УДК 004.932 **DOI** 10.15622/ia.23.4.5

М.Н. ФАВОРСКАЯ, А.И. ПАХИРКА

ВОССТАНОВЛЕНИЕ АЭРОФОТОСНИМКОВ СВЕРХВЫСОКОГО РАЗРЕШЕНИЯ С УЧЕТОМ СЕМАНТИЧЕСКИХ ОСОБЕННОСТЕЙ

Фаворская М.Н., Пахирка А.И. **Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей.**

Аннотация. В настоящее время происходит активное развитие технологий обработки изображений дистанционного зондирования, включающих как спутниковые снимки, так и аэрофотоснимки, полученные от видеокамер беспилотных летательных аппаратов. Зачастую такие снимки имеют артефакты, связанные с низким разрешением, размытостью фрагментов изображения, наличием шумов и т.д. Одним из способов преодоления таких ограничений является применение современных технологий для восстановления снимков сверхвысокого разрешения на основе методов глубокого обучения. Особенностью аэрофотоснимков является представление текстуры и структурных элементов более высокого разрешения, чем на спутниковых снимках, что объективно способствует лучшим результатам восстановления. В статье приводится классификация методов сверхвысокого разрешения с учетом основных архитектур глубоких нейронных сетей, а именно сверточных нейронных сетей, визуальных трансформеров и генеративно-состязательных сетей. В статье предлагается метод восстановления аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей SemESRGAN за счет использования на этапе обучения дополнительной глубокой сети для семантической сегментации. При этом минимизируется общая функция потерь, включающая состязательные потери, потери на уровне пикселов и потери воспирятия (сходства признаков). Для экспериментов использовались шесть наборов аннотированных аэрофотоснимков и спутниковых снимков CLCD, DOTA, LEVIR-CD, UAVid, AAD и AID. Было выполнено сравнение результатов восстановления изображений предложенным методом SemESRGAN с базовыми архитектурами сверточных нейронных сетей, визуальных трансформеров и генеративно-состязательных сетей. Получены сравнительные результаты восстановления изображений с применением объективных метрик PSNR и SSIM, что позволило оценить качество восстановления с использованием различных моделей глубоких сетей.

Ключевые слова: аэрофотоснимки, сверхвысокое разрешение, семантическая сегментация, сверточные нейронные сети, визуальные трансформеры, генеративно-состязательные сети.

1. Введение. Задача восстановления изображений сверхвысокого разрешения относится к методам предварительной обработки изображений. До появления методов глубокого обучения задача решалась традиционными методами интерполяции, включая билинейную интерполяцию, бикубическую интерполяцию, интерполяцию Ланцоша (Lanczos) (для снимков дистанционного зондирования Земли) и т.д. Развитие методов глубокого обучения существенно повлияло на качество восстановления изображений сверхвысокого разрешения.

Существуют два подхода к восстановлению снимков сверхвысокого разрешения: подход на основе одного исходного изображения (Single Image Super-Resolution, SISR) и подход с использованием нескольких исходных изображений или кадров видеопоследовательности (Multi Image Super-Resolution, MISR). Преимущественно используется первый подход (SISR), не требующий одновременного получения нескольких исходных снимков низкого разрешения (Low Resolution, LR) при разных ракурсах съемки. Следует отметить, что обе постановки обратных задач являются некорректными с математической точки зрения, поскольку существует множество способов восстановления снимка сверхвысокого разрешения, близкого по объективным метрикам к реальному снимку высокого разрешения (High Resolution, HR) [1].

Известны четыре категории SISR методов: на основе интерполяции, реконструкции, обучения и преобразований [2]. В настоящее время методы на основе интерполяции используются в качестве предварительной обработки при восстановлении SR снимков (при необходимости). Основное развитие получили методы реконструкции, обучения и преобразований. Методы реконструкции основаны на применении априорной информации. Они широко используются в медицинских приложениях. Методы на основе обучения анализируют взаимосвязи между снимками высокого и низкого разрешения (так называемые парные снимки) из предварительно подготовленного набора данных. При этом методы глубокого обучения считаются гибридными методами, основанными на реконструкции и обучении. Алгоритмы на основе преобразований, использующих сети-транформеры, имеют более сложную структуру, включая модули самокалибровки, внимания, иерархического разделения фрагментов и т.д.

Для решения задачи восстановления SR снимков подходят два типа архитектур глубоких нейронных сетей – сверточные нейронные сети (СНС) и генеративно-состязательные сети (ГСС). Они основаны на разных принципах обучения и, соответственно, восстановления LR снимков. Исторически первыми появились СНС, и их использование для решения задачи SISR превалировало до 2021 года. В свою очередь первая архитектура ГСС были разработана в 2014 году [3], и в настоящее время ГСС демонстрируют лучшие результаты восстановления за счет более сложных, но и более эффективных стратегий обучений [4]. Тем не менее, применяются оба подхода. Сети-трансформеры были разработаны в 2017 году как относительно простой способ улучшения результатов в языковом переводе.

Визуальные трансформеры для SISR задач позволили переосмыслить архитектуру классических СНС. Одним из последних достижений является разработка SwinV2 трансформера для восстановления SR изображений [5].

SISR технологии применяются во многих областях, включая удаленное зондирование Земли [6], медицинскую диагностику [7], биометрию [8], видеонаблюдение [9], метеорологию [10] и т.д. Каждая область применения предъявляет свои требования к результатам визуальной восстановления, и поскольку выполнить их одновременно не представляется возможным, формируются модификации SISR методов, учитывающие основные особенности той или иной сферы применения.

В данном исследовании предлагается метод восстановления аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей на основе усложненного обучения нейронных сетей различных видов. Помимо базовой архитектуры дополнительно применяется глубокая сеть для семантической сегментации. При этом минимизируется общая функция потерь, включающая состязательные потери, потери на уровне пикселов и потери восприятия (сходства признаков). Были исследованы различные базовые архитектуры, в частности, модели SRCNN, ESRGAN, а также модели СНС для семантической сегментации. Для экспериментов использовались шесть наборов аннотированных аэрофотоснимков и спутниковых снимков CLCD, DOTA, LEVIR-CD, UAVid, AAD и AID, что позволило провести анализ полученных результатов на основе объективных метрик оценки восстановленных изображений.

Снимки дистанционного зондирования характеризуются следующими особенностями: они получены с большого расстояния, изображения содержат множество малоразмерных объектов, изображения сцены зависят от времени года, различных атмосферных условий и разной геометрии обзора датчиков. Спутниковые снимки, как правило, являются мультиспектральными или гиперспектральными в отличие от аэрофотоснимков, которые преимущественно формируются в оптическом диапазоне. Поэтому можно говорить о разных методах восстановления спутниковых снимков и аэрофотоснимков. Далее приведем краткий обзор существующих в настоящее время SISR методов для восстановления аэрофотоснимков сверхвысокого разрешения.

2. Обзор SISR методов для восстановления аэрофотоснимков сверхвысокого разрешения. В настоящее время в реализации SISR методов для задач дистанционного зондирования превалируют методы

на основе глубокого обучения. Однако следует отметить, что существуют и менее распространенные методы обучения, такие как встраивание соседей (neighbor embedding) или разреженное кодирование (sparse coding). Методы глубокого обучения для восстановления снимков сверхвысокого разрешения можно разделить на девять категорий по типам связей:

– линейные связи;
– остаточные связи;
– рекурсивные связи;
– связи, основанные на внимании;
– многопоточные соединения;
– соединения высокой плотности;
– связи, обрабатывающие множественные искажения, на основе так называемой технологии обучения с нулевым выстрелом (zero-shot learning);
– связи, используемые в ГСС;
– связи прогрессивной реконструкции.

Каждая из категорий представлена несколькими моделями глубоких сетей, среди которых имеются как широко известные модели, так и редко применяемые.

Помимо связей важны виды глубокого обучения (с учителем или без учителя) и типы архитектуры. В настоящее время лучшие результаты демонстрируют архитектуры, основанные на использовании наборов данных, состоящих из парных LR–HR снимков, а методы на основе накапливаемой при обучении статистики, преобразовании фрагментов (визуальные трансформеры) или словарей фрагментов требуют дальнейшего развития.

