



Explainable Diffusion Model via Schrödinger Bridge in Multimodal Image Translation

Zhengyi Lu¹, Robert Ewing², Erik Blasch³, and Jia Li¹

¹ Oakland University, Rochester, MI 48309, USA
{zhengyilu, li4}@oakland.edu

² Air Force Research Laboratory, WPAFB, Dayton, OH 45433, USA

³ Air Force Research Laboratory, Rome, NY 13441, USA

Abstract. In this paper, we introduce a novel framework for multimodal image translation utilizing the Diffusion Schrödinger Bridge (DSB). By integrating diffusion models with the Schrödinger Bridge problem, our approach addresses stability and interpretability challenges in image translation tasks. Explainable Diffusion Model via Schrödinger Bridge Multimodal Image Translation (xDSBMIT) leverages the unique distribution characteristics of different modalities, achieving high-quality translations with limited datasets. The DSB framework enhances interpretability by elucidating the diffusion process between paired images, providing insights into how the translation is achieved. Specifically, we apply our framework to the translation of Synthetic Aperture Radar (SAR) images to Infrared (IR) and Electro-Optical (EO) images, demonstrating its effectiveness in remote sensing applications. Experimental results show that xDSBMIT framework outperforms established methods such as pix2pix, significantly enhancing image translation performance and model interpretability while requiring fewer training data.

Keywords: Image-to-Image Translation · Diffusion Models · Schrödinger Bridge · DDDAS · Dynamic Data Driven Applications Systems · InfoSymbiotic Systems

1 Introduction

Image-to-Image (I2I) translation involves converting images from one domain to another, leveraging techniques like style transfer, image colorization, super-resolution, and image synthesis [14, 15, 17]. Recently, this field has extended to multimodal learning, enabling translations across different modalities by training on extensive datasets [23]. These advancements have applications in artistic creation, medical imaging, and satellite image analysis, broadening the spectrum of image translation and enhancing the interpretation of visual data across various contexts.

Dynamic Data Driven Applications Systems (DDDAS) integrate instrumentation data with models in real-time, allowing these models to dynamically manage the use and acquisition of data. DDDAS-based methods adapt to the ever-changing nature of real-world systems, providing a robust and flexible framework for various applications that demand real-time data integration and dynamic system adaptability. Generative Adversarial Networks (GANs) are extensively used for image translation in dynamic data-driven applications systems (DDDAS) to generate augmented data for near-real support to deployed systems [1]. Notable techniques include pix2pix for paired image translation and CycleGAN for unpaired image translation [10, 24]. GAN-based methods use continuous feedback from the discriminator to produce realistic images closely approximating the ground truth, significantly advancing the field of image translation [2, 4, 20]. However, GANs face challenges like training difficulties, vanishing gradients, and poor interpretability.

To overcome these issues, diffusion-based methods for image translation have been explored [25]. Traditional diffusion models progressively denoise an image until it transforms into Gaussian noise, learning the reverse process to generate new images [7, 8]. However, these models struggle with paired image translation. To address this, the Diffusion Schrödinger Bridge (DSB) has been proposed as an innovative generative model, combining the Schrödinger Bridge framework with the diffusion process to transition smoothly between distributions and generate high-quality images.

DSB has shown promising results in various applications. De Bortoli explored its use in score-based generative modeling, demonstrating its versatility [5]. Liu demonstrated its effectiveness in I2I translation [13], and Chen highlighted its superiority in text-to-speech synthesis compared to traditional diffusion models [3]. However, previous applications of DSB have focused on image restoration and colorization, not on cross-modal image translation tasks. Additionally, there has been no practical implementation of DSB in the field of satellite imagery.

