

doi: 10.17586/2226-1494-2026-26-2-420-427

УДК 004.896

Применение гибридных методов искусственного интеллекта для практических производственных задач в условиях труднодоступности обучающих данных Глеб Витальевич Соловьев¹, Михаил Сергеевич Соколов²

^{1,2} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация¹ glebsolo46@gmail.com, <https://orcid.org/0009-0005-5479-2482>² mikhail.sokolov@itmo.ru, <https://orcid.org/0000-0003-4413-3676>

Аннотация

Введение. Современные производственные задачи, такие как контроль качества лазерной сварки и локализация геометрических признаков в промышленных процессах, требуют применения инновационных подходов машинного обучения. Недостаток размеченных данных и сложность геометрической аннотации являются критическими барьерами при разработке автоматизированных систем контроля. Научная новизна предлагаемого подхода заключается в комплексном использовании гибридных методов, объединяющих эволюционную оптимизацию, диффузионные модели и сверточные нейронные сети для эффективного решения практических инженерных задач с ограниченными ресурсами данных. **Метод.** Предложенный подход состоит из двух интегрированных компонентов. Первый компонент реализует гибридный алгоритм генерации синтетических данных, объединяющий эволюционную оптимизацию для генерации разнообразных геометрических вариантов с использованием диффузионных моделей для синтеза фотореалистичных изображений. Второй компонент включает специализированную архитектуру глубокого обучения, оптимизированную для точной локализации и классификации геометрических признаков в производственных контекстах. Обучение реализуется с применением комбинированной функции потерь, интегрирующей регрессионные и классификационные критерии. **Основные результаты.** На примере задачи контроля качества лазерной сварки синтетический набор данных расширен с 120 оригинальных изображений до 4537 реалистичных примеров, что позволило повысить точность сегментации швов с 2,4 до 0,75 по метрике потерь box loss . Для задачи локализации координат шва достигнута ошибка предсказания 31,8 пикселей по оси Y и 3,3 пикселя по оси X на исходном разрешении 1024×2448 пикселей. Экспериментальное сравнение показало превосходство сверточных архитектур над трансформерными моделями при сопоставимом количестве параметров, а также лучшую точность при регрессии с одного кадра, чем при использовании последовательности кадров. **Обсуждение.** Предложенные методы демонстрируют значительное превосходство по сравнению с классическими подходами масштабирования данных (mixup, cutmix) и чистыми подходами к диффузионному синтезу, которые требуют интенсивной подготовки наборов данных. Интеграция эволюционной оптимизации обеспечивает контролируемое разнообразие геометрических вариантов, а диффузионные модели гарантируют фотореалистичность синтезированных образцов. Данный гибридный подход имеет широкие перспективы применения в других промышленных секторах с ограниченной доступностью размеченных данных благодаря возможности построения полного конвейера синтеза труднодоступных промышленных данных, а затем их использования для обучения прикладных методов искусственного интеллекта, решающих целевые промышленные задачи.

Ключевые слова

гибридные методы машинного обучения, эволюционная оптимизация, диффузионные модели, синтетические данные, сверточные нейронные сети, локализация геометрических признаков, контроль качества производства, лазерная сварка, искусственный интеллект

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 25-11-00367, <https://rscf.ru/project/25-11-00367/>.

Ссылка для цитирования: Соловьев Г.В., Соколов М.С. Применение гибридных методов искусственного интеллекта для практических производственных задач в условиях труднодоступности обучающих данных // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 2. С. 420–427. doi: 10.17586/2226-1494-2026-26-2-420-427

© Соловьев Г.В., Соколов М.С., 2026

Application of hybrid artificial intelligence methods to practical industrial tasks under conditions of scarce training data

Gleb V. Solovev¹✉, Mikhail S. Sokolov²

^{1,2} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ glebsolo46@gmail.com✉, <https://orcid.org/0009-0005-5479-2482>

² mikhail.sokolov@itmo.ru, <https://orcid.org/0000-0003-4413-3676>

Abstract

Modern industrial tasks, such as quality control in laser welding and the localization of geometric features in industrial processes, require the application of innovative machine learning approaches. The scarcity of annotated data and the complexity of geometric annotation are critical barriers to the development of automated inspection systems. The scientific novelty of the proposed approach lies in the comprehensive use of hybrid methods, combining evolutionary optimization, diffusion models, and convolutional neural networks to effectively address practical engineering tasks with limited data resources. The proposed framework consists of two integrated components. The first component implements a hybrid algorithm for synthetic data generation, merging evolutionary optimization for generating diverse geometric variants with diffusion models for synthesizing photorealistic images. The second component involves a specialized deep learning architecture optimized for the precise localization and classification of geometric features in industrial settings. Training is performed using a combined loss function that integrates regression and classification criteria. In the case of laser welding quality control, the synthetic dataset was expanded from 120 original images to 4,537 realistic samples. This augmentation improved weld seam segmentation accuracy, reducing the box loss metric from 2.4 to 0.75. For the task of localizing weld seam coordinates, a prediction error of 31.8 pixels along the Y -axis and 3.3 pixels along the X -axis was achieved at the original resolution of $1,024 \times 2,448$ pixels. Experimental comparisons showed that convolutional architectures outperformed transformer-based models with a comparable number of parameters, and that regression from a single frame yielded higher accuracy than using a sequence of frames. The proposed methods demonstrate significant superiority over classical data augmentation techniques (e.g., mixup, cutmix) and pure diffusion-based synthesis approaches which require intensive dataset preparation. The integration of evolutionary optimization ensures controlled diversity in geometric variants, while diffusion models guarantee the photorealism of synthesized samples. This hybrid approach holds broad potential for application in other industrial sectors with limited availability of annotated data, owing to its capability to construct a complete pipeline for synthesizing hard-to-obtain industrial data and subsequently using it to train applied Artificial Intelligence methods for solving targeted industrial problems.