Рассмотрим более подробно тенденции развития архитектур СНС и ГСС как наиболее часто используемых для восстановления аэрофотоснимков сверхвысокого разрешения. В работе [11] используется мультимасштабное представление аэрофотоснимков на основе вейвлет-анализа. Каждое кратномасштабное LR представление (прямое вейвлет-преобразование) обрабатывается своей предварительно обученной СНС, а затем для получения SR представления выполняется вейвлет-синтез (обратное вейвлет-преобразование). Отметим, что восстановление SR снимков с использованием частотных преобразований (обычно вейвлет-преобразований) широко применяется для восстановления контуров малоразмерных объектов. Для обнаружения малоразмерных объектов была спроектирована нейронная сеть с глубокой памятью, имеющая архитектуру, подобную U-Net модели (Deep Memory Connected

Network) [12]. Улучшенная глубокая рекурсивная остаточная сеть (Improved Deep Recursive Residual Network) была представлена в работе [13]. Основная идея состояла в уменьшении сложности обучения базовой модели ResNet на основе глобального остаточного обучения и локального остаточного обучения с использованием рекурсивного блока. Точность восстановления существенно повышалась за счет подключения нескольких вторичных фильтров с адаптацией для параллельной обработки. Восстановление аэрофотоснимков проводилось в масштабах 2×, 3× и 4×. Однако значения пикового отношения сигнал–шум не превышали 30-35 дБ. Оригинальная архитектура в виде двухуровневой взаимно дополняемой аффинной СНС предложена в работе [14]. С точки зрения авторов взаимная дополняемость между уровнями позволяет получать более информативные признаки, используя стратегию адаптивного множественного внимания к разномасштабным визуальным объектам. Таким образом, реализуется эффективное объединение низкоуровневых и высокоуровневых признаков с помощью операции взаимной аффинной свертки.

Одним из направлений исследований SISR задачи является создание моделей с меньшим количеством параметров, не уступающих, а иногда и превосходящих полные модели СНС. Так, например, в работе [15] предложен новый сверточный слой, названный слоем контекстного преобразования (contextual transformation layer), который с одной стороны упрощает традиционный сверточный слой 3×3, а с другой стороны извлекает эффективные контекстные функции на разных иерархических уровнях обработки. Однако более существенным улучшением традиционных моделей СНС для задачи восстановления снимков дистанционного зондирования сверхвысокого разрешения является применение визуальных трансформеров.

Модели СНС имеют два существенных недостатка. Во-первых, они используют одно и то же ядро свертки для обработки разномасштабных областей изображения. Во-вторых, модели СНС имеют ограниченное поле восприятия, восстановление которого зависит только от локальной информации. В то же время сходные по контексту фрагменты могут предоставить дополнительную информацию для восстановления текущего фрагмента. Архитектура визуальных трансформеров, возникшая на основе моделей СНС, обеспечивает механизм самовнимания (self-attention mechanism) для сбора глобальной информации и использует свойство самоподобия изображения. В настоящее время семейство визуальных

трансформеров расширяется, причем, не только для решения рассматриваемой задачи.

Большинство моделей CHC использует слои повышающей дискретизации для получения SR представления, которые игнорируют извлечение признаков в многомерном пространстве, что влияет на качество выходных данных. Для устранения этой проблемы предложена улучшенная сеть на основе транформеров (Transformer-based Enhancement Network, TransENet) [16]. Ядром модели TransENet является иерархическая структура, объединяющая традиционные SR структуры с мультимасштабными функциями высокой или низкой размерности. Гибридная иерархическая сеть-транформер (Hybrid-Scale Hierarchical Transformer Network, HSTNet) [17] решает аналогичную задачу за счет использования трансформера кросс-масштабного улучшения, позволяющего улавливать долгосрочные зависимости в визуальных данных.

Контекстно-зависимая облегченная сеть (Context-Aware Lightweight Super-Resolution Network, CALSRN) для восстановления изображений дистанционного зондирования сверхвысокого разрешения была предложена в работе [18]. Сеть, имеющая U-Net архитектуру, состоит из последовательных контекстно-зависимых блоков-трансформеров, которые извлекают как локальный контекст, так и глобальный контекст. Ветвь генерации динамических весов позволяет динамически регулировать процесс агрегации локальных и глобальных функций.

Мотивацией применения ГСС для задачи восстановления изображений дистанционного зондирования сверхвысокого разрешения является множество наземных сцен с разномасштабными объектами, характеризующимися различными спектральными характеристиками. Модели CHC, как правило, игнорируют такие особенности. Лучшим вариантом является применение моделей ГСС. Модель ГСС, основанная на механизмах локального и глобального внимания (Attention-based Generative Adversarial Network), была представлена в работе [19]. Механизм локального внимания позволяет сосредоточиться на структурных компонентах земной поверхности, а механизм глобального внимания используется для выявления долгосрочных пространственных взаимозависимостей. Используемый в любых моделях ГСС процесс состязательного обучения позволяет улучшить дискриминационную способность и применять градиентный штраф к комплексной функции потерь, включающей потери пикселов, потери восприятия и состязательные потери.

Восстановление аэрофотоснимков сверхвысокого разрешения зачастую привносит искажения в текстурные области. Для устранения подобных артефактов была разработана плотная генеративно-состязательная сеть (Novel Dense Generative Adversarial Network for real aerial imagery Super-Resolution reconstruction, NDSRGAN), которая объединяла многоуровневую плотную сеть и матричный дискриминатор, учитывающий средние значения пикселов [20]. Для обучения были созданы наборы данных с парными реальными изображениями дистанционного зондирования высокого и низкого разрешения. Для ускорения сходимости модели вместо функции потерь $L1$ была применена функция потерь $smoothL1$ для лучшего визуального восприятия текстур.

Было замечено, что обучение на «идеальных» наборах данных приводит к резкому падению производительности модели на реальных снимках дистанционного зондирования, поскольку качество реальных LR снимков зависит от множества факторов, таких как освещение, состояние атмосферы, используемые датчики и т.д. В работе [21] подобные искажения моделировались с помощью ядер размытия и шумов. В качестве генератора была разработана сеть остаточного сбалансированного внимания (Residual Balanced Attention Network) для оценки результатов сверхвысокого разрешения на основе входных LR снимков. В качестве состязательного обучения был применен дискриминатор на основе U-Net модели для генерации более реалистичных текстур.

Методология трансферного обучения для восстановления SR аэрофотоснимков была применена в работе [22]. В качестве базовой архитектуры использовалась ГСС сверхвысокого разрешения (Super Resolution Generative Adversarial Network). При этом набор данных DIV2K применялся для предварительного обучения генеративной модели, а затем метод трансферного обучения использовался для обучения отдельных моделей на наборах проверочных данных xView и DOTA. Судя по значениям индекса восприятия и среднеквадратической ошибки, метод трансферного обучения показал хорошие результаты восстановления аэрофотоснимков. В дальнейшем такая технология была применена теми же авторами для улучшения обнаружения объектов на аэрофотоснимках [23].

Краткий обзор показал, что восстановление аэрофотоснимков сверхвысокого разрешения на основе методов глубокого обучения является актуальной задачей, требующей дальнейшего развития и поиска неординарных решений в связи со сложностью и многообразием реальных визуальных объектов.

3. Постановка задачи. Пусть $I_{HR} \in R^{H\times W\times C}$ и $I_{SR} \in R^{H\times W\times C}$ – исходный снимок высокого разрешения I_{HR} и восстановленный снимок сверхвысокого разрешения I_{SR}, которые представлены в многомерном пространстве R, имеющем размерности высоты H, ширины W и цветовых каналов C. Для моделирования снимка низкого разрешения I_{LR} строится модель искажений Ψ на основе снимка высокого разрешения I_{HR}, которая имеет вид:

$$I_{LR} = \Psi\left(I_{HR}, \theta_{\eta}\right),$$

где θ_{η} – параметры искажений, например, коэффициент масштабирования, шум, размытие и т.д.

В простейшем случае доступны наборы данных с парными снимками I_{HR} и I_{LR}, а целью является получение параметров искажений θ_{η}. Таким образом, задача восстановления снимка сверхвысокого разрешения заключается в устранении искажений и восстановления снимка, похожего на исходный снимок высокого разрешения I_{HR}.

$$I_{SR} = \Psi^{-1}\left(I_{LR}, \theta_{\xi}\right),$$

где θ_{ξ} – параметры модели сверхвысокого разрешения Ψ^{-1}.

В случае неизвестных параметров искажений задача восстановления SR изображений усложняется. При этом на процесс восстановления, как правило, влияют несколько факторов: шум (белый гауссов шум), размытие (расфокусировка, движение), сжатие и другие артефакты. Процесс обучения заключается в оптимизации параметров θ_{ξ} для модели Ψ^{-1}:

$$\hat{\theta}_{\xi} = \arg\min\left(Loss\left(I_{SR}, I_{HR}\right)\right),$$

где *Loss* – функция потерь.

