A Preprint
«Повышение пространственного разрешения спутниковых данных с применением нейронных сетей для определения характеристик лесных пород».

Выполнил: Валеев Арслан Рустамович
студент 417 группы ММП ВМК МГУ.

Научный руководитель: к.ф-м.н., доцент
Гуров Сергей Исаевич

декабрь 2024.

Содержание

1 Постановка задачи

Целью данной работы является сравнение производительности нескольких SR архитектур: ESRGAN, SwinIR, ESRGAN+LDL на имеющемся наборе данных. Эти архитектуры используют разные подходы для генерации изображений высокого разрешения. Основыне задачи исследования:

- Подготовка набора данных: выбрать лучший способ загрубления и аугментации данных для спутниковых снимков
- Реализация и обучение моделей: обучение архитектур ESRGAN, ESRGAN+LDL, SwinIR на наборе данных с целью повышения пространственного разрешения изображений.
- Анализ гиперпараметров: оценка влияния различных параметров и разработок (варьирование лосса) на обобщающую способность модели.

2 Введение

В настоящее время использование спутниковых данных для мониторинга и анализа объектов и явлений в окружающей среде становится все более распространенным и актуальным. Однако получение изображений высокого качества требует значительных ресурсов и использования высокотехнологичного оборудования. Кроме того, существует проблема передачи данных между различными устройствами, которые могут работать с разными допустимыми разрешениями изображений. Именно в связи с этими проблемами возникает потребность в разработке эффективных методов повышения пространственного разрешения изображений. Рассмотрим более подробно, почему нейронные сети становятся необходимыми для решения этой задачи и почему стандартные методы, такие как линейная или бикубическая интерполяция, могут быть недостаточно эффективными.

2.1 Введение в область super resolution

Далее используются обозначения:

- LR (image) - изображение низкого разрешения
- HR (image) - изображение высокого разрешения
- SR (image) - восстановленное изображение высокого разрешения из изображения низкого разрешения

В основном, сейчас все методы повышения разрешения можно грубо поделить на два направления: *Single Image Super Resolution* (SISR) и *Multi Image Super Resolution* (MISR). В SISR стоит задача реконструкции HR изображения по одному экземпляру LR. В MISR дано несколько LR снимков одной локации, по которым нужно восстановить HR изображение этой же локации. Хотя задача MISR имеет больше априорной информации (благодаря снимкам с разных ракурсов можно восстановить большую часть высокочастотной информации), но в реальных задачах, зачастую сложно получить несколько разных снимков одной и той же локации. Поэтому далее работа будет посвящена SISR. В этой сфере, задача реконструкции изображения поставлена некорректно - в LR изображении может потеряться часть высокочастотной информации, которая может быть восстановлена/дополнена множеством способов - корректных решений у изображения множество, и все они по-своему правильные.

В SISR методы преимущественно делятся на два семейства - Interpolation-based и Reconstruction-based.

Interpolation-based - это интерполяция изображений методами nearest neighbors, bicubic, bilinear e.t.c. эти методы вычисляют значения пикселей нового изображения, используя расположенные в окрестности пиксели LR снимка. Эти алгоритмы довольно быстрые, но не в состоянии восстановить высокочастотные детали изображения.

Reconstruction-based использует априорную информацию в домене, чтобы задать ограничения на генерацию HR изображения. Известные алгоритмы этого семейства - проекция на выпуклые множества (POCS) [17], maximum-a-posteriori (MAP) [2] подход. К этому семейству относится большинство нейросетевых алгоритмов.

2.2 Интуиция стоящая за нейросетевыми алгоритмами

Пусть рассматривается произвольное изображение P, тогда количество информации на нем обозначается как $I(P)$. В контексте алгоритмической трансформации (интерполяции) $Tr(.)$, справедливо, что $I(P) => I(Tr(P))$. Иными словами, любое детерминированное преобразование не увеличивает количество информации на изображении. Это открывает двери для применения нейронных сетей, которые способны «генерировать» или, можно сказать, «галлюцинировать» дополнительную информацию на изображениях, опираясь на закономерности, выявленные в тренировочных данных. Это позволяет нейронным сетям эффективно повышать разрешение изображений, что является критически важным в сфере анализа спутниковых данных.