Keywords

hybrid machine learning methods, evolutionary optimization, diffusion models, synthetic data, convolutional neural networks, geometric feature localization, industrial quality control, laser welding, artificial intelligence

Acknowledgements

The study was supported by the grant No. 25-11-00367 from the Russian Science Foundation, <https://rscf.ru/project/25-11-00367/>.

For citation: Solovev G.V., Sokolov M.S. Application of hybrid artificial intelligence methods to practical industrial tasks under conditions of scarce training data. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 2, pp. 420–427 (in Russian). doi: 10.17586/2226-1494-2026-26-2-420-427

Введение

Современное производство сталкивается с вызовом, требующим балансировки между автоматизацией и доступностью данных для обучения систем искусственного интеллекта. В частности, задачи контроля качества в лазерной сварке, локализации критических геометрических признаков в производственных процессах требуют высокой точности и надежности, чего невозможно достичь без больших объемов качественно размеченных данных [1]. Однако накопление таких наборов данных в промышленности сопряжено с существенными трудозатратами и технологическими сложностями.

Традиционные подходы к расширению наборов данных, такие как геометрические трансформации и *mixup*-подобные методы [2, 3], неприменимы к задачам, где сохранение точной геометрии критично. Методы на основе диффузионных моделей, например *Stable Diffusion* [4], требуют тонкой подстройки на целевом домене, что при ограниченности данных может привести к галлюцинациям и искажениям [5].

Предлагаемый в настоящей работе гибридный подход объединяет сильные стороны эволюционной оптимизации для генерации контролируемого разнообразия геометрических вариантов и диффузионных моделей для синтеза фотореалистичных изображений. Исходное предположение базируется на том факте, что в промышленных приложениях, в частности при оценке качества сварных швов, критична как правильная геометрия объекта, так и его визуальное представление. Данный комплексный подход позволяет минимизировать задачу галлюцинаций через явное управление геометрией.

Кроме синтеза данных, выполнена разработка и валидация архитектуры глубокого обучения для решения практических промышленных задач. Практическая валидация методологии проведена на двух задачах: сегментация лазерных сварных швов с целью измерения их геометрических параметров; точная локализация координат лазерной головки сварочного оборудования.

Научная новизна предложенного подхода заключается в синергетическом объединении методов эволюционной оптимизации, генеративных моделей и архитектур глубокого обучения в единую производ-

ственно-ориентированную систему, где каждый компонент компенсирует недостатки других и усиливает их преимущества.

Методы

Предлагаемая комплексная методология состоит из двух взаимодополняющих компонентов. Гибридный алгоритм синтеза обучающих данных с контролируемой геометрией, который необходим для создания требуемого объема обучающих данных для методов компьютерного зрения. Специализированная архитектура нейронных сетей используется для анализа геометрических признаков и классификации промышленных дефектов. Данная комбинация обеспечивает эффективное решение задач с ограниченным объемом размеченных данных в промышленных условиях.

Основной вызов при разработке систем контроля качества — критический дефицит размеченных промышленных данных. Традиционные методы аугментации изображений неприменимы, так как нарушают точность геометрической информации, необходимой для анализа качества [6]. Предлагаемый подход решает эту проблему путем применения эволюционных операторов к координатам полигонов, описывающих форму объектов (рис. 1).

Начальные геометрические приближения извлекаются из реальных данных с использованием алгоритмов компьютерного зрения (библиотека OpenCV) [7]. К полигональным представлениям применяются следующие эволюционные операторы: мутация координат (смещение вершин полигона в случайном направлении на величину, контролируемую параметром эволюции); вставка точек (добавление новых вершин между существующими для создания более сложных геометрий); удаление точек (удаление вершин для получения упрощенных вариантов); кроссовер (обмен сегментами полигональных точек между двумя родительскими формами).

Для оптимизации используется алгоритм многокритериальной эволюционной оптимизации (Multi-objective Evolutionary Algorithm Based on Decomposition, MOEA/D) [8] с двумя целевыми функциями: минимизация расхождения Intersection over Union (IoU) [9] от оригинальной геометрии и максимизация Structural Similarity Index Measure (SSIM) [10] между синтезированными и реальными изображениями. Число итераций установлено на 25, размер популяции — 50 особей. Отбор особей производится согласно критериям: $0,6 \leq \text{IoU} \leq 0,9$ и $\text{SSIM} > 0,5$.

После получения набора эволюционно-оптимизированных полигональных масок применяется диффузионная модель Stable Diffusion для синтеза фотореалистичных изображений, сочетающих исходный стиль реальных изображений сварных швов с новой геометрией [11]. Процесс основан на техниках Style transfer [12], управляемой маской сварного шва. Данный подход минимизирует риск возникновения артефактов и галлюцинаций благодаря указанию конкретной геометрии сварного шва, близкой к реальной. Из 120 оригинальных изображений сгенерировано 5000 синтетических образцов, из которых экспертами отобрано 4537 визуально реалистичных изображений. Для достижения данных результатов использовались стандартные настройки параметров, кроме Denoising Strength в диапазоне [0,65; 0,85].

