

УДК 004.932.4

DOI 10.24412/2413-7383-2025-4-39-254-265

И. В. Чернядьев, Т. В. Ермоленко  
Федеральное государственное бюджетное научное учреждение  
«Институт проблем искусственного интеллекта» г. Донецк,  
283048, г. Донецк, ул. Артема, 118б

## УЛУЧШЕНИЕ ИЗОБРАЖЕНИЙ С ПРИМЕНЕНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ: АНАЛИЗ И ПЕРСПЕКТИВНЫЕ РАЗРАБОТКИ\*

I. V. Chernyadev, T. V. Yermolenko  
Federal State Budgetary Scientific Institution  
"Institute of Artificial Intelligence Problems" Donetsk,  
283048, Donetsk, Artyoma str., 118b

## IMAGE IMPROVEMENT USING DEEP LEARNING: ANALYSIS AND PROSPECTIVE DEVELOPMENTS

В статье рассматривается актуальная проблема улучшения качества изображений с использованием методов глубокого обучения, с акцентом на super-resolution, шумоподавление и устранение артефактов. Анализируются ключевые архитектуры на основе Transformer и сверточных нейронных сетей, их преимущества и недостатки, такие как высокая ресурсоемкость и узкая специализация. Предлагаются улучшения в виде рандомизированной аугментации данных для синтеза обучающей выборки и замены стандартной свертки на быструю свертку Фурье для захвата глобального контекста. Целью работы является повышение качества восстановления изображений без высоких вычислительных затрат.

**Ключевые слова:** super-resolution, нейронные сети, компьютерное зрение, глубокое обучение, обработка изображений.

The article discusses the current problem of improving image quality using deep learning methods, with a focus on super-resolution, noise reduction, and artifact removal. It analyzes key architectures based on transformers and convolutional neural networks, their advantages and disadvantages, such as high resource consumption and narrow specialization. The article proposes improvements in the form of randomized data augmentation for synthesizing the training set and replacing the standard convolution with Fast Fourier Convolution to capture global context. The goal of the work is to improve the quality of image restoration without high computational costs.

**Keywords:** super-resolution, neural networks, computer vision, deep learning, image processing.

---

\* Исследование проводится при поддержке гранта РНФ (проект № 25-21-20602).

## Введение

На сегодняшний день существует потребность в автоматизации технических процессов, связанных с улучшением качества изображений, в том числе, с повышением разрешения и сохранением детализации. Изображения низкого разрешения часто страдают от таких явлений, как артефакты, шум, размытие, что ограничивает их практическую ценность.

Задача повышения качества изображений, включая super-resolution, приобретает все большую актуальность ввиду высокого роста объемов цифрового контента и требований к его качеству в различных отраслях. В спутниковой съемке методы повышения разрешения востребованы для мониторинга сельского хозяйства, реагирования на стихийные бедствия, где низкокачественные изображения могут быть преобразованы в высокодетализированные для более эффективного анализа. В индустрии видеоигр super-resolution оптимизирует производительность графических процессоров, позволяя рендерить кадры в низком разрешении с последующим улучшением до высокого качества.

В отличие от классических алгоритмов, нейронные сети обучаются адаптироваться к разнообразным искажениям, демонстрируя превосходство в задачах super-resolution, шумоподавления, восстановления утраченных деталей. Однако, несмотря на прогресс, остаются нерешенные проблемы. Во-первых, большинство моделей требуют огромных вычислительных ресурсов и объемов данных для обучения. Во-вторых, существующие алгоритмы часто демонстрируют нестабильность при работе с изображениями, содержащими сложные комбинации искажений, например, шум, размытие, низкая освещенность. В-третьих, вопросы обеспечения устойчивости моделей к артефактам генерации остаются открытыми. Вышеперечисленное дает простор для новых исследований и улучшения существующих прикладных решений.

**Целью данной работы** является определение основных проблем в задаче улучшения качества изображений, а также в выявлении ключевых особенностей применения глубокого обучения в этой предметной области. В работе были поставлены следующие задачи:

1. Провести анализ основных подходов к улучшению качества изображений.
2. Выявить основные проблемы в данной задаче.
3. Выделить ключевые особенности применения глубокого обучения к улучшению качества изображений и предложить методы устранения выявленных недостатков.

## Постановка задачи улучшения качества изображений

Обозначим  $I_{LR} \in \mathbb{R}^{h \times w \times c}$  – входное изображение низкого разрешения (LR) с высотой  $h$ , шириной  $w$  и количеством каналов  $c$ ;  $I_{HR} \in \mathbb{R}^{H \times W \times c}$  – выходное изображение высокого разрешения (HR) с высотой  $H$  и шириной  $W$ .

