

УДК 004.932:528.8

Обратимая многошаговая диффузионная модель для преобразования изображений на основе радиолокационных и оптических данных

© 2025 г. А. С. Минаев^{1,2}, В. В. Федоров^{1,2}

¹ Всероссийский научно-исследовательский институт технической физики имени академика Е.И. Забабахина, Снежинск, 456770, Россия

² Снежинский физико-технический институт – филиал ФГАОУ ВО «Национального исследовательского ядерного университета «МИФИ», Снежинск, 456776, Россия

В условиях активного развития дистанционного зондирования Земли, роста доступности спутниковых данных различных типов и необходимости оперативного анализа геопространственной информации особую актуальность приобретает эффективное преобразование и объединение данных, полученных в различных диапазонах электромагнитного спектра, в частности спутниковых радиолокационных (SAR) и оптических (RGB) изображений. В данной работе предложена обратимая диффузионная модель на основе моста Шредингера для двунаправленного преобразования SAR- и RGB-снимков без использования парной выборки. Многошаговый стохастический процесс постепенно зашумляет данные, а нейросеть архитектуры U-Net выполняет восстановление на каждом шаге. Двунаправленная схема обеспечивает сохранение структуры сцены: прямой генератор переводит оптику в радар, а обратный – радар в оптику, обучаясь итеративно с учетом энтропийно-регуляризованного стохастического управления. Тестирование на наборах данных SEN1-2 и SN6-SAROPT показало, что по метрикам PSNR, SSIM и FID метод превосходит классические GAN-подходы (CycleGAN) и однонаправленные диффузионные модели. В полном цикле (RGB → SAR → RGB) погрешность < 5–10 % (SSIM) относительно исходного снимка, при этом достоверно воспроизводятся спекл-шум и яркие отражения, а также восстанавливается детализированная оптическая информация. Обратимая диффузия без парной выборки позволяет избежать систематических искажений и облегчает генерацию недостающих данных в ДЗЗ, мультисенсорный анализ и устранение облачности в оптическом домене.

Ключевые слова: радиолокационные изображения (SAR), оптические изображения (RGB), диффузионные модели, генеративное моделирование, мост Шредингера, двунаправленная трансформация, глубокое обучение.

Введение

Объединение спутниковых SAR- и RGB-данных является актуальной задачей в дистанционном зондировании Земли [1, 2]. SAR слабо зависят от погодных условий и дают структурную информацию благодаря спекл-шуму и ярким отражениям, в то время как оптические снимки передают визуальные и цветовые признаки [3]. Вместе они способны обеспечить более полное описание земной поверхности, однако совместная обработка затруднена из-за различных физических принципов формирования данных и несопоставленных выборок [1].

В последние годы активно развиваются генеративные подходы к переводу SAR в оптику и обратно [4–6]. Наибольшее распространение получили GAN-сети, такие как CycleGAN для непарного перевода

✉ А.С. Минаев: asminaev@vniitf.ru
В.В. Федоров: vvfyodorov@vniitf.ru

Поступила в редакцию: 28.04.2025
После доработки: 28.05.2025
Принята к публикации: 10.06.2025

изображений [4] и UNIT [15], но они подвержены нестабильности обучения, коллапсу мод и потере важных деталей [5, 6]. Современные диффузионные модели демонстрируют более высокую стабильность и гибкость [8–10], однако нередко ориентированы лишь на однонаправленный перевод либо требуют больших объемов парных данных.

В данной работе предлагается обратимая диффузионная модель для двунаправленного преобразования SAR ↔ RGB. В отличие от предыдущих решений, она основана на мосте Шредингера [11–14] и не требует парной выборки. Ниже изложены математическая постановка, алгоритм обучения и результаты экспериментов, показывающие, что данный метод превосходит аналоги по качеству генерируемых изображений.

