

Секция «17.3 Искусственный интеллект и анализ данных в космических исследованиях»

Очистка мультиспектральных спутниковых снимков от облачности и повышение их эффективного разрешения генеративными моделями ИИ

Научный руководитель – Илларионова Светлана Владимировна

Беляков Никита Викторович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет
космических исследований, Москва, Россия

E-mail: MSUBelyakovNV@yandex.ru

Аннотация. В работе рассматривается задача одновременного удаления облачности и повышения пространственного разрешения мультиспектральных спутниковых снимков. Предлагается подход, основанный на использовании геопространственных эмбедингов мультимодальной модели AlphaEarth Foundations в качестве дополнительной априорной информации для генеративной состязательной нейросети ESRGAN. Для обучения и тестирования создана модифицированная версия датасета S2NAIP, включающая фильтрацию облачных сцен (10–40% облачности) и сопоставление им 64-канальных карт эмбедингов AlphaEarth Foundations. Проведено качественное сравнение модели ESRGAN с архитектурами для удаления облачности PMAA и CTGAN. Результаты показывают, что предложенный унифицированный подход позволяет одновременно восстанавливать безоблачную сцену и повышать разрешение снимка в 4 раза, достигая компромисса между пиксельной точностью и комплексностью решаемой задачи.

1. Введение

Безоблачные спутниковые снимки высокого разрешения необходимы для множества прикладных задач: мониторинг окружающей среды, точное земледелие, городское планирование и др. Однако получение таких изображений затруднено из-за ограничений съёмочной аппаратуры и постоянного наличия облачного покрова. Традиционно задачи удаления облачности (cloud removal) и повышения разрешения (super-resolution, SR) решаются независимо с помощью узкоспециализированных моделей. Первые восстанавливают пропущенные под облаками участки, используя многоспектральную или многовременную информацию, вторые увеличивают разрешение, генерируя высокочастотные детали. Оба подхода страдают от нехватки информации в сильно зашумлённых областях.

В последнее время появились геопространственные мультимодальные модели, способные предоставлять высокоуровневые эмбединги, кодирующие семантику и физические характеристики подстилающей поверхности. Одной из таких моделей является AlphaEarth Foundations (AEF) [1], обученная на мультимодальных данных дистанционного зондирования Земли (оптические, радарные, лидарные, климатические) за несколько лет. Её 64-мерные эмбединги в разрешении 10 м содержат информацию о типе местности, сезонных изменениях и могут служить мощным геопространственным приором.

Цель данной работы — исследовать возможность использования эмбедингов AEF для совместного решения задач очистки спутниковых снимков от облачности и повышения разрешения в рамках единой модели на базе ESRGAN.

2. Обзор существующих работ

2.1. Модели очистки от облачности спутниковых снимков

Среди современных подходов к очистке облачности выделяются архитектуры CTGAN и PMAA. CTGAN (Cloud Transformer GAN) [2] использует генеративно-состязательную сеть со специальным экстрактором признаков, который генерирует маску облачности и модулирует карты признаков, а также конформер для выделения критичных пространственно-временных зависимостей.

PMAA (Progressive Multi-scale Attention Autoencoder) [3] предлагает лёгкий автокодировщик с многомасштабным вниманием и локальным взаимодействием, что позволяет при крайне малом количестве параметров достигать качества сравнимо и даже выше CTGAN на ряде бенчмарков.

2.2. Генеративные модели для повышения разрешения

ESRGAN (Enhanced Super-Resolution GAN) [4] изначально разработан для односнимкового повышения разрешения, но его архитектура с Residual in Residual Dense Block (RRDB) блоками легко адаптируется для мультитимевого режима (MISR - Multi Image Super Resolution). В работе [5] показано, что ESRGAN способен синтезировать правдоподобные текстуры даже из зашумлённых временных последовательностей, что открывает возможность его применения для восстановления сцен под облаками. Однако стандартный ESRGAN опирается только на оптические каналы и не использует дополнительную метаинформацию.

Идея дополнения входных данных устойчивыми признаками из других источников (например, SAR) исследовалась ранее [6]. Новизна предлагаемого подхода заключается в использовании эмбедингов AEF, которые объединяют информацию различных сенсоров и инвариантны к кратковременным помехам.

3. Методология

3.1. Подготовка датасета

В качестве для всех экспериментов использован датасет S2NAIP [4], содержащий пары снимков Sentinel-2 (10 м/пиксель) и NAIP (2.5 м/пиксель) на территорию США. Для создания датасета для возможности одновременного удаления облачности и повышения разрешения была выполнена следующая обработка исходного набора S2NAIP:

- Отбор спектральных каналов RGB, NIR, SWIR1, SWIR2 для повышения чувствительности детектора облачности.
- Фильтрация «чёрных» образцов (испорченные тайлы).
- Применение обученной модели U-Net++ с энкодером EfficientNet-B0 [7] для сегментации облачности на каждом снимке. Порог облачности установлен в интервале 10–40%, что обеспечивает достаточно сложные, но решаемые сценарии.