В качестве входных данных для задач улучшения качества изображений, как правило, используются пары изображений  $I_{HR}$  и  $I_{LR}$  для обучения моделей в режиме с учителем. Выходные данные представлены в виде изображений высокого разрешения с восстановленными деталями, сниженным шумом и минимизированными артефактами, обычно в формате с увеличенным количеством пикселей. Задачи в улучшении качества изображений условно можно разделить на два типа:

1. Super-resolution с сохранением деталей. Принцип работы состоит в том, чтобы на основе LR-изображения сгенерировать HR-версию, максимально близкую к реальному высококачественному аналогу, с акцентом на восстановление текстур, границ и мелких элементов.

2. Комплексное устранение искажений. По входному LR-изображению с комбинацией шумов, размытия и других артефактов модель стремится восстановить чистую HR-версию, учитывая разнообразные сценарии реальных искажений.

## 1 Архитектуры глубоких сетей, используемые в super-resolution

### 1.1 Анализ глубоких архитектур для задачи повышения дискретизации изображений

Повышение разрешения является распространенной процедурой в цифровой обработке изображений и видео. Она выполняется всякий раз, когда осуществляется масштабирование изображения на экране мобильного устройства или воспроизведение видеопотока с исходным разрешением 480p на дисплее с разрешением 1080p. Традиционные методы повышения дискретизации не ориентированы на получение высокоточных предсказаний внешнего вида изображения при увеличенном разрешении. Их основное назначение заключается в обеспечении корректного отображения визуальных данных после масштабирования. Подобные алгоритмы применяются, например, при необходимости адаптации изображения к размерам экрана монитора или при подготовке его к печати, когда требуется изменение пространственного разрешения. Несмотря на то, что такое повышение разрешения технически увеличивает количество пикселей, это по-прежнему картинка с низкой детализацией, которая просто просматривается в более высоком формате с некоторым сглаживанием. В данной работе основное внимание уделяется анализу эффективности методов повышения дискретизации изображений с использованием моделей глубоких нейронных сетей.

Среди существующих архитектурных подходов особое внимание привлекают модели SRCNN, SRResNet, SRGAN, а также решения, основанные на архитектуре Transformer. Каждая из них характеризуется специфическими особенностями и принципами построения, определяющими эффективность восстановления высокочастотных деталей изображения. Несмотря на хорошие результаты, эти подходы сталкиваются с различными ограничениями: от проблем масштабируемости и вычислительной сложности до артефактов генерации и зависимости от больших объемов обучающих данных. Ниже приведено описание технологических особенностей существующих решений, а также основные недостатки этих архитектур:

– SRCNN (*Super-Resolution Convolutional Neural Network*). Ключевым преимуществом SRCNN является ее простота – сеть содержит всего три сверточных слоя, что обеспечивает быстроту обработки. Эксперименты показывают превосходство над бикубической интерполяцией [1]. Из недостатков стоит отметить, что структура модели не позволяет захватывать сложные иерархические особенности изображений, что проявляется в размытии текстур при масштабировании [2].

– SRResNet (*Super-Resolution Residual Network*) представляет собой глубокую нейронную сеть, специально разработанную для сверхвысокого разрешения. Она включает в свою архитектуру остаточные блоки и пропускные соединения из архитектуры ResNet, позволяющие эффективно обучать глубокие нейронные сети, избегая проблемы затухающих градиентов [3]. К недостаткам нейросети можно отнести однородную обработку всех регионов, независимо от их сложности, приводит к недостаточной детализации текстур в областях с высокой вариативностью, тогда как однородные зоны восстанавливаются лучше. В результате изображения, особенно в местах с высокой детализацией, теряют четкость и выглядят размытыми.

– SRGAN (*Super-Resolution Generative Adversarial Network*) является инновационной архитектурой, сочетающей super-resolution и генеративные состязательные сети. Генератор пытается создавать изображения с увеличенным разрешением, а дискриминатор обучается отличать реальные от сгенерированных [4]. Модель способна создавать изображения с более естественными текстурами и деталями, что делает результаты более приятными для человеческого глаза по сравнению с предыдущими методами [5]. К недостаткам можно отнести: 1) чрезмерный акцент на текстурах создает ложные элементы на изображении, которых могло не быть в оригинале; 2) нестабильность: если одна из сетей начинает доминировать, это может привести к дисбалансу в обучении. Если дискриминатор становится слишком эффективным, он подавляет генератор, что приводит к замедлению или даже полному прекращению его обучения. Если же доминирует генератор, дискриминатор теряет способность направлять обучение, и качество выходных изображений резко падает.