Математическая постановка задачи

Пусть заданы два пространства: X – множество RGB-изображений и Y – пространство SAR-изображений. Обозначим через p_X и p_Y истинные распределения в этих доменах. Задача двунаправленного преобразования состоит в том, чтобы выучить отображения $T: X \rightarrow Y$ и $T^{-1}: Y \rightarrow X$, способные при отсутствии парных образцов удовлетворять условиям $T(\mathbf{x}) \sim p_Y$ при $\mathbf{x} \sim p_X$ и $T^{-1}(\mathbf{y}) \sim p_X$ при $\mathbf{y} \sim p_Y$. В идеале, T^{-1} должно быть обратным отображением для T (т.е. $T^{-1}(T(\mathbf{x})) \approx \mathbf{x}$), однако строго обратимой функция может не быть из-за стохастического характера генерации (одно и то же SAR-изображение может соответствовать множеству допустимых оптических, и наоборот). Тем не менее, требуется, чтобы преобразование было статистически обратимым, т.е. сохраняло основную структуру сцены и не вносило систематических искажений при проходе туда и обратно.

Формализуем задачу через непрерывный диффузионный процесс X_t , $t \in [0, 1]$, со следующими свойствами: при $t = 0$ вектор X_0 имеет распределение p_X , при $t = 1$ – p_Y . Динамика X_t задается стохастическим дифференциальным уравнением Ито

$$dX_t = f(X_t, t)dt + g(t)dW_t, \quad t \in [0, 1], \quad (1)$$

где W_t – винеровский процесс (белый шум), $g(t)$ – заданная функция шума (постоянная или зависящая от t), а $f(X_t, t)$ – искомая функция дрейфа.

Чтобы удовлетворить граничным условиям p_X и p_Y , функция f должна «направлять» процесс от распределения p_X к p_Y . Для нахождения f используем подход моста Шредингера [7], вводя простой референтный процесс (например, «чистое» добавление шума) и минимизируя расхождение Кульбака–Лейблера относительно него при фиксированных начальном и конечном распределениях:

$$P^* \text{Pargmin}(KL(P||P^*), P(X_0 \sim p_X, X_1 \sim p_Y)). \quad (2)$$

Задача (2) приводит к решению в виде набора прямых и обратных диффузионных уравнений, которые можно приближать итеративными алгоритмами типа IPF (Iterative Proportional Fitting) [7–9]. Для нашей задачи функция f параметризуется нейросетью, обучаемой методом максимального правдоподобия по данным обоих доменов. Обучение чередует уточнение прямой (оптика → радар) и обратной (радар → оптика) цепочек диффузии до согласования их на промежуточных шагах.

Описание архитектуры и метода

Разработанный подход реализуется как серия дискретных шагов прямой и обратной диффузии, каждый из которых аппроксимируется сверточной U-Net-сетью. При прямом преобразовании $X \rightarrow Y$ каждый шаг преобразует $x_k \rightarrow x_{k+1}$, а при обратном $Y \rightarrow X$ – $y_k \rightarrow y_{k+1}$. Таким образом, итеративно строится стохастический мост [7, 8], обеспечивающий обратимость.

Чтобы повысить реалистичность синтеза SAR (спекл-шум, яркие отражения) и детализированность оптики, в архитектуру могут добавляться механизмы внимания (Axial Attention) и модули многосторонней сверточной обработки, а также использоваться gradient checkpointing для снижения затрат памяти при обучении.

Обучение ведется по принципу состязательных сетей (GAN). Дискриминаторы, работающие на нескольких масштабах (оригинал и уменьшенное изображение), проверяют реалистичность выходных кадров. Они используют сверточные блоки с вниманием, что повышает распознавание характерных паттернов. Потери вычисляются в схеме LSGAN [9], а генераторы минимизируют сумму adversarial-потерь и регуляризаторов: cycle-consistency, identity, перцептивных (с опорой на VGG [5]).

Модель обучается на несопоставленных выборках SAR- и оптических изображений. Для каждого шага прямой/обратной диффузии оптимизируются отдельные параметры, а также учитываются циклические реконструкции. В результате, полный каскад при прямом проходе генерирует радарное изображение из оптического, а при обратном восстанавливает исходную оптику, что дает статистическую обратимость.

На стадии тестирования достаточно последовательно пропустить изображение через n генераторов прямой либо обратной диффузии. Благодаря двунаправленной постановке [7, 8] результат при применении обратной цепочки согласован со входными данными, что подтверждает возможность обратного преобразования и сохранение структурных признаков.