На созданном наборе синтетических данных можно обучать полноценные глубокие нейронные сети для задач компьютерного зрения. В рамках промышленных задач, рассматриваемых в настоящей работе, представлены примеры из области лазерной сварки, а именно сегментация поперечных разрезов сварных швов, и детекция геометрических координат лазерной головки в процессе выполнения сварки. Для задач сегментации используется You Only Look Once (YOLO) [13] — модель на основе сверточных нейронных сетей.

Для решения задач детекции положения сварочной лазерной головки в процессе сварки и корректировки

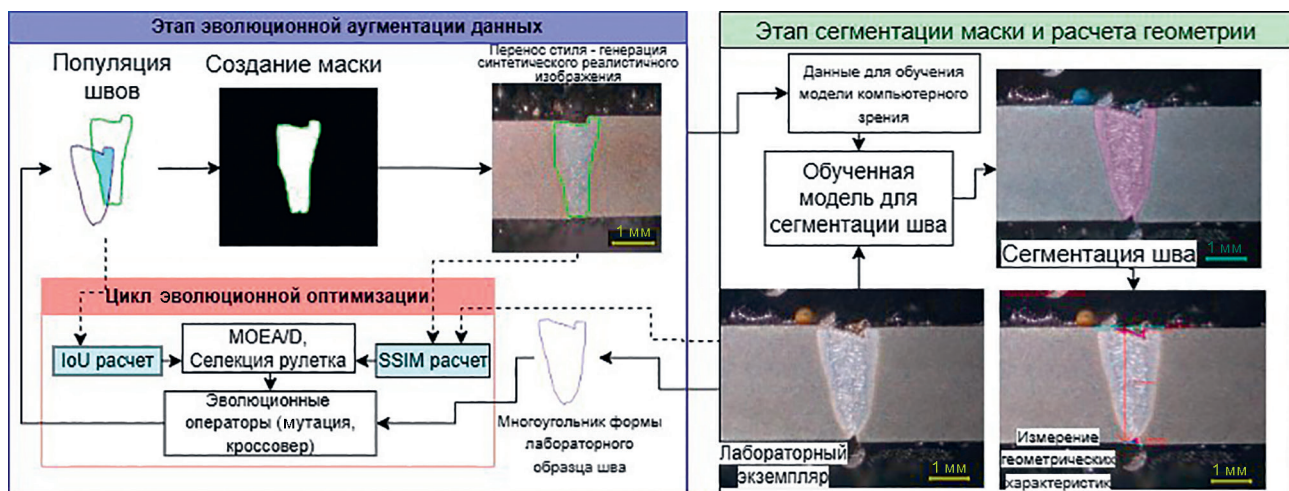


Рис. 1. Процесс обработки реальной фотографии сварного шва для создания синтетического набора данных, необходимого для обучения моделей компьютерного зрения

Fig. 1. Pipeline for processing a real weld-seam photograph to generate a synthetic dataset required for training computer vision models

ее положения была разработана комплексная нейронная сеть. Основной целью в данной задаче является детекция разрыва в лазерной разметке, где находится лазерная головка сварочного оборудования (зеленый крест, рис. 2). В процессе сварки вероятны ошибки оборудования и возможно нарушение траектории сварки. В результате была разработана автоматизированная система для исправления траектории сварки и повышения ее качества.

Предложенная архитектура нейронной сети состоит из следующих компонентов: backbone (базовая сеть) (сверточная нейронная сеть, преобразующая входное изображение в признаковое представление); глобальное усреднение (трансформация признакового представления в одномерный вектор); полносвязные слои (два линейных слоя, где финальный слой выдает три значения: координаты x , y координаты шва и оценка уверенности).

Рассматриваемая архитектура реализует гибридный подход к предсказанию: нормализованные координаты x , y шва в диапазоне $[0; 1]$ и вероятностная оценка наличия шва в кадре. Координаты определяются на основании значения пикселей изображения и масштаба изображения, позволяя реализовать систему обратной связи с точной передачей координат в метрической системе измерений.

Экспериментальные исследования

Сегментация сварных швов формулируется как классификация пикселей, принадлежащих телу шва. Реальные данные: 120 высокоточных изображений поперечных сечений лазерных сварных швов с ручной экспертной разметкой масок.

Для синтетических данных применялся гибридный конвейер: эволюционная оптимизация и Stable Diffusion. Stable Diffusion выполняла перенос стиля с исходных изображений швов на новую геометрию полигона, полученную эволюционной оптимизацией. Например (рис. 1), исходный полигон формы шва,

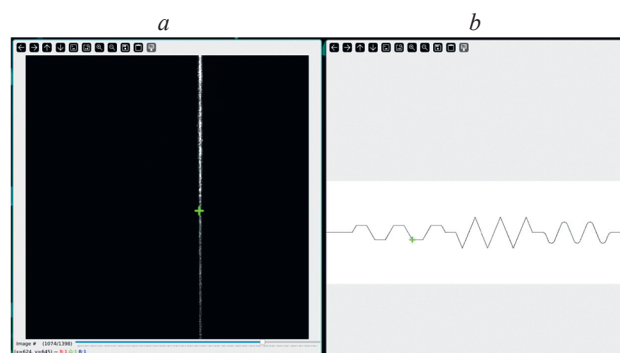


Рис. 2. Кадр с проекцией лазерной линии, перпендикулярной сварному шву, (а) и заданная траектория сварки (б)

Fig. 2. Frame showing the projection of a laser line perpendicular to the weld seam (a) and the predefined welding trajectory (b)

извлеченный с помощью OpenCV, поступает в цикл эволюционной оптимизации, где генерируется новый полигон, близкий к исходному, после чего по нему создается реалистичное синтетическое изображение.