– TTSR (*Texture Transformer Super-Resolution*) представляет собой архитектуру, основанную на Transformer. TTSR предлагает альтернативный подход к задачам компьютерного зрения, отказываясь от сверточных нейронных сетей (*Convolutional Neural Network, CNN*) в пользу Transformer, которые обычно используются в задачах обработки естественного языка. Использование механизмов внимания позволяет глобально учитывать взаимосвязи между различными частями изображения, что полезно для восстановления текстур и деталей [6]. Однако, TTSR требует значительных вычислительных ресурсов. К тому же, данная нейросеть, по сравнению с CNN, где свёрточные слои изначально ориентированы на локальные особенности, хуже восстанавливает такие компоненты, как тонкие линии, текстуры, волосы.

Исходя из вышеперечисленных методов, их достоинств и недостатков, можно сделать вывод, что каждая архитектура решает конкретные проблемы, но приносит новые ограничения, что подчеркивает узкую специализацию подходов. Например, SRGAN, отлично справляется с визуальной реалистичностью, но при этом порождает артефакты из-за стремления к максимальной детализации. Для улучшения качества изображения необходимо одновременно улавливать мелкие локальные детали и сохранять целостность изображения, учитывая глобальный контекст. CNN сильны в первом, Transformer во втором, GAN в имитации реализма, но ни один подход не охватывает оба аспекта полноценно. Каждый из методов имеет свои сильные и слабые стороны, что подчеркивает необходимость разработки новых подходов сочетающих достоинства различных архитектур для достижения оптимального баланса между качеством и вычислительной эффективностью.

Архитектуры для реконструкции, основанные на Transformer, обычно делят входное изображение на фрагменты фиксированного размера (патчи) и обрабатывают каждый патч независимо. Такой метод имеет один недостаток: граничные пиксели не могут использовать соседние пиксели, находящиеся вне патча для восстановления изображения, в следствии чего восстановленное изображение может содержать артефакты по краям вокруг каждого фрагмента. Эту проблему можно решить путем перекрытия патчей, но это приведет к дополнительной вычислительной нагрузке. Архитектура SwinIR, основанная на Swin Transformer [7] на данный момент является перспективным решением, поскольку объединяет в себе преимущества CNN и Transformer.

## 1.2 Аналитический обзор архитектуры SwinIR

SwinIR состоит из трёх модулей [8]: 1) извлечения поверхностных признаков, 2) извлечения глубоких признаков, 3) реконструкции изображений (рисунок 1).



Рисунок 1 – Модули архитектуры SwinIR

На вход нейросети подается изображение в низком разрешении LQ. Для извлечения поверхностных признаков  $F_{SF}$  используется сверточный слой  $L_{SF}$ :

$$F_{SF} = L_{SF}(LQ)$$

Затем извлекаются глубокие признаки  $F_{DF}$ :

$$F_{DF} = L_{DF}(F_{SF}),$$

где  $L_{DF}$  – модуль извлечения глубоких признаков, который содержит N блоков RSTB и сверточный слой. Промежуточные признаки  $T_1, \dots, T_N$  и выходной глубокий признак  $F_{DF}$  извлекаются по блокам:

$$F_{DF} = L_{conv} \left( L_{RSTB_i}(T_{i-1}) \right), \quad i = 1, 2, \dots, N,$$

где  $L_{RSTB_i}$  означает i-тый блок RSTB, а  $L_{conv}$  – сверточный слой в конце.

Поверхностные и глубокие признаки объединяются и подаются в модуль реконструкции  $F_{RHQ}$ :

$$F_{RHQ} = L_{REC}(F_{SF} + F_{DF}),$$

где  $L_{REC}$  – функция модуля реконструкции. Для реализации модуля реконструкции используется субпиксельный сверточный слой для повышения дискретизации изображения.

Как показано на рисунке 2, блок RSTB состоит из нескольких слоев *Swin Transformer Layer* (STL) и сверткой в конце.

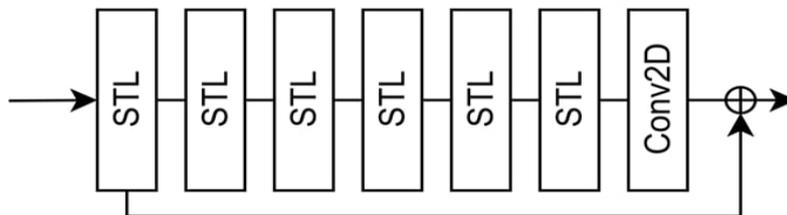


Рисунок 2 – Строение блока RSTB

Учитывая входной признак  $T_{i,0}$  i-го RSTB блока, сначала извлекаются промежуточные признаки  $T_{i,0}, T_{i,2}, \dots, T_{i,K}$  с помощью K слоёв STL. Выход RSTB выглядит следующим образом:

$$T_{i,out} = F_{conv_i} \left( F_{STL_{i,j}}(T_{i,j-1}) \right) + T_{i,0}, \quad j = 0, 1, \dots, K,$$

где  $F_{STL_{i,j}}$  – это j-й слой STL в i-том блоке RSTB,  $F_{conv_i}$  – сверточный слой в i-том RSTB. STL основан на стандартном механизме внимания из оригинального Transformer [9]. Изображения сначала разбиваются на неперекрывающиеся локальные окна, после чего вычисляется внутреннее внимание отдельно для каждого окна. Далее границы окон сдвигаются ровно на половину размера окна. Сдвинутое окно включает в себя части других, это создает связь между соседними областями, позволяя модели учитывать глобальный контекст для восстановления изображения, избегая при этом артефактов на границах.