В силу сложности SR задача для изображений дистанционного зондирования обычно формулируется как задача обучения с учителем. При наличии парных HR–LR снимков обучение упрощается, однако HR снимки не всегда доступны. Если исходные снимки можно считать HR изображениями, то LR изображения в простейшем случае генерируются следующим образом [24]:

$$I_{LR} = \downarrow_{s}\left(I_{HR}\right),$$

где $\downarrow_s$ – понижающая выборка с коэффициентом масштабирования *s*.

Однако на практике предпочитают использовать усложненную модель с ядрами размытия *k* и наложенным шумом *n*:

$$I_{LR} = \downarrow_s \left(I_{HR} \otimes k \right) + n,$$

где символ $\otimes$ означает оператор свертки.

Если исходные снимки имеют низкое разрешение, то при применении СНС из них синтезируются LR изображения еще более низкого разрешения (методом понижающей дискретизации), а при применении ГСС они сравниваются с восстановленными SR снимками, также подвергнутыми понижающей дискретизации. Данный случай является самым сложным, демонстрирующим наихудшие результаты.

Таким образом, целью восстановления SR снимка методами обучения является минимизация функции потерь *Loss*(·):

$$Loss\left(I_{SR}, \theta_\eta, k \right) = \left\| \left(I_{HR} \otimes k \right) - I_{LR} \right\| + \alpha \Psi \left(I_{HR}, \theta_\eta \right),$$

где α – параметр регуляризации. При этом первое слагаемое определяет точность модели, а второе слагаемое отвечает за регуляризацию.

4. Восстановление аэрофотоснимков сверхвысокого разрешения с применением СНС. Базовой моделью восстановления изображений сверхвысокого разрешения на основе СНС считается модель SRCNN [25], разработанная в 2014 году. Она использовала бикубическую интерполяцию для повышения дискретизации входного LR изображения до целевого SR изображения и существенно улучшила качество восстановления по сравнению с традиционными методами. Далее для улучшения базовой модели исследовались различные концепции, например, концепции субпиксельных сверточных слоев, остаточных блоков, плотных блоков, механизмов рекурсии и внимания, пирамидальной обработки, каскадной обработки и т.д. Более того, известны модели, которые совмещают сразу несколько механизмов улучшения. Тем не менее, основной концепцией является последовательное применение трех операций:

– извлечение и представление фрагментов (patches) исходного LR изображения, когда каждый фрагмент преобразуется в многомерный вектор;

– нелинейное отображение, когда многомерный вектор нелинейно отображается на другой многомерный вектор большей размерности;

– реконструкция, когда выполняется объединение фрагментов сверхвысокого разрешения, преобразованных из многомерных векторов большей размерности, для генерации выходного изображения сверхвысокого разрешения, похожего на реальное HR изображение.

В данном исследовании был проведен подробный анализ моделей СНС для восстановления аэрофотоснимков. Для демонстрации преимущества учета семантических особенностей была выбрана одна из последних моделей СНС для восстановления снимков дистанционного зондирования, а именно, гибридная U-образная сеть, основанная на внимании HAUNet (Hybrid Attention-based U-shaped Network) [26]. Модель HAUNet, имеющая достаточно сложную структуру, извлекает и адаптивно агрегирует мультимасштабную информацию с помощью двух модулей извлечения одномасштабных признаков на основе сверточного внимания (Single-scale feature Extraction Modules, SEM): модуль извлечения глобального пространственного контекста (Spatial-aware Context feature Extraction Module, S-CEM) и модуль извлечения абстрактного контента (Content feature Extraction Module, CEM). Модуль межмасштабного взаимодействия (Cross-scale Interaction Module, CIM), расположенный между энкодерами и декодерами на разных уровнях масштабирования, предназначен для устранения семантических разрывов на одном уровне масштабирования, а также разрывов в разрешении между различными уровнями масштабирования. Энкодеры и декодеры состоят из модулей S-CEM и CEM. Исходное изображение низкого разрешения $I_{HR} \in R^{H\times W\times 3}$ поступает на сверточный слой 3×3 для преобразования данных из RGB цветового пространства в представление низкоуровневых признаков $F_0 \in R^{H\times W\times C}$, где $H \times W$ – пространственная размерность, а C – количество каналов. Затем представление F_0 пропускается через три модуля, последовательно уменьшающих пространственный размер до $C \times H \times W$ (F_1), $C \times H/2 \times W/2$ (F_2) и $C \times H/4 \times W/4$ (F_3):

$$\begin{cases} F_1 = Enc_1\left(F_0\right) \\ F_i = Enc_i\left(Conv^{\downarrow}_{2\times 2}\left(F_{i-1}\right)\right) \quad i = 2,3 \end{cases},$$

где $Conv_{2\times2}^{\downarrow}$ означает сверточный слой 2×2 с шагом 2 для понижения частоты дискретизации.

После вычисления представления F_3 модуль CIM адаптивно объединяет и устраняет разрывы между тремя разномасштабными представлениями [F_1, F_2, F_3], получая выходные данные [O_1, O_2, O_3]:

$$\left[O_1, O_2, O_3\right] = CIM\left[F_1, F_2, F_3\right].$$

Выходные данные самого низкого разрешения поступают на вход декодера последнего уровня Dec_3, формирующего выход P_1. Аналогично происходит повышающая дискретизация на 2-ом и 1-ом уровнях с постепенным обогащением представлений P_2 и P_3:

$$\begin{cases} P_1 = DeConv_{2\times2}^{\uparrow}\left(Dec_3\left(O_3\right)\right) \\ P_2 = DeConv_{2\times2}^{\uparrow}\left(Dec_2\left(O_2 + P_1\right)\right), \\ P_3 = Dec_1\left(O_1 + P_2\right) \end{cases}$$

где $DeConv_{2\times2}^{\uparrow}$ означает транспонированный сверточный слой 2×2 с шагом 2 для повышения частоты дискретизации.

Далее выходные данные декодера P_3 реконструируются и подвергаются повышающей дискретизации с помощью сверточного слоя 3×3 и операций перемешивания пикселей для получения окончательных результатов сверхвысокого разрешения. Сформированные таким образом контекстно-зависимые данные суммируются с данными после бикубической интерполяции исходного снимка, в результате чего и происходит формирование снимка I_{SR}. При обучении сети HAUNet используется функция потерь $L1$:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\| I_{HR}^{(i)} - I_{SR}^{(i)} \right\|,$$

где θ – параметры модели, N – количество обучающих экземпляров.

Модель HAUNet способна увеличивать масштаб в 2, 3 и 4 раза. Обучение выполняется по методу обучения с учителем. Входными изображениями являются случайные фрагменты (patches) LR изображений размерностью 48×48 пикселов. Результат восстановления оценивается с помощью соответствующих доступных HR фрагментов изображений по известным метрикам PSRN и SSIM.

Интересно, что настройка сети в процессе обучения выполняется с использованием диагностической технологии LAM (Local Attribution Maps) [27].

Однако ограниченное поле восприятия СНС не способствует эффективному обнаружению малоразмерных объектов, а также ограничивает производительность моделей и возможность их установки на терминальные устройства из-за их высокой вычислительной сложности и большого количества параметров. Для решения подобных проблем были разработаны так называемые визуальные трансформеры, которые в настоящее время являются наиболее интересной и востребованной модификацией СНС. Визуальные трансформеры используются в разных задачах машинного зрения и, в частности, при восстановлении снимков дистанционного зондирования сверхвысокого разрешения. В данном исследовании для сравнительного анализа была выбрана контекстно-зависимая облегченная сеть сверхвысокого разрешения (Context-Aware Lightweight Super-Resolution Network, CALSRN) [18]. Модель CALSRN, в основном, состоит из блоков контекстно-зависимых преобразователей (Context-Aware Transformer Block, CATB), предназначенных как для извлечения локального контекста (Local Context Extraction Branch, LCEB), так и глобального контекста (Global Context Extraction Branch, GCEB) изображений. При этом модуль LCEB использует механизм перекрестного внимания на основе СНС для извлечения локальной информации, а модуль GCEB имеет структуру Swin Transformer для получения глобальной информации. Генерация весов агрегации модулей LCEB и GCEB осуществляется модулем генерации динамических весов (Dynamic Weight Generation Branch, DWGB), что повышает качество восстановления изображений сверхвысокого разрешения. Усложненная архитектура глубокой сети CALSRN требует меньшего количества параметров и меньшей вычислительной сложности по сравнению с существующими методами.