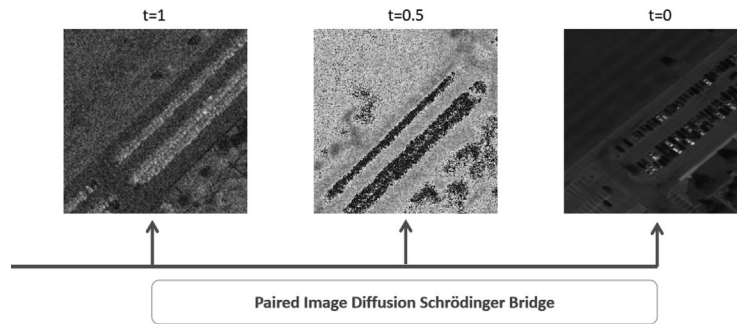


Fig. 1. Diffusion Process of Paired Images

To address the cross-modal image translation gap, and achieve stable, highly interpretable image translation, we propose the application of the diffusion Schrödinger bridge in multimodal image translation tasks. As shown in Fig. 1,

the red arrows in the figure represent the sampling process, the Paired Image Diffusion Schrödinger Bridge (DSB) can transform a noisy image at $t = 1$ into a clear image at $t = 0$ through intermediate states. In addition to denoising, we have accomplished translation between paired SAR and IR, as well as SAR and EO images, based on the diffusion Schrödinger bridge framework. These tasks are crucial because they enable the fusion of complementary information from different sensor modalities, improving overall situational awareness and enhancing the quality of remote sensing applications. Remarkably, with only 500 pairs of data, yielding results comparable to those of pix2pix trained on a significantly larger dataset. Our main contributions are as follows:

1. **Introducing the Diffusion Schrödinger Bridge to Multimodal Image Translation:** Our framework uniquely combines diffusion models with the Schrödinger Bridge, enhancing the stability and interpretability of image translation across complex modalities.
2. **Efficient Learning from Limited Data:** By focusing on the distribution characteristics between different modalities, the Diffusion Schrödinger Bridge enables our method to achieve high-quality image translation with significantly fewer data pairs, thereby reducing dependency on large datasets.
3. **Validation Across Modalities:** Our approach is robustly tested across diverse imaging modalities such as SAR, IR, and EO, showcasing its adaptability and effectiveness.

This paper is organized as follows. The next section reviews the existing literature on diffusion models and the Schrödinger Bridge problem, providing context and background for our proposed method. The Method section introduces the Explainable Diffusion Model via Schrödinger Bridge in Multimodal Image Translation framework, detailing its formulation, optimization, and implementation. The Experiment section describes the datasets used, experimental setup, and results of our image translation tasks, comparing the performance of our approach with existing methods. The last section concludes with a summary of the findings, highlighting the contributions and potential future work in this domain.

2 Related Work

The field of image translation has seen significant advancements with the development of various generative models. Among these, diffusion models and score-based generative models have gained prominence due to their effectiveness in producing high-fidelity images. This section reviews the key contributions in these areas, setting the stage for our proposed Explainable Diffusion Model via Schrödinger Bridge in Multimodal Image Translation framework.

2.1 Diffusion Model

A diffusion model is a class of generative models that simulates the process of gradually adding noise to data until it becomes indistinguishable from random

noise, and then learns to reverse this process to generate data from noise [20]. The concept is inspired by the physical process of diffusion, where particles move from areas of higher concentration to lower concentration, leading to an equilibrium state. In the context of machine learning, diffusion models are trained through a sequence of forward steps that progressively corrupt the data with noise, followed by a learned reverse process that aims to reconstruct the original data from the noisy state. The reverse process is typically achieved through a neural network that is trained to predict earlier, less noisy states of the data given its current state. Diffusion models have shown remarkable success in generating high-quality, diverse samples in various domains such as images, audio, and text, distinguishing themselves by their ability to model complex data distributions and produce outputs with high fidelity and variation [7, 8].