Чтобы собрать набор данных из 4537 примеров, проведено обучение сегментационной модели YOLO на 120 реальных изображениях, что необходимо для автоматизированной разметки масок синтетических данных; без этого автоматизация метода невозможна.

Семейство полигонов получено за 1–3 шага эволюционной оптимизации, чтобы сохранить сходство с исходной формой и повысить реалистичность. Затем требовалась настройка гиперпараметров диффузионной модели, поскольку они влияют на уровень галлюцинаций. Как показано на рис. 3, после итеративной настройки удалось достичь $SSIM > 0,5$, что позволило генерировать реалистичные изображения, пригодные для дальнейшего обучения нейросети.

Для сравнения метода аугментации выполнено обучение архитектуры CycleGAN: на медицинских наборах данных она достигает $SSIM 0,70$ (магнитно-резонанс-

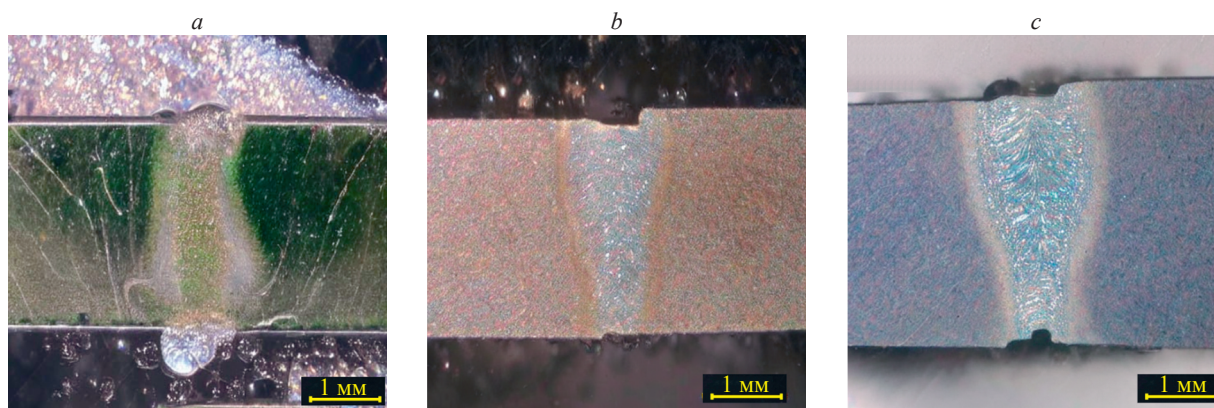


Рис. 3. Примеры создания синтетических изображений поперечных разрезов сварных швов при помощи разработанного метода на основе гибридизации эволюционных алгоритмов и диффузионных нейронных сетей. Результаты для дальнейшего обучения нейронной сети для реальной задачи: нереалистичный $SSIM = 0,43$ (а); реалистичные $SSIM = 0,68$ (б) и $SSIM = 0,83$ (с)

Fig. 3. Examples of synthetic image generation for weld cross-sections using the proposed method based on hybridizing evolutionary algorithms and diffusion neural networks. Results for further training of the neural network for mathematical problems: unrealistic $SSIM = 0.43$ (a); realistic $SSIM = 0.68$ (b) and $SSIM = 0.83$ (c)

ная томография, MRI) и 0,72 (компьютерная томография, CT) [14]. Однако при обучении на 120 реальных примерах из данного исследования не удалось получить ни одного сгенерированного изображения с SSIM > 0,5, что указывает на необходимость применения гибридных методов, не зависящих от промышленных задач с труднодоступностью данных.

Вторая экспериментальная задача — точная локализация шва в видеопотоках: на каждом кадре предсказываются координаты (x, y) и бинарная переменная score (наличие шва). Метрика ошибки — евклидово расстояние между предсказанной точкой и истинной. Исследованы backbone-архитектуры: MobileNetV4 [15], LeViT [16], ShuffleNetV2 [17], EfficientViT [18]. Для сравнения обучены YOLOv11-pose [19] и ResNet18 с Conv3D слоями [20]. Обучение: PyTorch [21], NVIDIA GeForce RTX 2060, оптимизатор AdamW [22], скорость обучения lr = 0,0001, размер порции данных для одного шага обучения нейронной сети batch = 32. Функции потерь: Huber loss для координат [23] и функция бинарной кросс-энтропии (binary cross-entropy) для переменной score.

Основные результаты

Применение синтетических данных привело к значительному улучшению качества: без синтетических данных (обучение только на 120 оригинальных изображениях): потеря метрики сегментации контура достигала минимум box loss = 2,4 [24], однако при обучении YOLO с синтетическими данными (на 4537 примерах) удалось снизить ошибку сегментации сварного шва до box loss = 0,75. Итоговое улучшение качества сегментации сварного шва представляет собой снижение потери на 68,75 %.