Архитектура модели CALSRN состоит из двух базовых ветвей: ветвь билинейной интерполяции и ветвь реконструкции. В свою очередь ветвь реконструкции состоит из трех основных частей: модуля извлечения низкоуровневых признаков, модуля извлечения высокоуровневых признаков и слоя реконструкции. Модуль извлечения низкоуровневых признаков использует сверточный слой 3×3 и параметрическую функцию активации PReLU (Parametric Rectification Linear Unit). Таким образом, на выходе данного модуля формируются признаки $F_0 \in \mathrm{R}^{H \times W \times C}$ в соответствии с выражением:

$$F_0 = PReLU\left(conv_{3\times3}\left(I_{LR}\right)\right),$$

где $conv_{3\times3}$ – сверточный слой с ядром 3 × 3.

Высокоуровневые функции извлекаются с помощью n каскадных блоков CATB. На вход первого каскада поступают низкоуровневые признаки F_0, а выходные признаки n-го каскада $F_n \in \mathrm{R}^{H\times W\times C}$ вычисляются следующим образом:

$$F_n = f_{CATB}^{n}\left(f_{CATB}^{n-1}\ldots\left(f_{CATB}^{1}\left(F_0\right)\right)\right),$$

где f_{CATB}^{n} – функция n-го блока CATB.

Выходы всех блоков CATB объединяются и последовательно пропускаются через два сверточных слоя $conv_{1\times1}$ и $conv_{3\times3}$ для извлечения высокоуровневых признаков. На выходе модуля извлечения высокоуровневых признаков выполняется суммирование низкоуровневых и высокоуровневых признаков, в результате чего формируются признаки $F_{add} \in \mathrm{R}^{H\times W\times C}$:

$$F_{add} = conv_{3\times3}\left(conv_{1\times1}\left(\left[F_1, F_2, \cdot, F_n\right]\right)\right) + F_0,$$

где символ [·,·] обозначает операцию конкатенации.

Таким образом, выходные функции блоков CATB разных уровней объединяются для получения изображения I_{SR}. Каждый блок CATB имеет сложную структуру, включающую небольшую CHC с механизмом внимания и трансформер с возможностью генерации динамических весов модулем DWGB.

Слой реконструкции состоит из сверточного слоя 3 × 3 с выходной функцией F_{add} и оператора перемешивания пикселов (Pixel Shuffle). Далее выходные функции двух базовых ветвей (ветвь билинейной интерполяции и ветвь реконструкции) суммируются, образуя изображение сверхвысокого разрешения I_{SR}:

$$I_{SR} = H_P\left(conv_{3\times3}\left(F_{add}\right)\right) + H_B\left(I_{LR}\right),$$

где H_P – оператор перемешивания пикселов, H_B – оператор билинейной интерполяции.

Интересно отметить, что применение структуры смещенных окон (Swin Transformer) в модуле GCEB позволяет лучше

анализировать пространственные контексты, одновременно улучшая восприятие локальных пространственных особенностей и снижая вычислительную сложность.

При обучении сети CALSRN используется функция потерь $L1$, аналогичная функции потерь сети HAUNet. Модель CALSRN также способна увеличивать масштаб в 2, 3 и 4 раза.

5. Восстановление аэрофотоснимков сверхвысокого разрешения с применением ГСС. Для учета семантических особенностей аэрофотоснимков разработана новая модель ГСС (SemESRGAN), состоящая из базовой сети на основе известной модели восстановления изображений сверхвысокого разрешения ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) [28] и сети сегментации изображений, настроенной на конкретную прикладную задачу. Архитектура предлагаемой модели ГСС на этапе обучения представлена на рисунке 1. Генератор модели ESRGAN основан на архитектуре плотной сети «остаток в остатке» (Residual-in-Residual Dense Network, RRDN), сочетающей многоуровневую остаточную сеть и плотные соединения. Для повышения производительности и снижения сложности вычислений были удалены все слои пакетной нормализации (batch normalization layers). Дискриминатор модели ESRGAN предсказывает вероятность того, что реальное изображение более реалистично, чем поддельное изображение (в данном случае I_{SR}). Сеть сегментации была предварительно обучена для генерации семантических масок таких типичных классов, встречающихся на аэрофотоснимках, как автомобили, дороги, растительность, крыши и т.д., с использованием общедоступных наборов данных CLCD, DOTA LEVIR-CD и UAVid. Сеть сегментации представляет собой СНС с архитектурой энкодер-декодер, на выходе которой формируется изображение в виде псевдоцветов. На этапе обучения требуются две идентичные сети сегментации (по сути, сиамская архитектура). На вход одной сети подается изображение из обучающего набора данных I_{HR}, в то время, как на вход другой сети поступает реконструированное генератором изображение I_{SR}. Такое решение позволяет оценить сходство признаков как часть общей функции потерь и, следовательно, улучшить обучение генератор в семантическом аспекте.

Функция состязательных потерь ГСС L_{GAN} формулируется как задача min-max оптимизации, когда генератор обучен минимизировать потери, а дискриминатор обучен их максимизировать:

$$\min_{G}\max_{D} L_{GAN}\left(G,D\right)=\mathrm{E}_{I_{HR}}\left[\log D_{\theta_D}\left(I_{HR}\right)\right]+\mathrm{E}_{I_{LR}}\left[\log\left(1-D_{\theta_D}\left(G_{\theta_G}\left(I_{lR}\right)\right)\right)\right],$$

где θ_G и θ_D – параметры генератора G и дискриминатора D соответственно.

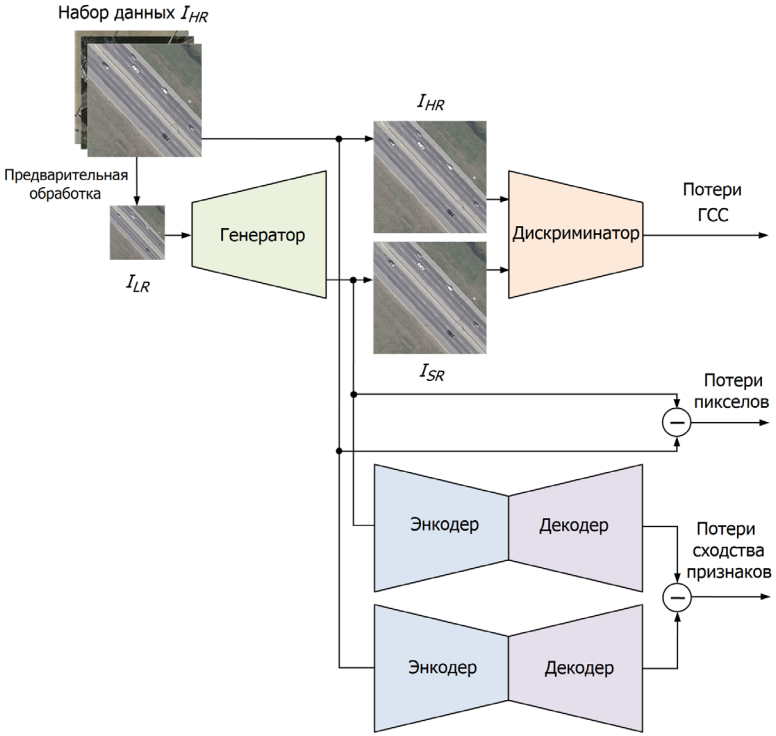


Рис. 1. Архитектура предложенной модели ГСС на этапе обучения

Обучение ГСС останавливается, когда дискриминатор достигает значение вероятности 0,5 для всех сгенерированных изображений $\{I_{SR}\}$. Следует отметить, что после обучения дискриминатор удаляется из архитектуры ГСС.

Как показано в работе [29], при восстановлении изображений сверхвысокого разрешения средняя абсолютная ошибка (Mean Absolute Error, MAE) выше среднеквадратической ошибки (Mean Square Error, MSE) с точки зрения различий на уровне пикселов. Применим функцию потерь пикселов вида MAE:

$$L_{MAE} = \frac{1}{HWC}\left\| I_{HR} - I_{SR} \right\|_1,$$

где H, W, C – высота, ширина и количество каналов изображения.

Недостаток функции потерь пикселов заключается в том, что эти потери рассчитываются на уровне пикселов при условии минимизации средней абсолютной ошибки и, следовательно, создают размытые фрагменты изображения. Поэтому функцию потерь L_{MAE} необходимо применять с понижающим коэффициентом в выражении для функции общих потерь.

Потери сходства признаков, известные как потери восприятия, показывают различия семантических особенностей, вычисленных энкодерами-декодерами СНС:

$$L_{feat} = \frac{1}{HWC}\left\| M\left(I_{HR}\right) - M\left(I_{sR}\right)\right\|_2^2,$$

где $M(I_{HR})$ и $M(I_{SR})$ – маски, сгенерированные СНС, предварительно обученными на аэрофотоснимках высокого разрешения.

