Повышение пространственного разрешения спутниковых данных с применением нейронных сетей для определения характеристик лесных пород

А. Р. Валеев

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

27 декабря 2024 г.

Постановка задачи

Спутниковые данные всё шире используются для мониторинга окружающей среды, однако получение и передача изображений высокого качества требуют значительных ресурсов.

В связи с этим возрастает спрос на эффективные методы повышения разрешения, и здесь на первый план выходят нейронные сети.

Они превосходят стандартные интерполяционные подходы, такие как линейная и бикубическая интерполяция, обеспечивая более точные результаты.

Single Image Super Resolution

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), maximum-a-posteriori (MAP) подход. К этому семейству относится большинство нейросетевых алгоритмов.

Обзор области

CNN-based подходы. Первой заметной работой в этом направлении стала SRCNN, после чего появилась архитектура SRResNet с блоками residual connections, ставшая популярной базовой моделью. Более свежие решения, такие как WindSR и SRS3, используют усовершенствованные блоки (Residual-in-Residual Dense Blocks, механизм внимания и т.п.) и показывают прирост метрик при работе со спутниковыми данными.

Generative-based подходы. GAN-модели (SRGAN, ESRGAN и их модификации) способны генерировать более детализированные изображения по сравнению с классическими CNN-based методами. Они находят применение в улучшении качества спутниковых снимков и помогают повышать точность задач, таких как сегментация незарегистрированного населения в городах, благодаря дополнительным модулям (EASR, LDL), снижающим артефакты.

Обзор области

Transformer-based подходы. Здесь в фокусе архитектуры SwinIR и её варианты, которые демонстрируют высокую эффективность при работе с задачей SISR. Также существуют гибридные модели, совмещающие CNN и Transformer для снижения вычислительных затрат и сохранения высокого качества.

Diffusion models. Диффузионные модели (DMDC, EDiffSR) привлекают внимание способностью генерировать визуально более приятные результаты, чем многие SOTA-решения. Их применение в спутниковых задачах пока ограничено из-за длительного инференса, однако интерес к ним быстро растёт.

Датасет

В данной работе для сравнения производительности ведущих архитектур использовались спутниковые снимки **Массачуссетских дорог** 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.

Предложенный метод: ESRGAN+LDL

Архитектура ESRGAN представляет собой генеративно-состязательную модель, в которой по сравнению с SRGAN исключён BatchNorm и заменена функция потерь дискриминатора на релятивистскую:

$$D_R(x_r) = \sigma(D(x_r) - \mathbb{E}_{x_f}[D(x_f)]) \to 1, \quad D_R(x_f) = \sigma(D(x_f) - \mathbb{E}_{x_r}[D(x_r)]) \to 0, \quad (1)$$

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

Данная архитектура дополняется модулем LDL, уменьшающим вероятность появления артефактов. Для изображения I_{SR} вычисляется карта остатков

$$R = I_{HR} - I_{SR}, \quad (2)$$

а локальная дисперсия определяется как

$$S(i,j) = \frac{1}{(n+1)^2} \sum_{x=i-\frac{n}{2}}^{i+\frac{n}{2}} \sum_{y=j-\frac{n}{2}}^{j+\frac{n}{2}} \big(R(x,y) - \mu\big), \quad \mu = \frac{1}{(n+1)^2} \sum_{x=i-\frac{n}{2}}^{i+\frac{n}{2}} \sum_{y=j-\frac{n}{2}}^{j+\frac{n}{2}} R(x,y). \quad (3)$$

Глобальная дисперсия выражается через

$$\delta = \big(\mathrm{var}(R)\big)^{\frac{1}{\alpha}}, \quad \alpha = \tfrac{1}{4}. \quad (4)$$

Предложенный метод: ESRGAN+LDL

EMA-модель

Для стабильности обучения используется экспоненциальная скользящая средняя (EMA):

$$W_k^{\mathrm{EMA}} = \alpha \cdot W_{k-1}^{\mathrm{EMA}} + (1-\alpha) \cdot W_k, \quad \alpha = 0.999, \tag{5}$$

где W_k — модель на k-ом шаге, а W_k^{EMA} — сглаженная версия модели. Оценка артефактов проводится путём сравнения двух карт остатка:

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

Если $|R_1(i,j)| \geq |R_2(i,j)|$, пиксель считается артефактом:

$$M(i,j) = \begin{cases} 0, & \text{если } |R_1(i,j)| < |R_2(i,j)|, \\ \delta \cdot S(i,j), & \text{иначе.} \end{cases} \tag{6}$$