Результаты

Выполнено обучение и тестирование модели на двух наборах данных: SEN1-2 [14], содержащем пары участков Sentinel-1 (SAR) и Sentinel-2 (RGB), а также на датасете SN6-SAROPT (SAR- и оптические изображения высокого разрешения). При этом использовались два графических процессора Tesla T4, размер входного изображения устанавливался равным 512, общее число шагов диффузии 4, а в качестве планировщика изменения скорости обучения применялась косинусная схема. Оптимизатор – Adam с начальным $lr = 10^{-4}$ и параметрами $\beta = (0.5, 0.999)$.

Для оценки точности перевода между доменами SAR и RGB использовались следующие метрики:

PSNR (дБ) – пиковое отношение сигнала к шуму (большие значения означают лучшее восстановление);

SSIM (0 ... 1) – показатель структурного сходства;

FID – Frechet Inception Distance, характеризующий статистическое расхождение распределений (меньше – лучше).

В табл. 1 приведено сравнение предложенной модели (обозначена как SB-Diffusion) с другими известными подходами. Для теста мы взяли независимый набор пар изображений из SEN1-2, а также аналогичную выборку из SN6-SAROPT. Отметим, что Pix2Pix использует парное обучение, поэтому в несопоставленном сценарии (unpaired) приведен лишь в качестве ориентира.

Как видно, наш метод показывает лучшие результаты по всем ключевым метрикам в обоих направлениях (SAR \rightarrow RGB и RGB \rightarrow SAR), а также сохраняет обратимость: при полном цикле (например, RGB \rightarrow SAR \rightarrow RGB) более 90 % пикселей восстанавливаются с точностью по яркости ± 5 единиц. Пример перевода представлен на рис. 1: слева исходный оптический снимок, рядом синтезированный радарный, затем вновь восстановленный оптический и справочно-реальный SAR. Видно, что характерные особенности (спекл-шум, яркие отражения, цвета) генерируются достоверно.

Таблица 1. Сравнение качества перевода изображений различными методами (SEN1-2, тестовый набор)

Модель	SAR \rightarrow RGB			RGB \rightarrow SAR		
	PSNR	SSIM	FID	PSNR	SSIM	FID
CycleGAN [4]	11.4	0.42	120.3	12.0	0.38	135.6
UNIT [15]	10.7	0.40	130.5	11.3	0.36	140.1
Pix2Pix	12.9	0.47	95.8	13.5	0.45	110.2
GAN Casc. [1]	12.1	0.44	84.9	13.0	0.45	90.5
BBDM [8]	13.2	0.55	82.1	14.0	0.50	88.7
SB-Diffusion (ours)	14.0	0.57	72.4	15.1	0.53	79.3