Таким образом, общая функция потерь имеет вид:

$$L = \alpha L_{GAN} + \beta L_{MAE} + \gamma L_{feat},$$

где α, β и γ – эмпирически подобранные коэффициенты.

6. Экспериментальные результаты. Для обучения и тестирования предложенной модели глубокой сети (SemESRGAN) использовались шесть открытых наборов данных (рисунок 2):

– CropLand Change Detection (CLCD) [30] состоит из 600 изображений (512×512 пикселов) пахотных земель, собранных при помощи спутника Gaofen-2 в Китае, с пространственным разрешением от 0,5 до 2 м;

– Dataset of Object deTection in Aerial (DOTA) [31] включает коллекцию изображений с разрешением от 800×800 до 20000×20000 пикселов собранных из различных источников (сервис Google Earth, спутник GF-2, другие аэрофотоснимки). Содержит 18 категорий объектов и состоит из 11268 изображений;

– LEVIR building Change Detection (LEVIR-CD) [32] состоит из 637 изображений, взятых из сервиса Google Earth высокого разрешения (50 см/пиксел) размером 1024×1024 пикселов. Набор данных охватывает различные типы зданий жилого частного сектора;

– UAVid [33] содержит изображения уличных сцен высокого разрешения (3840×2160 пикселов) полученных с применением БПЛА. В общей сложности набор состоит из 300 изображений, на которых

размечены 8 классов объектов: здания, дороги, деревья, растительность, движущиеся автомобили, статичные автомобили, люди, фон;

– Airbus Aircraft Detection (AAD) [34] содержит 103 изображения с разрешением 2560×2560 пикселов (пространственное разрешение 50 см). Набор данных включает изображения аэропортов по всему миру, на некоторых изображениях присутствуют туман или облака;

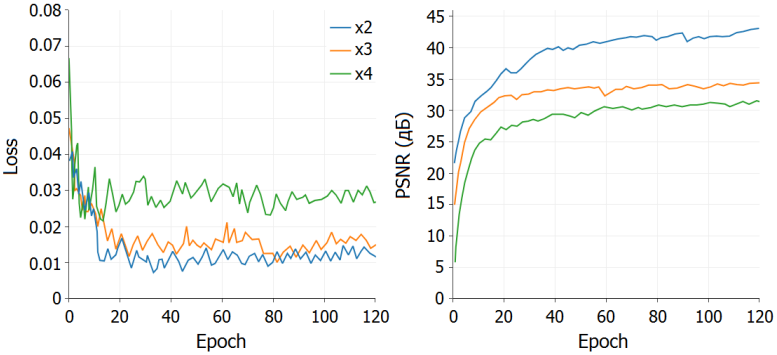
– Aerial Image Dataset (AID) [35] включает 10000 изображений разрешением 600×600 пикселов, принадлежащих 30 классам (аэропорты, стадионы, поля, пляжи, мосты, коммерческая застройка, пустыня, посевные площади, лес, горы, парки, стоянки, детские площадки, порты, железнодорожные станции, реки, школы, жилые массивы, площади, виадуки и т.д.).

а б в г д е

Рис. 2. Примеры изображений из используемых наборов данных:
а) AAD, б) CLCD, в) DOTA, г) LEVIR-CD, д) UAVid, е) AID

Каждый набор данных, за исключением наборов данных AAD и AID, был разделен на обучающую, проверочную и тестовую выборки в соотношении 70/20/10 соответственно. В качестве аугментации применялось вертикальное и горизонтальное отражение изображений. Предложенная модель SemESRGAN была реализована на языке Python с использованием Pytorch repository. В экспериментах использовались графические процессоры NVIDIA Geforce RTX 2080 Ti (11 Гб), операционная система – MS Windows 10.

На рисунке 3 показаны графики функции потерь и валидации в процессе обучения предложенной модели сети SemESRGAN для получения изображений в масштабах ×2, ×3 и ×4.



а б

Рис. 3. Результаты обучения предложенной модели SemESRGAN: а) графики потерь, б) графики валидации

Для оценки качества восстановления изображений применялись две метрики: пиковое отношение сигнала к шуму (PSNR, Peak Signal-to-Noise Ratio) и индекс структурного сходства (SSIM, Structure Similarity).

В таблице 1 показаны средние значения метрик PSNR и SSIM для тестовых изображений из наборов данных CLCD, DOTA, LEVIR-CD и UAVid, а также сравнительные показатели предложенного метода SemERSGAN с другими методами (HAUNet, CALSRN и ESRGAN). При этом эксперименты проводились для 2-, 3- и 4-кратного увеличения исходных LR изображений.

Наборы данных AAD и AID не участвовали в обучении, а применились для тестирования разработанной модели на обобщаемость (таблица 2). Полученные оценки свидетельствуют о том, что предложенная модель SemESRGAN способна обеспечить качественную генерацию изображений сверхвысокого разрешения. В ходе экспериментов были выявлены некоторые артефакты при восстановлении SR изображений, связанные с нарушением текстуры при сильном размытии исходных LR изображений. Однако модель SemESRGAN лучше других моделей восстанавливает контуры объектов, что видно из PSNR показателей в таблицах 2 и 3, а также на рисунках 4-9.

Таблица 1. Средние значения PSNR (дБ)/SSIM

Метод	CLCD	DOTA	LEVIR-CD	UAVid
Увеличение ×2				
HAUNet [26]	32,15/0,895	33,04/0,914	25,78/0,714	28,34/0,833
CALSRN [18]	39,12/0,948	35,98/0,925	26,77/0,740	31,71/0,897
ESRGAN [28]	41,14/0,934	39,28/0,929	29,88/0,841	35,38/0,935
SemESRGAN	44,42/0,978	40,64/0,962	30,62/0,848	37,54/0,958
Увеличение ×3				
HAUNet [26]	30,91/0,839	28,63/0,824	23,88/0,678	26,30/0,716
CALSRN [18]	30,22/0,865	33,82/0,892	25,27/0,692	28,41/0,779
ESRGAN [28]	38,10/0,897	32,21/0,874	25,98/0,684	30,98/0,843
SemESRGAN	40,24/0,942	34,56/0,902	26,65/0,693	32,30/0,868
Увеличение ×4				
HAUNet [26]	29,36/0,783	30,45/0,847	22,98/0,528	25,08/0,632
CALSRN [18]	27,58/0,793	31,79/0,844	24,91/0,592	26,94/0,696
ESRGAN [28]	31,25/0,822	32,93/0,884	25,12/0,638	28,27/0,734
SemESRGAN	34,00/0,868	34,87/0,895	26,66/0,653	29,60/0,774

Таблица 2. Средние значения PSNR (дБ)/SSIM

Метод	AAD	AID
Увеличение ×2		
HAUNet [26]	29,76/0,892	27,12/0,720
CALSRN[18]	33,21/0,932	29,52/0,860
ESRGAN [28]	35,55/0,932	34,94/0,920
SemESRGAN	37,81/0,967	36,44/0,953
Увеличение ×3		
HAUNet [26]	27,53/0,811	26,57/0,708
CALSRN [18]	29,26/0,849	28,59/0,750
ESRGAN [28]	30,73/0,870	32,26/0,901
SemESRGAN	32,17/0,899	33,52/0,916
Увеличение ×4		
HAUNet [26]	25,88/0,697	26,22/0,698
CALSRN [18]	27,07/0,768	27,93/0,737
ESRGAN [28]	27,27/0,798	30,12/0,850
SemESRGAN	29,07/0,817	31,36/0,872

Дополнительно для оценки визуального качества изображений использовалась метрика LPIPS (Learned Perceptual Image Patch Similarity). Метрика LPIPS [36] применяется для измерения сходства восприятия между изображениями, созданными нейросетевыми моделями. В таблицах 3-4 показаны среднее значение метрики LPIPS для различных наборов данных. Более высокие значения означают большие различия в изображениях, а более низкие значения – большее

сходство оригинальных и восстановленных изображений. Если оценка по метрике LPIPS принимает нулевое значение, то изображения идентичны с точки зрения восприятия человеком.

