

Cross-Domain Generalization: Enhancing Rare Disease Data Representation using Diffusion Model

Wonseok Oh
Electrical and Computer Engineering
University of Michigan
okong@umich.edu

Abstract

We introduce a Schrödinger Bridge method to generate datasets from diverse domains. This enables the collection of data for rare diseases and limited datasets. Therefore, in this paper, we introduce **Domain Knowledge Diffusion Models (DKDM)** using Schrödinger Bridge to generate rare disease images and limited data in medical imbalance datasets. Our methods demonstrate the capability to generate cross-domain generalized images for rare diseases beyond utilizing a single-domain dataset by training the separated model with domain datasets. Also, our method surpasses the existing Schrödinger Bridge diffusion models by using domain phase loss. Furthermore, we show that utilizing images generated from our method is more competitive than training with existing methods.

1. Introduction

Deep learning has achieved remarkable progress in computer-aided diagnosis.[22, 28, 29, 34, 43] Recently, numerous studies utilizing diffusion methods have demonstrated successive success. There are two main ways to use diffusion methods in medical imaging. Firstly, anomaly detection utilizes methods that combine deterministic iterative noising and denoising schemes.[2, 16, 40, 45] Secondly, by using rare disease data, the diffusion model generates images of these domain.[6, 19]. However, these methods show significant performance yet rely on training with large datasets. Many attempts have been made to train the generative model with limited data[6, 26, 40], but it still requires large datasets or only single domain datasets. By utilizing limited data or multi-domain datasets, training a generative model with good quality and high fidelity seems limited. Domain adaption and cross-domain generalization approaches show their generalization ability across diverse datasets.[5, 15, 41] Specifically, domain adaption is a commonly used technique that fine-tunes gen-

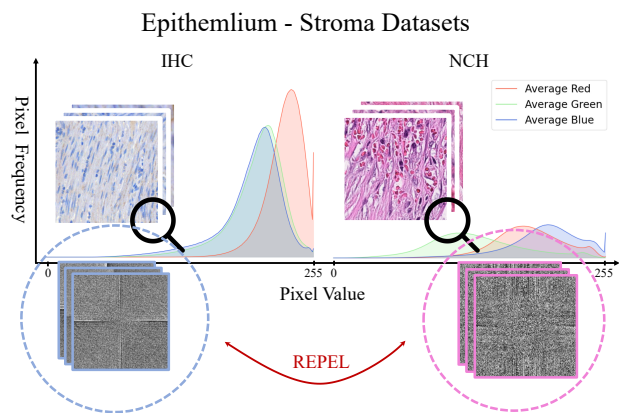


Figure 1. Following RGB statistics, each histopathology dataset shows diverse datasets/domains. Visual differences, different collected methods, intensity differences, and diverse patterns can be observed from RGB statistics. We also observed that the phase terms of data within the same domain are quite similar during Fourier transformation. However, when comparing data from different domains but belonging to the same class, we found significant differences in the phase terms.

erators and discriminators on target datasets. These are usually based on the generative models. In particular, after training the model with the source domain, a few parameters are fine-tuned, and regularization methods are added to transfer knowledge about the target domain.[6, 40] However, resulting diversity is typically less semantic and produces only very similar images. Therefore, more advanced methods are required for good quality and high fidelity. For this reason, we conducted training models with various domain datasets using the Schrödinger Bridge. This is used to generate cross-domain generalized images to balance datasets for imbalanced datasets. Here, we find that transferring the trained model with the source domain directly to target-domain knowledge results in a loss of prior knowledge. To address problems, we utilize domain phase loss to minimize the cosine similarity of the reconstructed

image and target domain. Since similar phases are observed in datasets in the same domain, the method of reducing loss in the domain transfer process is meaningful in creating cross-domain data. Also, we explore why diversity drops when utilizing diversity datasets and how to produce highly diverse images in multi-domain datasets. To that end, we introduce **Domain Knowledge Diffusion Models(DKDM)** using cross-domain generalization technique to generate rare disease images with limited data in medical imbalance datasets. The idea behind **DKDM** is to learn various domain representations. Compared to the conventional strategy, since utilizing design loss for domain generalization, **DKDM** can produce high-fidelity images without catastrophic forgetting problems for source domain datasets. Specifically, we demonstrate that **DKDM** is competitive for generating images compared to existing methods. Also, When the training model utilizes the existing dataset with generated images, we show competitive performance. Through extensive experiments, our contributions follow that:

- Our methods demonstrate the capability to generate cross-domain generalized images for rare diseases beyond utilizing a single-domain dataset by training the model with multiple diverse domain datasets.
- Domain phase loss resolves the issue of losing prior knowledge when directly transferring a trained model from the source domain to the target domain.
- We show competitive performance among the existing unpaired image-to-image translation models.