2.3 Обзор методов в рассматриваемой области

На сегодняшний день существует несколько нейросетевых подходов к решению задачи SISR (Single Image Super-Resolution). Среди них **CNN-based** подход, использующий свёрточные нейронные сети в своей архитектуре. Первые шаги в этом направлении были сделаны в статье [3], предложившей архитектуру SRCNN. Далее была разработана модель SRResNet, использующая Residual connections, которая является адаптацией модели ResNet для задачи SISR. Данная модель активно используется исследователями в качестве бейзлайна или базовой архитектуры для более продуктивных решений.

2.3.1 CNN-based подходы

Из недавних работ в этой сфере можно выделить архитектуру WindSR [5], основанную на SRResNet, где residual blocks заменены на Residual-in-Residual Dense Blocks. Данная архитектура применялась для повышения пространственного разрешения карты скорости ветра. Данные были взяты из датасета GEOS-5 Nature

Run с разрешением 7 км на ячейку сетки и бикубической интерполяцией понижались до 28 км на ячейку сетки для получения низкого разрешения (LR) изображений. Авторам удалось достичь прироста 11.35% в метрике RMSE по сравнению с передовыми GAN-методами.

Архитектура SRS3 [4] использует механизм внимания для выделения наиболее информативных каналов в картах признаков. В качестве данных использовались снимки Sentinel-2 с MSI сенсором для получения высокоразрешённых (HR) изображений и Sentinel-3 с OLCI сенсором для низкоразрешённых (LR) изображений. Авторам удалось достичь прироста +0.33 dB для увеличения разрешения в 2 раза по сравнению с SRCNN.

Работа SARNet [21] использует SRResNet с добавлением channel-attention механизма в блоки residual connections. Модель обучалась на данных с Sentinel-2 с разрешением 10 м и данных спутника PlanetScope с разрешением 5 м и 2.5 м.

2.3.2 Generative-based подходы

Данный подход предполагает обучение генератора и дискриминатора для создания изображений высокого разрешения (HR) и их классификации соответственно. GAN (Generative Adversarial Networks) широко используется в SISR, так как способен генерировать больше мелких деталей по сравнению с CNN-based подходами, которые преимущественно сглаживают изображение. Первопроходцем считается архитектура SRGAN [6], однако часто используется ESRGAN [16], улучшенная версия SRGAN, использующая relativistic discriminator.

Интересна работа [1], где SRGAN применялся для улучшения качества сегментации незарегистрированного населения в городах Китая. Данные брались со спутников GaoFen-2 (1 м) и Sentinel-2 (10 м). Авторам удалось добиться улучшения качества сегментации по метрике IoU более чем в два раза по сравнению с изображениями низкого разрешения, что сопоставимо с оригинальными изображениями.

Также стоит отметить работу [19], использующую SRGAN архитектуру с EASR модулем, который использует градиент реконструированного изображения для улучшения генерации деталей. Это позволило улучшить качество сегментации на 10% по метрике IoU.

Отдельно можно отметить работу [8], в которой был представлен LDL модуль, улучшающий качество изображений путём штрафования модели за генерацию артефактов. Модуль тестировался на датасетах DIV2K и DF2K, но в [20] он используется для спутниковых изображений GaoFen-2 и GaoFen-7.

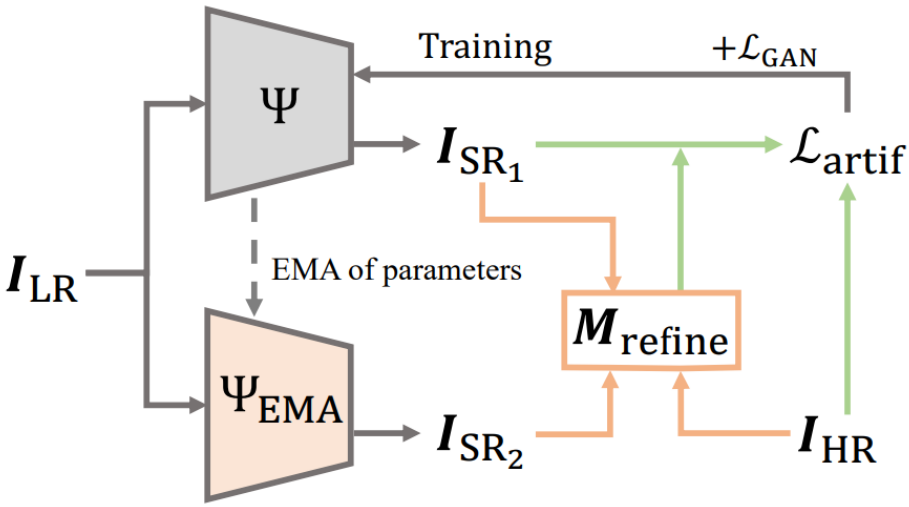

Рис. 1: Схема работы модуля LDL

2.3.3 Transformer-based подходы

Transformer-based подходы позволяют более эффективно переводить изображения в формат высокого разрешения. В работах [9, 14] представлены две модели. В первой из них предложена SOTA модель SwinIR использующая Swin, а во второй предложено объединение CNN-based и Transformer-based подходов с использованием легковесных моделей.