Таблица 3. Средние значения LPIPS

Метод	CLCD	DOTA	LEVIR-CD	UAVid
Увеличение ×2				
HAUNet [26]	0,07011	0,05234	0,02916	0,01781
CALSRN [18]	0,06591	0,04879	0,03672	0,03314
ESRGAN [28]	0,00487	0,00523	0,00846	0,00440
SemESRGAN	0,00162	0,00146	0,00088	0,00043
Увеличение ×3				
HAUNet [26]	0,16280	0,06919	0,04812	0,02732
CALSRN [18]	0,10083	0,05515	0,04331	0,03752
ESRGAN [28]	0,02496	0,00566	0,00764	0,00583
SemESRGAN	0,01978	0,00266	0,00183	0,00086
Увеличение ×4				
HAUNet [26]	0,24689	0,09020	0,08371	0,04773
CALSRN [18]	0,15654	0,06528	0,05594	0,04411
ESRGAN [28]	0,08494	0,00956	0,01101	0,00932
SemESRGAN	0,07709	0,00627	0,00620	0,00395

Таблица 4. Средние значения LPIPS

Метод	AAD	AID
Увеличение ×2		
HAUNet [26]	0,01430	0,03062
CALSRN[18]	0,01993	0,03619
ESRGAN [28]	0,00240	0,00443
SemESRGAN	0,00033	0,00110
Увеличение ×3		
HAUNet [26]	0,02614	0,05373
CALSRN [18]	0,02554	0,04700
ESRGAN [28]	0,00255	0,00807
SemESRGAN	0,00062	0,00381
Увеличение ×4		
HAUNet [26]	0,04701	0,08912
CALSRN [18]	0,03474	0,06234
ESRGAN [28]	0,00454	0,03341
SemESRGAN	0,00267	0,01420

На рисунках 4-9 показаны примеры применения SR методов для 4-кратного увеличения изображения. Поскольку оригинальные HR изображения из разных наборов данных имеют разное разрешение, они были предварительно приведены к единому масштабу.

Рис. 4. Примеры восстановления SR фрагментов для набора данных CLCD в четырехкратном увеличении, файл – 00526.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

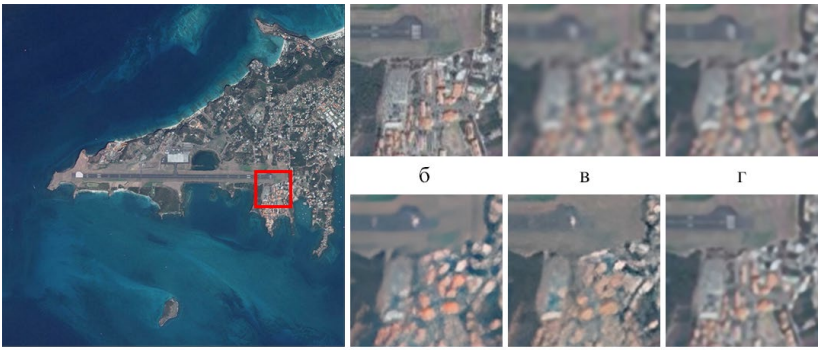


Рис. 5. Примеры восстановления SR фрагментов для набора данных DOTA в четырехкратном увеличении, файл – P4219.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

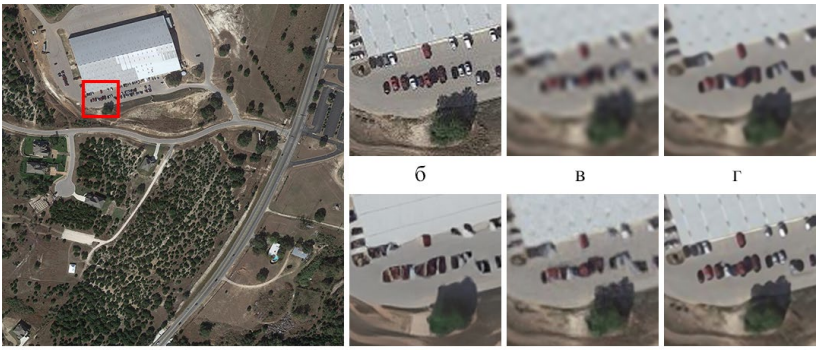


Рис. 6. Примеры восстановления SR фрагментов для набора данных LEVIRCD в четырехкратном увеличении, файл – test_16.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

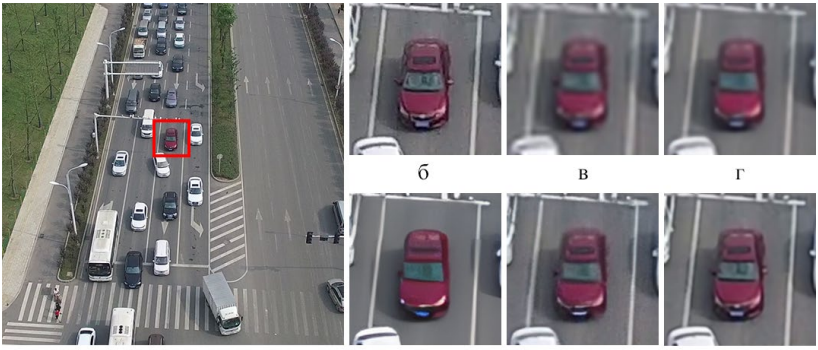


Рис. 7. Примеры восстановления SR фрагментов для набора данных UAVid в четырехкратном увеличении, файл – file15-1.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

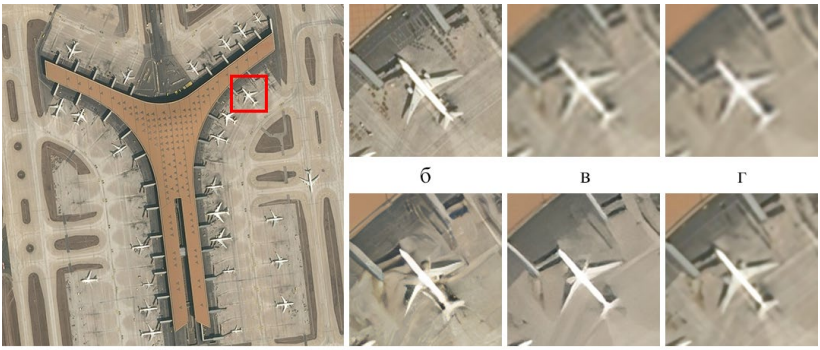


Рис. 8. Примеры восстановления SR фрагментов для набора данных AAD в четырехкратном увеличении, файл – 12210ad7-83f8-4b54-bb4b-e93f8ff6ac1f.jpg: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

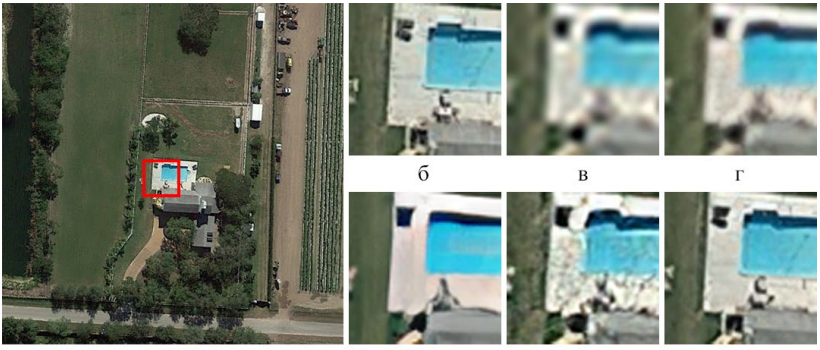


Рис. 9. Примеры восстановления SR фрагментов для набора данных AID в четырехкратном увеличении, файл sparseresidential_51.jpg: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

7. Заключение. Предложенная нейросетевая модель SemESRGAN на основе ГСС позволяет генерировать достаточно реалистичные снимки дистанционного зондирования. Данный подход характеризуется повышенной чувствительностью к деталям изображения, которую обычно не могут обеспечить нейросетевые модели на основе СНС, что приводит к чрезмерной размытости

и исчезновению текстурных особенностей. Отметим, что потеря текстурных особенностей является общей проблемой любых методов восстановления снимков сверхвысокого разрешения, а качество восстановления сильно зависит от параметров исходных изображений, входящих в обучающие наборы данных.