К числу ограничений архитектуры SwinIR относятся недостаточная способность к реалистичному моделированию искажений, возникающих в реальных условиях формирования низкокачественных изображений, а также ограниченная эффективность в учёте глобального контекста сцены. Это обусловлено особенностями синтеза пар изображений с низким и высоким разрешением для обучения (LR-HR), которые не всегда отражают реальные виды искажений, встречающихся на практике. Кроме того, механизмы обработки информации в SwinIR преимущественно ориентированы на локальные области, что может привести к потере глобальных взаимосвязей и снижению достоверности восстановления сложных структур на изображении. Ниже изложены перспективы развития этой архитектуры и процедуры обучения для задачи super-resolution.

## 2 Актуальные подходы к совершенствованию архитектуры и процесса обучения

### 2.1 Методы синтеза изображений для формирования обучающих выборок

Для обучения модели используется набор данных, состоящий из пар изображений высокого разрешения (HR) и соответствующих низкоразрешённых (LR) вариантов, которые обычно синтезируются путём последовательного понижения дискретизации и добавления шума к исходным изображениям. Такой подход формирует однородно искажённый тренировочный набор, что не всегда адекватно отражает разнообразие повреждений, встречающихся в реальных условиях съёмки. Вследствие этого эффективность модели снижается, если реальные искажения не соответствуют тем, на которых она обучалась [10], [11]. Для решения этой проблемы в перспективном представляется рандомизированный подход к аугментации, заключающийся в случайном перемешивании различных видов искажений в процессе синтеза LR изображений (перемешиваемые размытия, интерполяции и шумы). В данном контексте следует отметить работу [12], в которой впервые введена концепция «mixture of augmentations» для задачи super-resolution. Авторы показали, что комбинирование и случайный выбор нескольких аугментаций приводит к устойчивому улучшению качества восстановления. Данный подход закрепил идею случайного смешения аугментаций как эффективной стратегии повышения обобщающей способности моделей super-resolution.

Методы данного подхода для генерации пар изображений с низким и высоким разрешением основаны на трёхступенчатой рандомизированной аугментации. Подобный рандомизированный синтез позволяет моделировать широкий спектр реальных искажений, что способствует лучшей обобщающей способности обучаемой модели. Предусмотрено последовательное применение (в случайном порядке и с рандомизацией параметров) следующих этапов искусственного искажения исходного изображения:

1. Комбинация изотропного и анизотропного размытия. Изотропное размытие имитирует расфокусировку, а анизотропное размытие движение камеры или объекта [13]. Их комбинация даёт широкий спектр реальных искажений.
2. Случайный выбор типа интерполяционного алгоритма (метода понижения дискретизации):
  - Метод К-ближайших соседей.
  - Билинейная интерполяция.
  - Бикубическая интерполяция.

Сначала изображение сильно уменьшается, а затем увеличивается до стандартного размера LR. Этот метод позволяет смоделировать артефакты, возникающие при многократном пережатии изображения [14].

3. Добавление шума со случайной вероятностью с вариацией его свойств и параметров:

- Гауссов шум. Уровень шума выбирается случайным образом.
- Шум JPEG-сжатия. Изображение сжимается с помощью JPEG со случайным коэффициентом качества.
- Шум Пуассона. Этот тип шума актуален для изображений, снятых в условиях низкой освещенности [15]. Уровень шума варьируется случайным образом, чтобы имитировать реальные фотографические сценарии.

Иллюстрация синтеза обучающих данных с помощью трёхступенчатой рандомизированной аугментации представлена на рисунке 3. Приведённые результаты демонстрируют внешний вид синтетических изображений LR, формируемых из исходных HR посредством случайных комбинаций размытия, интерполяции и добавления шума. Подобная визуализация позволяет наглядно оценить реализуемое разнообразие искусственных искажений, приближённых к реальным условиям формирования низкокачественных данных.



