ENGINEERING Information Technology & Electronic Engineering
https://www.academax.com/EITEE; https://www.engineering.org.cn
https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=11344065
E-mail: eng.itee@zju.edu.cn

Engineering Journals
Available on IEEE Xplore®

Research Article
https://doi.org/10.1631/ENG.ITEE.2026.0118

High-fidelity unsupervised virtual staining for digital pathology via progressive Schrödinger bridge dynamics

Jianqiao YE[1,2], Jiahong SUN[3], Jiandong WANG[3], Xinyang LI[4], Puti YAN[2], Xingye CHEN[2]✉

[1]*National College for Excellent Engineers, Beihang University, Beijing 100191, China*
[2]*School of Electronics and Information Engineering, Beihang University, Beijing 100191, China*
[3]*Senior Department of General Surgery, Chinese PLA General Hospital, Beijing 100853, China*
[4]*College of AI, Tsinghua University, Beijing 100084, China*

Abstract: Virtual staining converts label-free autofluorescence (AF) images into hematoxylin and eosin (H&E)-like brightfield contrast for nondestructive pathology. We adapt the unpaired neural Schrödinger bridge (UNSB) framework to AF-to-H&E virtual staining and develop an enhanced UNSB (E-UNSB) model for pathology-oriented image translation. E-UNSB retains the unpaired, unidirectional, and progressive stochastic transport formulation of UNSB while introducing pathology-specific structural regularization to preserve cellular morphology and tissue architecture. In particular, saliency constraints and global feature consistency provide multi-scale structural guidance to reduce distortion during staining generation. Experiments demonstrate improved distributional similarity and morphometric consistency over the evaluated unsupervised baselines. Quantitative downstream analysis further indicates that virtual H&E images retain morphometric features relevant to pathological assessment, supporting their potential for high-throughput multimodal tissue analysis.

Key words: Virtual staining; Unsupervised learning; Schrödinger bridge; Digital pathology; Image processing

1 Introduction

Histopathological examination using hematoxylin and eosin (H&E) staining is a cornerstone of clinical diagnosis, providing essential morphological information (Feldman and Wolfe, 2014). This is often complemented by immunohistochemistry (IHC), which provides insights into molecular expression patterns. However, conventional acquisition of these stains presents several challenges: it is labor-intensive, time-consuming, and, most importantly, destructive. This destructive nature precludes tissue reuse for multimodal analysis, necessitating the use of adjacent sections that rarely align perfectly. Such spatial mismatch complicates the precise co-localization of morphological features with their underlying molecular profiles, a key objective in modern pathology. The contrast between this conventional destructive workflow and a nondestructive computational alternative is illustrated in Fig. 1.

Virtual staining has emerged as a transformative, nondestructive alternative that leverages optical imaging and artificial intelligence to computationally generate histological stains from digital images (Yang et al., 2022; Bai et al., 2023; Kreiss et al., 2023; Zhang RR et al., 2024; Zhang SW et al., 2024; Huang JH et al., 2025). Pioneering work by Rivenson et al. (2019) demonstrated the feasibility of translating label-free autofluorescence (AF) images into diagnostically useful H&E-stained images. This concept has since evolved into a broad field of research encompassing diverse label-free modalities, including fluorescence lifetime imaging, quantitative phase imaging, and stimulated Raman scattering (SRS) (Abraham et al., 2020, 2023; Wang Q et al., 2022; Cao et al., 2023; Liu et al., 2024; Zhang SW et al., 2024). Recent advances in wide-field interferometric scattering microscopy have further enabled label-free mapping of subcellular dynamics through spectral exponent analysis, highlighting the potential of computational label-free imaging to reveal biologically relevant structures and dynamic processes without exogenous staining (Anyi et al., 2026). Beyond label-free synthesis, the field has also explored stain-to-stain transformations, including

✉ Xingye CHEN, chenxingye@buaa.edu.cn
Jianqiao YE, https://orcid.org/0009-0007-0187-5799
Xingye CHEN, https://orcid.org/0000-0001-9482-9319
CLC number: TP391

Received: Apr. 24, 2026; Revision accepted: Aug. 6, 2026;
Crosschecked: Sept. 7, 2026; Published online: Oct. 6, 2026

© The Authors 2026. Published by Zhejiang University Press Co., Ltd. This is an open access article distributed under the terms of the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

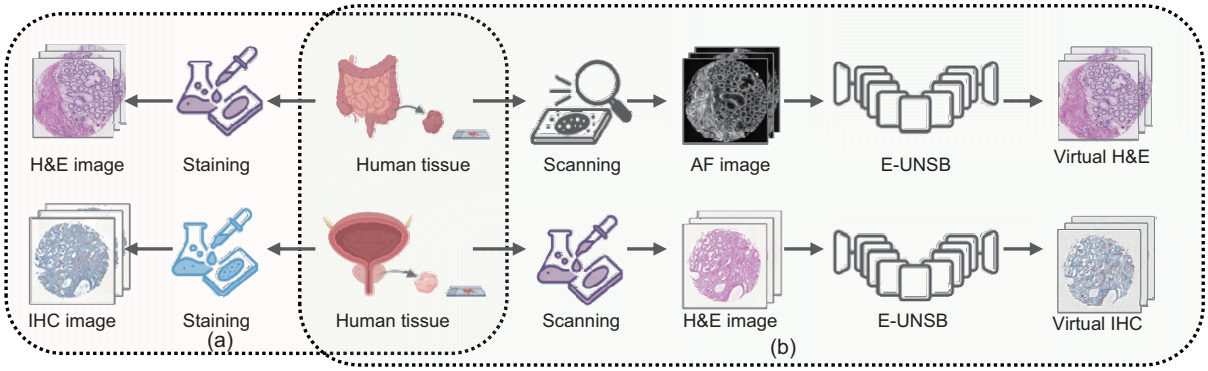


Fig. 1 Traditional staining requiring destructive chemical processing (a) and the proposed E-UNSB framework generating nondestructive virtual H&E and exploratory IHC-like images (b)

the generation of virtual special stains and IHC-like marker visualizations conditioned on H&E morphology. Although these approaches may provide complementary cues for exploratory analysis, they require rigorous reliability assessment and cell-level validation before clinical use (de Haan et al., 2021; Bai et al., 2022; Pati et al., 2024).