2.3.4 Diffusion models

Диффузионные модели активно используются для SISR благодаря своей способности генерировать визуально более приятные изображения, чем другие SOTA решения [13, 7]. Однако в сфере спутниковых изображений они начали применяться относительно недавно из-за низкой скорости инференса модели. Среди примеров можно выделить модели DMDC [10] и EDiffSR [18].

3 Рассматриваемые подходы

3.1 ESRGAN+LDL

3.1.1 ESRGAN

В работе [16] предоставлена архитектура ESRGAN, являющаяся генеративно-состязательной моделью. Она состоит из генератора 4 и дискриминатора. По сравнению с базовой моделью SRGAN [6] ESRGAN избавили от BatchNorm и изменили функцию потерь для дискриминатора - на relativistic loss 3, суть которого в следующем: заложить в модель сравнивать фальшивые изображения с настоящими, а

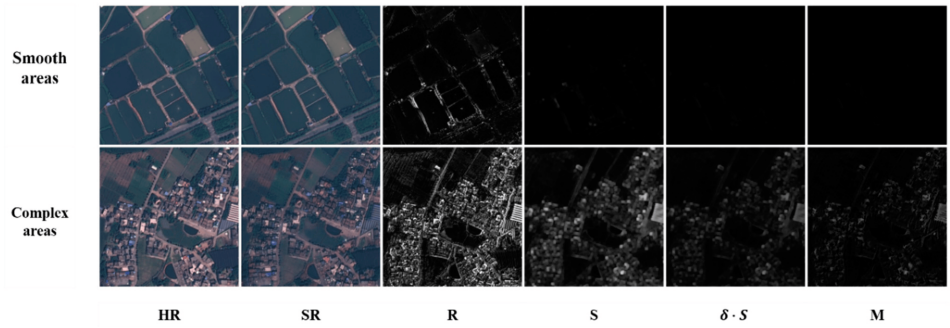

Рис. 2: Визуализация процесса создания карт вероятности артефактов. HR, SR, R, S, $\delta \cdot$ S и M представляют собой выходные данные модели GAN, изображение HR, карту остатков для расчетов HR и SR, карту локальной дисперсии, скорректированную карту локальной дисперсии и карту вероятности для артефактов соответственно.

не рассматривать каждое изображение отдельно. В работе [8] используется архитектура ESRGAN вместе с модулем LDL (см. рис. 1), направленным на снижение частоты генерации артефактов. Рассмотрим его структуру ниже.

Целью данного исследования является создание попиксельной карты артефактов для изображения I_{SR}, полученного методом дистанционного зондирования. Предлагаемый метод использует как локальную, так и глобальную дисперсию, чтобы учесть влияние артефактов на высокочастотные компоненты изображения.

3.1.2 Locally Discriminative Learning (LDL)

Для изображения I_{SR}, полученного с помощью дистанционного зондирования, целью является создание попиксельной карты $M \in \mathbb{R}^{H \times W \times 1}$, где $M(i,j) \in [0,1]$ указывает на вероятность того, что пиксель (i,j) в I_{SR} является артефактом.

3.1.3 Выделение высокочастотной составляющей

Так как артефакты и детали являются высокочастотными компонентами изображения, и сглаженная область позволяет лучше восстановить сеть, вычисляется разница между изображением высокого разрешения I_{HR} и результатом суперразрешения I_{SR} для выделения высокочастотной составляющей:

$$R = I_{HR} - I_{SR}$$

Как показано в 3 столбце рис. 2 гладких областях невязки относительно малы, в то время как в областях с высокой детализацией они велики.