Рисунок 3 – Пример синтетического LR-изображения, полученного в результате трёхэтапной аугментации (справа) из исходного HR-изображения (слева)

Дальнейшим развитием подхода рандомизированного синтеза искажений является внедрение формализованной трёхэтапной схемы с вероятностной композицией и параметризацией операций. Такая организация процесса синтеза позволяет моделировать широкий спектр возможных искажений, характерных для реальных условий формирования низкоразрешённых изображений, и способствует повышению обобщающих свойств обучаемой модели по сравнению с традиционными методами генерации обучающих данных.

## 2.2 Интеграция блока быстрого преобразования Фурье в архитектуру SwinIR

В задачах восстановления изображений и повышения их качества с использованием архитектур на основе Transformer основное внимание уделяется преодолению разрыва между извлечением локальных признаков и интеграцией глобального

контекста. В частности, в модели SwinIR каждый *Swin Transformer Block* (RSTB) завершается стандартной сверточной операцией, что обеспечивает выделение локальных признаков благодаря ограниченному рецептивному полю. Однако в задаче super-resolution глобальный контекст, включая распределение цветов, повторяющиеся текстурные структуры и прочие интегративные характеристики, имеет существенное значение для получения изображений высокого качества без артефактов. Следует отметить, что стандартные сверточные операции, обладая ограниченным рецептивным полем, недостаточно эффективны для захвата глобального контекста, а попытки увеличить эффективное поле за счёт глубины сети ведут к росту вычислительной сложности [16].

Внедрение блока *Fast Fourier Convolution* (FFC) позволяет организовать обработку входных признаков в двух параллельных ветвях: локальной и глобальной. Локальная ветвь использует стандартные сверточные операции для анализа деталей в малых областях изображения. Глобальная ветвь применяет быстрое преобразование Фурье, переводя признаки в частотную область, где изображение описывается спектром коэффициентов, отражающих вклад различных пространственных частот. Низкочастотные компоненты характеризуют такие свойства, как общий цветовой фон и крупные объекты, а высокочастотные – детализацию, границы и текстурные элементы. Точечная свертка обновляет коэффициенты по отдельности, что влияет на всё изображение как единое целое. После обратного преобразования Фурье восстанавливается пространственное представление признаков.

Объединение обработанных признаков из локальной и глобальной ветвей формирует обогащённое представление, включающее как локальные детали, так и глобальный контекст [17]. Это объединение происходит перед передачей признаков на вход следующего блока RSTB или модуля реконструкции изображения. Измененный выход RSTB выглядит следующим образом:

$$T'_{i,out} = F_{FFC_i} \left( F_{STL_{i,j}}(T_{i,j-1}) \right) + T_{i,0}, \quad j = 0, 1, \dots, K,$$

где  $F_{FFC_i}$  – это быстрая свертка Фурье в  $i$ -том блоке RSTB, которая состоит из двух ветвей:

$$F_{FFC_i} = L_{conv} \left( F_{STL_{i,j}}(T_{i,j-1}) \right) + L_{global} \left( F_{STL_{i,j}}(T_{i,j-1}) \right), \quad j = 0, 1, \dots, K,$$

где  $L_{global}$  – глобальная ветвь,  $L_{conv}$  – локальная.

Исследования в смежных областях, таких как классификация [18], сегментация [19] и детекция объектов [20], [21], показывают, что использование FFC приводит к улучшению значений количественных метрик (PSNR, SSIM) на 5–10% по сравнению с моделями, в которых используются исключительно стандартные сверточные операции. На основании полученных результатов можно ожидать сопоставимого прироста качества и в задаче super-resolution.

## Заключение

В настоящей работе проведён сравнительный анализ современных архитектур, применяемых для повышения качества изображений в задаче super-resolution, включая SRCNN, SRResNet, SRGAN, TTSR и SwinIR. На основании проведённого анализа особенностей архитектур можно сделать вывод, что наиболее перспективным направлением дальнейшего развития является использование архитектуры SwinIR, обладающей высокой эффективностью и гибкостью.

Ключевые возможности для повышения эффективности такой архитектуры включают использование рандомизированной аугментации при формировании обучающих данных, что обеспечивает более реалистичное моделирование искажений, а также интеграцию блока FFC, позволяющую заменить традиционные сверточные слои и эффективно учитывать глобальный контекст изображения. Совокупное применение указанных подходов способствует улучшению качества реконструкции изображений при оптимальных вычислительных затратах, определяя потенциал дальнейшего совершенствования методов глубокого обучения в сфере повышения разрешения визуальных данных. Такие разработки не только повысят точность анализа изображений, но и расширят область их применимости в критически важных сферах, требующих анализа изображений высокого качества.