A central challenge for many of these approaches, particularly supervised methods, is their reliance on precisely registered, pixel-paired images, which are laborious and often impractical to acquire. This limitation has motivated the adoption of unsupervised image-to-image translation methods. Frameworks such as CycleGAN (Zhu et al., 2017) and its recent variant for image restoration (Guo et al., 2026) establish a foundation for unpaired translation by enforcing cycle consistency, while subsequent models such as contrastive unpaired translation (CUT) (Park et al., 2020) introduce contrastive learning for more efficient one-sided translation. More broadly, unsupervised contrastive representation learning has demonstrated strong performance in histopathological image classification, supporting the use of contrastive objectives to extract pathology-relevant features from unlabeled images (Wang XY et al., 2022). To address the limitations of one-to-one mappings, multimodal unsupervised image-to-image translation (MUNIT) (Huang X et al., 2018) disentangles content and style to enable multimodal outputs. Other methods have enhanced structural fidelity: geometry-consistent generative adversarial network (GcGAN) (Fu et al., 2019) imposes geometry consistency under spatial transformations, unsupervised content-preserving transformation for optical microscopy (UTOM) (Li XY et al., 2021) incorporates saliency consistency for microscopy images, and Santa (Xie et al., 2023) regularizes translation using a shortest-path hypothesis to reduce artifacts. These generative adversarial network (GAN)-based approaches have also been applied to virtual staining, including H&E generation from fresh-tissue SRS data and more advanced frameworks such as value mapping generative adversarial network (VM-GAN) (Wang JJ et al., 2025). Despite their utility, GAN-based methods are susceptible to training instability, mode collapse, and limited fidelity. Consequently, the field has increasingly explored more robust alternatives. Probabilistic generative models, particularly diffusion models, have emerged as promising alternatives with improved stability and fidelity, and have recently been applied to tasks such as pixel-super-resolved virtual staining (Qiu et al., 2024; Zhang YJ et al., 2025).

A notable recent development in this direction is the Schrödinger bridge (SB) model (Kim et al., 2024). Grounded in optimal transport theory, SB provides a principled alternative to the rigid, one-step mappings commonly used in GAN-based image translation. Rather than enforcing an abrupt domain shift, SB learns a progressive stochastic transport process that iteratively transforms a source image toward the target distribution. This gradual evolution may improve stability and is particularly appealing for histopathology, where preservation of fine biological structures during domain transformation is essential.

This motivation is particularly relevant to AF-to-H&E translation, which is not a generic style-transfer problem but a morphology-preserving cross-modality transport task. AF and H&E images can be viewed as distinct observations of the same underlying tissue microarchitecture, captured through different contrast-generation mechanisms. Accordingly, the desired transformation should alter domain-specific appearance while preserving diagnostically critical structures, including glandular boundaries, nuclear organization, and stromal architecture. From this perspective, SB provides a well-matched formulation because it explicitly models translation as progressive stochastic transport between source and target distributions rather than as a direct mapping. However, despite its theoretical appeal, this framework has not yet been fully adapted to the specific requirements of virtual pathology, where faithful preservation of diagnostically relevant morphology is paramount.

To address this gap, we develop an enhanced unpaired neural Schrödinger bridge (E-UNSB) for AF-to-H&E virtual staining. We adapt the UNSB framework to virtual staining by leveraging its unpaired, unidirectional, and progressive stochastic transport mechanism for pathology-oriented image translation and incorporating morphology-preserving regularization tailored to histopathological structures. These pathology-specific constraints are designed to reduce structural distortion and pathological inconsistencies that may arise during conventional GAN-based translation and diffusion-based generation while preserving diagnostically relevant features, including nuclear morphology, glandular boundaries, and stromal organization.

Our main contributions are threefold. (1) We adapt the

UNSB framework to unsupervised AF-to-H&E virtual staining, demonstrating its suitability for morphology-preserving cross-modality translation in digital pathology. (2) We introduce pathology-oriented structural regularization, including saliency guidance and global feature consistency, to improve the preservation of nuclei, glandular boundaries, and stromal organization during virtual staining. (3) We provide comprehensive empirical validation against representative unsupervised baselines, demonstrating improvements in perceptual quality, structural fidelity, and morphometric consistency, together with an exploratory evaluation of IHC-like prediction.

2 Methods

2.1 Framework overview

The E-UNSB framework (Fig. 2) performs unsupervised image-to-image translation by learning the dynamics of an SB. It progressively transforms a source-domain image into a target-domain representation by iteratively refining the image state along a stochastic transport path. This process is governed by a time-conditional generator and a PatchGAN discriminator (Isola et al., 2017). Architecturally, E-UNSB employs a unidirectional design comprising a single time-conditional generator, one discriminator, and an auxiliary energy network, whereas CycleGAN uses bidirectional translation with two generators and two discriminators. Consequently, E-UNSB has fewer trainable parameters during training (20.98×10^6 vs. 28.29×10^6 trainable parameters) while directly targeting the AF-to-H&E mapping.

2.2 Implementation details

2.2.1 Time-conditional generator

The generator is implemented as a U-Net comprising eight downsampling and eight upsampling blocks connected by skip connections. Each encoder block uses 4×4 convolutions with a stride of 2, followed by LeakyReLU activation and instance normalization, with the number of channels increasing from 64 to 512. The decoder symmetrically upsamples the feature maps using transposed convolutions.

Time conditioning maps each time step $t_i \in [0, 1]$ to a high-dimensional embedding using Fourier features followed by a multilayer perceptron (MLP). This embedding is injected into multiple U-Net layers to condition the generation process on the transport time. Noise conditioning maps a Gaussian noise vector $\boldsymbol{z}$ to affine parameters (γ, β), which modulate residual blocks to enhance output diversity. The generator is defined as

$$G_\phi(\boldsymbol{X}_{t_i}, t_i, \boldsymbol{z}), \tag{1}$$

where ϕ denotes the generator parameters, $\boldsymbol{X}_{t_i}$ is the intermediate image state at time t_i, and $\boldsymbol{z}$ is a sampled noise vector.