2. Related Works

2.1. Unpaired image-to-image translation

Image-to-image (I2I) translation aims to learn how to map between source and target domains. Initially, the method, utilizing paired I2I, maps an image from the input to the output domain by utilizing an adversarial loss. These methods are augmented by measuring the similarity between the generated result and the target image. Furthermore, the Cycle consistency approach, which enforces the translation to preserve important properties of the source domain data, has emerged. Inspired by the CycleGAN[46], which undertook the task with unpaired training images, several methods are proposed by utilizing unpaired adversarial learning strategy[8, 18, 20]. However, these methods are vexed by asymmetrical[3] in domain knowledge between hazy and clear images and training due to the necessity of additional generator training. To address these problems, these methods[18, 27] based on I2I models have shifted their focus to one-sided I2I translation, which utilizes Schrödinger Bridge and ODEsolver. Recently, Contrastive Unpaired Translation[21] proposed a multi-layer patch-based contrastive learning approach and demonstrated improvements

in I2I. While DRIT and SwapAE[13, 23] proposed disentangling representations into style and content. These GAN-based methods have been extended to the medical domain. To balance the data for all classes, various works have been tried, such as MRI images [8], cell staining[20], and COVID-19 images[7]. Despite the continued success of GAN-based methods in the medical I2I domain, Our proposed method enhances existing methods by iterative refinement through the Schrödinger Bridge, addressing the collapse problem.

2.2. Diffusion Models

Recently, diffusion methods have achieved successive success. Those methods encompass a broad family, including VAEs[12, 25], Markov Chains[24, 31], and score matching[32, 33], etc. The diffusion model aims to gradually recover ground truth signal $x_0 \sim p(x_0)$ added random noise $\epsilon_t \sim N(0, I)$ to desired images. To be more specific, the forward diffusion process $p(x_T|x_0)$ utilizes a Markov Chain that gradually mitigates x_0 to x_T with random Gaussian noise.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (1)$$

where $\beta_t \in (0, 1)$ is the noise scale. Following the noise scheduler, β_t increases as the timestep grows, and finally, ground truth images are completely covered with noise.

$$\begin{aligned} q(x_t|x_0) &= \prod_{t=1}^T \mathcal{N}(x_t|\sqrt{1-\beta_t}x_{t-1}, \beta_t I) \\ &= \mathcal{N}(x_t|\sqrt{\alpha_t}x_0, \sqrt{1-\alpha_t}I) \end{aligned} \quad (2)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. The diffusion model $\epsilon_\theta(x_t, t)$ is training to estimate ϵ_t from x_t , by gradually remove noise from the x_t . The backward process, commonly known as

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}|\mu(x_t, t), \Sigma_\theta(x_t, t)) \quad (3)$$

However, existing diffusion methods don't consider domain shifts in datasets. Thus, it is limited to improving model training with generated images from existing methods to fill rare disease data and imbalanced data. To address the problem, we utilize a designed loss function for generating domain-generalized images.

2.3. Schrödinger Bridge (SB)

A Schrodinger bridge extension of score-based generative models (SGMs) has been introduced to transfer from an initial distribution to a terminal distribution over time. It is closely related to probability theory and stochastic control. To approximate score-based generative models (SGMs), [4, 38] introduce leveraging Iterative Proportional Fitting