## Список литературы

1. Сай С.В. Способ оценки четкости фотореалистичных изображений с высоким разрешением // Компьютерная оптика. 2022. Т. 46, № 1. С. 121-129. ISSN 2412-6179. DOI: 10.18287/2412-6179-CO-899.
2. Коновалов, В.Ф. Единый нейросетевой метод повышения разрешения для разнородных цифровых изображений дистанционного зондирования Земли / В.Ф. Коновалов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. 2024. Т. 48, № 6. С. 944-955. DOI: 10.18287/2412-6179-CO-I610.
3. Singh P., Ganotra D. Brain MRT image super resolution using discrete cosine transform and convolutional neural network. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no. 4, pp. 734-742. doi: 10.17586/2226-1494-2023-23-4-734-742.
4. Сай С.В., Кудяшов А.А. Оценка качества изображений природного ландшафта с повышенным разрешением на основе GAN // Вестник ТОГУ. 2024. № 1 (72). С. 45-52. УДК 004.896.
5. Шелковникова, Т.Е. Применение генеративно-сопоставительных нейросетей для формирования баз данных в сканирующей туннельной микроскопии / Т.Е. Шелковникова, С.Ф. Егоров, П.В. Гуляев // Компьютерная оптика. 2023. Т. 47, № 2. С. 314-322. DOI: 10.18287/2412-6179-CO-1144.
6. Yang F., Yang H., Fu J., Lu H., Guo B. Learning Texture Transformer Network for Image Super-Resolution. 2020. P. 5790-5799. DOI: 10.1109/CVPR42600.2020.00583.
7. Liu, Z. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [Text] / Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo // arxiv preprint. 2021. arxiv:2103.14030.
8. Liang, J. SwinIR: Image Restoration Using Swin Transformer / J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte // arxiv preprint. 2021. arxiv:2108.10257.
9. Бережнов, Н.И. Совершенствование механизмов внимания для архитектуры трансформер в задачах повышения качества изображений / Н.И. Бережнов, А.А. Сирота // Компьютерная оптика. 2024. Т. 48, № 5. С. 726-733. DOI: 10.18287/2412-6179-CO-1393.
10. Ермоленко Т. В., Самородский И. Е. Анализ эффективности архитектур глубоких нейросетей для классификации // Журнал «Проблемы искусственного интеллекта». 2022. № 1(24). С. 54-64. ISSN 2413-7383.
11. Денисов А.К., Быковский С.В., Кустарев П.В. Метод увеличения разрешения изображения с использованием референсных изображений на основе диффузионной модели // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 321-327. doi: 10.17586/2226-1494-2025-25-2-321-327.
12. Yoo, J.S., Kim, M., & Kim, S.J. (2020). Rethinking Data Augmentation for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA, pp. 10735-10744. DOI: 10.1109/CVPR42600.2020.01075.
13. Чаганова О.Б., Григорьев А.С., Николаев Д.П., Николаев И.П. Прикладные аспекты современных методов неослепительного восстановления изображений [Текст] // Компьютерная оптика. – 2024. Т. 48, № 4. С. 562–572. ISSN 2412-6179. DOI: 10.18287/2412-6179-CO-1409.
14. Zhang X., Wu X. Dual-layer Image Compression via Adaptive Downsampling and Spatially Varying Upsampling. 2023. P. 1-10. DOI: arXiv:2302.06096.
15. Котенко, И. В. Метод противодействия состязательным атакам на системы классификации изображений [Текст] / И. В. Котенко, И. Б. Саенко, О. С. Лаута, Н. А. Васильев, В. Е. Садовников // Вопросы кибербезопасности. 2025. № 2(66). С. 114-123. ISSN 2311-3456. DOI: 10.21681/2311-3456-2025-2-114-123.

16. Пикалёв Я.С., Ермоленко Т.В. О нейронных архитектурах извлечения признаков для задачи распознавания объектов на устройствах с ограниченной вычислительной мощностью // Журнал «Проблемы искусственного интеллекта». 2023. № 3. С. 44-54. ISSN 2413-7383. DOI 10.34757/2413-7383.2023.30.3.004.
17. Чернядьев, И. В. Системный анализ программного обеспечения концептуальной нейросетевой модели системы классификации эмоций на изображениях / И. В. Чернядьев В. В. Бондарчук // Евразийский Союз Ученых. Серия: технические и физико-математические науки. 2024. № 08(123). С. 6-15. ISSN 2413-9335. DOI: 10.31618/ESU.2413-9335.
18. Rao, Y. Global Filter Networks for Image Classification / Y. Rao, W. Zhao, Z. Zhu, J. Lu, J. Zhou // arxiv preprint. 2021. arxiv:2107.00645.
19. Quattrini, F. Volumetric Fast Fourier Convolution for Detecting Ink on the Carbonized Herculaneum Papyri [Text] / F. Quattrini, V. Pippi, S. Cascianelli, R. Cucchiara // arxiv preprint. 2023. arxiv:2308.05070.
20. Gu, L. Fast Fourier Convolution Based Remote Sensor Image Object Detection for Earth Observation [Text] / L. Gu, R. Qiao, X. Chen // arxiv preprint. 2022. arxiv:2209.00551.
21. Miao W., Shen J., Xu Q., Hamalainen T., Xu Y., Cong F. SpikingYOLOX: Improved YOLOX Object Detection with Fast Fourier Convolution and Spiking Neural Networks / W. Miao, J. Shen, Q. Xu, T. Hamalainen, Y. Xu, F. Cong. 2025. Vol. 39, iss. 2. P. 1465-1473. DOI: 10.1609/aaai.v39i2.32137.