2.2.2 Discriminator

The discriminator is implemented as a PatchGAN comprising stacked 4×4 convolutions with a stride of 2, LeakyReLU activations, and instance normalization. It produces a score map over local receptive fields. The discriminator is denoted

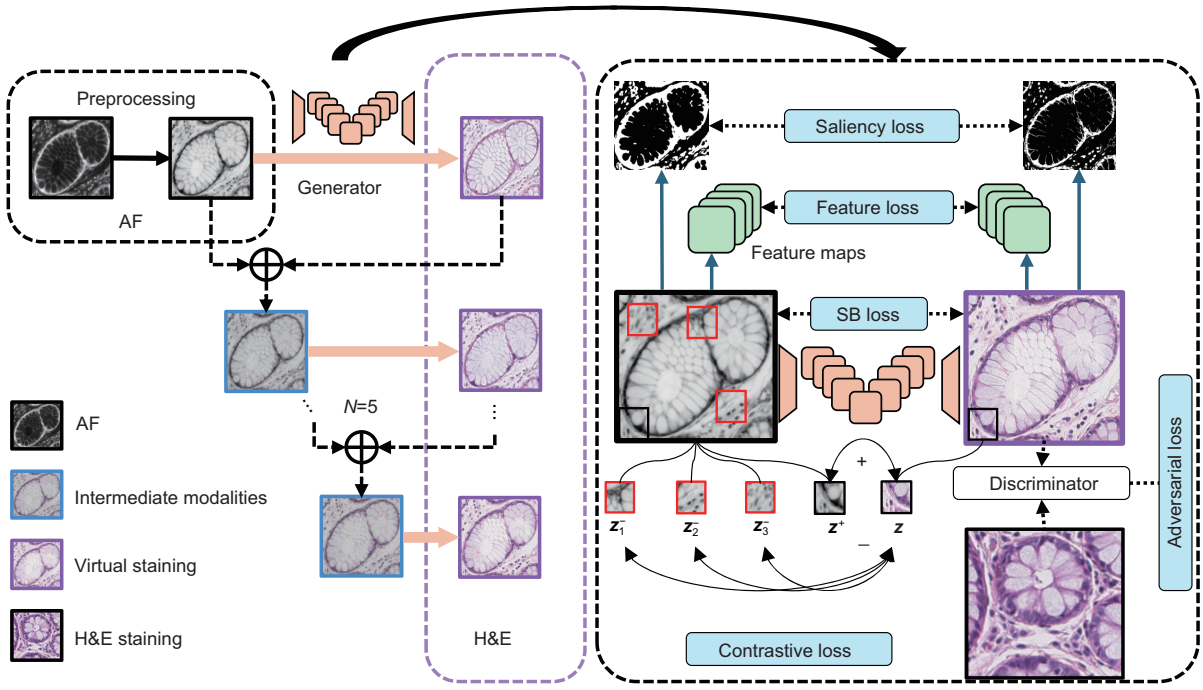


Fig. 2 Overview of the E-UNSB framework. E-UNSB learns Schrödinger bridge dynamics to progressively transform images through iterative refinement. In the contrastive-loss branch, z^+ denotes the positive patch sampled at the corresponding spatial location, whereas z_1^-, z_2^-, and z_3^- denote negative patches randomly sampled from other spatial locations of the large image. These patch labels are distinct from the Gaussian noise vector z used to condition the generator

by D_ψ. The adversarial loss is defined as

$$\mathcal{L}_{\mathrm{Adv}}(\phi,\psi,t_i) = \mathbb{E}_{y\sim p_{\mathrm{B}}}[\log D_\psi(y)] + \mathbb{E}_{x\sim p_{\mathrm{A}},\boldsymbol{z}}[\log(1-D_\psi(G_\phi(\boldsymbol{X}_{t_i},t_i,\boldsymbol{z})))], \tag{2}$$

where ψ denotes the trainable parameters of the discriminator, $\mathbb{E}[\cdot]$ denotes mathematical expectation with respect to the indicated sampling distribution, and log denotes the natural logarithm. p_{A} and p_{B} denote the probabilities of samples in the source domain and target domain, respectively.

2.2.3 Auxiliary energy network

Following the original UNSB formulation (Kim et al., 2024), E-UNSB uses an auxiliary time-conditional energy network E_η during training to estimate the entropy term in the Schrödinger bridge objective. At each bridge time t_i, it applies a log-sum-exp objective to contrast a positive transition pair, formed by the current state and its generated prediction, with a negative pair from an independently-sampled training patch. The network is optimized using Adam and is omitted during inference, which requires only the generator.

2.2.4 Loss function

1. Schrödinger bridge path loss. To account for the unpaired setting, the path regularization is defined at the distribution level using independently-sampled source- and target-domain images, $x\sim p_{\mathrm{A}}$ and $y\sim p_{\mathrm{B}}$, rather than a pixel-paired ground-truth target for each source image:

$$\mathcal{L}_{\mathrm{SB}}(\phi,t_i) = \mathbb{E}_{x\sim p_{\mathrm{A}},\, y\sim p_{\mathrm{B}},\, \boldsymbol{z}}\left[\|G_\phi(\boldsymbol{X}_{t_i},t_i,\boldsymbol{z})-y\|_2^2\right] - C(t_i)\mathcal{H}(G_\phi(\boldsymbol{X}_{t_i},t_i,\boldsymbol{z}))\,, \tag{3}$$

where $\mathcal{H}(\cdot)$ denotes the entropy term, $C(t_i)=t_i(1-t_i)$, and $\|\cdot\|_2^2$ denotes the squared L_2 (Euclidean) norm.

2. Contrastive regularization loss. To improve local structural fidelity, we introduce a contrastive regularization loss based on the PatchNCE objective used in CUT. The underlying principle is to treat features at corresponding spatial locations as positive pairs, while features from different sampled spatial locations are treated as negative pairs.