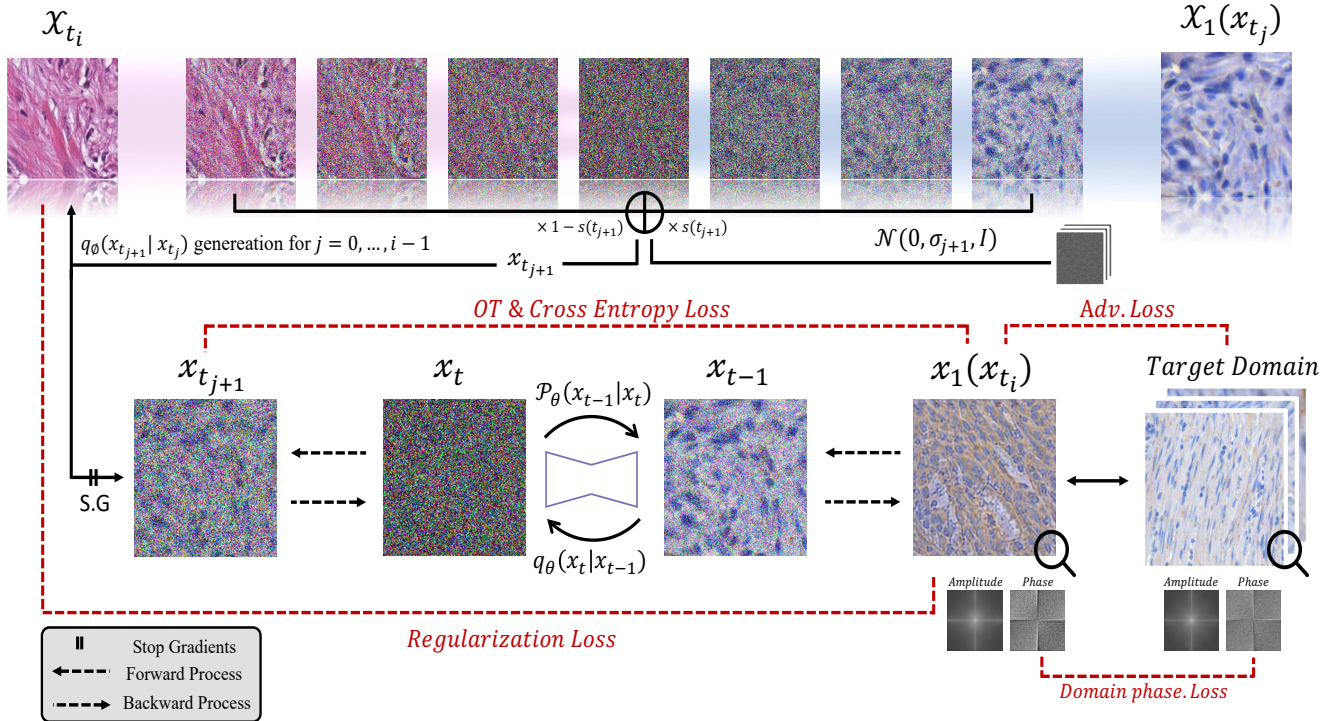


Figure 2. The process begins with an initial set of sample images x_{t_i} . These images are subjected to a series of transformations involving the addition of noise through a process defined by the function q_ϕ . As the images pass through each step, they become increasingly noisy, simulating a Markov chain. The images are then processed through a stochastic gradient (S.G) operation and a backward process that aims to recover the original image features from the noise, indicated by the function $p_\theta(x_{t-1}|x_t)$. Throughout this transformation process, several loss functions are employed.

(IPF) algorithm and [14, 33] introduced similar algorithms. With the successive success of research utilizing SB, several variants have come on to the diverse stage, such as Probabilistic Lambert Problem, inverse problems[30], Mean-Field Games[17], constrained transport problems[36], Riemannian manifolds[37], and path samplers[42, 44]. [39] investigate entropy interpolation between Dirac delta and noisy data with SB in the unsupervised setting. Otherwise, in the supervised setting, I2SB[18] and InDI[11] used paired data to learn SBs between Dirac delta and data while finding the continuous path. Recently, DDIB[35] tried to not unpair data but concentrate two SBs between different domain pairs data and demonstrated utility in a wide variety of translation tasks. However, to the best of our knowledge in the Medical community, the SB problem has not been investigated between Unpaired image-to-image translation. Thus, our work endeavors to demonstrate that inherent optimal transport properties are achievable in various aspects by utilizing paired or unpaired images. Also, our work tried to adapt various datasets to demonstrate scalability.

3. Methods

Our model is based on an elaborate Schrödinger Bridge, which elucidates that SB can be articulated as a sequential integration of generators determined via adversarial learning paradigms. More precisely, considering a partition $\{t_i\}_{i=0}^N$ of the unit interval $[0, 1]$ with $t_0 = 0$, $t_N = 1$, and $\tilde{x}_0 = x_{t_0}$, $\tilde{x}_N = x_{t_N}$, we can represent sb via the Markov chain decomposition.

$$p(\{\tilde{x}_n\}) = p(\tilde{x}_N|\tilde{x}_{N-1})p(\tilde{x}_{N-1}|\tilde{x}_{N-2}) \cdots p(\tilde{x}_1|\tilde{x}_0)p(\tilde{x}_0) \quad (4)$$

Through decomposition, we learn $p(\tilde{x}_{i+1}|\tilde{x}_i)$ presuming we can sample from $p(\tilde{x}_i)$ (for $i = 0, \dots, N-1$). Thus, we can learn $p(\tilde{x}_{i+k}|\tilde{x}_{i+(k+1)})$ and so forth. Consider $q_{\phi_i}(\tilde{x}_1|\tilde{x}_i)$, a conditional distribution orchestrated by a DNN with parameters ϕ_i . This is also the denoising and generation step, which estimates the target domain image x_i . Therefore, here, we optimize ϕ_i with the arbitrary step i .