## References

1. Sai S.V. A method for assessing photorealistic image quality with high resolution [Text] // Computer Optics. – 2022. – Vol. 46, No. 1. – P. 121-129. – ISSN 2412-6179. – DOI: 10.18287/2412-6179-CO-899.
2. Konovalov, V.F. Unified Neural Network Method for Enhancing Resolution for Heterogeneous Digital Images of Earth Remote Sensing / V.F. Konovalov, V.V. Myasnikov, and V.V. Sergeev // Computer Optics. - 2024. - Vol. 48, No. 6. - Pp. 944-955. - DOI: 10.18287/2412-6179-CO-I610.
3. Singh P., Ganotra D. Brain MRT image super resolution using discrete cosine transform and convolutional neural network. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no. 4, pp. 734-742. doi: 10.17586/2226-1494-2023-23-4-734-742.
4. Sai S.V., Kudyashov A.A. Assessment of the quality of natural landscape images with increased resolution based on GAN [Text] // Bulletin of TOGU. – 2024. – No. 1 (72). – P. 45-52. – UDC 004.896.
5. Shelkovnikova TE, Egorov SF, Gulyaev PV. Application of generative adversarial neural networks for the formation of databases in scanning tunneling microscopy. Computer Optics 2023; 47(2): 314-322. DOI: 10.18287/2412-6179-CO-1144.
6. Yang F., Yang H., Fu J., Lu H., Guo B. Learning Texture Transformer Network for Image Super-Resolution / F. Yang, H. Yang, J. Fu, H. Lu, B. Guo. – 2020. – P. 5790-5799. – DOI: 10.1109/CVPR42600.2020.00583.
7. Liu, Z. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [Text] / Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo // arxiv preprint. – 2021. – arxiv:2103.14030.
8. Liang, J. SwinIR: Image Restoration Using Swin Transformer [Text] / J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte // arxiv preprint. – 2021. – arxiv:2108.10257.
9. Berezhnov NI, Sirota AA. Improving attention mechanisms in transformer architecture in image restoration. Computer Optics 2024; 48(5): 726-733. DOI: 10.18287/2412-6179-CO-1393.
10. Yermolenko T.V., Samorodsky I.Ye. Analysis of the efficiency of deep neural networks architectures for the classification of products images [Text]//Yermolenko T.V., Samorodsky I.Ye. Journal "Problems of Artificial Intelligence". – 2022. – № 1(24). – С. 54-64. – ISSN 2413-7383.
11. Denisov A.K., Bykovskii S.V., Kustarev P.V. Reference-based diffusion model for super-resolution. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2025, vol. 25, no. 2, pp. 321-327 (in Russian). doi: 10.17586/2226-1494-2025-25-2-321-327.
12. Yoo, J.S., Kim, M., & Kim, S.J. (2020). Rethinking Data Augmentation for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA, pp. 10735-10744. DOI: 10.1109/CVPR42600.2020.01075.
13. Chaganova O.B., Grigoryev A.S., Nikolaev D.P., Nikolaev I.P. Applied aspects of modern non-blind image deconvolution methods [Text] // Computer Optics. – 2024. – Vol. 48, no. 4. – P. 562–572. – ISSN 2412-6179. – DOI: 10.18287/2412-6179-CO-1409.
14. Zhang X., Wu X. Dual-layer Image Compression via Adaptive Downsampling and Spatially Varying Upconversion / X. Zhang, X. Wu. – 2023. – P. 1-10. – DOI: arXiv:2302.06096.