For feature layer l, let $\boldsymbol{f}_l^{\mathrm{q}}(\boldsymbol{s})$ denote the projected query feature at spatial location $\boldsymbol{s}$, and $\boldsymbol{f}_l^{\mathrm{k}}(\boldsymbol{s}')$ denote the projected key feature at spatial location $\boldsymbol{s}'$. Both feature vectors are ℓ_2-normalized before similarity computation. The exponential similarity score is defined as

$$a_l(\boldsymbol{s},\boldsymbol{s}') = \exp\left(\frac{(\boldsymbol{f}_l^{\mathrm{q}}(\boldsymbol{s}))^{\mathrm{T}}\,\boldsymbol{f}_l^{\mathrm{k}}(\boldsymbol{s}')}{\tau_{\mathrm{NCE}}}\right), \tag{4}$$

where τ_{NCE} is the dimensionless temperature parameter used in the noise-contrastive estimation (NCE) objective.

For each sampled spatial location $\boldsymbol{s}$, the PatchNCE loss at feature layer l is defined as

$$\ell_{\mathrm{NCE}}^{l,\boldsymbol{s}} = -\log\frac{a_l(\boldsymbol{s},\boldsymbol{s})}{a_l(\boldsymbol{s},\boldsymbol{s}) + \sum\limits_{\boldsymbol{s}'\in S_l,\,\boldsymbol{s}'\neq\boldsymbol{s}} a_l(\boldsymbol{s},\boldsymbol{s}')}. \tag{5}$$

The final contrastive regularization loss is obtained by averaging the PatchNCE loss over all selected feature layers and sampled spatial locations:

$$\mathcal{L}_{\mathrm{NCE}} = \frac{1}{|L_{\mathrm{layers}}|}\sum_{l\in L_{\mathrm{layers}}}\frac{1}{|S_l|}\sum_{\boldsymbol{s}\in S_l}\ell_{\mathrm{NCE}}^{l,\boldsymbol{s}}. \tag{6}$$

Here, L_{layers} denotes the set of selected feature layers, and S_l denotes the set of sampled spatial locations at layer l. The term $a_l(\boldsymbol{s},\boldsymbol{s})$ corresponds to the positive pair, consisting of the query feature and key feature extracted from the same spatial location. In contrast, $a_l(\boldsymbol{s},\boldsymbol{s}')$, where $\boldsymbol{s}'\neq\boldsymbol{s}$, represents a negative pair in which the query feature is compared with a key feature from a different sampled location. By maximizing the relative similarity of positive pairs to negative pairs, this loss encourages local structural consistency and helps preserve fine image details.

3. Saliency consistency loss. To preserve structural consistency under the unidirectional E-UNSB formulation, we impose a forward-only saliency consistency constraint between the source AF image and its translated H&E image. Following UTOM (Li XY et al., 2021), saliency maps are constructed from channel-averaged image intensities using differentiable intensity-threshold operators:

$$T_{\mathrm{a}}(a) = \sigma(k_{\mathrm{a}}\,(\bar{a}-\alpha))\,, \tag{7}$$

$$T_{\mathrm{p}}(G(a)) = 1-\sigma\big(k_{\mathrm{p}}\,(\overline{G}(a)-\beta)\big)\,, \tag{8}$$

where a denotes the source AF image, $G(a)$ denotes its generated H&E prediction, and $\bar{a}$ and $\overline{G}(a)$ denote their channel-averaged intensities. The operators T_{a} and T_{p} denote the saliency mappings for the AF input and prediction, respectively. $\alpha=-0.451$ and $\beta=0.725$ are the intensity thresholds, $\sigma(\cdot)$ denotes the sigmoid function, and $k_{\mathrm{a}}=k_{\mathrm{p}}=100$ control the transition sharpness for the AF and prediction. The opposite sigmoid directions account for the difference in contrast polarity between the two imaging domains. The saliency consistency loss is defined as

$$\mathcal{L}_{\mathrm{saliency}} = \mathbb{E}_{a\sim p_{\mathrm{A}}}\left[\|T_{\mathrm{a}}(a)-T_{\mathrm{p}}(G(a))\|_1\right], \tag{9}$$

where $\|\cdot\|_1$ denotes the L_1 norm. The threshold and sharpness parameters, as well as the saliency-weighting strategy, follow the empirical implementation of UTOM (Li XY et al., 2021). These parameters are fixed for all training and test images. During training, the contribution of the saliency term is gradually reduced using the same exponentially decaying weighting strategy.

4. Global feature consistency loss. To preserve global tissue organization during image translation, we enforce feature consistency between the source image and its translated output using a pretrained feature space. Let $\phi_l(\cdot)=V_{\mathrm{pt}}^l(\cdot)$ denote the feature maps extracted from layer l of a pretrained VGG16 network. Let $N_l=C_lH_lW_l$ denote the total number of elements in the feature map at layer l, with C_l, H_l, and W_l denoting the number of channels, height, and width, respectively. The normalized feature discrepancy at layer l is defined as

$$d_l(a) = \frac{1}{N_l}\,\|\phi_l(a)-\phi_l(G(a))\|_2^2\,. \tag{10}$$

The global feature consistency loss is then formulated as

$$\mathcal{L}_{\mathrm{Feat}}(G) = \mathbb{E}_{a\sim p_{\mathrm{A}}}\left[\frac{1}{|L_{\mathrm{feat}}|}\sum_{l\in L_{\mathrm{feat}}} d_l(a)\right], \tag{11}$$

where L_{feat} denotes the set of VGG16 layers used for feature extraction. The ImageNet-pretrained VGG16 network is used solely as a structural regularizer rather than for pathological

diagnosis. Its intermediate features are intended to encourage preservation of global tissue organization, while pathology-pretrained encoders such as CTransPath will be investigated in future work.

5. Total objective. The overall training objective combines all loss components:

$$\begin{aligned}\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Adv}} + \lambda_{\text{SB}}\mathcal{L}_{\text{SB}} + \lambda_{\text{NCE}}\mathcal{L}_{\text{NCE}} \\ + \lambda_{\text{Sal}}\mathcal{L}_{\text{saliency}} + \lambda_{\text{Feat}}\mathcal{L}_{\text{Feat}}.\end{aligned} \tag{12}$$

Here, λ_{SB}, λ_{NCE}, λ_{Sal}, and λ_{Feat} denote the weights assigned to the corresponding loss terms. The relative weights of the SB and NCE terms follow the original UNSB setting (Kim et al., 2024). The feature and saliency loss weights are selected such that their weighted magnitudes are approximately comparable to those of the other structural regularization terms during training. The final weights are set to $\lambda_{\text{SB}} = 10$, $\lambda_{\text{NCE}} = 1$, $\lambda_{\text{Sal}} = 2$, and $\lambda_{\text{Feat}} = 10$, and are kept fixed across all experiments.