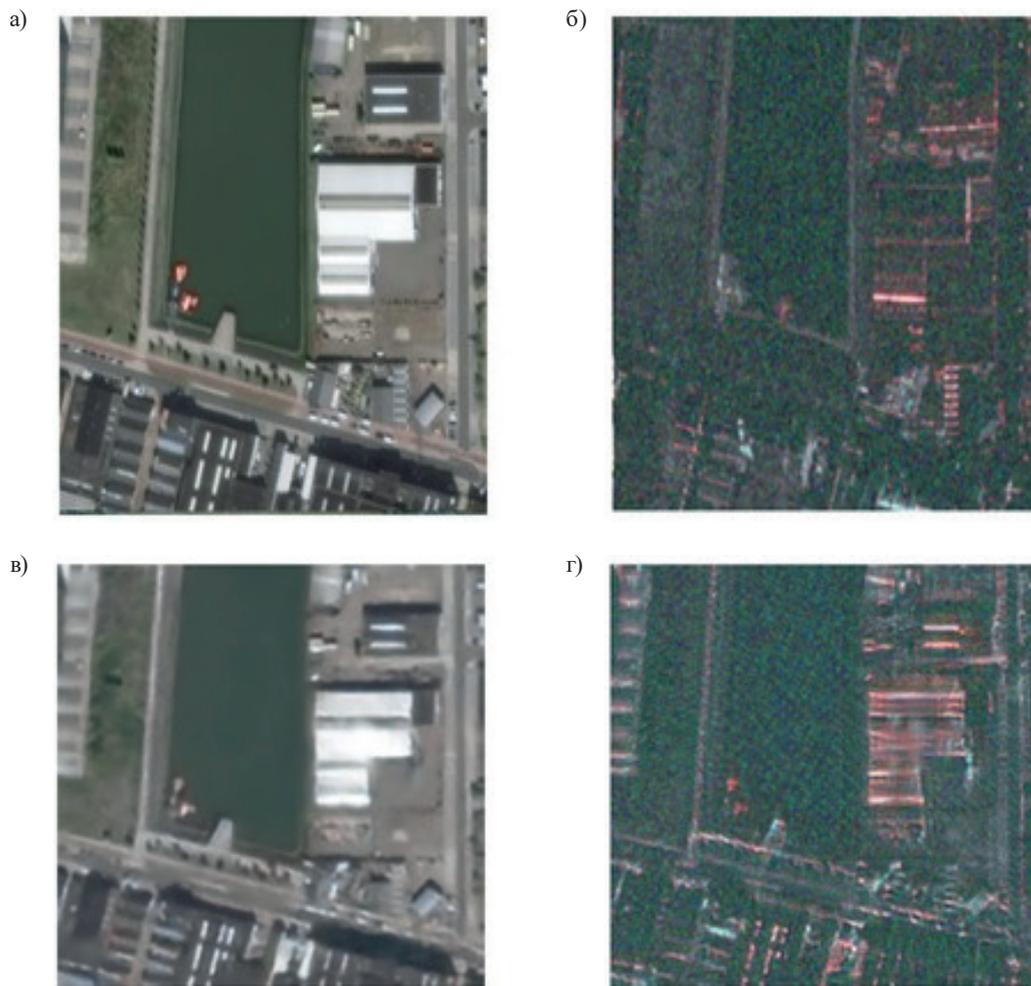


Рис. 1. Пример преобразования: а) исходный RGB; б) реальный SAR; в) восстановленный RGB; г) синтезированный SAR

По вычислительным затратам четырехшаговая диффузия на изображении 512×512 занимает $\sim 3-5$ с. в среднем на связке из двух $T4$. Ускорение возможно за счет методов прореживания шагов, дающих ~ 0.5 сек/кадр при незначительном снижении SSIM. Таким образом, модель пригодна для офлайн-задач и может масштабироваться на более обширные сцены.

Обсуждение

Результаты подтверждают, что обратимая диффузионная модель надежно решает двунаправленную задачу $SAR \leftrightarrow RGB$ даже при отсутствии парных данных. В отличие от классических GAN, здесь подчеркивается согласованность распределений, а не минимизация пиксельной ошибки, что позволяет корректно воспроизводить спекл-шум и сохранять детали. Использование моста Шредингера дает статистическую обратимость: при полном цикле переводов исходная сцена восстанавливается без существенных искажений.

Ограничением остается потенциальное усреднение деталей в очень сложных сценах (плотная городская застройка). Перспективное направление – добавление детерминированных подсказок (например, контуров объектов), а также гибридизация с уже существующими подходами типа CycleGAN для улучшения мелких деталей.

Выводы

Представлен новый метод двунаправленного преобразования спутниковых изображений (SAR и оптических), основанный на мосте Шредингера и идее обратимой диффузии. Модель не требует парных данных и обучается так, чтобы при проходе в прямом и обратном направлениях сохранять структурные особенности сцены. Эксперименты на реальных наборах данных (SEN1-2, SN6-SAROPT) продемонстрировали превосходство по PSNR, SSIM и FID над классическими GAN и односторонними диффузионными методами. Метод особенно полезен для мультисенсорного анализа, пополнения недостающих снимков при дефиците радарных (или оптических) данных, а также для восстановления оптики при облачной обстановке.

Благодарности

Авторы выражают глубокую благодарность Всероссийскому научно-исследовательскому институту технической физики имени академика Е.И. Забабахина за финансовую и организационную поддержку исследования. Особая признательность Национальному исследовательскому ядерному университету «МИФИ» за предоставление вычислительных ресурсов и доступ к специализированному программному обеспечению. Авторы также благодарят коллег из научно-исследовательского отдела за ценные замечания и обсуждения, которые способствовали улучшению методологии и интерпретации результатов. Отдельное спасибо выражается международным научным сообществам, предоставившим открытые наборы данных SEN1-2 и SN6-SAROPT, что сделало возможным проведение сравнительных экспериментов.