Литература

1. Фаворская М.Н. Аналитическое исследование моделей глубокого обучения для создания снимков ДЗЗ сверхвысокого разрешения // Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2023): Сб. тр. Всероссийской конф. с междунар. участ. 2023. С. 17–25.
2. Lepcha D.C., Goyal B., Dogra A., Goyal V. Image super-resolution: A comprehensive review, recent trends, challenges and applications // Information Fusion. 2023. vol. 91. pp. 230–260.
3. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets. Advances in Neural Information Processing Systems (NIPS 2014). 2014. vol. 27. pp. 1–9.
4. Фаворская М.Н., Пахирка А.И. Улучшение разрешения снимков ДЗЗ на основе глубоких генеративно-состязательных сетей // Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2023): Сб. тр. Всероссийской конф. с междунар. участ. 2023. С. 163–168.
5. Conde M.V., Choi U.J., Burchi M., Timofte R. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration // Computer Vision – ECCV 2022 Workshops. LNCS. Springer, Cham. 2023. vol. 13802. pp. 669–687.
6. Wang P., Bayram B., Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods // Earth-Science Reviews. 2022. vol. 232(15). DOI: 10.1016/j.earscirev.2022.104110.
7. Qiu D., Cheng Y., Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey // Computer Methods and Programs in Biomedicine. 2023. vol. 238. DOI: 10.1016/j.cmpb.2023.107590.
8. Jiang J., Wang C., Liu X., Ma J. Deep learning-based face super-resolution: A survey // ACM Computing Surveys. 2021. vol. 55. no. 1. pp. 1–36.
9. Liu H., Ruan Z., Zhao P., Dong C., Shang F., Liu Y., Yang L., Timofte R. Video super-resolution based on deep learning: A comprehensive survey // Artificial Intelligence Review. 2022. vol. 55. no. 8. pp. 5981–6035.
10. Sun Y., Deng K., Ren K., Liu J., Deng C., Jin Y. Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review // ISPRS Journal of Photogrammetry and Remote Sensing. 2024. vol. 208. pp. 14–38.
11. Wang T., Sun W., Qi H., Ren P. Aerial image super resolution via wavelet multiscale convolutional neural networks // IEEE Geoscience and Remote Sensing Letters. 2018. vol. 15. no. 5. pp. 769–773.
12. Xu W.-J., Xu G.-L., Wang Y., Sun X., Lin D.-Y., Wu Y.-R. High quality remote sensing image super-resolution using deep memory connected network. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). 2018. pp. 8889-8892.
13. Tang J., Zhang J., Chen D., Al-Nabhan N., Huang C. Single-frame super-resolution for remote sensing images based on improved deep recursive residual network // EURASIP J Image Video Proc. 2021. vol. 2021. DOI: 10.1186/s13640-021-00560-8.

14. Tang S., Liu J., Xie X., Yang S., Zeng W., Wang X. A stage-mutual-affine network for single remote sensing image super-resolution // Chinese Conference on Pattern Recognition and Computer Vision (PRCV). 2022. pp. 249–261.
15. Wang S., Zhou T., Lu Y., Di H. Contextual transformation network for lightweight remote-sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–13. DOI: 10.1109/TGRS.2021.3132093.
16. Lei S., Shi Z., Mo W. Transformer-based multistage enhancement for remote sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–11. DOI: 10.1109/TGRS.2021.3136190.
17. Shang J., Gao M., Li Q., Pan J., Zou G., Jeon G. Hybrid-scale hierarchical transformer for remote sensing image super-resolution // Remote Sens. 2023. vol. 15. no. 13. pp. 1–20.
18. Peng G., Xie M., Fang L. Context-aware lightweight remote-sensing image super-resolution network // Frontiers in Neurorobotics. 2023. vol. 17. DOI: 10.3389/fnbot.2023.1220166.
19. Li Y., Mavromatis S., Zhang F., Du Z., Sequeira J., Wang Z., Zhao X., Liu R. Single-image super-resolution for remote sensing images using a deep generative adversarial network with local and global attention mechanisms // IEEE Transactions on Geoscience and Remote Sensing. 2021. vol. 60. pp. 1–24. DOI: 10.1109/TGRS.2021.3093043.
20. Guo M., Zhang Z., Liu H., Huang Y. NDSRGAN: A novel dense generative adversarial network for real aerial imagery super-resolution reconstruction // Remote Sens. 2022. vol. 14. no. 7. pp. 1–23. DOI: 10.3390/rs14071574.
21. Zhang J., Xu T., Li J., Jiang S., Zhang Y. Single-image super resolution of remote sensing images with real-world degradation modeling // Remote Sens. 2022. vol. 14. no. 12. pp. 1–22. DOI: 10.3390/rs14122895.
22. Haykır A.A., Oksuz I. Transfer learning based super resolution of aerial images // 2022 30th Signal Processing and Communications Applications Conference (SIU). 2022. pp. 1–4.
23. Haykir A.A., Öksuz I. Super-resolution with generative adversarial networks for improved object detection in aerial images // Information Discovery and Delivery. 2023. vol. 51. no. 4. pp. 349–357.
24. Tuna C., Unal G., Sertel E. Single-frame super resolution of remote-sensing images by convolutional neural networks // Int. J. Remote Sens. 2018. vol. 39. no. 8. pp. 2463–2479.
25. Dong C., Loy C.C., He K., Tang, X. Learning a deep convolutional network for image super-resolution // Computer Vision – ECCV 2014: 13th European Conference. 2014. pp. 184–199.
26. Wang J., Wang B., Wang X., Zhao Y., Long T. Hybrid attention-based U-shaped network for remote sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2023. vol. 61. pp. 1–15.
27. Gu J., Dong C. Interpreting super-resolution networks with local attribution maps // Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. pp. 9199–9208.
28. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C.C. ESRGAN: Enhanced super-resolution generative adversarial networks // Computer Vision – ECCV 2018 Workshops. 2019. pp. 63–79.
29. Johnson J., Alahi A., Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution // Computer Vision – ECCV 2016: 14th European Conference. 2016. pp. 694–711.
30. Liu M., Chai Z., Deng H., Liu R. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection // IEEE Journal of Selected

Topics in Applied Earth Observations and Remote Sensing. 2022. vol. 15. pp. 4297–4306.
31. Xia G., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 3974–3983.
32. Chen H., Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection // Remote Sens. 2020. vol. 12. no. 10. DOI: 10.3390/rs12101662.
33. Lyu Y., Vosselman G., Xia G-S., Yilmaz A., Yang M.Y. UAVid: A semantic segmentation dataset for UAV imagery // ISPRS Journal of Photogrammetry and Remote Sensing. 2020. vol. 165. pp. 108–119.
34. Airbus Aircraft Detection. URL: www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset (дата обращения: 04.03.2024).
35. Xia G.-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L. AID: A benchmark dataset for performance evaluation of aerial scene classification // IEEE Transactions on Geoscience and Remote Sensing. 2017. vol. 55. no. 7. pp. 3965–3981.
36. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: Salt Lake City, UT, USA. 2018. pp. 586–595.

Фаворская Маргарита Николаевна — д-р техн. наук, профессор, заведующий кафедрой, кафедра информатики и вычислительной техники, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева (СибГУ им. М.Ф. Решетнева). Область научных интересов: компьютерное зрение, обработка изображений и видеопоследовательностей, машинное обучение, глубокое обучение, распознавание образов. Число научных публикаций — 300. favorskaya@sibsau.ru; проспект им. газеты Красноярский Рабочий, 31, 660037, Красноярск, Россия; р.т.: +7(391)213-9622.

Пахирка Андрей Иванович — канд. техн. наук, доцент, кафедра информатики и вычислительной техники, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева (СибГУ им. М.Ф. Решетнева). Область научных интересов: компьютерное зрение, обработка изображений и видеопоследовательностей, машинное обучение, глубокое обучение, распознавание образов. Число научных публикаций — 50. pahirka@sibsau.ru; проспект им. газеты Красноярский Рабочий, 31, 660037, Красноярск, Россия; р.т.: +7(391)213-9622.

DOI 10.15622/ia.23.4.5

M. FAVORSKAYA, A. PAKHIRKA

RESTORATION OF SEMANTIC-BASED SUPER-RESOLUTION AERIAL IMAGES

Favorskaya M., Pakhirka A. **Restoration of Semantic-Based Super-Resolution Aerial Images.**

Abstract. Currently, technologies for remote sensing image processing are actively developing, including both satellite images and aerial images obtained from video cameras of unmanned aerial vehicles. Often such images have artifacts such as low resolution, blurred image fragments, noise, etc. One way to overcome such limitations is to use modern technologies to restore super-resolution images based on deep learning methods. The specificity of aerial images is the presentation of texture and structural elements in a higher resolution than in satellite images, which objectively contributes to better results of restoration. The article provides a classification of super-resolution methods based on the main architectures of deep neural networks, namely convolutional neural networks, visual transformers and generative adversarial networks. The article proposes a method for reconstructing super-resolution aerial images SemESRGAN taking into account semantic features by using an additional deep network for semantic segmentation during the training stage. The total loss function, including adversarial losses, pixel-level losses, and perception losses (feature similarity), is minimized. Six annotated aerial and satellite image datasets CLCD, DOTA, LEVIR-CD, UAVid, AAD, and AID were used for the experiments. The results of image restoration using the proposed SemESRGAN method were compared with the basic architectures of convolutional neural networks, visual transformers and generative adversarial networks. Comparative results of image restoration were obtained using objective metrics PSNR and SSIM, which made it possible to evaluate the quality of restoration using various deep network models.