We employ an alternating optimization scheme for training on unpaired datasets. The model is optimized using the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, an initial learning rate of 2×10^{-4}, and a 100-epoch training schedule. The learning rate is linearly decayed over the final 50 epochs. Training uses a batch size of 4 and a bridge step number of $N = 5$ corresponding to five neural network function evaluations, with a PatchNCE temperature of $\tau = 0.07$.

For large-image inference, each input image is processed using an overlapping sliding-window strategy. Images are divided into 256×256-pixel tiles with a stride of 128 pixels in both spatial directions, resulting in 50% overlap between adjacent tiles. Each tile is independently translated by the trained generator. During reconstruction, predictions from overlapping tiles are aggregated using a center-weighted blending window, assigning greater weights to pixels near the tile centers and lower weights near the tile boundaries. The overlapping predictions are then normalized by the accumulated weights to obtain the final output image, which is cropped to the original dimensions when padding is applied. This strategy reduces abrupt transitions and grid-like seams associated with direct concatenation of non-overlapping patches, although residual artifacts may persist in regions with strong intensity variations or highly heterogeneous tissue structures.

2.3 Evaluation metrics

We use two standard distributional metrics for image generation: Fréchet inception distance (FID) and kernel inception distance (KID) (Heusel et al., 2017; Bińkowski et al., 2018). Unlike pixel-level metrics such as the structural similarity index measure (SSIM), which require paired ground-truth images, FID and KID enable evaluation of perceptual similarity and distributional alignment in unsupervised virtual staining. Both metrics compare feature distributions of real and generated images extracted using a pretrained Inception-v3 network. All images are normalized to $[-1, 1]$ before feature extraction. FID measures the Fréchet distance between Gaussian approximations of the real and generated feature distributions:

$$\text{FID} = \|\boldsymbol{\mu}_{\text{r}} - \boldsymbol{\mu}_{\text{g}}\|_2^2 + \text{Tr}\left(\boldsymbol{\Sigma}_{\text{r}} + \boldsymbol{\Sigma}_{\text{g}} - 2(\boldsymbol{\Sigma}_{\text{r}}\boldsymbol{\Sigma}_{\text{g}})^{1/2}\right), \tag{13}$$

where $\boldsymbol{\mu}_{\text{r}}$ and $\boldsymbol{\Sigma}_{\text{r}}$ ($\boldsymbol{\mu}_{\text{g}}$ and $\boldsymbol{\Sigma}_{\text{g}}$) denote the feature means and covariance matrices for the real (generated) features, respectively. Lower FID values indicate closer alignment between the generated and reference feature distributions.

KID measures the squared maximum mean discrepancy (MMD) between the real and generated feature distributions:

$$\begin{aligned}\text{KID} = \mathbb{E}_{x,x'\sim p_{\text{r}}}[k(x,x')] + \mathbb{E}_{y,y'\sim p_{\text{g}}}[k(y,y')] \\ - 2\mathbb{E}_{x\sim p_{\text{r}},y\sim p_{\text{g}}}[k(x,y)],\end{aligned} \tag{14}$$

where $k(\cdot,\cdot)$ is a kernel function and p_{r} and p_{g} denote the real and generated feature distributions, respectively. KID provides an unbiased estimator of MMD and is less sensitive to sample size than FID.

3 Results

To comprehensively evaluate E-UNSB, we structured the experiments into two stages. First, we evaluated the primary AF-to-H&E (AF→H&E) translation task through benchmarking and ablation studies. Second, we conducted an exploratory feasibility study of H&E-conditioned IHC-like appearance prediction.

3.1 Datasets and baselines

The primary AF-to-H&E experiments were conducted using the UTOM dataset (Li XY et al., 2021; Li X, 2023). The dataset comprised 20 tissue cores, each represented by one AF image and one H&E image, yielding 40 modality-specific images, corresponding to 20 AF-H&E tissue–core pairs. Before patch extraction, the 20 tissue–core pairs were divided into 14 training pairs and 6 test pairs. AF and H&E images from the same tissue core were assigned to the same subset, ensuring that no patches derived from the same tissue core were shared between the training and test sets. Pair information was used solely for dataset partitioning and was not provided as paired supervision during model training. All 20 tissue–core pairs were included without additional inclusion or exclusion criteria. After patch extraction, 5372 AF patches and 5568 H&E patches were obtained. For the exploratory H&E-to-IHC-like translation task, we used the EMPaCT prostate datasets (Karkampouna and Kruithof-de Julio, 2023; Pati et al., 2024) for the CD146 and NKX3.1 markers. For each marker dataset, 20 parent images were similarly divided into 14 training images and 6 test images before tiling into 256 × 256-pixel patches. Random flipping and cropping were used for data augmentation.

E-UNSB was compared with SynDiff (Özbey et al., 2023), Santa, CUT, UTOM, CycleGAN, GcGAN, and MUNIT. All methods used the same data splits, input resolution, preprocessing, augmentation, and test sets, while each baseline was trained using its official or recommended configuration.

3.2 Analysis on AF-to-H&E translation

1. Qualitative performance. We first evaluated E-UNSB on the challenging task of translating AF images into virtual H&E images. As shown qualitatively in Fig. 3, the generated images exhibited high visual fidelity while preserving diagnostically relevant morphological features relative to

the reference H&E images. E-UNSB preserved overall tissue architecture and fine-grained cellular morphology, including nuclear–cytoplasmic contrast and well-defined glandular and stromal interfaces.