Финансирование

Исследование выполнено при финансовой поддержке Российского федерального ядерного центра – Всероссийского научно-исследовательского института технической физики имени академика Е.И. Забабахина (РФЯЦ-ВНИИТФ).

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов

Авторы внесли эквивалентный вклад в подготовку публикации.

Список литературы

1. Прокофьев К.В., Иванов С.В. Генерация радиолокационных снимков местности из оптических // Информатика, телекоммуникации и управление. 2022. Т. 15. № 4. С. 7–21.
2. Zhu J., Park T., Isola P. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks // Proceedings of IEEE International Conference on Computer Vision (ICCV), 2017. P. 2242–2251.
3. Fu S., Xu F., Jin Y.Q. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks // Science China Information Sciences, 2021. V. 64, 112401. DOI: 10.1007/s11432-020-3077-5
4. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // Advances in Neural Information Processing Systems (NeurIPS), 2020. V. 33. P. 6840–6851.
5. Song Y., Sohl-Dickstein J., Kingma D. et al. Score-Based generative modeling through stochastic differential equations // Proceedings of International Conference on Learning Representations (ICLR), 2021. 22 p. DOI: 10.48550/arXiv.2011.13456
6. De Bortoli V., Thornton J., Heng J., Doucet A. Diffusion Schrodinger bridge with applications to score-based generative modeling // Advances in Neural Information Processing Systems, 2021. V. 34. P. 17695–17709.
7. Su X., Song J., Meng C., Ermon S. Dual diffusion implicit bridges for image-to-image translation // Eleventh International Conference on Learning Representations (ICLR), 2023. 18 p. DOI: 10.48550/ARXIV.2203.08382.
8. Li B., Xue K., Liu B. et al. BBDM: Image-to-image translation with Brownian bridge diffusion models // Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2023. P. 1952–1961.

9. Liu G.H., Vahdat A., Huang D.A. et al. I²SB: Image-to-image Schrodinger bridge // International Conference on Machine Learning (ICML), 2023. 18 p. DOI: 10.48550/ARXIV.2302.05872.
10. Kim B., Kwon G., Kim K. et al. Unpaired image-to-image translation via neural Schrodinger bridge // Proceedings of International Conference on Learning Representations (ICLR), 2024. DOI: 10.48550/arXiv.2305.15086.
11. Shi H., Wang Y., Xu M. et al. A brain-inspired approach for SAR-to-optical image translation based on diffusion models // Frontiers in Neuroscience, 2024. V. 18, 1352841. 14 p.
12. Seo M., Oh Y., Kim D. et al. Improved flood insights: Diffusion-based SAR to EO image translation // arXiv preprint arXiv:2307.07123. 2023. 10 p.
13. Wang X., Liao H., Yang Z. et al. Optical-ISAR image translation via denoising diffusion implicit model // Proc. IEEE Int. Conf. Image Processing and Computer Vision Applications (ICIPCA), 2023. P. 1437–1441.
14. Schmitt M., Hughes L.H., Zhu X.X. The SEN1-2 dataset for deep learning in SAR-optical data fusion // ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018. V. IV-1. P. 141–146.
15. Liu M.-Y., Breuel T., Kautz J. Unsupervised image-to-image translation networks // Proceedings of 31st Conference on Neural Information Processing Systems (NIPS), 2017, P. 700–708.

Vestnik Natsional'nogo Issledovatel'skogo Yadernogo Universiteta «MIFI», 2025, vol. 14, no. 3, pp. 249–255

Reversible multi-step diffusion model for image translation between radar and optical data

A. S. Minaev^{1,2,✉}, V. V. Fedorov^{1,2,✉}

¹ All-Russian Scientific Research Institute of Technical Physics named after Academician E.I. Zababakhin, Snezhinsk, 456776, Russia

² Snezhinsk Physics and Technology Institute of the National Research Nuclear University MEPhI, Snezhinsk, 456776, Russia