3.1.4 Локальная дисперсия

Для расчета вероятности наличия артефактов вводится локальная дисперсия S остатков R:

$$S(i,j) = \frac{1}{(n+1)^2} \sum_{x=i-\frac{n}{2}}^{i+\frac{n}{2}} \sum_{y=j-\frac{n}{2}}^{j+\frac{n}{2}} (R(x,y) - \mu)$$

Где μ - это среднее значение в окрестности размера n:

$$\mu = \frac{1}{(n+1)} \sum_{x=i-\frac{n}{2}}^{i+\frac{n}{2}} \sum_{y=j-\frac{n}{2}} R(x,y)$$

3.1.5 Глобальная дисперсия

Как показано в 4 столбце рис. 2, локальная дисперсия не учитывает глобальную информацию, поэтому дополнительно вычисляется глобальная дисперсия:

$$\delta = (\mathrm{var}(R))^{\frac{1}{\alpha}}$$

Где α - весовой коэффициент. В экспериментах установлено, что $\alpha = 1/4$.

3.1.6 Использование экспоненциальной скользящей средней модели (EMA)

Как видно в 5 столбце рис. 2, вероятность появления артефактов уже почти равна нулю для гладких участков, таких как сельскохозяйственные угодья. Для повышения стабильности обучения сети GAN используется подход экспоненциальной скользящей средней (EMA):

$$W_k^{EMA} = \alpha \cdot W_{k-1}^{EMA} + (1-\alpha) \cdot W_k$$

Где W_k - модель на k-ом шаге, а W_k^{EMA} - скользящее среднее модели. Параметр α был установлен на уровне 0.999.

Модель W_k^{EMA} является более стабильной и генерирует меньше артефактов по сравнению с W_k. Она используется для корректировки направления градиентного спуска:

$$I_{SR}^{EMA} = W_k^{EMA}(I_{LR}) \quad \text{и} \quad I_{SR} = W_k(I_{LR}),$$

где I_{LR} — исходное изображение низкого разрешения, I_{SR} и I_{SR}^{EMA} — результаты работы моделей. Для анализа ошибок вычисляются остаточные карты R_1 и R_2:

$$R_1 = I_{HR} - I_{SR}, \quad R_2 = I_{HR} - I_{SR}^{EMA}.$$

Если R_1 больше R_2, это указывает на неверное обновление модели. Части R_1, превышающие R_2, считаются артефактами. Итоговая карта артефактов M формируется следующим образом:

$$M(i,j) = \begin{cases} 0, & \text{если } |R_1(i,j)| < |R_2(i,j)|; \\ \delta \cdot S(i,j), & \text{если } |R_1(i,j)| \geq |R_2(i,j)|. \end{cases}$$

3.1.7 Функция потерь

Потери артефактов L_{art} вычисляются по формуле:

$$L_{art} = \|M \cdot R_1\|_1$$

Для повышения качества изображений использовалась сеть VGG для выделения признаков и вычисления потерь L_p между картами характеристик изображений SR и HR:

$$L_p = \sum_i \alpha_i \|VGG_i(I_{HR}) - VGG_i(I_{SR})\|$$

где i — номер функциональной карты сети VGG, а α_i — вес. Были использованы карты объектов слоев 3, 4 и 5, веса которых составили $\frac{1}{4}$, $\frac{1}{4}$ и $\frac{1}{2}$ соответственно.

В отличие от SRGAN, был применен релятивистский дискриминатор, оценивающий вероятность того, что реальное изображение более реалистично, чем SR:

$$D_R(x_r) = \sigma(D(x_r) - \mathbb{E}_{x_f}(D(x_f))) \to 1$$

$$D_R(x_f) = \sigma(D(x_f) - \mathbb{E}_{x_r}(D(x_r))) \to 0$$

где D — дискриминатор, x_r — изображение HR, x_f — изображение SR, σ — сигмоида.

функция потерь разделена на две части: L_G для генератора и L_D для дискриминатора:

$$L_G = -\mathbb{E}_{x_r}[\log(1 - D_R(x_r))] - \mathbb{E}_{x_f}[\log(D_R(x_f))]$$

$$L_D = -\mathbb{E}_{x_r}[\log(D_R(x_r))] - \mathbb{E}_{x_f}[\log(1 - D_R(x_f))]$$

Для восстановления изображения использовалась потеря L_1:

$$L_1 = \mathbb{E}_I \|I_{HR} - I_{SR}\|_1$$

Итоговые потери сети представлены следующим образом:

$$L = \lambda_1 L_1 + \lambda_2 L_p + \lambda_3 L_G + \lambda_4 L_D + \lambda_5 L_{art}$$

где λ — весовые коэффициенты. В данной работе использовались значения $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.05$, $\lambda_4 = 1$, $\lambda_5 = 1$.