Benchmark comparisons (Fig. 4) show that E-UNSB produced visually realistic images with few apparent artifacts. Santa achieved the second-best qualitative performance but exhibited slightly reduced textural sharpness. SynDiff generated visually realistic cellular structures; however, its cytoplasmic staining was less consistent with the reference H&E images, resulting in weaker nuclear–cytoplasmic contrast in some regions. CUT produced generally plausible results but exhibited subtle blurring. Other methods showed more pronounced deficiencies; notably, MUNIT exhibited an apparent failure in color mapping, with an inversion of the characteristic nuclear and cytoplasmic colors. Pixel-density analysis (Fig. 4 bottom) further suggests that the E-UNSB output distribution was more closely aligned with those of the reference H&E images.

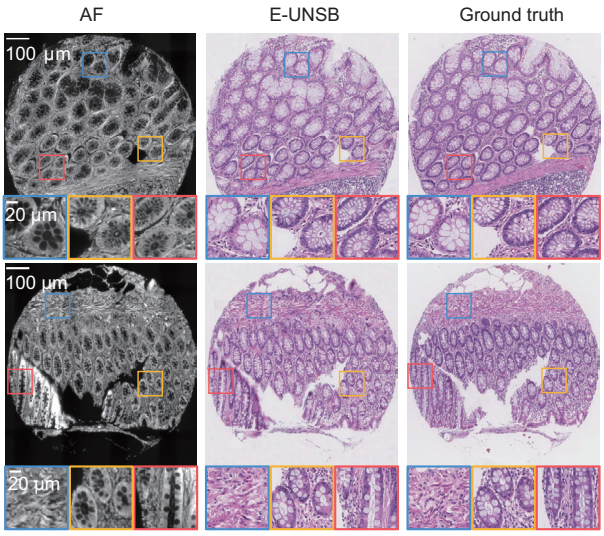


Fig. 3 Experimental results of E-UNSB for AF-to-H&E translation. Representative AF inputs (left), E-UNSB outputs (middle), and reference H&E images (right) are shown. E-UNSB preserves tissue architecture and cellular morphology

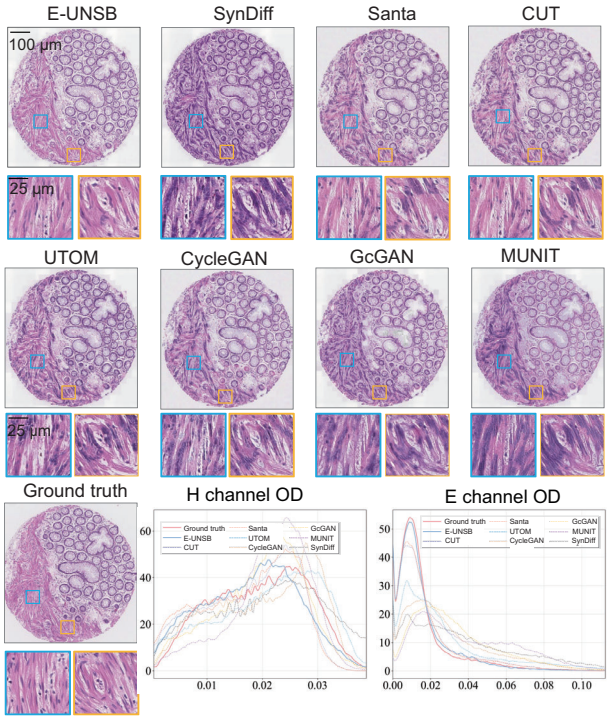


Fig. 4 Performance comparison. Top: representative images with enlarged regions of interest; Bottom: pixel-density curves for hematoxylin (H) and eosin (E) OD, showing that the E-UNSB distribution is more closely aligned with those of the reference H&E images. The vertical axis represents the probability density, and the horizontal axis represents the pixel OD

2. Quantitative performance. As shown in Table 1, E-UNSB achieved the lowest mean FID and KID among the competing methods, with an FID of 75.27 and a KID of 18.1×10^{-3}. The diffusion-based SynDiff baseline achieved an FID of 79.54 and a KID of 20.8×10^{-3}. Compared with SynDiff, E-UNSB reduced the mean FID by 5.37% and the mean KID by 12.98%, indicating closer distributional alignment with the reference H&E images under the specified evaluation protocol. Santa achieved the second-lowest mean FID and KID, whereas methods exhibiting greater visual degradation, such as MUNIT, also showed larger distributional distances.

Table 1 Quantitative comparison on the AF-to-H&E translation task

Method	FID ↓	KID ($\times10^{-3}$) ↓
E-UNSB	**75.27 ± 3.76**	**18.1 ± 0.9**
Santa	76.26 ± 3.12	19.5 ± 0.7
UTOM	78.65 ± 2.94	20.3 ± 0.8
CycleGAN	79.11 ± 2.85	21.6 ± 0.8
SynDiff	79.54 ± 3.89	20.8 ± 1.3
CUT	81.02 ± 2.73	26.6 ± 1.0
GcGAN	82.81 ± 2.67	28.1 ± 1.1
MUNIT	85.77 ± 3.25	31.2 ± 1.2

Lower FID and KID indicate better distributional similarity to reference images. Values are reported as mean ± standard deviation across three evaluation subsets. The best results are in bold

3. Downstream task validation. To assess morphometric consistency, we performed downstream analysis using QuPath (Bankhead et al., 2017; Belaldavar et al., 2024). As shown in Fig. 5, the distributions of quantitative morphological features extracted from E-UNSB virtual H&E images showed the closest overall alignment with those extracted from the reference H&E images. E-UNSB achieved the smallest mean absolute Cohen's d (0.0953) and the lowest mean Wasserstein distance (0.7360) among the compared methods. Here, the absolute Cohen's d is a standardized effect-size measure of the difference between the virtual and reference H&E feature distributions; a value closer to zero indicates closer agreement. Notably, E-UNSB showed close agreement with the reference H&E images for key morphological metrics, including mean cell hematoxylin optical density (OD) and the nucleus-to-cell area ratio.