2.2 Schrödinger Bridge

The Schrödinger Bridge (SB), dating back to Schrödinger [18] and revisited by Léonard [12], is conceptualized as an entropy-regularized variant of the classical optimal transport problem. This framework encompasses the subsequent stochastic differential equations (SDEs):

$$dX_t = [f_t + \beta_t \nabla \log v(X_t, t)] dt + \sqrt{\beta_t} dW_t \quad (1)$$

$$d\tilde{X}_t = [\tilde{f}_t - \beta_t \nabla \log u(\tilde{X}_t, t)] dt + \sqrt{\beta_t} d\tilde{W}_t \quad (2)$$

where f is the drift term, W is the Wiener process, β is a constant, $X_0 \sim p_A$ and $X_1 \sim p_B$ are distributed according to the boundary conditions in two discrete domains. The potentials v and u , belonging to the space $C^{1,2}(\mathbb{R}^d \times [0, 1])$, are dynamic entities governed by the ensuing coupled partial differential equations (PDEs). Here, the superscript 1, 2 indicates that the functions have continuous first-order derivatives with respect to time and continuous second-order derivatives with respect to spatial variables.

$$\frac{\partial v(x, t)}{\partial t} = -\nabla \cdot (f_t v) + \frac{1}{2} \beta_t \Delta v \quad (3)$$

$$\frac{\partial u(x, t)}{\partial t} = -\nabla \cdot (\tilde{f}_t u) + \frac{1}{2} \beta_t \Delta u \quad (4)$$

subject to the initial and terminal distribution conditions:

$$v(x, 0) = p_A(x), \quad u(x, 1) = p_B(x) \quad (5)$$

In the formulation Eqs. (1) to (5), the path probability measures induced by the SDEs in (3) and (4) coincide almost surely, reminiscent of the equivalence established in the earlier SDEs. Consequently, the marginal probability densities, hereinafter denoted by $q(\cdot, t)$, are correspondingly equivalent.

The insights from existing diffusion models, score-based generative models, and the Schrödinger Bridge problem will inform the development of xDSB-MIT. By integrating the stability and interpretability features of the Schrödinger Bridge with the robust image generation capabilities of diffusion models, our approach aims to enhance multimodal image translation. The related works provide foundational principles and techniques that are crucial in formulating and optimizing the Explainable Diffusion Model via Schrödinger Bridge in Multimodal Image Translation framework, which we detail in the subsequent method section.

3 Method

In this section, we describe the proposed Explainable Diffusion Model via Schrödinger Bridge (xDSB) for multimodal image translation. xDSBMIT integrates the diffusion process with the Schrödinger Bridge framework to enhance the stability and interpretability of image translations across different modalities. We will incorporate linking statements between sections and include relevant theories on multimodal image distributions. Because the Schrödinger Bridge framework provides a clear mathematical formulation of the transport process, making the transformation from source to target distributions more transparent, so we refer this as an explainable diffusion model. This approach allows for better understanding and interpretability of how the model transitions between different imaging modalities, addressing concerns about the “black-box” nature of traditional deep learning models.

3.1 Preliminaries

We begin by introducing the fundamental principles underlying our approach. The xDSB leverages the probabilistic nature of diffusion models to transition between image distributions. Specifically, xDSB models the distribution of paired images in the source and target domains, ensuring that essential features of the source images are preserved in the translation process. This is achieved through the Schrödinger Bridge problem, which formulates a continuous path between two probability distributions. The path minimizes the Kullback-Leibler (KL) divergence between the distributions, leading to an optimal transport solution.

To formalize this, let $\mu \in \mathcal{P}_{N+1}$ represent the distribution sequence of diffusion paths, with $\mu_0 = p_A$ and $\mu_N = p_B$ indicating the source and target distributions, respectively. The objective is to find μ^* that satisfies:

$$\mu^* = \arg \min_{\mu} \{ \text{KL}(\mu \| \mu_{\text{ref}}) : \mu \in \mathcal{P}_{N+1}, \mu_0 = p_A, \mu_N = p_B \}. \quad (6)$$

In multimodal contexts, the distribution of paired images involves diverse modalities [9,11], such as SAR and IR, ensuring the translation preserves the modality-specific features.