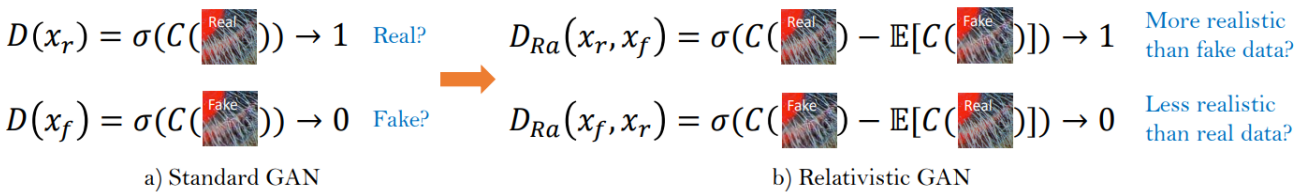

Рис. 3: Отличие стандартного дискриминатора от относительного

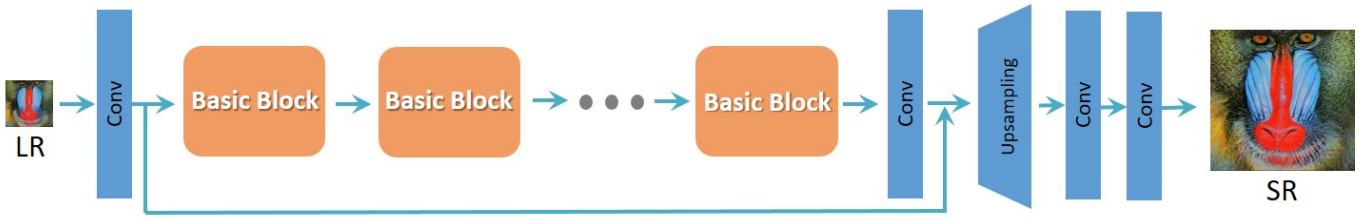

Рис. 4: Архитектура генератора ESRGAN

3.2 SwinIR

Как показано на рис. 5, архитектура SwinIR [9] состоит из трёх основных модулей: извлечение признаков на поверхностном и глубоком уровне (Shallow and Deep Feature Extraction), а также высококачественная (HQ) реконструкция изображений (HQ Image Reconstruction). Один и тот же модуль используется для извлечения признаков, но для разных задач применяются разные модули реконструкции.

3.2.1 Извлечение признаков

На этапе извлечения признаков входное изображение низкого качества I_{LR} с разрешением $R^{H\times W\times 3}$ обрабатывается через свёрточные слои $HSF(\cdot)$ с ядром 3×3 для получения поверхностных признаков $F_0 \in R^{H\times W\times C}$:

$$F_0 = HSF(I_{LQ}),$$

где C — количество каналов. После этого глубокие признаки $F_{DF} \in R^{H\times W\times C}$ извлекаются через модуль глубокого извлечения $H_{DF}(\cdot)$, содержащий несколько блоков Swin-трансформеров (RSTB):

$$F_{DF} = H_{DF}(F_0).$$

Извлечение происходит блок за блоком, где каждый i-й RSTB обозначается как $H_{RSTB_i}(\cdot)$:

$$F_i = H_{RSTB_i}(F_{i-1}), \quad F_{DF} = H_{CONV}(F_K),$$

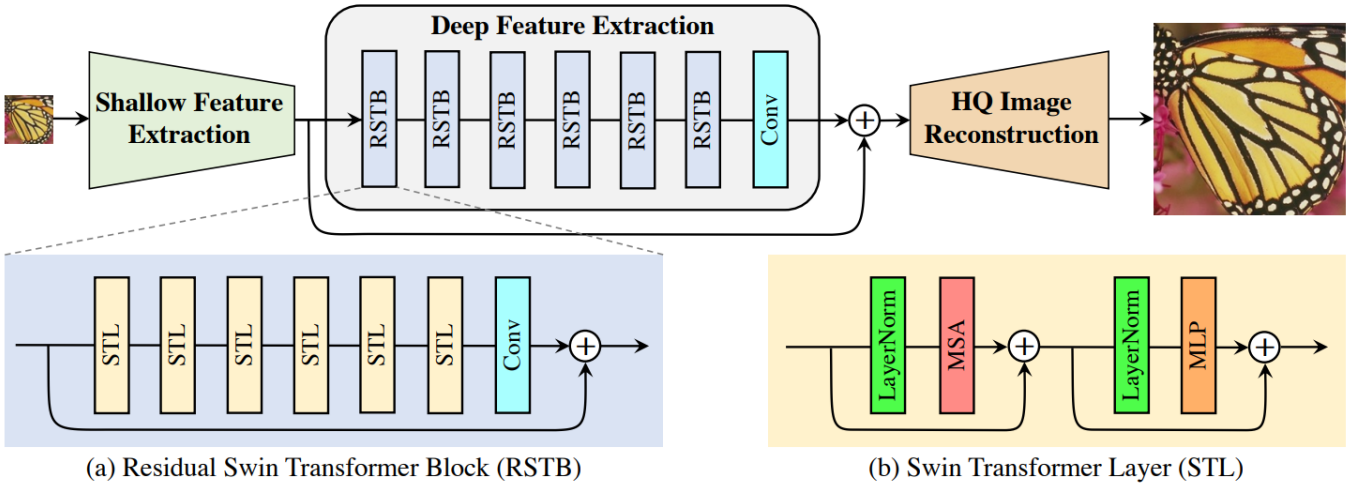

Рис. 5: Архитектура SwinIR

где H_{CONV} — завершающий свёрточный слой, обеспечивающий объединение мелких и глубоких признаков для последующей реконструкции.

3.2.2 Блок Swin-трансформера

Блок Swin-трансформера (RSTB) содержит несколько слоёв трансформера (STL) и свёрточных слоёв. Внутри каждого блока признаки $F_{i,j}$ извлекаются через слои Swin-трансформера:

$$F_{i,j} = H_{STL_{i,j}}(F_{i,j-1}),$$

где $H_{STL_{i,j}}(\cdot)$ — j-й уровень Swin-трансформера. Затем добавляется свёрточный слой и остаточное соединение:

$$F_{i,\text{out}} = H_{CONV_i}(F_{i,L}) + F_{i,0}.$$

Эта структура позволяет объединять признаки на разных уровнях и улучшает эквивариантность трансляции SwinIR.

3.2.3 Swin Transformer Layer

Swin Transformer Layer (STL) [11] основан на механизме многоголового самовнимания, который впервые применен в Transformer [15]. Основные отличия STL заключаются в локальном внимании и механизме смещенного окна.

3.2.4 Механизм работы

Для входного сигнала размером $H \times W \times C$ Swin Transformer разделяет его на неперекрывающиеся локальные окна размером $M \times M$, где количество окон

— $\frac{HW}{M^2}$. Для каждого окна $X \in \mathbb{R}^{M^2 \times C}$ вычисляются запросы, ключи и значения как:

$$Q = XP_Q, \quad K = XP_K, \quad V = XP_V,$$

где P_Q, P_K, P_V — проекционные матрицы, общие для всех окон. Самовнимание для локальных окон рассчитывается как:

$$Attention(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d} + B)V,$$

где B — обучаемая относительная позиционная кодировка. Механизм многозадачного самовнимания (MSA) применяется параллельно для h голов.

3.2.5 Обработка и нормализация

После self-attention применяется многослойный персептрон (MLP) с двумя полностью соединенными слоями и нелинейностью GELU. Также добавляется слой нормализации (LN) перед MSA и MLP. Остаточные соединения включают:

$$X = \text{MSA}(\text{LN}(X)) + X,$$

$$X = \text{MLP}(\text{LN}(X)) + X.$$

Для обеспечения взаимодействия между окнами используется смещенное оконное разбиение, при котором окна сдвигаются на $(\frac{M}{2}, \frac{M}{2})$ пикселей перед разбиением [11].

3.2.6 Реконструкция изображений

Для задачи повышения разрешения (SR) изображение высокого качества I_{SR} восстанавливается путем объединения поверхностных и глубоких признаков:

$$I_{SR} = H_{REC}(F_0 + F_{DF}),$$

где $H_{REC}(\cdot)$ — функция реконструкции. Мелкие признаки передают низкочастотную информацию, в то время как глубокие концентрируются на восстановлении высоких частот, что способствует стабилизированию процесса обучения.

3.2.7 Функция потерь

Для задачи SR параметры SwinIR оптимизируются путём минимизации потерь следующей функции:

$$L = \lambda_1 L_1 + \lambda_2 L_p$$

где L_1 - L1 разница между восстановленным и исходным изображением, L_p - perceptual loss описанный выше, $\lambda_1 = 1$, $\lambda_2 = 0.5$.