Результат демонстрирует эффективность гибридного подхода и подтверждает, что синтетические данные с контролем геометрии могут адекватно замещать натурные образцы при обучении моделей глубокого обучения. Применение предложенного метода может быть практически не ограничено к другим промышленным задачам, где требуется решать задачу компьютерного

зрения. Практически во всех подобных областях существуют проблемы с данными, их сборе в необходимом количестве, качестве и разнообразии.

В таблице приведены результаты обучения нескольких моделей с разными backbone, результаты показывают, что лучшей моделью для данной задачи является MobileNetV4.

Предложенный подход к определению координат сварного шва по разрыву лазерной линии показал лучшие результаты в сравнении с классическим методом (поиск минимума по одномерному профилю интенсивности — ошибка 78,3 пикселей по оси Y), детектором ключевых точек YOLOv11-pose-nano (ошибка 39,5 пикселей по оси Y) и 3D ResNet18 (ошибка 42,1 пикселей по оси Y). На рис. 4 видно, что ошибка между предсказанными и реальными значениями координат нахождения сварочной головки лазера минимальна, что говорит о возможной автоматизированной корректировке положения оборудования в зависимости от запланированного маршрута сварки. Однако, для исключения шумов, необходима постобработка выхо-

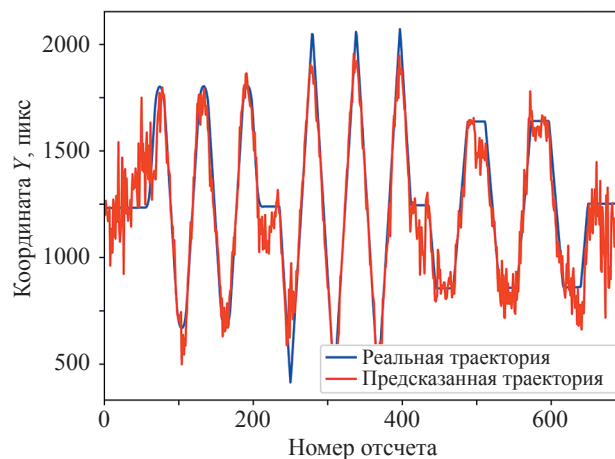


Рис. 4. Пример предсказанного и реального маршрутов сварочного процесса

Fig. 4. Example of the predicted and ground-truth welding process paths

Таблица. Результат оценки различных моделей для задачи предсказания положения лазерной головки в процессе сварки
Table. Evaluation results of different models for the task of predicting the laser head position during welding

Модель	Разрешение, пиксели	Средняя абсолютная ошибка (MAE) по горизонтальной оси X, пикс/мм	MAE по вертикальной оси Y, пикс/мм
ShuffleNet	320 × 768	4,1/0,17	44,5/1,87
ShuffleNet	640 × 1536	3,9/0,16	35,6/1,50
MobileNetV4	320 × 768	3,9/0,16	42,1/1,76
MobileNetV4	640 × 1536	3,3/0,14	31,8/1,33
EfficientViT	320 × 768	4,5/0,19	46,4/1,95
EfficientViT	640 × 1536	4,2/0,18	45,3/1,90
LeViT	320 × 768	4,4/0,18	46,2/1,94
LeViT	640 × 1536	4,1/0,17	41,0/1,72
YOLOv11-pose-nano	320 × 640	3,4/0,14	39,5/1,66
3D ResNet18	320 × 768	3,5/0,14	42,1/1,77

дов нейронной сети, например фильтрация с помощью фильтра Калмана [25] или фильтра скользящего среднего. Большая разница между ошибками вдоль оси X и оси Y может объясняться соотношением сторон исходного кадра (1024×2448 пикселей), а также шумной и неоднородной разметкой.

Обсуждение

Исследование подтверждает жизнеспособность комплексной методологии для промышленных задач при критически малом объеме размеченных данных. Ключевое преимущество перед классической аугментацией (mixup, cutmix, геометрические трансформации) — сохранение и контролируемое варьирование точной геометрии, критичной для измерений и контроля качества.

Включение эволюционных алгоритмов в синтез данных снижает типичную проблему «чистых» диффузионных моделей — артефакты и геометрические искажения при малых наборах данных. Управление формой через эволюционную оптимизацию полигонов уменьшает риск «галлюцинаций» и обеспечивает разнообразие в реалистичных для домена сценариях. МОЕА/D балансирует геометрическую близость (IoU) и визуальное качество (SSIM), давая компромисс между достоверностью и разнообразием синтетики.

Сравнение архитектур показало: для локализации геометрических признаков сверточная нейронная сеть превосходит трансформеры при сопоставимой вычислительной сложности. Метод регрессии координат сварного шва оказался лучше YOLOv11-pose-nano, а обработка одного кадра дала меньшую ошибку, чем последовательности кадров с 3D-сверткой.

Ограничение работы гибридного метода — валидация на домене лазерной сварки. Перенос на другие задачи (дефектоскопия, контроль износа, мониторинг линий) потребует настройки гиперпараметров эволюционной оптимизации и диффузионных моделей, но модульная архитектура и разделение управления геометрией и синтеза изображений упрощают адаптацию.