To directly assess the contribution of the proposed structural constraints, we included the original UNSB baseline in the downstream validation, as shown in Table 2. Compared with UNSB, E-UNSB reduced the mean absolute Cohen's d from 0.1215 to 0.0953 and the mean Wasserstein distance from 1.7282 to 0.7360. E-UNSB also achieved lower mean absolute

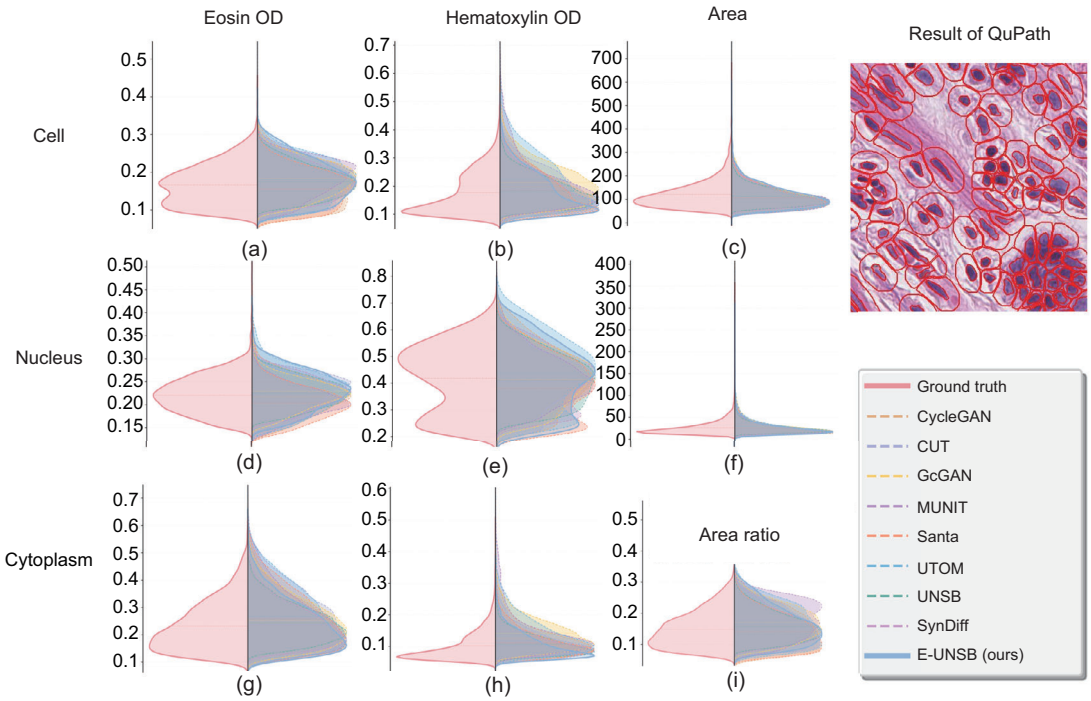


Fig. 5 Downstream morphometric validation using QuPath. Violin plots show the distributions of nine quantitative features and their agreement between virtual and reference H&E images: (a) cell eosin OD mean; (b) cell hematoxylin OD mean; (c) cell area; (d) nucleus eosin OD mean; (e) nucleus hematoxylin OD mean; (f) nucleus area; (g) cytoplasm eosin OD mean; (h) cytoplasm hematoxylin OD mean; (i) nucleus-to-cell area ratio. The vertical axes show the corresponding feature values

Table 2 QuPath-based downstream morphometric validation

Method	Cohen's d ↓	Wasserstein distance ↓
E-UNSB	**0.0953**	**0.7360**
UNSB	0.1215	1.7282
SynDiff	0.1340	1.3310
CycleGAN	0.1374	1.3379
CUT	0.1741	1.3887
Santa	0.2039	1.4415
GcGAN	0.2610	0.7820
UTOM	0.2793	1.3571
MUNIT	0.3006	1.5338

Absolute Cohen's d and Wasserstein distance were calculated separately for each of the nine quantitative features and then averaged across features. Lower values (↓) indicate better agreement with the reference images. The best results are in bold

Cohen's d and mean Wasserstein distance than the diffusion-based SynDiff baseline, which yielded values of 0.1340 and 1.3310, respectively. These results suggest that the proposed pathology-specific structural regularization improves not only distributional similarity as measured by FID and KID but also quantitative morphological consistency with the reference H&E images.

3.3 Pathologist subjective evaluation

To assess perceived staining quality, we conducted a blinded reader study using an independently acquired breast cancer dataset collected in our study and publicly released on Zenodo (Ye and Sun, 2025). The dataset and its pathological application were further described by Sun et al. (2026). The study included 20 breast cancer tissue sections. For each section, four corresponding images were prepared: one reference H&E image and three virtual staining images generated by different methods, yielding 80 images for evaluation.

Before evaluation, all images were randomly reordered, and three experienced pathologists were blinded to image source. Each image was independently rated on a five-point Likert scale (1–5) across four criteria: nuclear staining detail, intercellular matrix staining detail, cytoplasmic staining detail, and overall staining quality.

Representative comparisons are shown in Fig. 6. Compared with CycleGAN and Santa, E-UNSB produced virtual H&E images with more realistic staining appearance, clearer nuclear structures, and better-preserved tissue morphology, resulting in greater visual similarity to the reference H&E images.

The quantitative scoring results are summarized in Table 3. Reference H&E images received the highest mean scores across all four criteria and served as the real-image comparator. Among the virtual staining methods, E-UNSB consistently received the highest ratings and showed the closest scores to those of the reference H&E images across all evaluated criteria.

For each tissue section and staining method, scores from the three pathologists were averaged. Differences among the four image groups were assessed using the Friedman test.

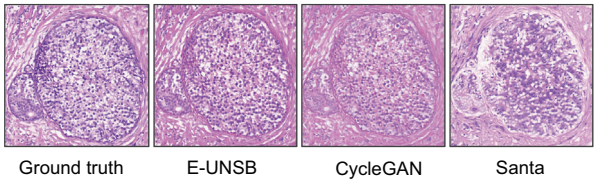


Fig. 6 Representative visual comparison on the independently acquired breast cancer dataset used for blinded pathologist evaluation. From left to right: reference H&E image and virtual staining results generated by E-UNSB, CycleGAN, and Santa

Table 3 Blinded evaluation of staining quality on 20 breast cancer tissue sections by three pathologists (mean scores)

Method	Nuclear	Matrix	Cytoplasm	Overall
Reference H&E	3.07	3.05	3.10	3.06
E-UNSB	2.86	2.73	2.78	2.80
CycleGAN	2.24	2.20	2.27	2.24
Santa	2.11	2.08	2.14	2.10

When the overall test was significant, pairwise comparisons between E-UNSB and each other virtual staining method were performed using paired Wilcoxon signed-rank tests with Holm correction. E-UNSB received the highest scores among the evaluated virtual staining methods and significantly outperformed CycleGAN and Santa across all four criteria (all adjusted $p < 0.01$). Inter-rater agreement was assessed using the intraclass correlation coefficient.