3.1. Model operations of DKDM

Model Description Our model embarks on optimizing a loss function for a randomly selected time step t_i during

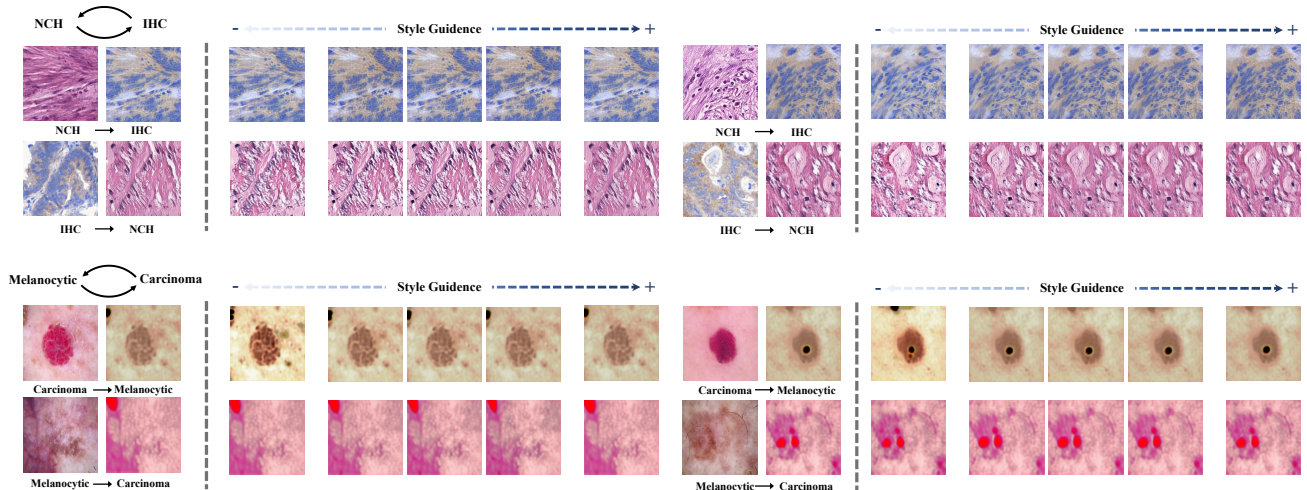


Figure 3. For qualitative evaluation of image-to-image translation results from our DKDM, we transfer input images into different medical domain images and style images. We also show successive refinements of the predicted target domain image.

the training stage. The inception of the model involves sampling an image x_{t_i} from the initial domain and a corresponding image from the target distribution denoted by x_1 . The target distribution is signified by π_1 , and the details of the sampling procedure for x_t will be elucidated shortly.

The sampled image x_t is subsequently processed through a transformation function $q_\phi(x_1|x_t)$, yielding $x_1(x_t)$, which is an estimation of the target domain data given x_t . The pairs $(x_t, x_1(x_t))$ and $(x_1, x_1(x_t))$ are utilized to compute the $\mathcal{L}_{SB}(\phi, t_i)$ and the Adversarial Loss $\mathcal{L}_{Adv}(\phi, t_i)$. Here, $\mathcal{L}_{SB}(\phi, t_i)$ measures the discrepancy between the learned distribution q_ϕ and the true data distribution in the context of a time step t_i in a Markov chain. Specifically, the \mathcal{L}_{SB} is defined as an expectation over the data distribution of the squared Euclidean distance between x_{t_i} and $x_{t_{i+1}}$, regularized by a term involving the entropy of the learned distribution q_ϕ . The entropy component in \mathcal{L}_{SB} is estimated through a mutual information estimator, leveraging the relationship $I(X, X) = H(X)$ for a random variable X , where I represents mutual information. The divergence in \mathcal{L}_{Adv} is evaluated utilizing adversarial learning techniques, with x_1 and $x_1(x_t)$ serving as the "real" and "fake" inputs to the discriminator.

Sampling Procedure The intermediate and final samples are generated through a procedure that simulates a Markov chain, as described above, using the transformation function q_ϕ . Commencing with x_t , we predict the image in the target domain $x_1(x_t)$ through iterative sampling and the application of Gaussian noise, facilitating the generation of $x_{t_{j+1}}$. This iterative procedure enhances the prediction of the target domain sample, thereby refining it through the trajectory $\{x_1(x_t) : i = 0, \dots, N - 1\}$.

Process Illustration Figure 2 encapsulates the generation stage of the model, elucidating the transformation from the original domain X_{t_i} to the target domain $X_1(x_{t_j})$. The illustration conveys the integration of various machine learning paradigms, including stochastic processes, adversarial training, regularization, and potentially Fourier analysis, to effectuate this domain transformation.

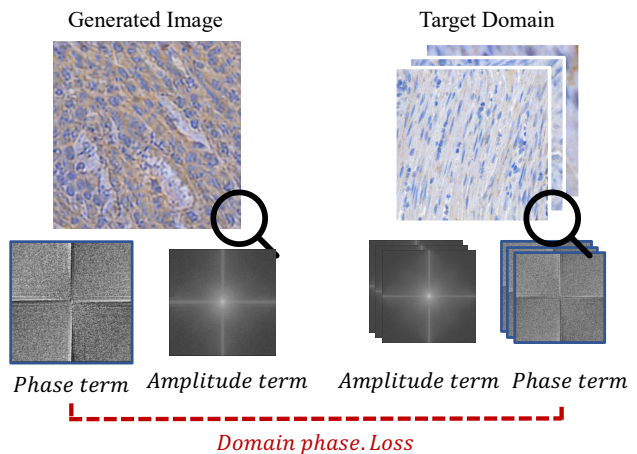


Figure 4. We employ Fourier transformation on both the generated and target domain images to analyze texture. The loss function applied to the model utilizes cosine similarity in the Phase term, representing the texture. However, the Amplitude term is not utilized.

