

**Улучшение качества повышения разрешения спутниковых снимков
генеративными моделями ИИ под задачу сегментации**

Научный руководитель – Беляков Никита Викторович

Сизов Михаил Михайлович

Студент (магистр)

Национальный исследовательский университет «Высшая школа экономики», Факультет
компьютерных наук, Москва, Россия

E-mail: misha1284465@gmail.com

Аннотация. В данной работе исследуется возможность улучшения качества сегментации спутниковых снимков путем использования генеративных моделей для Super-Resolution (SR) с добавлением сегментационного downstream task loss в функцию потерь. Эксперименты проводятся на датасете LandCover.ai, содержащем 41 тысячу изображений размером 256×256 пикселей с бинарными масками 5 классов для задачи сегментации. Базовый сегментатор на исходных изображениях (64×64) достигает качества $\text{IoU}=0.5979$. Применение модели повышения разрешения ESRGAN повышает качество до $\text{IoU}=0.6771$, а дообучение ESRGAN с downstream task loss дает дополнительное улучшение до $\text{IoU}=0.6810$. Результаты подтверждают эффективность предложенного подхода.

1. Введение

Спутниковые снимки широко используются в задачах мониторинга земной поверхности, планирования территорий и анализа изменений ландшафта. Однако ограниченное разрешение снимков часто затрудняет точную сегментацию объектов. Технологии Super-Resolution (SR) позволяют повысить разрешение изображений, что потенциально может улучшить качество последующей сегментации.

В данной работе исследуется гипотеза: обучение SR-модели с добавлением компоненты потерь от downstream задачи (сегментации) позволяет получить изображения, более полезные для сегментатора, чем стандартные SR-модели.

2. Обзор существующих работ

Современные подходы к Super-Resolution можно разделить на три категории:

- **GAN-based методы:** ESRGAN [1] использует Residual-in-Residual Dense Blocks (RRDB)¹ и улучшенный перцептуальный loss. Real-ESRGAN [2] расширяет этот подход, добавляя моделирование реальных деградаций.
- **Трансформеры:** SwinIR [3] и HAT [4] используют механизмы внимания для улучшения качества реконструкции.
- **Диффузионные модели:** SR3 [5] и ResDiff [6] применяют итеративный процесс диффузии, обеспечивая высокое качество, но требуя значительных вычислительных ресурсов.

¹RRDB — Residual-in-Residual Dense Block, архитектурный блок с плотными связями и остаточными соединениями, позволяющий эффективно обучать глубокие сети для задачи Super-Resolution.

В области сегментации спутниковых снимков наиболее популярны архитектуры U-Net [7] и DeepLabV3+ [8], которые показывают высокие результаты на различных датасетах [11].

Однако вопрос о влиянии качества SR-изображений на downstream задачи исследован недостаточно. В работе [9] авторы исследуют SR для спутниковых снимков, но не рассматривают влияние на сегментацию. Настоящая работа восполняет этот пробел.

3. Методология

3.1. Датасет

Для экспериментов использован датасет LandCover.ai [10], содержащий 41,632 патча размером 256×256 пикселей в каналах RGB с исходным разрешением 0.5 м. Датасет размечен на 5 классов:

- 0 — фон
- 1 — здания
- 2 — лес
- 3 — вода
- 4 — дороги

Данные разделены на train/val/test в пропорции 70/15/15 соответственно.

3.2. Подготовка данных

Для создания изображений низкого разрешения 2 м (Low Resolution, LR) применялся закругление методом бикубической интерполяции до размера 64×64 пикселя. Соответствующие маски уменьшались с интерполяцией методом ближайшего соседа для сохранения целостности классов.

3.3. Архитектуры моделей

SR-модель: ESRGAN с 6 RRDB-блоками и 48 фильтрами в сверточных слоях. Upsampling $\times 4$ осуществляется через PixelShuffle. Функция потерь для SR-модели с downstream task loss имеет вид:

$$\mathcal{L}_{total} = \mathcal{L}_{L1} + \lambda \cdot \mathcal{L}_{seg} \quad (1)$$

где \mathcal{L}_{L1} — L1-компонента (средняя абсолютная ошибка между восстановленным и оригинальным изображением), а \mathcal{L}_{seg} — кросс-энтропия между предсказанием замороженного сегментатора на SR-изображении и истинной маской. Коэффициент $\lambda = 0.1$ подобран эмпирически.