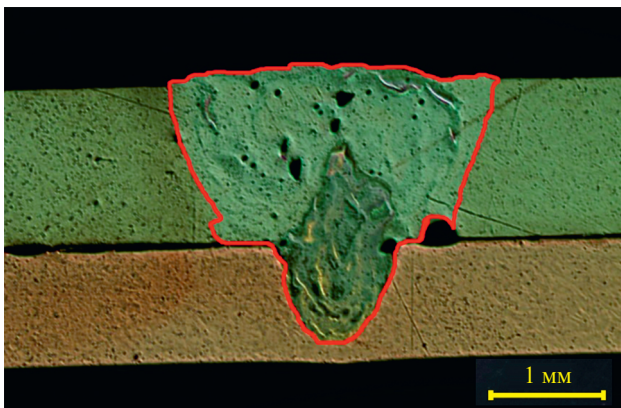


Рис. 5. Пример адаптации разработанного метода к альтернативному типу сварки

Fig. 5. Example of adapting the method to an alternative welding process

Практическая применимость метода подтверждена точностью: абсолютная ошибка локализации составила 31,8 пиксела по оси Y и 3,3 пиксела по оси X при разрешении 1024×2448 пикселей, что соответствует допускам систем автоматического позиционирования и коррекции траектории сварки. Контролируемый синтез данных делает возможным внедрение компьютерного зрения там, где сбор размеченных наборов данных экономически или технически затруднен.

Одним из важных результатов стоит отметить масштабируемость и обобщаемость результатов. Так, например, предложенный анализ сварных швов может быть применен и при работе с другими материалами, и при другой конфигурации сварочного процесса, и конфигурации свариваемых деталей. Например, на примере сварки меди и алюминия алгоритм также смог выделить необходимый контур шва, что необходимо для создания семейства масок швов и генерации новых примеров, как показано на рис. 5. Заметим, что подобные типы швов являются нетривиальными, и не могут быть сформированы при помощи алгоритмического построения сплайна, что затрудняет еще больше аугментацию подобных типов данных без применения методов описанных в данной работе.

Заключение

В работе предложена и экспериментально валидирована методология применения гибридных методов искусственного интеллекта для практических производственных задач с ограниченной доступностью обучающих данных. Синергетическое объединение эволюционной оптимизации полигональных представлений, диффузионных моделей синтеза изображений и специализированных архитектур глубокого обучения обеспечивает создание полного конвейера — от генерации труднодоступных данных до обучения прикладных моделей компьютерного зрения.

Экспериментальная апробация на задачах контроля качества лазерной сварки продемонстрировала практическую эффективность подхода: расширение исходного набора данных в 37 раз (от 120 до 4537 примеров) обеспечило снижение ошибки сегментации швов на 68,75 % (box loss от 2,4 до 0,75). Для задачи высокоточной локализации координат лазерной головки достигнута ошибка 31,8 пикселей по вертикальной оси Y и 3,3 пиксела по горизонтальной оси X на исходном разрешении 1024×2448 пикселей (0,13 и 1,33 мм ошибки соответственно при разрешении изображения 600 точек на дюйм), и максимальные значения 0,23 мм по оси X и 1,87 мм по оси Y , что удовлетворяет требованиям промышленных систем автоматической коррекции траектории сварки.

Предложенная методология открывает новые возможности для развития автоматизированных систем технического зрения в секторах промышленности, сталкивающихся с дефицитом размеченных данных: от микроэлектроники и аддитивного производства до медицинской диагностики и робототехники. Направления дальнейших исследований включают систематическую адаптацию конвейера синтеза к дополнительным про-

изводственным доменам, оптимизацию баланса вычислительных затрат и качества генерации, а также разработку методов активного обучения для минимизации

необходимого объема экспертной разметки при инициализации системы.

Литература

- Xiao S., Liu Z., Yan Z., Wang M. Grad-MobileNet: A gradient-based unsupervised learning method for laser welding surface defect classification // *Sensors*. 2023. V. 23. N 9. P. 4563. <https://doi.org/10.3390/s23094563>
- Vu T., Sun B., Yuan B., Ngai A., Li Y., Frahm J.-M. Supervision interpolation via lossmix: generalizing mixup for object detection and beyond // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024. V. 38. N 6. P. 5280–5288. <https://doi.org/10.1609/aaai.v38i6.28335>
- Jin X., Zhu H., Li S., Wang Z., Liu Z., Tian J., et al. A survey on mixup augmentations and beyond // *arXiv*. 2024. arXiv:2409.05202. <https://doi.org/10.48550/arXiv.2409.05202>
- Capogrosso L., Girella F., Taioli F., Chiara M., Aqeel M., Fummi F., et al. Diffusion-based image generation for in-distribution data augmentation in surface defect detection // *Proc. of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 2024. V. 2. P. 409–416. <https://doi.org/10.5220/0012350400003660>
- Tai Y., Yang K., Peng T., Huang Z., Zhang Z. Defect image sample generation with diffusion prior for steel surface defect recognition // *IEEE Transactions on Automation Science and Engineering*. 2024. V. 22. P. 8239–8251. <https://doi.org/10.1109/tase.2024.3482362>
- Alzarooni A., Iqbal E., Ullah Khan, S. Javed S., Moyo B., Abdulrahman Y. Anomaly detection for industrial applications: challenges, solutions, and future directions // *International Journal of Computer Integrated Manufacturing*. 2025. <https://doi.org/10.1080/0951192X.2025.2599548>
- Bradski G. The openCV library // *Dr. Dobb's Journal: Software Tools for the Professional Programmer*. 2000. V. 25. N 11. P. 120–123.
- Zhang Q., Li H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition // *IEEE Transactions on Evolutionary Computation*. 2007. V. 11. N 6. P. 712–731. <https://doi.org/10.1109/tevc.2007.892759>
- Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression // *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019. P. 658–666. <https://doi.org/10.1109/cvpr.2019.00075>
- Bakurov I., Buzzelli M., Schettini R., Castelli M., Vanneschi L. Structural similarity index (SSIM) revisited: a data-driven approach // *Expert Systems with Applications*. 2022. V. 189. P. 116087. <https://doi.org/10.1016/j.eswa.2021.116087>
- Zhang H., Liu Y., Yang J., Guo W., Wang X., Fua P. DiffAtlas: GenAI-Flying atlas segmentation via Image-Mask Diffusion // *Lecture Notes in Computer Science*. 2025. V. 15975. P. 161–172. https://doi.org/10.1007/978-3-032-05325-1_16
- Gatys L.A., Ecker A.S., Bethge M. Image style transfer using convolutional neural networks // *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 2414–2423. <https://doi.org/10.1109/CVPR.2016.265>
- Jiang P., Ergu D., Liu F., Cai Y., Ma B. A Review of Yolo algorithm developments // *Procedia Computer Science*. 2022. V. 199. P. 1066–1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- Chen Y., Lin H., Zhang W., Chen W., Zhou Z., Heidari A.A., et al. ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation // *Biomedical Signal Processing and Control*. 2024. V. 92. P. 106100. <https://doi.org/10.1016/j.bspc.2024.106100>
- Qin D., Leichner C., Delakis M., Fornoni M., Luo S., Yang F., et al. MobileNetV4: universal models for the mobile ecosystem // *Lecture Notes in Computer Science*. 2025. V. 15098. P. 78–96. https://doi.org/10.1007/978-3-031-73661-2_5
- Graham B., El-Nouby A., Touvron H., Stock P., Joulin A., Jégou H., et al. LeViT: a vision transformer in convnet's clothing for faster inference // *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021. P. 12239–12249. <https://doi.org/10.1109/ICCV48922.2021.01204>