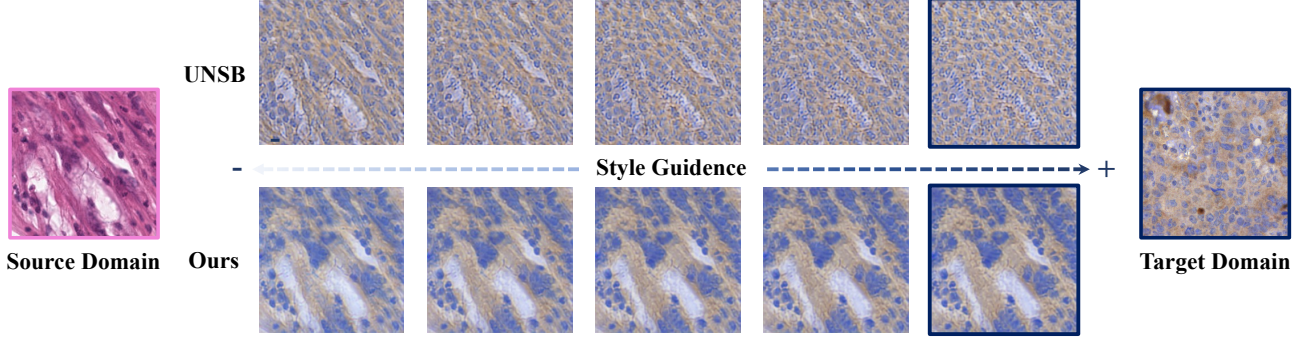


Figure 5. DKDM can be possible to iteratively enhance the anticipated image of the target domain, allowing the model to adjust intricate elements while maintaining the texture.

3.2. Resolving Prior Knowledge Loss

If datasets with a very large difference between domains are generalized, plenty of existing information on images is lost when translating input data from the source domain to the target domain. This phenomenon occurs when new images are filled only with information from other domains while information is put in a specific domain. Our goal was to solve this problem. Since medical data have biases for each data, there is a big difference between domains. We used the Fourier transform for each dataset to show amplitude and phase. As a result, as can be seen in the 4, it was confirmed that data belonging to the same domain have similar amplitude and phase, and in the case of data in different domains, the difference was found to be very large. Based on this, we designed a domain phase loss function for optimization.

$$\mathcal{L}_{DP} = \frac{\sum_{i,j} \text{Phase}(\mathcal{X})_{i,j} \cdot \text{Avg}(\hat{X})_{i,j}}{\sqrt{\sum_{i,j} (\text{Phase}(\mathcal{X})_{i,j})^2} \cdot \sqrt{\sum_{i,j} (\text{Avg}(\hat{X})_{i,j})^2}} \quad (5)$$

where \mathcal{X} means generated image and \hat{X} means Average for getting phase term numerical value from target domains. In short, the domain phase loss function measures the cosine similarity of the phase of the generated image and the phase of the average target domain.

We also add regularization loss to make the final objective for **DKDM**. To further refine the **DKDM** objective, regularization is introduced to compel the generator network q_ϕ to uphold a consistency between the predicted outcome x_1 and the initial state x_0 :

$$\mathcal{L}_{Reg}(\phi, t_i) = \mathbb{E}_{p(x_0, x_t)} \mathbb{E}_{q_\phi(x_1|x_t)} [S(x_0, x_1)] \quad (6)$$

In this context, S denotes a scalar, differentiable function that encapsulates a domain-specific measure of resemblance between its two inputs. Essentially, S encodes our preconceived notion of similarity across image pairs. We can de-

rive the final loss function here by incorporating the previously introduced Adversarial Loss and Schrödinger Bridge Loss term. Consequently, the amended **DKDM** objective at time t_i can be stated as:

$$\mathcal{L}_{DKDM}(\phi, t_i) = \mathcal{L}_{Adv}(\phi, t_i) + \lambda_{DP, t_i} \mathcal{L}_{DP}(\phi, t_i) + \lambda_{SB, t_i} \mathcal{L}_{SB}(\phi, t_i) + \lambda_{Reg, t_i} \mathcal{L}_{Reg}(\phi, t_i) \quad (7)$$

This is the definitive goal within our **DKDM** scheme.