3.2 Training Algorithm

The xDSB training algorithm follows the iterative proportional fitting procedure (IPF) [6] to refine the distribution sequence μ . The goal is to adjust μ iteratively until convergence, ensuring that $\mu_0 \approx p_A$ and $\mu_N \approx p_B$. The IPF updates are given by:

$$\mu^{2n+1} = \arg \min_{\mu} \{ \text{KL}(\mu \| \mu^{2n}) : \mu \in \mathcal{P}_{N+1}, \mu_N = p_B \} \quad (7)$$

$$\mu^{2n+2} = \arg \min_{\mu} \{ \text{KL}(\mu \| \mu^{2n+1}) : \mu \in \mathcal{P}_{N+1}, \mu_0 = p_A \}. \quad (8)$$

Each iteration alternates between optimizing the distribution at the source and target ends, gradually refining the transition path to minimize the overall divergence. This method accounts for the multimodal nature of image distributions by considering the unique characteristics of each modality in the optimization process.

3.3 Static Schrödinger Bridge

For a static version of the problem, we consider the entropy-regularized optimal transport, which links the Schrödinger Bridge problem with traditional optimal transport theory. The SB approach ensures the convergence of the distribution sequence by balancing the entropy terms with the transport cost:

$$\mu_{\text{static}}^* = \arg \min_{\mu} \{ E_{\mu} [\|x_0 - x_N\|^2] - 2\sigma^2 H(\mu) : \mu \in \mathcal{P}_2, \mu_0 = p_A, \mu_N = p_B \}. \quad (9)$$

3.4 Diffusion Schrödinger Bridge

The Diffusion Schrödinger Bridge (DSB) combines the dynamic aspects of diffusion processes with the optimal transport properties of the Schrödinger Bridge. In DSB, the forward and backward transition probabilities are updated iteratively to ensure convergence towards the equilibrium state [22]. The forward and backward passes are given by:

$$p_{t+1|t}(x_{t+1}|x_t) = N(x_{t+1}; x_t + \gamma_t f_t(x_t), 2\gamma_t I), \quad (10)$$

$$q_{t|t+1}(x_t|x_{t+1}) = N(x_t; x_{t+1} + \gamma_t b_{t+1}(x_{t+1}), 2\gamma_t I). \quad (11)$$

where γ_t is a constant and I is the identity matrix. The training loss functions for DSB are defined to minimize the discrepancies between the forward and backward transitions:

$$L_{t+1}^B = E_{(x_{t+1}, x_t) \sim p_{t+1,t}^n} [\|B_{t+1}^n(x_{t+1}) - (x_{t+1} + F_t^n(x_t) - F_t^n(x_{t+1}))\|^2], \quad (12)$$

$$L_{t+1}^F = E_{(x_t, x_{t+1}) \sim q_{t,t+1}^n} [\|F_t^n(x_t) - (x_t + B_{t+1}^n(x_{t+1}) - B_{t+1}^n(x))\|^2]. \quad (13)$$

where B and F are two learnable neural networks. The diffusion process of DSB optimizes towards a static goal.

3.5 Algorithm Design

The Explainable Diffusion Model via Schrödinger Bridge (xDSB) integrates diffusion models with the Schrödinger Bridge framework to enhance the stability and interpretability of multimodal image translation. Diffusion models simulate the gradual addition of noise and learn to reverse this process through denoising steps, generating high-quality images. Score-based generative models iteratively refine samples using the score function, the gradient of the log probability density. The Schrödinger Bridge provides an optimal transport solution by creating a continuous path between two probability distributions, minimizing the Kullback-Leibler divergence. In xDSB, iterative proportional fitting (IPF) refines the distribution sequence to ensure convergence towards an optimal transport path [16]. The model combines static entropy-regularized optimal transport with dynamic diffusion processes, optimizing transitions using neural networks. This approach ensures equilibrium states, enhancing the quality and efficiency of image translations across different modalities. The algorithm design is as follows:

Algorithm 1. Training the xDSB Model

- 1: **Input:** $p_A(\cdot)$ and $p_B(\cdot|X_0)$ datasets
 - 2: **Initialization:** Initialize μ and model parameters θ
 - 3: **repeat**
 - 4: Sample $t \sim \mathcal{U}([0, 1])$, $X_0 \sim p_A(X_0)$, $X_1 \sim p_B(X_1|X_0)$
 - 5: Compute $X_t \sim q(X_t|X_0, X_1)$ according to the Schrödinger Bridge formulation
 - 6: Update μ using iterative proportional fitting (IPF)
 - 7: Perform gradient descent step on $\epsilon(X_t, t; \theta)$
 - 8: **until** convergence
-