References

- Xiao S., Liu Z., Yan Z., Wang M. Grad-MobileNet: A gradient-based unsupervised learning method for laser welding surface defect classification. *Sensors*, 2023, vol. 23, no. 9, pp. 4563. <https://doi.org/10.3390/s23094563>
- Vu T., Sun B., Yuan B., Ngai A., Li Y., Frahm J.-M. Supervision interpolation via lossmix: generalizing mixup for object detection and beyond. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 6, pp. 5280–5288. <https://doi.org/10.1609/aaai.v38i6.28335>
- Jin X., Zhu H., Li S., Wang Z., Liu Z., Tian J., et al. A survey on mixup augmentations and beyond. *arXiv*, 2024. arXiv:2409.05202. <https://doi.org/10.48550/arXiv.2409.05202>
- Capogrosso L., Girella F., Taioli F., Chiara M., Aqeel M., Fummi F., et al. Diffusion-based image generation for in-distribution data augmentation in surface defect detection. *Proc. of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2024, vol. 2, pp. 409–416. <https://doi.org/10.5220/0012350400003660>
- Tai Y., Yang K., Peng T., Huang Z., Zhang Z. Defect image sample generation with diffusion prior for steel surface defect recognition. *IEEE Transactions on Automation Science and Engineering*, 2024, vol. 22, pp. 8239–8251. <https://doi.org/10.1109/tase.2024.3482362>
- Alzarooni A., Iqbal E., Ullah Khan, S. Javed S., Moyo B., Abdulrahman Y. Anomaly detection for industrial applications: challenges, solutions, and future directions. *International Journal of Computer Integrated Manufacturing*, 2025, <https://doi.org/10.1080/0951192X.2025.2599548>
- Bradski G. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 2000, vol. 25, no. 11, pp. 120–123.
- Zhang Q., Li H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 2007, vol. 11, no. 6, pp. 712–731. <https://doi.org/10.1109/tevc.2007.892759>
- Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 658–666. <https://doi.org/10.1109/cvpr.2019.00075>
- Bakurov I., Buzzelli M., Schettini R., Castelli M., Vanneschi L. Structural similarity index (SSIM) revisited: a data-driven approach. *Expert Systems with Applications*, 2022, vol. 189, pp. 116087. <https://doi.org/10.1016/j.eswa.2021.116087>
- Zhang H., Liu Y., Yang J., Guo W., Wang X., Fua P. DiffAtlas: GenAI-Flying atlas segmentation via Image-Mask Diffusion. *Lecture Notes in Computer Science*, 2025, vol. 15975, pp. 161–172. https://doi.org/10.1007/978-3-032-05325-1_16
- Gatys L.A., Ecker A.S., Bethge M. Image style transfer using convolutional neural networks. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414–2423. <https://doi.org/10.1109/CVPR.2016.265>
- Jiang P., Ergu D., Liu F., Cai Y., Ma B. A Review of Yolo algorithm developments. *Procedia Computer Science*, 2022, vol. 199, pp. 1066–1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- Chen Y., Lin H., Zhang W., Chen W., Zhou Z., Heidari A.A., et al. ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation. *Biomedical Signal Processing and Control*, 2024, vol. 92, pp. 106100. <https://doi.org/10.1016/j.bspc.2024.106100>
- Qin D., Leichner C., Delakis M., Fornoni M., Luo S., Yang F., et al. MobileNetV4: universal models for the mobile ecosystem. *Lecture Notes in Computer Science*, 2025, vol. 15098, pp. 78–96. https://doi.org/10.1007/978-3-031-73661-2_5
- Graham B., El-Nouby A., Touvron H., Stock P., Joulin A., Jégou H., et al. LeViT: a vision transformer in convnet's clothing for faster inference. *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 12239–12249. <https://doi.org/10.1109/ICCV48922.2021.01204>

