

ГЕНЕРАЦИЯ РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ

А. С. Минаев, В. В. Фёдоров

ФГУП «РФЯЦ – ВНИИТФ им. академ. Е. И. Забабахина», Снежинск, Россия

Снежинский физико-технический институт – филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ», Снежинск, Россия

Радиолокационная съемка (SAR) широко востребована в задачах картографирования и мониторинга Земли благодаря способности обеспечивать всепогодное, круглосуточное наблюдение и получать информацию о рельефе и структуре поверхности. Оптические спутниковые снимки, в свою очередь, предлагают высокую детализацию форм и цветов объектов, но подвержены ограничениям из-за погодных условий и освещенности сцены. Комбинация SAR и оптических данных позволяет более полно характеризовать территорию, однако объем доступных радиолокационных данных зачастую бывает недостаточен: высокая стоимость съемки, ограниченное временное покрытие и сложные физические эффекты затрудняют сбор больших наборов. В этой связи актуальной становится задача синтеза правдоподобных SAR-изображений на основе оптических снимков с помощью генеративно-сопоставительных сетей (GAN), что дает возможность восполнять пропущенные радарные данные и дополнять реальные наблюдения.

В настоящей работе предлагается модель двунаправленного перевода между доменами оптики и радиолокации, объединяющая идеи Pix2Pix/UGATIT и механизмы внимания [1, 2]. Система включает два генератора (перевод оптики в SAR и обратно) и четыре дискриминатора: глобальные (CAM-дискриминаторы) и локальные (PatchGAN-подобные). Включение глобальных и локальных дискриминаторов позволяет контролировать как общий стиль сцены, так и достоверность мелких деталей. Для повышения устойчивости обучения применяется комбинированная функция потерь, включающая сопоставительный и L1-критерии, что помогает лучше сохранять структуру объектов и характерные эффекты на SAR-сценах (яркие отражения, спекл-шум). В генераторах применяются сверточные энкодер-декодерные блоки с остаточными слоями (ResNet), а также механизмы внимания на основе CAM (Class Activation Mapping) и Swin Transformer [3]. CAM-модуль позволяет выделять наиболее значимые фрагменты изображения, используя глобальные и максимальные пулинги (GAP/GMP) для формирования карт активаций, тогда как Swin Transformer способствует учету пространственного контекста на разных масштабах. Нормализации ILN и AdaLN (Instance/Layer Interpolation Normalization) обеспечивают гибкость при переходе между статистиками каналов и всего слоя. Отдельное внимание уделяется возможности компенсации сдвигов между оптическим и радарным снимками. Для этого в модель опционально включен регистрационный модуль на основе сверточной сети (ResUnet), вычисляющей поле деформаций с последующим выравниванием (Transformer_2D) сгенерированного и эталонного изображений.

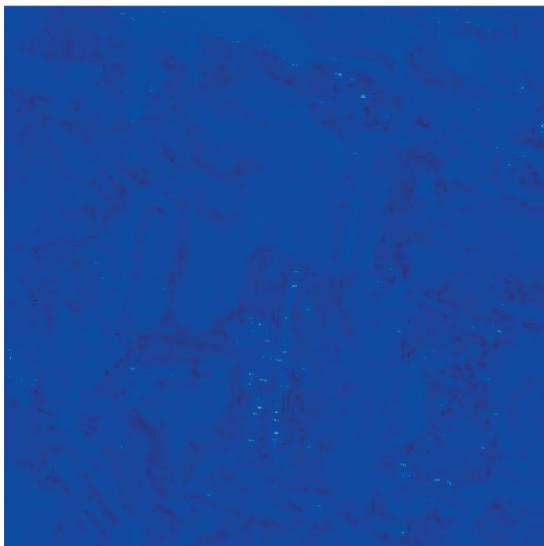
Обучение и тестирование проводилось на наборе данных OpenEarthMap-SAR, содержащем парные сцены оптики и радиолокационных данных (амплитудные каналы VV или HH), охватывающем городские и сельские ландшафты в США, Японии, Франции, с разрешением 0,15–0,5 м [4]. Из 17 тысяч пар размером 512×512 около 10% было выделено для тестовой выборки. По результатам экспериментов, модель демонстрирует высокую реалистичность сгенерированных SAR-изображений: среднее значение PSNR достигает 19 дБ, а SSIM – 0,78. Для метрики распределения FID получено значение порядка 42, что говорит о статистической близости синтетических данных к реальным. Дополнительные эксперименты с использованием взвешенных метрик (Weighted SSIM и др.) подтверждают стабильное качество генерации на различных типах ландшафтов. На рисунке 1 представлены визуальные результаты для одной из тестовых сцен, включающие исходное оптическое изображение, реальный SAR-снимок, соответствующую CAM-карту и сгенерированный радарный кадр, что наглядно демонстрирует возможности предлагаемого подхода.



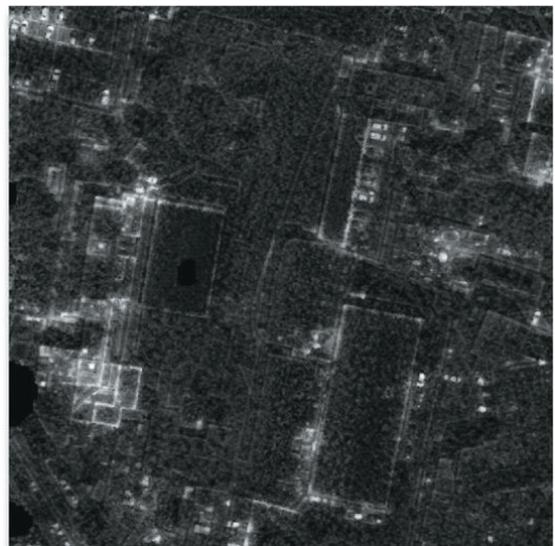
Оптическое изображение



Радиолокационное изображение



Карта сопоставления активационных классов



Сгенерированное изображение

Рис. 1. Визуальные результаты одной из тестовых сцен

Литература

1. **Isola, P.** Image-to-image translation with conditional adversarial networks [Text] / P. Isola et al. // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 1125–1134.
 2. **Kim, J.** U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation [Text] // arXiv preprint arXiv:1907.10830. – 2019.
 3. **Liu, Z.** Swin transformer: Hierarchical vision transformer using shifted windows [Text] / Z. Liu et al. // Proceedings of the IEEE/CVF international conference on computer vision. – 2021. – P. 10012–10022.
 4. **Xia, J.** OpenEarthMap-SAR: A Benchmark Synthetic Aperture Radar Dataset for Global High-Resolution Land Cover Mapping [Text] / J. Xia et al. // arXiv preprint arXiv:2501.10891. – 2025.
-