Algorithm 2. Generating Images with the Trained xDSB Model

- 1: **Input:** $X_N \sim p_B(X_N)$, trained $\epsilon(\cdot, \cdot; \theta)$
 - 2: **for** $n = N$ to 1 **do**
 - 3: Predict X_0^ϵ using $\epsilon(X_n, t_n; \theta)$
 - 4: Sample $X_{n-1} \sim p(X_{n-1}|X_0^\epsilon, X_n)$ according to the trained model
 - 5: **end for**
 - 6: **Output:** X_0
-

4 Experiment

4.1 Model Training and Testing

Using an NVIDIA RTX8000 GPU, xDSBMIT demonstrated significant efficiency compared to DDPM. During the training phase, xDSBMIT had notably lower runtimes, as shown in Fig. 2a, it took 4.3 h to train on 1000 data points, whereas DDPM required 7.5 h. Similarly, in the testing phase, as shown in Fig 2b, xDSBMIT completed testing in 2.5 h for 1000 data points, compared to 3.8 h for DDPM. Overall, xDSBMIT exhibited superior efficiency in both training and testing, highlighting its advantages for real-world applications.

4.2 Dataset

In our study, we utilize a comprehensive collection of datasets spanning different imaging modalities to support our research on multi-modal image translation. Synthetic Aperture Radar (SAR) images capture radar signal representations of the Earth’s surface, providing valuable data under conditions where optical sensors might fail, such as in adverse weather. Infrared (IR) images offer critical insights into thermal properties of the landscape, important for environmental monitoring. Electro-optical (EO) sensors contribute images in the visible spectrum, including both RGB and grayscale images, which are predominantly used in computer vision applications due to their detailed representation of visible light information. Lastly, the RGB images provide high-resolution, color imagery of agricultural and urban landscapes. The data in our experiments are sourced from the UNICORN dataset and the PBVS 2024 public competition. The UNICORN dataset comprises paired SAR and EO data, while the dataset provided in the PBVS 2024 open competition includes paired SAR, IR, and RGB data, as illustrated in Fig. 2c. These datasets form a robust foundation for exploring and enhancing techniques in image translation across various modalities.

4.3 SAR2IR

The SAR2IR task utilizes the dataset provided in the PBVS 2024 open competition. As shown in Fig. 2, we demonstrated the feasibility of translating Synthetic Aperture Radar (SAR) images into Infrared (IR) imagery. The translated IR images effectively reconstructed the primary contours and structural details present in the original SAR data, as evidenced by the image sequence in the middle column of Fig. 2(d). However, despite these promising results, the translated images exhibited diminished brightness and were unable to capture some finer details compared to the original IR images. These observations suggest that while the approach is effective in capturing major features, further refinement is needed to enhance the detail fidelity and brightness levels of the translated images, thus improving their utility for practical applications in remote sensing.