ESRGAN+LDL: Функция потерь

Для борьбы с артефактами вводятся потери $L_{\text{art}} = \|M \cdot R_1\|_1$. Дополнительно применяется перцептуальная потеря L_p с использованием признаков сети VGG:

$$L_p = \sum_i \alpha_i \left\| VGG_i(I_{HR}) - VGG_i(I_{SR}) \right\|. \tag{7}$$

Для генератора и дискриминатора в релятивистской постановке вводятся потери:

$$L_G = -\mathbb{E}_{x_r}[\log(1-D_R(x_r))] - \mathbb{E}_{x_f}[\log(D_R(x_f))], \quad L_D = -\mathbb{E}_{x_r}[\log(D_R(x_r))] - \mathbb{E}_{x_f}[\log(1- \tag{8}$$

а для восстановления используется L_1:

$$L_1 = \mathbb{E}_I \left\| I_{HR} - I_{SR} \right\|_1. \tag{9}$$

Итоговая функция потерь имеет вид:

$$L = \lambda_1 L_1 + \lambda_2 L_p + \lambda_3 L_G + \lambda_4 L_D + \lambda_5 L_{\text{art}}, \tag{10}$$

где λ_i — весовые коэффициенты.

Предложенный метод: SwinIR

SwinIR состоит из трёх модулей: поверхностного извлечения признаков $HSF(\cdot)$, глубокого извлечения признаков $H_{DF}(\cdot)$ и реконструкции $H_{REC}(\cdot)$. На первом шаге из входного изображения низкого разрешения I_{LQ} выделяются поверхностные признаки:

$$F_0 = HSF(I_{LQ}), \tag{11}$$

затем глубокие признаки:

$$F_{DF} = H_{DF}(F_0). \tag{12}$$

Итоговое изображение высокого разрешения получается путём объединения этих признаков:

$$I_{SR} = H_{REC}(F_0 + F_{DF}). \tag{13}$$

Предложенный метод: SwinIR

Каждый блок Swin-трансформера (RSTB) включает несколько слоёв Swin Transformer Layer (STL). STL использует механизм локального самовнимания в неперекрывающихся окнах $M \times M$. Для входа $X \in \mathbb{R}^{M^2 \times C}$ вычисляются:

$$Q = XP_Q, \quad K = XP_K, \quad V = XP_V, \tag{14}$$

а внимание для каждого окна:

$$\mathit{Attention}(Q, K, V) = \mathrm{SoftMax}\big(QK^T/\sqrt{d} + B\big)V, \tag{15}$$

где B — обучаемая относительная позиционная кодировка. После многоголового самовнимания (MSA) и MLP выполняются остаточные соединения:

$$X = \mathrm{MSA}\big(\mathrm{LN}(X)\big) + X, \quad X = \mathrm{MLP}\big(\mathrm{LN}(X)\big) + X.$$

SwinIR: функция потерь

Для задачи SISR SwinIR обучается путём минимизации:

$$L = \lambda_1 L_1 + \lambda_2 L_p, \tag{16}$$

где L_1 — L1-потери между I_{SR} и I_{HR}, L_p — перцептуальная потеря.
Коэффициенты $\lambda_1 = 1$ и $\lambda_2 = 0.5$.

Метрики

Для оценки качества изображений использовались PSNR, SSIM, FID и LPIPS. PSNR измеряет степень искажения:

$$PSNR = 10 \log_{10} \frac{MAX}{MSE}, \quad MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{SR} - I_{HR}]^2. \tag{17}$$

SSIM оценивает сходство по яркости, контрастности и структуре:

$$SSIM(x, y) = [l(x, y)]^{\alpha} [c(x, y)]^{\beta} [s(x, y)]^{\gamma}. \tag{18}$$

FID вычисляет расстояние между реальными и сгенерированными изображениями:

$$FID(x, y) = ||\mu_x - \mu_y||_2^2 + Tr(\Sigma_x + \Sigma_y - 2(\Sigma_x \Sigma_y)^{1/2}), \tag{19}$$

где μ_x, μ_y — средние значения, Σ_x, Σ_y — ковариации признаков.
LPIPS измеряет визуальное сходство, используя сеть VGG:

$$\sum_l \frac{1}{H_l W_l} \sum_{h,w} ||\omega_l \odot (x_{hw})_l - (y_{hw})_l||_2^2, \tag{20}$$

где x, y — признаки изображений, ω_l — вес слоя. Меньшие значения LPIPS соответствуют более высокому качеству.

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.

Заключение

В работе проведён сравнительный анализ современных нейросетевых методов повышения пространственного разрешения спутниковых изображений. Продемонстрировано, что использование модуля LDL улучшает качество моделей.

Рассмотрены архитектуры ESRGAN и SwinIR с интеграцией LDL для уменьшения частоты появления артефактов при восстановлении изображений.