4 Эксперименты

4.1 Датасет

В данной работе для сравнения производительности ведущих архитектур использовались спутниковые снимки Массачуссетских дорог [12] с разрешением 1 пиксель на квадратный метр. Для обучения модели использвались 1170 изображений размером 1500x1500 пискелей, для валидации 50 снимков того же разрешения. Для задачи повышения разрешения в 4 раза в качестве HR изображений брались вырезанные фрагменты 256x256, LR изображения - загрубленные при помощи интерполяции и гауссовского шума изображения 64x64 пикселей.

4.2 Детали обучения

Архитектура модели ESRGAN состоит из генератора и дискриминатора. В данной работе использовался генератор состоящий из 23 RRDB блоков с количеством входящих каналов = 64. В модели SwinIR используется 6 блоков RSTB и 6 слоев STL (как в иллюстрации). Во время обучения к LR-HR изображениям применялись вертикальный и горизонтальный поворот и вращение на 90 градусов. Модели обучались 400000 итераций при помощи оптимизатора Adam с шагом 1e-4. Метрики замерялись с частотой 10000 шагов.

4.3 Метрики

Для оценки качества сгенерированных изображений использовались следующие метрики: peak signal noise-to-ratio (PSNR), structural similarity (SSIM), Frechet inception distance (FID) и learned perceptual image patch similarity (LPIPS). PSNR определяет степень искажения, основываясь на разнице между изображениями с низким и высоким разрешением. Чем выше значение PSNR, тем выше качество изображения:

$$PSNR = 10 \log_{10} \frac{MAX}{MSE},$$

где MAX — максимальное значение пикселя, а MSE определяется как:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{SR} - I_{HR}]^2.$$

SSIM оценивает сходство по яркости, контрастности и структуре:

$$SSIM(x, y) = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma}.$$

Высокие значения PSNR и SSIM не всегда коррелируют с визуальным качеством, что требует использования дополнительных метрик, таких как FID и LPIPS.

FID измеряет расстояние между реальными и сгенерированными изображениями, используя KL дивергенцию между признаками сгенерированых и исходных изображений, которые предположительно имеют нормальное распределение:

$$FID(x, y) = ||\mu_x - \mu_y||_2^2 + Tr(\Sigma_x + \Sigma_y - 2(\Sigma_x\Sigma_y)^{1/2}),$$

где μ_x, μ_y — средние значения признаков реальных и сгенерированных изображений, Σ_x, Σ_y — их ковариации. Для замера метрики используется весь датасет или его подмножество.

LPIPS оценивает визуальное сходство, используя сеть VGG для выделения особенностей:

$$\sum_l \frac{1}{H_l W_l} \sum_{h,w} ||\omega_l \odot (x_{hw})_l - (y_{hw})_l||_2^2,$$

где x, y — признаки сгенерированных и реальных изображений, ω_l — весовой коэффициент для слоя l. Чем меньше LPIPS, тем выше качество.

4.4 Результаты

metrics comparsion 4x				
architectures	PSNR	SSIM	LPIPS	FID
Interpolation	18.2	0.623	0.412	61.8
ESRGAN	23.3	0.696	0.302	51.2
ESRGAN + LDL	**24.5**	**0.743**	0.291	49.4
SwinIR	23.7	0.73	0.212	43.4
SwinIR + LDL	24.2	0.735	**0.176**	**36.7**

Как видно из результатов ESRGAN показывает себя лучше в стандартных метриках SSIM и PSNR, ориентированных на попиксельное сходство, в то время как SwinIR выигрывает в метриках LPIPS и FID, выражающих более глубокое сходство восстановленного и исходного изображения.

5 Заключение

В данной работе был проведен сравнительный анализ передовых нейросетевых методов в области повышения пространственного разрешения для спутниковых снимков. Было показано, что LDL модуль повышает производительность модели. Рассмотрены архитектуры ESRGAN и SwinIR с применением LDL модуля для снижения частоты генерации артефактов при восстановлении изображения.

Список литературы

[1] Chunzhu Wei Alessandro Crivellari Hong Wei и Yuhui Shi. «Super-resolution GANs for upscaling unplanned urban settlements from remote sensing satellite imagery – the case of Chinese urban village detection». В: *International Journal of Digital Earth* 16.1 (2023), с. 2623—2643. DOI: 10.1080/17538947.2023.2230956. eprint: https://doi.org/10.1080/17538947.2023.2230956. URL: https://doi.org/10.1080/17538947.2023.2230956.