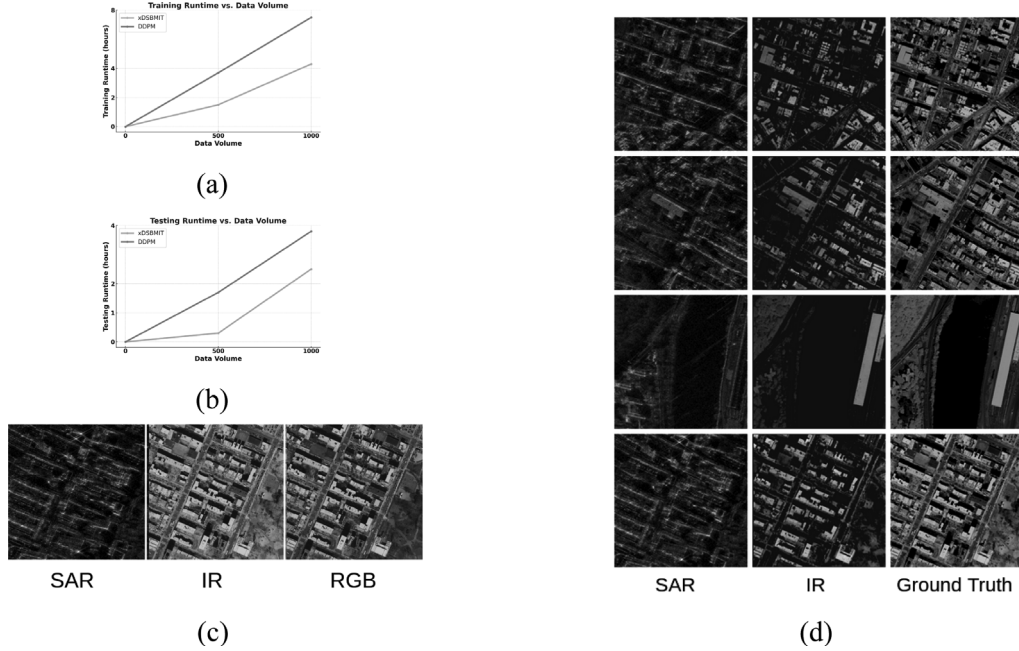


Fig. 2. Combined Figures: (a) Training runtime, (b) Testing runtime, (c) Paired Images: SAR, IR and RGB, (d) SAR2IR translation. Left: SAR. Middle: IR generated via translation. Right: Ground Truth of IR.

4.4 SAR2EO

In the subsequent SAR2EO translation experiment, notable advancements were achieved using a relatively modest dataset of only 500 training images from UNICORN dataset. Despite the limited data, our approach surpassed the performance benchmarks set by well-established frameworks such as pix2pix, pix2pixHD, and traditional GAN models. The results indicate a significant improvement not only in the accuracy of the translated EO images but also in the clarity and color fidelity. This breakthrough demonstrates the potential of our model to efficiently learn and generalize from sparse datasets, outperforming existing methods in both qualitative and quantitative evaluations. The successful application with minimal training data underscores our model’s robustness and efficiency, suggesting it as a highly effective tool for enhancing EO image generation in remote sensing technologies. Table 1 provides a performance comparison of different image translation methods for SAR2EO. Our model, referred to as EDSB-500, exhibits superior performance across both LPIPS and FID metrics. Specifically, EDSB-500 achieved an LPIPS score of 0.35 and an FID score of 0.10, outperforming the GAN-500, pix2pix-500, and pix2pixHD-500 models, which recorded higher LPIPS and FID scores. LPIPS (Learned Perceptual Image Patch Similarity) measures the perceptual similarity between generated and real images. The LPIPS is constructed based on the VGG-16 [19] architecture. Lower LPIPS scores indicate higher perceptual quality, as the generated images are closer to the real ones in terms of human visual perception. In our experiment,

the EDSB-500 model achieved the lowest LPIPS score, suggesting that it produces more perceptually accurate images compared to the other models. FID (Fréchet Inception Distance) evaluates the quality of generated images by comparing the distributions of real and generated image features extracted by a pre-trained Inception network [21]. Lower FID scores indicate that the generated images have a distribution closer to the real images, thus reflecting higher quality. The EDSB-500 model achieved the lowest FID score, indicating a significant improvement in image quality and fidelity over the other models. These results highlight the effectiveness of our model in generating high-quality EO images from SAR data, even with a limited training dataset.

Table 1. Performance comparison of different image translation methods for SAR2EO

	LPIPS	FID
GAN-500	0.52	0.44
pix2pix-500	0.48	0.27
pix2pixHD-500	0.45	0.18
xDSBMIT-500	0.35	0.10

5 Conclusion

This study introduces the Explainable Diffusion Model via Schrödinger Bridge for Multimodal Image Translation (xDSBMIT). Our method integrates diffusion models with the Schrödinger Bridge to enhance the stability and interpretability in image translation, achieving superior performance with limited datasets. Our experimental results demonstrate the effectiveness of xDSBMIT in translating between various image modalities, such as Synthetic Aperture Radar (SAR), Infrared (IR), and Electro-Optical (EO) images.