Сегментатор: DeepLabV3+ с энкодером ResNet34 [12], предобученным на ImageNet [13]. Количество классов — 5.

3.4. Метрики

Для оценки качества сегментации использовались:

- **IoU (Intersection over Union):**

$$IoU = \frac{TP}{TP + FP + FN}$$

- **F1-score:**

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Для оценки качества реконструкции снимков с помощью SR-моделей использовались:

- **MSE (Mean Squared Error):**

$$MSE = \frac{1}{n} \sum_{i=1}^n (I_{SR}(i) - I_{HR}(i))^2$$

- **PSNR (Peak Signal-to-Noise Ratio):**

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

- **SSIM (Structural Similarity Index)**

где $MAX_I = 1$ (изображения нормализованы к диапазону $[0, 1]$).

4. Эксперименты и результаты

4.1. Базовый сегментатор на Low Resolution

Первый эксперимент — обучение сегментатора на LR-изображениях (64×64). Результаты представлены в таблице 2.

4.2. Обучение Super Resolution моделей

Были обучены две SR-модели на основе архитектуры ESRGAN:

- **ESRGAN (baseline)** — обучена с использованием только L1-компоненты.
- **ESRGAN + DS** — обучена с добавлением downstream task loss (\mathcal{L}_{seg}).

Метрики качества представлены в таблице 1.

Модель	PSNR	SSIM	IoU (на SR)
ESRGAN (baseline) (0.5 м → 2 м)	30.74	0.892	0.6771
ESRGAN + DS (0.5 м → 2 м)	30.50	0.887	0.6810

Таблица 1: Сравнение SR-моделей и качества сегментации на полученных изображениях

4.3. Сегментация на SR-изображениях

С помощью обученных SR-моделей были улучшены все LR-изображения, после чего на них обучены сегментаторы. Результаты представлены в таблице 2.

Метод	IoU	F1	Улучшение
LR baseline (2 м)	0.5979	0.7066	—
SR baseline (0.5 м → 2 м)	0.6771	0.7675	+13.2%
SR + downstream task loss (0.5 м → 2 м)	0.6810	0.7735	+13.9%

Таблица 2: Сравнительная таблица результатов сегментации

5. Заключение

В ходе исследования была подтверждена гипотеза о том, что использование SR-моделей улучшает качество сегментации спутниковых снимков, а добавление downstream task loss в обучение SR позволяет получить дополнительное улучшение.

Ключевые результаты:

- Базовый сегментатор на LR: $\text{IoU} = 0.5979$
- Сегментатор на SR-изображениях: $\text{IoU} = 0.6771$ (+13.2%)
- Сегментатор на SR-DS-изображениях: $\text{IoU} = 0.6810$ (+13.9%)

Таким образом, предложенный подход может быть использован для практического применения в задачах анализа спутниковых снимков.

Список литературы

- [1] Wang X., et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. ECCV Workshops, 2018.
- [2] Wang X., et al. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. ICCV Workshops, 2021.
- [3] Liang J., et al. SwinIR: Image Restoration Using Swin Transformer. ICCV, 2021.
- [4] Chen X., et al. HAT: Hybrid Attention Transformer for Image Restoration. arXiv:2309.05239, 2023.
- [5] Saharia C., et al. Image Super-Resolution via Iterative Refinement. arXiv:2104.07636, 2021.
- [6] Shang S., et al. ResDiff: Combining CNN and Diffusion Model for Image Super-Resolution. AAAI, 2024.
- [7] Ronneberger O., et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI, 2015.
- [8] Chen L.C., et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. ECCV, 2018.
- [9] Wolters P., et al. Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing. arXiv:2311.18082, 2023.
- [10] Boguszewski A., et al. LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands, Water and Roads from Aerial Imagery. CVPR Workshops, 2021.
- [11] Nikita V. Belyakov, Svetlana Illarionova. CSIA: Climate structures inpainting augmentations for multispectral remote sensing imagery segmentation. Advances in Space Research, 2025.
- [12] He K. et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
- [13] Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012.