17. Ma N., Zhang X., Zheng H.-T., Sun J. ShuffleNet V2: practical guidelines for efficient CNN architecture design // *Lecture Notes in Computer Science*. 2018. V. 11218. P. 122–138. https://doi.org/10.1007/978-3-030-01264-9_8
18. Cai H., Li J., Hu M., Gan C., Han S. EfficientViT: lightweight multi-scale attention for high-resolution dense prediction // *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023. P. 17256–17267. <https://doi.org/10.1109/ICCV51070.2023.01587>
19. Khanam R., Hussain M. YOLOv11: an overview of the key architectural enhancements // *arXiv*. 2024. arXiv:2410.17725. <https://doi.org/10.48550/arXiv.2410.17725>
20. Hara K., Kataoka H., Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition // *Proc. of the IEEE International Conference on Computer Vision Workshops (ICCVW)*. 2017. P. 3154–3160. <https://doi.org/10.1109/iccvw.2017.373>
21. Imambi S., Prakash K.B., Kanagachidambaresan G.R. PyTorch // *Programming with TensorFlow: Solution for Edge Computing Applications*. 2021. P. 87–104. https://doi.org/10.1007/978-3-030-57077-4_10
22. Loshchilov I., Hutter F. Fixing weight decay regularization in Adam // *arXiv*. 2017. arXiv:1711.05101v2.
23. Meyer G.P. An alternative probabilistic interpretation of the Huber loss // *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021. P. 5257–5265. <https://doi.org/10.1109/cvpr46437.2021.00522>
24. Huang P., Tian S., Su Y., Tan W., Dong Y., Xu W., et al. IA-CIOU: an improved IOU bounding box loss function for SAR ship target detection methods // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2024. V. 17. P. 10569–10582. <https://doi.org/10.1109/jstars.2024.3402540>
25. Wan E.A., Van Der Merwe R. The unscented Kalman filter for nonlinear estimation // *Proc. of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. 2000. P. 153–158. <https://doi.org/10.1109/ASSPCC.2000.882463>
17. Ma N., Zhang X., Zheng H.-T., Sun J. ShuffleNet V2: practical guidelines for efficient CNN architecture design. *Lecture Notes in Computer Science*, 2018, vol. 11218, pp. 122–138. https://doi.org/10.1007/978-3-030-01264-9_8
18. Cai H., Li J., Hu M., Gan C., Han S. EfficientViT: lightweight multi-scale attention for high-resolution dense prediction. *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 17256–17267. <https://doi.org/10.1109/ICCV51070.2023.01587>
19. Khanam R., Hussain M. YOLOv11: an overview of the key architectural enhancements. *arXiv*, 2024. arXiv:2410.17725. <https://doi.org/10.48550/arXiv.2410.17725>
20. Hara K., Kataoka H., Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition. *Proc. of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 3154–3160. <https://doi.org/10.1109/iccvw.2017.373>
21. Imambi S., Prakash K.B., Kanagachidambaresan G.R. PyTorch. *Programming with TensorFlow: Solution for Edge Computing Applications*, 2021, pp. 87–104. https://doi.org/10.1007/978-3-030-57077-4_10
22. Loshchilov I., Hutter F. Fixing weight decay regularization in Adam. *arXiv*, 2017. arXiv:1711.05101v2.
23. Meyer G.P. An alternative probabilistic interpretation of the Huber loss. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, P. 5257–5265. <https://doi.org/10.1109/cvpr46437.2021.00522>
24. Huang P., Tian S., Su Y., Tan W., Dong Y., Xu W., et al. IA-CIOU: an improved IOU bounding box loss function for SAR ship target detection methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, vol. 17, pp. 10569–10582. <https://doi.org/10.1109/jstars.2024.3402540>
25. Wan E.A., Van Der Merwe R. The unscented Kalman filter for nonlinear estimation. *Proc. of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, 2000, pp. 153–158. <https://doi.org/10.1109/ASSPCC.2000.882463>

Авторы

Соловьев Глеб Витальевич — аспирант, инженер-исследователь, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 59409934700](https://orcid.org/0009-0005-5479-2482), <https://orcid.org/0009-0005-5479-2482>, glebsolo46@gmail.com

Соколов Михаил Сергеевич — кандидат технических наук, ведущий научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56994304500](https://orcid.org/0000-0003-4413-3676), <https://orcid.org/0000-0003-4413-3676>, mikhail.sokolov@itmo.ru

Authors

Gleb V. Solovlev — PhD Student, Engineer Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 59409934700](https://orcid.org/0009-0005-5479-2482), <https://orcid.org/0009-0005-5479-2482>, glebsolo46@gmail.com

Mikhail S. Sokolov — PhD, Leading Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56994304500](https://orcid.org/0000-0003-4413-3676), <https://orcid.org/0000-0003-4413-3676>, mikhail.sokolov@itmo.ru

Статья поступила в редакцию 24.11.2025
Одобрена после рецензирования 22.01.2026
Принята к печати 22.03.2026

Received 24.11.2025
Approved after reviewing 22.01.2026
Accepted 22.03.2026



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»