4. Experiments

4.1. Detail of datasets

We evaluate the proposed method with multiple Epithelium-Stroma, CheXpert, and ISIC2019 datasets. Epithelium-Stroma datasets(e.g IHC[1] and NCH[10]) are common publicly available datasets. The datasets were collected from different methods and institutions, which caused the domain shift among them, and they were labeled as epithelium or stroma. The IHC[1] datasets contain 10015 histopathological images, and NCH[1] datasets contain a total of 8015 images. We generate stylized images of the epithelium class in the IHC dataset within the NCH dataset, corresponding to the same class. We also apply the reverse scenario. The CheXpert dataset is composed of 14 classes with 224,319 images. Due to limitations in incorporating all available data into the training, we opted to sample the image for experimentation purposes. The ISIC2019 dataset has 8 classes with 25,331 images and can be found publicly. For qualitative comparison, we also utilize the ISIC2019 dataset to demonstrate the leverage for transfer among images of different classes within pairs that belong to the same domain.

4.2. Implementation details

Our proposed methods use Markovian discriminator in \mathcal{L}_{Adv} and utilize patch-wise contrastive matching in \mathcal{L}_{Reg}

Datasets	class	samples	imbalance ratio	domain
IHC	2	820	1.5	•
NCH	2	24,763	1.3	•
ISIC2019	8	25,331	53.8	
CheXpert	14	224,316	37.9	

Table 1. The details of medical datasets. We utilized IHC and NCH for the unpaired image-to-image transfer task, while applying class-wise transfer to CheXpert and ISIC2019.

For training DKDM network for 10 epochs with batch size 1 and Adam optimizer with $\mathcal{B}_1 = 0.5, \mathcal{B}_2 = 0.9999$. We also set $\lambda = 0.0002$ for the initial learning rate, learning decay 500, and $\mathcal{L}_{dp} = 0.1$ for Domain phase loss. We resized the input image 256×256 and normalized it into the range $[-1, 1]$. For SB training and simulation, we divided the unit interval $[0, 1]$ into 5 uniform space intervals with uniform spacing. We adopt entropy estimation methods [Belghazi et al., 2018] to our model. We apply the data augmentation method following previous work. However, unlike previous studies, we applied the mentioned settings consistently to both the Epithelium-Stroma and ISIC2019 datasets. We used the PyTorch library with a single V100 GPU to train Epithelium-Stroma and ISIC2019 datasets, which took around 3 and 5 hours.

5. Ablation Studies

5.1. Qualitative Analysis

To substantiate the continual improvement of our proposed methodology in achieving a seamless transformation of texture to the target domain, avoiding over-fitting compared to conventional methods, we have undertaken the visualization of consecutively transformed images. This visualization serves as evidence for the ongoing enhancement and stability of our approach as the texture progressively evolves into the target domain without succumbing to over-fitting issues. 3 illustrates the transformation results between the NCH and IHC datasets, and vice versa, from IHC to NCH. Similarly, the transformations between Melanocytic and Carcinoma datasets and the reverse from Carcinoma to Melanocytic are depicted. The degree of Style Guidance indicates the number of iterations our model has undergone to produce the outcomes; the images on the far left contain the most information from the original data, progressing towards the right; they increasingly resemble the target domain. Notably, during the testing phase, the images resembling the target domain were generated without any target domain information, based solely on randomly sampled images from the source domain.

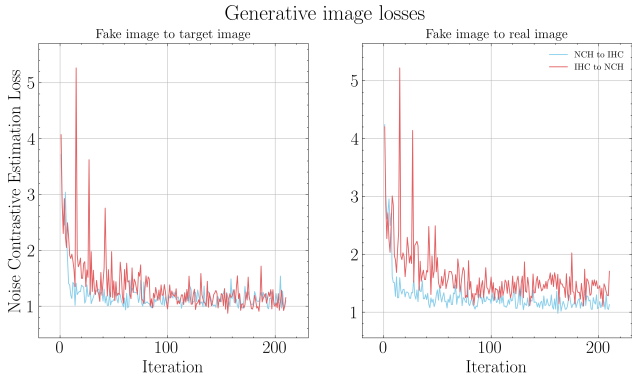


Figure 6. Quantitative graph of the losses while training the models. This shows the model almost converges from 100 iterations.

5.2. Quantitative Analysis

To assess quantitative metrics, we compared the Fréchet Inception Distance (FID) [9] scores with those of existing methods. 2 The evaluation focused on the FID scores during the conversion process from NCH to IHC and vice versa. Given the nature of the FID score, which compares the images themselves, we observed that the scores were relatively high, which is attributable to our dataset’s characteristic of being comprised of unpaired images across different domains. Upon comparison, it was evident that our model achieved lower FID scores than the pre-existing UNSB model. This trend of reduced FID scores was consistent even when the domain conversion was reversed in our experiments. These findings corroborate the efficacy of our domain phase loss in mitigating the disparities between domains, thereby validating its role in bridging the domain gap. Additionally, as depicted in Figure 6, it is observed that our model converges after approximately 100 iterations of the training steps. This applies to both scenarios: transforming fake images into target images and converting fake images into real images.