15. Kotenko, I. V. A Method for Countering Competitive Attacks on Image Classification Systems [Text] / I. V. Kotenko, I. B. Saenko, O. S. Lauta, N. A. Vasilyev, and V. E. Sadovnikov // *Cybersecurity Issues*. – 2025. – No. 2(66). – Pp. 114-123. – ISSN 2311-3456. – DOI: 10.21681/2311-3456-2025-2-114-123.
16. Pikalyov Ya.S., Yermolenko T.V. About Neural Architectures of Feature Extraction for the Problem of Object Recognition on Devices with Limited Computing Power [Text] // *Journal "Problems of Artificial Intelligence"*. – 2023. – № 3. – С. 44-54. – ISSN 2413-7383. – DOI 10.34757/2413-7383.2023.30.3.004.
17. Chernyadev, I. V. System Analysis of the Software of the Conceptual Neural Network Model of the Emotion Classification System in Images [Text] / I. V. Chernyadyev, V. V. Bondarchuk // *Eurasian Union of Scientists. Series: Technical and Physical and Mathematical Sciences*. – 2024. – No. 08(123). – Pp. 6-15. – ISSN 2413-9335. – DOI: 10.31618/ESU.2413-9335.
18. Rao, Y. Global Filter Networks for Image Classification [Text] / Y. Rao, W. Zhao, Z. Zhu, J. Lu, J. Zhou // *arxiv preprint*. – 2021. – arxiv:2107.00645.
19. Quattrini, F. Volumetric Fast Fourier Convolution for Detecting Ink on the Carbonized Herculaneum Papyri [Text] / F. Quattrini, V. Pippi, S. Cascianelli, R. Cucchiara // *arxiv preprint*. – 2023. – arxiv:2308.05070.
20. Gu, L. Fast Fourier Convolution Based Remote Sensor Image Object Detection for Earth Observation [Text] / L. Gu, R. Qiao, X. Chen // *arxiv preprint*. – 2022. – arxiv:2209.00551.
21. Miao W., Shen J., Xu Q., Hamalainen T., Xu Y., Cong F. SpikingYOLOX: Improved YOLOX Object Detection with Fast Fourier Convolution and Spiking Neural Networks / W. Miao, J. Shen, Q. Xu, T. Hamalainen, Y. Xu, F. Cong. – 2025. – Vol. 39, iss. 2. – P. 1465-1473. – DOI: 10.1609/aaai.v39i2.32137.

## RESUME

*I. V. Chernyadev, T. V. Yermolenko*  
*Image improvement using deep learning:*  
*analysis and prospective developments*

The article is devoted to the analysis and development of image quality improvement methods based on deep learning aimed at eliminating the disadvantages of existing solutions, such as artifacts, high resource intensity.

Improvements are proposed in the work: 1) randomized augmentation to create LR-HR pairs: a combination of isotropic and anisotropic blurring, random selection of interpolation; adding noise with varying levels and sequences. This simulates real-world distortions, increasing the generalization of the model; 2) replacing convolution with FFC, which divides features into local and global branches, which enhances the global context, improving the quality of image recovery. An increase in metrics is expected, as in related tasks.

As a result, the following conclusions can be drawn: 1) during neural network training, special attention should be paid to LR image synthesis methods, because they significantly affect the quality of the model; 2) Transformer-based architectures are a promising basis for further work on image quality improvement; 3) the proposed modifications will increase the efficiency and applicability of the neural network.

## РЕЗЮМЕ

*И. В. Чернядьев, Т. В. Ермоленко*  
*Улучшение изображений с применением*  
*глубокого обучения: анализ и перспективные разработки*

Статья посвящена анализу и разработке методов улучшения качества изображений на основе глубокого обучения, направленных на устранение недостатков существующих решений, таких как артефакты и высокая ресурсоемкость.

В работе предлагаются улучшения: 1) рандомизированная аугментация для создания пар LR-HR: комбинация изотропного и анизотропного размытия, случайный выбор понижения дискредитации, добавление шумов с вариацией уровней и

последовательности. Это моделирует реальные искажения, повышая обобщение модели. 2) замена свертки на FFC, разделяющий признаки на локальную и глобальную ветви, что усиливает глобальный контекст, повышая качество восстановления изображений. Ожидается прирост в метриках, как в смежных задачах.

Выводы: 1) во время обучения нейросети необходимо уделять особое внимание на методы синтеза LR изображений, потому что они значительно влияют на качество работы модели; 2) архитектуры, основанные на Transformer, являются перспективной основой для дальнейшей работы над повышением качества изображений; 3) предложенные модификации повысят эффективность и применимость нейросети.

**Чернядьев Иван Валерьевич** – инженер-исследователь ФГБНУ "Институт проблем искусственного интеллекта", бакалавр физико-технического факультета ФГБОУ ВО "ДонГУ". *Область научных интересов:* компьютерное зрение, машинное обучение, нейронные сети.

**Ермоленко Татьяна Владимировна** – к.т.н., научный сотрудник отдела системного анализа и интеллектуальных интерфейсов ФГБНУ "Институт проблем искусственного интеллекта", кандидат технических наук, доцент. *Область научных интересов:* машинное обучение, обработка естественного языка, распознавание образов, интеллектуальный анализ данных.

Статья поступила в редакцию 03.10.2025.