[2] Giannis K. Chantas, Nikolaos P. Galatsanos и Nathan A. Woods. «Super-Resolution Based on Fast Registration and Maximum a Posteriori Reconstruction». В: *IEEE Transactions on Image Processing* 16.7 (2007), с. 1821—1830. DOI: 10.1109/TIP.2007.896664.

[3] Chao Dong и др. «Image Super-Resolution Using Deep Convolutional Networks». В: (2015). arXiv: 1501.00092 [cs.CV].

[4] Rafael Fernandez и др. «Sentinel-3 Super-Resolution Based on Dense Multireceptive Channel Attention». В: 14 (2021), с. 7359—7372. DOI: 10.1109/JSTARS.2021.3097410.

[5] Ashutosh Kumar и др. «WindSR: Improving Spatial Resolution of Satellite Wind Speed Through Super-Resolution». В: *IEEE Access* 11 (2023), с. 69486—69494. DOI: 10.1109/ACCESS.2023.3292966.

[6] Christian Ledig и др. «Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network». В: (2017). arXiv: 1609.04802 [cs.CV].

[7] Haoying Li и др. «SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models». В: *Neurocomputing* 479 (2021), с. 47—59.

[8] Jie Liang, Hui Zeng и Lei Zhang. «Details or artifacts: A locally discriminative learning approach to realistic image super-resolution». В: (2022), с. 5657—5666.

[9] Jingyun Liang и др. «SwinIR: Image Restoration Using Swin Transformer». В: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (2021), с. 1833—1844. URL: https://api.semanticscholar.org/CorpusID:237266491.

[10] Jinzhe Liu и др. «Diffusion Model with Detail Complement for Super-Resolution of Remote Sensing». В: *Remote Sensing* 14.19 (2022). DOI: 10.3390/rs14194834. URL: https://www.mdpi.com/2072-4292/14/19/4834.

[11] Ze Liu и др. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. 2021. arXiv: 2103.14030 [cs.CV]. URL: https://arxiv.org/abs/2103.14030.

[12] Volodymyr Mnih. «Machine Learning for Aerial Image Labeling». Дис. ... док. University of Toronto, 2013.

[13] Chitwan Saharia и др. «Image Super-Resolution via Iterative Refinement». В: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.4 (2023), с. 4713—4726. DOI: 10.1109/TPAMI.2022.3204461.

[14] Jingzhi Tu и др. «SWCGAN: Generative Adversarial Network Combining Swin Transformer and CNN for Remote Sensing Image Super-Resolution». В: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), с. 5662—5673. DOI: 10.1109/JSTARS.2022.3190322.

[15] Ashish Vaswani и др. *Attention Is All You Need*. 2017. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

[16] Xintao Wang и др. «ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks». В: (2018). arXiv: 1809.00219 [cs.CV].

[17] Frederick W. Wheeler, Ralph T. Hoctor и Eamon B. Barrett. «Super-resolution image synthesis using projections onto convex sets in the frequency domain». В: (2005). URL: https://api.semanticscholar.org/CorpusID:16894124.

[18] Yi Xiao и др. «EDiffSR: An Efficient Diffusion Probabilistic Model for Remote Sensing Image Super-Resolution». В: *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), с. 1—14. DOI: 10.1109/TGRS.2023.3341437.

[19] Lixian Zhang и др. «Making Low-Resolution Satellite Images Reborn: A Deep Learning Approach for Super-Resolution Building Extraction». В: *Remote Sensing* 13.15 (2021). DOI: 10.3390/rs13152872. URL: https://www.mdpi.com/2072-4292/13/15/2872.

[20] Jiayi Zhao и др. «SA-GAN: A Second Order Attention Generator Adversarial Network with Region Aware Strategy for Real Satellite Images Super Resolution Reconstruction». В: *Remote Sensing* 15.5 (2023). DOI: 10.3390/rs15051391. URL: https://www.mdpi.com/2072-4292/15/5/1391.

[21] Xi Zhu, Yang Xu и Zhihui Wei. «Super-Resolution of Sentinel-2 Images Based on Deep Channel-Attention Residual Network». В: (2019), с. 628—631. DOI: 10.1109/IGARSS.2019.8897860.