и безопасность моделей, а также обосновать необходимость перехода к генеративному синтезу сложных помех. Научная новизна работы заключается в комплексной оценке влияния аугментации одновременно на точность, надёжность и устойчивость к состязательным воздействиям, а также в обосновании преимуществ генеративного синтеза перед классическими подходами (таблица 1). Дальнейшие исследования целесообразно сосредоточить на разработке семантически контролируемых генеративных конвейеров для воспроизводимого расширения выборок в экстремальных условиях эксплуатации.

Примечания

[1] Контролируемая аугментация изображений и устойчивость моделей компьютерного зрения // GitHub: сайт. URL: https://github.com/GhosT-FlexAgen/augmentation-robustness-study (дата обращения: 17.05.2026).

Список литературы

Ворона А.А. Методы повышения устойчивости нейронных сетей к состязательным атакам в системах компьютерного зрения / А.А. Ворона, Е.А. Севастей // Молодой ученый. 2026. № 7(610). С. 3–8.

Герасимов В.М. Защита от состязательных атак на аудио и изображения в моделях искусственного интеллекта с применением метода SGEC / В.М. Герасимов, М.А. Маслова, Э.И. Халилаева // Научный результат. Информационные технологии. 2023. Т. 8, № 2. С. 53–60. DOI: 10.18413/2518-1092-2022-8-2-0-7.

Чехонина Е.А. Обзор состязательных атак и методов защиты для детекторов объектов / Е.А. Чехонина, В.В. Костюмов // International Journal of Open Information Technologies. 2023. Т. 11, № 7. С. 11-20.

Athalye C. Domain-guided data augmentation for deep learning on medical imaging / C. Athalye, R. Arnaout // PLOS ONE. 2023. Vol. 18, No. 3. DOI: 10.1371/journal.pone.0282532

Buslaev A. Albumentations: Fast and flexible image augmentations / A. Buslaev, V. Iglovikov, E. Khvedchenya, et al. // Information. 2020. Vol. 11, No. 2. P. 125. DOI: 10.3390/info11020125

Cubuk E.D. AutoAugment: Learning augmentation policies from data / E.D. Cubuk, B. Zoph, J. Shlens, Q.V. Le // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 113–123. DOI:

10.1109/CVPR.2019.00020

Cubuk E.D. RandAugment: Practical automated data augmentation with a reduced search space / E.D. Cubuk, B. Zoph, J. Shlens, Q.V. Le // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA, 2020. P. 3008–3017. DOI: 10.1109/CVPRW50498.2020.00359

DeVries T. Improved regularization of convolutional neural networks with cutout / T. DeVries, G.W. Taylor // arXiv preprint. Cornell University, 2017. DOI: 10.48550/arXiv.1708.04552

Dhariwal P. Diffusion models beat GANs on image synthesis / P. Dhariwal, A. Nichol // Advances in Neural Information Processing Systems (NeurIPS). 2021. Vol. 34. P. 8780–8794.

Goodfellow I.J. Explaining and harnessing adversarial examples / I.J. Goodfellow, J. Shlens, C. Szegedy // Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, CA, USA, 2015.

Li L. AROID: Improving adversarial robustness through online instance-wise data augmentation / L. Li, J. Qiu, M. Spratling // International Journal of Computer Vision. 2024. Vol. 132, No. 2. P. 929–950. DOI: 10.1007/s11263-023-01912-9

Madry A. Towards deep learning models resistant to adversarial attacks / A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu // Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, BC, Canada, 2018. DOI: 10.48550/arXiv.1706.06083

Shorten C. A survey on Image Data Augmentation for Deep Learning / C. Shorten, T.M. Khoshgoftaar // Journal of Big Data. 2019. Vol. 6, No. 1. P. 60. DOI: 10.1186/s40537-019-0197-0

Yun S. CutMix: Regularization strategy to train strong classifiers with localizable features / S. Yun, D. Han, S.J. Oh, et al. // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 6023–6032. DOI: 10.1109/ICCV.2019.00612

Zhang H. Mixup: Beyond empirical risk minimization / H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz // Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, BC, Canada, 2018. DOI: 10.48550/arXiv.1710.09412

Zhu J.-Y. Unpaired image-to-image translation using cycle-consistent adversarial networks / J.-Y. Zhu, T. Park, P. Isola, A.A. Efros // Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. P. 2223–2232. DOI: 10.1109/ICCV.2017.244

Maksim P. PAVLOV — master's degree, Petrozavodsk State University (Petrozavodsk, Russia), *maksim_pavlov_2003@list.ru*

CONTROLLED IMAGE TRANSFORMATIONS FOR EXPANDING AI MODEL TRAINING SETS IN COMPUTER VISION

Scientific adviser:
Alexander A. Rogov
Reviewer:
Dmitry G. Korzun
Paper submitted on: 05/17/2026;
Accepted on: 06/27/2026;
Published online on: 06/27/2026.