Для каждого отобранного сэмпла (трёх облачных снимков одного района) из модели AEF получены 64-канальные эмбединги с разрешением 10 м. Общий объём выборки составил около 230К. тайлов размером 32 × 32 пикселя до фильтрации по облачности, 140К - после фильтрации.

3.2. Архитектура ESRGAN с эмбедингами AEF

Предлагаемая модификация ESRGAN заключается в конкатенации трёх облачных снимков (9 каналов RGB или же 18 каналов RGB, NIR, SWIR1, SWIR2) с соответствующими 64 каналами эмбедингов AEF. Полученный тензор подаётся на вход генератора с блоками RRDB. Генератор обучается восстанавливать безоблачное изображение с разрешением 2.5 м. Дискриминатор оценивает реалистичность выходных фрагментов по патчам. Обучение ведётся с комбинацией L1, перцептуальных (VGG) и состязательных функций потерь.

4. Эксперименты и результаты моделей

Все модели (PMAA, CTGAN, ESRGAN) обучены на одинаковой тренировочной выборке. Для PMAA и CTGAN входом служат три облачных снимка 10 м, выходом — очищенный снимок 10 м. ESRGAN сразу выдаёт изображение 2.5 м. При обучении использован оптимизатор AdamW с начальной скоростью обучения 5×10^{-4} и косинусным планировщиком обучения. Обучение проводилось на видеокарте GPU NVIDIA и L40S в течение 100 эпох.

В таблице 1 приведены средние метрики PSNR и SSIM:

$$\text{MSE } amp; = amp; \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2, \quad (1)$$

$$\text{RMSE } amp; = amp; \sqrt{\text{MSE}}, \quad (2)$$

$$\text{PSNR } amp; = amp; 20 \log_{10} \left(\frac{MAX_I}{\text{RMSE}} \right), \quad (3)$$

$$\text{SSIM}(X, \hat{X}) \text{ } amp; = amp; \frac{(2\mu_X \mu_{\hat{X}} + C_1)(2\sigma_{X\hat{X}} + C_2)}{(\mu_X^2 + \mu_{\hat{X}}^2 + C_1)} \quad (4)$$

на тестовой части датасета. Для ESRGAN дополнительно показаны результаты после за-
грубления SR до 10 м для сравнения с моделями очистки от облачности на одинаковом
пространственном разрешении.

Таблица 1: Сравнение качества восстановления на тестовой выборке S2NAIP.

Модель	amp; PSNR ↑	amp; SSIM ↑
PMAA (RGB)	amp; 22.515	amp; 0.443
CTGAN (RGB)	amp; 18.796	amp; 0.314
ESRGAN + AEF (2.5 м)	amp; 19.812	amp; 0.288
ESRGAN + AEF (10 м)*	amp; 20.058	amp; 0.345

* выход ESRGAN за-грублен до 10 м для честного сравнения.

Видно, что модель PMAA показывает наилучшие метрики при работе на исходном разрешении. ESRGAN с эмбедингами, решая более сложную комплексную задачу, уступает по метрикам PSNR/SSIM на вповышенном разрешении, однако после приведения к 10 м приближается к показателям PMAA и превосходит результаты CTGAN. Это свидетельствует о том, что предложенный подход успешно восстанавливает семантику под облаками, а основная потеря качества связана с генерацией высокочастотных деталей.

5. Заключение

В работе продемонстрирована возможность совместного удаления облачности и повышения разрешения спутниковых снимков с помощью генеративной модели ESRGAN, дополненной геопространственными эмбедингами AlphaEarth Foundations. Создан модифицированный датасет S2NAIP с фильтрацией по облачности и эмбедингами. Проведено сравнение с современными подходами PMAA и CTGAN. Несмотря на более низкие метрики по сравнению с моделью PMAA, предложенный унифицированный метод обеспечивает практическое решение для задач, где необходимо одновременно очистить изображение от облаков и увеличить его разрешение. Дальнейшие исследования будут направлены на более сложные механизмы слияния эмбедингов и применение мультизадачного обучения.

Список литературы

- [1] Brown, C.F. et al. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv:2507.22291, 2025.
- [2] Huang, G.-L., Wu, P.-Y. CTGAN: Cloud transformer generative adversarial network. In: ICIP, pp. 511–515, 2022.
- [3] Zou, X. et al. PMAA: A Progressive Multi-scale Attention Autoencoder Model for High-Performance Cloud Removal. In: ECAI, pp. 3165–3172, 2023.
- [4] Wolters, P., Bastani, F., Kembhavi, A. Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing. arXiv:2311.18082, 2023.
- [5] Arefin, M.R. et al. Multi-image super-resolution for remote sensing using deep recurrent networks. In: CVPR Workshops, pp. 206–207, 2020.
- [6] Ebel, P. et al. SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal. IEEE TGRS, vol. 60, pp. 1–14, 2022.
- [7] Belyakov, N.V., Illarionova, S., Burnaev, E. Cross-Satellite Climate Structure Segmentation: A Neural Network Domain Adaptation Approach. IEEE Access, vol. 13, pp. 200497–200513, 2025.