✉ asminaev@vniitf.ru

✉ vvfedorov@vniitf.ru

Received April 28, 2025; revised May 28, 2025; accepted June 10, 2025

In the context of expanding Earth remote sensing capabilities, increasing availability of diverse satellite data, and the need for efficient geospatial analysis, effective transformation and integration of synthetic aperture radar (SAR) and optical (RGB) imagery has become highly relevant. We propose a reversible diffusion model based on the Schrodinger bridge for bidirectional transformation of unpaired SAR and RGB images. A multi-step stochastic process progressively perturbs the data with noise, while a U-Net-based neural network denoises each step. A bidirectional scheme ensures reversibility: the forward generator converts optical images into radar, and the backward generator does the opposite. Both components are trained iteratively under an entropy-regularized stochastic control framework, aligning intermediate distributions and preserving key scene structures. The model was tested on the SEN1-2 and SN6-SAROPT datasets. According to PSNR, SSIM, and FID metrics, it surpasses traditional GAN-based approaches (e.g., CycleGAN) and one-way diffusion models. Moreover, for the complete RGB → SAR → RGB cycle, the discrepancy from the original image remains under 5–10% (SSIM). The approach can be applied to generate missing imagery, support multisensor data analysis, and mitigate cloud coverage in optical domains.

Keywords: synthetic aperture radar (SAR), optical images (RGB), diffusion models, generative modeling, Schrödinger bridge, bidirectional transformation, deep learning.

References

1. Prokofiev K.V., Ivanov S.V. Generatsiya radiolokatsionnykh snimkov mestnosti iz opticheskikh [Generation of radar imagery from optical]. *Informatika, telekommunikatsii i upravlenie*, 2022. Vol. 15. No. 4. Pp. 7–21 (in Russian).
2. Zhu J., Park T., Isola P., Efros A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017. Pp. 2242–2251.
3. Fu S., Xu F., Jin Y.Q. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks. *Science China Information Sciences*, 2021. Vol. 64. 112401. DOI: 10.1007/s11432-020-3077-5.
4. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. Vol. 33. Pp. 6840–6851.
5. Song Y., Sohl-Dickstein J., Kingma D.P., Kumar M., Ermon S., Poole B. Score-based generative modeling through stochastic differential equations. *Proceedings of International Conference on Learning Representations (ICLR)*, 2021. 22 p. DOI: 10.48550/arXiv.2011.13456.
6. De Bortoli V., Thornton J., Heng J., Doucet A. Diffusion Schrodinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 2021. V. 34. P. 17695–17709.
7. Su X., Song J., Meng C., Wu Y., Zhu J. Dual diffusion implicit bridges for image-to-image translation. *Proceedings of Eleventh International Conference on Learning Representations (ICLR)*, 2023. 18 p. DOI: 10.48550/ARXIV.2203.08382.
8. Li B., Xue K., Liu B., Zhao J., Tao D. BBDM: Image-to-image translation with Brownian bridge diffusion models. *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023. P. 1952–1961.
9. Liu G.H., Vahdat A., Huang D.A., et al. I²SB: Image-to-image Schrodinger bridge. *International Conference on Machine Learning (ICML)*, 2023. 18 p. DOI: 10.48550/ARXIV.2302.05872.
10. Kim B., Kwon G., Kim K., et al. Unpaired image-to-image translation via neural Schrodinger bridge. *Proceedings of International Conference on Learning Representations (ICLR)*, 2024. DOI: 10.48550/arXiv.2305.15086.
11. Shi H., Wang Y., Xu M., et al. A brain-inspired approach for SAR-to-optical image translation based on diffusion models. *Frontiers in Neuroscience*, 2024. Vol. 18, 1352841. 14 p.
12. Seo M., Oh Y., Kim D., et al. Improved flood insights: Diffusion-based SAR to EO image translation. arXiv:2307.07123, 2023.
13. Wang X., Liao H., Yang Z., et al. Optical-ISAR image translation via denoising diffusion implicit model. *Proc. IEEE Int. Conf. Image Processing and Computer Vision Applications (ICIPCA)*, 2023. Pp. 1437–1441.
14. Schmitt M., Hughes L.H., Zhu X.X. The SEN1-2 dataset for deep learning in SAR-optical data fusion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2018. Vol. IV-1. Pp. 141–146.
15. Liu M.-Y., Breuel T., Kautz J. Unsupervised image-to-image translation networks. *Proceedings of 31st Conference on Neural Information Processing Systems (NIPS)*, 2017. Pp. 700–708.