Abstract. Manual preparation of images for computer vision tasks is labor-intensive and expensive, limiting an AI model's training set and leading to overfitting and instability under varying acquisition conditions. This work systematizes methods of controlled image transformations that augment the training set and examines their impact on model quality and security. The proposed approach improves accuracy, generalization ability, and robustness to adversarial attacks. The article also discusses transitioning to generative synthesis of complex noise, which enables the development of reliable systems operating under rare distortions or phenomena.
Keywords: image augmentation, controlled transformations, training set, model robustness, adversarial defense, generative models

For citation: Pavlov, M. P. Controlled Image Transformations for Expanding AI Model Training Sets in Computer Vision. StudArctic Forum. 2026, 11 (2): 24–32.

References

Vorona A.A., Sevastei E.A. Methods for increasing robustness of neural networks to adversarial attacks in computer vision systems. *Molodoy uchyoniy*, 2024, No. 7(610), pp. 3–8. (In Russ.)

Gerasimov V.M., Maslova M.A., et al. Protection against adversarial attacks on audio and images in artificial intelligence models using the SGEC method. *Research Result. Information Technologies* , 2023, Vol. 8, No. 2, pp. 53–60. DOI: 10.18413/2518-1092-2022-8-2-0-7 (In Russ.)

Chekhonina E.A., Kostyumov V.V. Overeview of adversarial attacks and defenses for object detectors. *International Journal of Open Information Technologies*, 2023, No. 7, pp. 11-20. (In Russ.)

Athalye C., Arnaout R. Domain-guided data augmentation for deep learning on medical imaging. *PLOS ONE*, 2023, Vol. 18, No. 3. DOI: 10.1371/journal.pone.0282532

Buslaev A., Iglovikov V., et al. Albumentations: Fast and flexible image augmentations. *Information*, 2020, Vol. 11, No. 2, p. 125. DOI: 10.3390/info11020125

Cubuk E.D., Zoph B., et al. AutoAugment: Learning augmentation policies from data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 113–123. DOI: 10.1109/CVPR.2019.00020

Cubuk E.D., Zoph B., et al. RandAugment: Practical automated data augmentation with a reduced search space. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA, USA, 2020, pp. 3008–3017. DOI: 10.1109/CVPRW50498.2020.00359

DeVries T., Taylor G.W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint*. Cornell University, 2017. DOI: 10.48550/arXiv.1708.04552

Dhariwal P., Nichol A. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021, Vol. 34, pp. 8780–8794.

Goodfellow I.J., Shlens J., et al. Explaining and harnessing adversarial examples. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. San Diego, CA, USA, 2015.

Li L., Qiu J., et al. AROID: Improving adversarial robustness through online instance-wise data augmentation. *International Journal of Computer Vision*, 2024, Vol. 132, No. 2, pp. 929–950. DOI: 10.1007/s11263-023-01912-9

Madry A., Makelov A., et al. Towards deep learning models resistant to adversarial attacks. *Proceedings of the 6th International Conference on Learning Representations (ICLR)*. Vancouver, BC, Canada, 2018. DOI: 10.48550/arXiv.1706.06083

Shorten C., Khoshgoftaar T.M. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 2019, Vol. 6, No. 1, p. 60. DOI: 10.1186/s40537-019-0197-0

Yun S., Han D., et al. CutMix: Regularization strategy to train strong classifiers with localizable features. *Proceedings o the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 6023–6032. DOI: 10.1109/ICCV.2019.00612

Zhang H., Cisse M., et al. Mixup: Beyond empirical risk minimization. *Proceedings of the 6th International Conference on Learning Representations (ICLR)*. Vancouver, BC, Canada, 2018. DOI: 10.48550/arXiv.1710.09412

Zhu J.-Y., Park T., et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2223–2232. DOI: 10.1109/ICCV.2017.244