A key contribution is the application of DSB to the multimodal domain, offering clear insights into the image translation process, crucial for remote sensing and medical imaging. Our method also demonstrates robust performance with fewer training pairs compared to methods like pix2pix, making it valuable where large datasets are impractical. Comprehensive evaluations reveal xDSBMIT’s adaptability across imaging modalities, producing high-fidelity, coherent results.

Furthermore, xDSBMIT aligns with the Dynamic Data Driven Applications Systems (DDDAS) paradigm by dynamically integrating runtime data into the models. This dynamic integration, which is applicable across different modalities, enhances the system’s adaptability and responsiveness, epitomizing the core principles of DDDAS. By supporting advanced systems-analytics capabilities, xDSBMIT addresses the limitations of traditional diffusion approaches that often overlook dynamic system conditions.

In conclusion, xDSBMIT combines stability, interpretability, and efficiency, showing potential for advancing remote sensing and image translation across various domains. Future work will explore further enhancements and applications of this approach.

Acknowledgments. This work was supported by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-21-1-0224.

References

1. Blasch, E.P., Darema, F., Ravela, S., Aved, A.J. (eds.): Handbook of Dynamic Data Driven Applications Systems, vol. 1, 2nd edn. Springer (2022)
2. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: International Conference on Learning Representations (2019)
3. Chen, Z., He, G., Zheng, K., Tan, X., Zhu, J.: Schrödinger bridges beat diffusion models on text-to-speech synthesis. arXiv:2312.03491 (2023)
4. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: an overview. *IEEE Signal Process. Mag.* **35**(1), 53–65 (2018)
5. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. In: Advances in Neural Information Processing Systems, vol. 34, pp. 17695–17709 (2021)
6. Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11**(4), 427–444 (1940). <https://doi.org/10.1214/aoms/1177731829>
7. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. In: Advances in Neural Information Processing Systems (2021)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (2020)
9. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 172–189 (2018)
10. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
11. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: International Conference on Machine Learning, pp. 1857–1865 (2017)
12. Léonard, C.: A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete Continuous Dyn. Syst.* (2013)
13. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E.A., Nie, W., Anandkumar, A.: I2SB: image-to-image schrödinger bridge. arXiv:2302.05872 (2023)
14. Liu, S., Gao, M., John, V., Liu, Z., Blasch, E.: Deep learning thermal image translation for night vision perception. *ACM Trans. Intell. Syst. Technol.* **12**(1), 1–18 (2020)

15. Low, S., Nina, O., Sappa, A.D., Blasch, E., Inkawhich, N.: Multi-modal aerial view image challenge: translation from synthetic aperture radar to electro-optical domain results-PBVS 2023. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 515–522 (2023)
16. Rüschen-dorf, L., Thomsen, W.: Note on the schrödinger equation and i-projections. *Stat. Probab. Lett.* **17**(5), 369–375 (1993)
17. Savakis, A., Taufique, A.M.N., Nagananda, N.: Advances in domain adaptation for aerial imagery. In: Darema, F., Blasch, E.P., Ravela, S., Aved, A.J. (eds.) *Handbook of Dynamic Data Driven Applications Systems*. Springer, Cham (2023)
18. Schrödinger, E.: über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften Physikalisch-Mathematische Klasse* (1932)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations* (2015)
20. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: *Proceedings of the International Conference on Machine Learning* (2015)
21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. *CoRR* (2015)
22. Tang, Z., Hang, T., Gu, S., Chen, D., Guo, B.: Simplified diffusion schrödinger bridge. *arXiv:2403.14623* (2024)
23. Zheng, Y., Blasch, E., Liu, Z.: *Multispectral Image Fusion and Colorization*. SPIE Press (2018)
24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision* (2017)
25. Özbey, M., et al.: Unsupervised medical image translation with adversarial diffusion models. *IEEE Trans. Med. Imaging* **42**(12), 3524–3539 (2023)