Keywords: aerial images, super-resolution, semantic segmentation, convolutional neural networks, visual transformers, generative adversarial networks.

References

1. Favorskaya M.N. [Analytical study of deep learning models for the problem of remote sensing single image super resolution] Obrabotka prostranstvennyx dannyx v zadachax monitoringa pripodnyx i antropogennyx prozessov (SDM-2023): Sb. tr. Vserossijskoj nkonf. s mezhdunar. Uchast. [Processing of spatial data in tasks of monitoring natural and anthropogenic processes: Collected papers]. 2023. pp. 17–25. (In Russ.).
2. Lepcha D.C., Goyal B., Dogra A., Goyal V. Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion. 2023. vol. 91. pp. 230–260.
3. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets. Advances in Neural Information Processing Systems (NIPS 2014). 2014. vol. 27. pp. 1–9.
4. Favorskaya M.N., Pakhirka A.I. [GAN-based remote sensing single-image resolution improvement] Obrabotka prostranstvennyx dannyx v zadachax monitoringa pripodnyx i antropogennyx prozessov (SDM-2023): Sb. tr. Vserossijskoj nkonf. s mezhdunar. Uchast. [Processing of spatial data in tasks of monitoring natural and anthropogenic processes: Collected papers]. 2023. pp. 163–168. (In Russ.).

5. Conde M.V., Choi U.J., Burchi M., Timofte R. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. Computer Vision – ECCV 2022 Workshops. LNCS. Springer, Cham. 2023. vol. 13802. pp. 669–687.
6. Wang P., Bayram B., Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth-Science Reviews. 2022. vol. 232(15). DOI: 10.1016/j.earscirev.2022.104110.
7. Qiu D., Cheng Y., Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey. Computer Methods and Programs in Biomedicine. 2023. vol. 238. DOI: 10.1016/j.cmpb.2023.107590.
8. Jiang J., Wang C., Liu X., Ma J. Deep learning-based face super-resolution: A survey. ACM Computing Surveys. 2021. vol. 55. no. 1. pp. 1–36.
9. Liu H., Ruan Z., Zhao P., Dong C., Shang F., Liu Y., Yang L., Timofte R. Video super-resolution based on deep learning: A comprehensive survey. Artificial Intelligence Review. 2022. vol. 55. no. 8. pp. 5981–6035.
10. Sun Y., Deng K., Ren K., Liu J., Deng C., Jin Y. Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review. ISPRS Journal of Photogrammetry and Remote Sensing. 2024. vol. 208. pp. 14–38.
11. Wang T., Sun W., Qi H., Ren P. Aerial image super resolution via wavelet multiscale convolutional neural networks. IEEE Geoscience and Remote Sensing Letters. 2018. vol. 15. no. 5. pp. 769–773.
12. Xu W.-J., Xu G.-L., Wang Y., Sun X., Lin D.-Y., Wu Y.-R. High quality remote sensing image super-resolution using deep memory connected network. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018). 2018. pp. 8889-8892.
13. Tang J., Zhang J., Chen D., Al-Nabhan N., Huang C. Single-frame super-resolution for remote sensing images based on improved deep recursive residual network. EURASIP J Image Video Proc. 2021. vol. 2021. DOI: 10.1186/s13640-021-00560-8.
14. Tang S., Liu J., Xie X., Yang S., Zeng W., Wang X. A stage-mutual-affine network for single remote sensing image super-resolution. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). 2022. pp. 249–261.
15. Wang S., Zhou T., Lu Y., Di H. Contextual transformation network for lightweight remote-sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–13. DOI: 10.1109/TGRS.2021.3132093.
16. Lei S., Shi Z., Mo W. Transformer-based multistage enhancement for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–11. DOI: 10.1109/TGRS.2021.3136190.
17. Shang J., Gao M., Li Q., Pan J., Zou G., Jeon G. Hybrid-scale hierarchical transformer for remote sensing image super-resolution. Remote Sens. 2023. vol. 15. no. 13. pp. 1–20.
18. Peng G., Xie M., Fang L. Context-aware lightweight remote-sensing image super-resolution network. Frontiers in Neurorobotics. 2023. vol. 17. DOI: 10.3389/fnbot.2023.1220166.
19. Li Y., Mavromatis S., Zhang F., Du Z., Sequeira J., Wang Z., Zhao X., Liu R. Single-image super-resolution for remote sensing images using a deep generative adversarial network with local and global attention mechanisms. IEEE Transactions on Geoscience and Remote Sensing. 2021. vol. 60. pp. 1–24. DOI: 10.1109/TGRS.2021.3093043.
20. Guo M., Zhang Z., Liu H., Huang Y. NDSRGAN: A novel dense generative adversarial network for real aerial imagery super-resolution reconstruction. Remote Sens. 2022. vol. 14. no. 7. pp. 1–23. DOI: 10.3390/rs14071574.

21. Zhang J., Xu T., Li J., Jiang S., Zhang Y. Single-image super resolution of remote sensing images with real-world degradation modeling. Remote Sens. 2022. vol. 14. no. 12. pp. 1–22. DOI: 10.3390/rs14122895.
22. Haykır A.A., Oksuz I. Transfer learning based super resolution of aerial images. 2022 30th Signal Processing and Communications Applications Conference (SIU). 2022. pp. 1–4.
23. Haykir A.A., Öksuz I. Super-resolution with generative adversarial networks for improved object detection in aerial images. Information Discovery and Delivery. 2023. vol. 51. no. 4. pp. 349–357.
24. Tuna C., Unal G., Sertel E. Single-frame super resolution of remote-sensing images by convolutional neural networks. Int. J. Remote Sens. 2018. vol. 39. no. 8. pp. 2463–2479.
25. Dong C., Loy C.C., He K., Tang, X. Learning a deep convolutional network for image super-resolution // Computer Vision – ECCV 2014: 13th European Conference. 2014. pp. 184–199.
26. Wang J., Wang B., Wang X., Zhao Y., Long T. Hybrid attention-based U-shaped network for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing. 2023. vol. 61. pp. 1–15.
27. Gu J., Dong C. Interpreting super-resolution networks with local attribution maps. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. pp. 9199–9208.
28. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C.C. ESRGAN: Enhanced super-resolution generative adversarial networks. Computer Vision – ECCV 2018 Workshops. 2019. pp. 63–79.
29. Johnson J., Alahi A., Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. Computer Vision – ECCV 2016: 14th European Conference. 2016. pp. 694–711.
30. Liu M., Chai Z., Deng H., Liu R. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2022. vol. 15. pp. 4297–4306.
31. Xia G., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 3974–3983.
32. Chen H., Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote Sens. 2020. vol. 12. no. 10. DOI: 10.3390/rs12101662.
33. Lyu Y., Vosselman G., Xia G-S., Yilmaz A., Yang M.Y. UAVid: A semantic segmentation dataset for UAV imagery. ISPRS Journal of Photogrammetry and Remote Sensing. 2020. vol. 165. pp. 108–119.
34. Airbus Aircraft Detection. Available at: www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset (accessed 04.03.2024).
35. Xia G.-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L. AID: A benchmark dataset for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing. 2017. vol. 55. no. 7. pp. 3965–3981.
36. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: Salt Lake City, UT, USA. 2018. pp. 586–595.

Favorskaya Margarita — Ph.D., Dr.Sci., Professor, Head of the department, Department of informatics and computer techniques, Reshetnev Siberian State University of Science and Technology (Reshetnev University). Research interests: computer vision, image and video sequence processing, machine learning, deep learning, pattern recognition. The number of publications — 300. favorskaya@sibsau.ru; 31, Krasnoyarsky Rabochy Ave., 660037, Krasnoyarsk, Russia; office phone: +7(391)213-9622.

Pakhirka Andrey — Ph.D., Associate professor, Department of informatics and computer techniques, Reshetnev Siberian State University of Science and Technology (Reshetnev University). Research interests: computer vision, image and video sequence processing, machine learning, deep learning, pattern recognition. The number of publications — 50. pahirka@sibsau.ru; 31, Krasnoyarsky Rabochy Ave., 660037, Krasnoyarsk, Russia; office phone: +7(391)213-9622.