Model	NCH→IHC ↓	IHC→NCH ↓
UNSB	365.38	157.16
DKDM(Ours)	307.52	102.08

Table 2. The quantitative FID score for the reconstructed images. Our model surpasses the FID score of the state-of-the-art model UNSB

5.3. Gray-scale image to image translation

Recently, few studies have proposed diffusion models for image transfer using gray-scale images. Therefore, we applied our proposed method to explore the efficacy of gray-scale image to image transfer, specifically on CheXpert im-

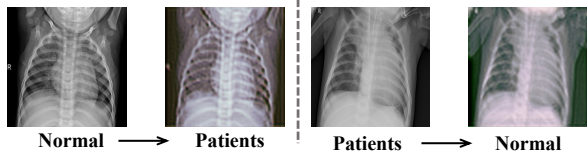


Figure 7. To demonstrate our proposed method well on the gray-scale image translation task, we utilize the CheXpert dataset. In particular, we tried to transfer images from patient data to normal data.

ages, aiming to assess its performance across unpaired images. However, Gray-scale image transfer is limited due to the absence of color information. It is also typically composed of fine-grained disease. The experiments are shown in Table 7.

5.4. Detailed imagery description of DKDM

DKDM can generate images that align with the target domain while preserving certain aspects of the source domain. As depicted in Figure 6, DKDM, unlike the state-of-the-art model UNSB, adeptly retains features of the source domain’s images while accurately producing images suitable for the target domain. This can be attributed to the domain phase loss, which effectively minimizes the differences between the domains. In the case of UNSB, it is noted that as the number of evaluations increases, the distinctive features of the original source domain gradually diminish. In contrast, our model maintains the characteristics of the source domain well while approaching the target domain.

6. Limitations.

Despite generating a generalized image and showing competitive performance, our proposed DKDM still has limitations in terms of twofold. Firstly, we could use diffusion pre-trained by large datasets. A pre-training model with large datasets is highly unlikely to cause catastrophic forgetting compared to not. Secondly, in datasets comprising fine-grained medical images (such as epithelium-stroma and ISIC2019, etc.), the image-to-image transfer did not clearly demonstrate the generation of distinct images. This issue arises due to variations in individual genetic characteristics, even for the same medical condition. Additionally, our experiments aim to generate generalized images. For these reasons, It is not sure that it will work well on other tasks(Classification, Detection, Segmentation, etc.). Specifically, since the generated image and existing datasets are different in terms of density, it is not certain whether they will improve density prediction. (e.g. segmentation). Also, when we adapt to gray-scale image datasets, transferring texture isn’t impossible. Thus, we believe that developing a diffusion model with universal applicability across

both unpaired gray-scale and RGB domains would be an intriguing avenue for future research.

7. Conclusion

This paper introduces DKDM, which enhances UNSB utilizing Schrödinger Bridge(SB) and combines SB with GAN training techniques for unpaired image-to-image translation by adding Domain phase term loss. We demonstrated the efficacy and prospects of DKDM by extensive experiments on the unpair domain data or Different classes from the same dataset. Unlike existing models, we have confirmed that DKDM preserves the appearances from the source domain well while generating images suitable for the target domain. We also highlight the significance and limitations of the gray-scale image transfer task, identifying it as a challenge that needs to be addressed in future research endeavors. We believe that our DKDM method will be useful in the future medical imaging community.

References

- [1] Andrew H Beck, Ankur R Sangoi, Samuel Leung, Robert J Marinelli, Torsten O Nielsen, Marc J Van De Vijver, Robert B West, Matt Van De Rijn, and Daphne Koller. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science translational medicine*, 3(108):108ra113–108ra113, 2011. 5
- [2] Finn Behrendt, Debayan Bhattacharya, Julia Krüger, Roland Opfer, and Alexander Schlaefer. Patched diffusion models for unsupervised anomaly detection in brain mri. *arXiv preprint arXiv:2303.03758*, 2023. 1
- [3] Chia-Ming Chang and Tsung-Nan Lin. Damix: A density-aware mixup augmentation for single image dehazing under domain shift. *arXiv preprint arXiv:2109.12544*, 2021. 2
- [4] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021. 2
- [5] Jiahua Dong, Yang Cong, Gan Sun, Bineng Zhong, and Xiaowei Xu. What can be transferred: Unsupervised domain adaptation for endoscopic lesions segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4023–4032, 2020. 1
- [6] Kuang Gong, Keith Johnson, Georges El Fakhri, Quanzheng Li, and Tinsu Pan. Pet image denoising based on denoising diffusion probabilistic model. *European Journal of Nuclear Medicine and Molecular Imaging*, pages 1–11, 2023. 1
- [7] Maryam Hammami, Denis Friboulet, and Razmig Kéchichian. Cycle gan-based data augmentation for multi-organ detection in ct images via yolo. In *2020 IEEE international conference on image processing (ICIP)*, pages 390–393. IEEE, 2020. 2
- [8] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv:1912.02781*, 2019. 2

- [9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 6
- [10] Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019. 5
- [11] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 3
- [12] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2
- [13] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018. 2
- [14] Christian Léonard. Some properties of path measures. *Séminaire de Probabilités XLVI*, pages 207–230, 2014. 3
- [15] Daiqing Li, Junlin Yang, Karsten Kreis, Antonio Torralba, and Sanja Fidler. Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8300–8311, 2021. 1
- [16] Chengyi Liu, Wenqi Fan, Yunqing Liu, Jiatong Li, Hang Li, Hui Liu, Jiliang Tang, and Qing Li. Generative diffusion models on graphs: Methods and applications. *arXiv preprint arXiv:2302.02591*, 2023. 1
- [17] Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos Theodorou. Deep generalized schrödinger bridge. *Advances in Neural Information Processing Systems*, 35:9374–9388, 2022. 3
- [18] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I²sb: Image-to-image schrödinger bridge. *arXiv preprint arXiv:2302.05872*, 2023. 2, 3
- [19] Ye Mao, Lan Jiang, Xi Chen, and Chao Li. Disc-diff: Disentangled conditional diffusion model for multi-contrast mri super-resolution. *arXiv preprint arXiv:2303.13933*, 2023. 1
- [20] Harshal Nishar, Nikhil Chavanke, and Nitin Singhal. Histopathological stain transfer using style transfer network with adversarial loss. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*, pages 330–340. Springer, 2020. 2
- [21] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 2
- [22] Wongi Park, Inhyuk Park, Sungeun Kim, and Jongbin Ryu. Robust asymmetric loss for multi-label long-tailed learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2711–2720, 2023. 1
- [23] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 2
- [24] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019. 2
- [25] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. *Advances in neural information processing systems*, 32, 2019. 2
- [26] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023. 1
- [27] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018. 2
- [28] Liyue Shen, Wentao Zhu, Xiaosong Wang, Lei Xing, John M Pauly, Baris Turkbey, Stephanie Anne Harmon, Thomas Hogue Sanford, Sherif Mehralivand, Peter L Choyke, et al. Multi-domain image completion for random missing input data. *IEEE transactions on medical imaging*, 40(4):1113–1122, 2020. 1
- [29] Liyue Shen, Lequan Yu, Wei Zhao, John Pauly, and Lei Xing. Novel-view x-ray projection synthesis through geometry-integrated deep learning. *Medical image analysis*, 77:102372, 2022. 1
- [30] Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion schrödinger bridges. In *Uncertainty in Artificial Intelligence*, pages 1792–1802. PMLR, 2022. 3
- [31] Jiaming Song, Shengjia Zhao, and Stefano Ermon. A-nice-mc: Adversarial training for mcmc. *Advances in Neural Information Processing Systems*, 30, 2017. 2
- [32] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 2
- [33] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2, 3
- [34] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021. 1
- [35] Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual diffusion implicit bridges for image-to-image translation. *arXiv preprint arXiv:2203.08382*, 2022. 3
- [36] Ella Tamir, Martin Trapp, and Arno Solin. Transport with support: Data-conditional diffusion bridges. *arXiv preprint arXiv:2301.13636*, 2023. 3

- [37] James Thornton, Michael Hutchinson, Emile Mathieu, Valentin De Bortoli, Yee Whye Teh, and Arnaud Doucet. Riemannian diffusion schrödinger bridge. *arXiv preprint arXiv:2207.03024*, 2022. 3
- [38] Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Solving schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021. 2
- [39] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge. In *International Conference on Machine Learning*, pages 10794–10804. PMLR, 2021. 3
- [40] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. Diffusion models for medical anomaly detection. In *International Conference on Medical image computing and computer-assisted intervention*, pages 35–45. Springer, 2022. 1
- [41] Xiaowei Xu, Yinan Chen, Jianghao Wu, Jiangshan Lu, Yuxiang Ye, Yechong Huang, Xin Dou, Kang Li, Guotai Wang, Shaoting Zhang, et al. A novel one-to-multiple unsupervised domain adaptation framework for abdominal organ segmentation. *Medical Image Analysis*, 88:102873, 2023. 1
- [42] Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. *arXiv preprint arXiv:2111.15141*, 2021. 3
- [43] Zhaonian Zhang and Richard Jiang. User-centric democratization towards social value aligned medical ai services. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6326–6334, 2023. 1
- [44] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. The spatially-correlative loss for various image translation tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16407–16417, 2021. 3
- [45] Dewei Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*, 2023. 1
- [46] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2