Overall, E-UNSB received the highest pathologist ratings among the evaluated virtual staining methods and produced images with the greatest perceived similarity to the reference H&E images.

3.4 Ablation and robustness studies

1. Ablation studies. We evaluated the contributions of the auxiliary structural constraints, as shown in Fig. 7. The original UNSB exhibited limited translation fidelity. Adding global feature consistency improved preservation of overall structural organization, whereas saliency guidance enhanced the preservation of fine-grained nuclear structures. The complete E-UNSB achieved the best visual alignment and the lowest FID and KID values among the evaluated variants. Notably, removing the NCE loss substantially slowed model convergence. During the early stages of training, cellular structures were poorly preserved, and substantially more epochs were required for the model to reach a stable and visually plausible solution. This observation suggests that NCE provides an important regularization effect that facilitates convergence and contributes to structural consistency during image translation.

As shown in Table 4, the complete E-UNSB achieved the lowest FID and KID among all ablation variants. Compared with the original UNSB baseline, E-UNSB reduced FID from 86.86 to 75.27 and KID from 26.5×10^{-3} to 18.1×10^{-3}. Removing global feature consistency or saliency guidance degraded performance, whereas removing NCE resulted in the largest performance decrease. These results support the effectiveness of the proposed structural regularization terms in improving the distributional similarity of UNSB-generated images to real H&E images.

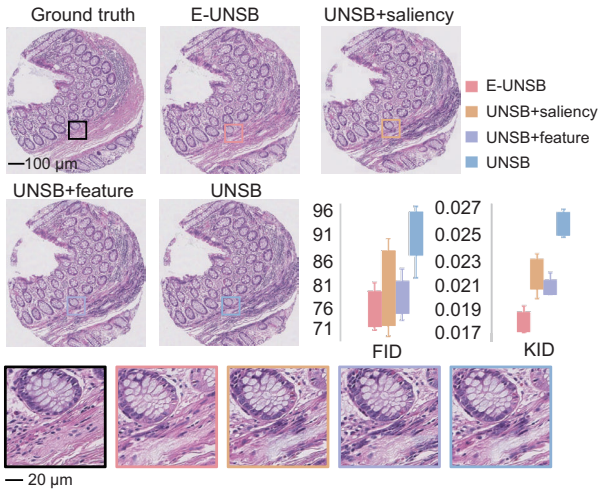


Fig. 7 Ablation study results: visual comparison of reference H&E images, ablation variants, and their corresponding FID and KID scores

Table 4 Quantitative ablation results for E-UNSB components

Method	FID ↓	KID ($\times10^{-3}$) ↓
E-UNSB	**75.27 ± 3.76**	**18.1 ± 0.9**
w/o feature consistency	79.78 ± 3.21	21.3 ± 1.0
w/o saliency guidance	78.10 ± 3.08	20.8 ± 0.9
w/o NCE loss	87.53 ± 3.64	27.4 ± 1.2
UNSB	86.86 ± 3.52	26.5 ± 1.1

Lower FID and KID indicate better distributional similarity to reference images. Values are reported as mean ± standard deviation across three evaluation subsets. The best results are in bold. w/o: without

2. Effect of bridge step number. We further investigated the effect of the bridge step number N on AF-to-H&E translation, as shown in Fig. 8. As N increased from 1 to 5, both FID and KID improved substantially, with FID decreasing from 80.95 to 75.27 and KID from 26.4×10^{-3} to 18.1×10^{-3}. This trend suggests that progressive transport can facilitate a smoother cross-domain transition and improve structural preservation relative to a one-step mapping.

When N was further increased to 7 (9), the performance gains became marginal. FID decreased only slightly, from 75.27 to 74.87 (74.69), while KID changed from 18.1×10^{-3} to 17.8×10^{-3} (17.9×10^{-3}). In contrast, inference time increased steadily, from 21.5 ms at $N = 5$ to 26.2 and 30.5 ms at $N = 7$ and $N = 9$. These results suggest that the benefits of increasing the number of bridge steps diminish beyond a moderate trajectory depth. Accordingly, $N = 5$ provided a favorable trade-off between translation quality and computational efficiency and was adopted as the default setting in all experiments.

3. Robustness analysis. We compared E-UNSB with CycleGAN and CUT across different training data scales (Table 5). On the tiny-scale dataset containing 1000 patches, E-UNSB achieved FID and KID values that were 6.51% and 10.06% lower than those of CycleGAN, respectively, indicating greater robustness to reduced training data availability.

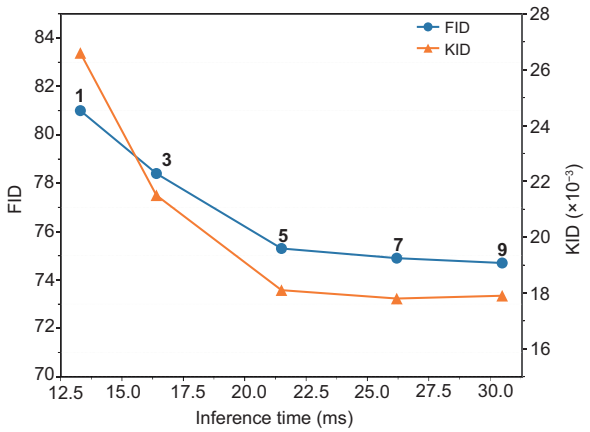


Fig. 8 Effect of bridge step number N on AF-to-H&E translation. Performance improved with N up to 5, while larger N provided marginal gains at higher inference cost

Table 5 Robustness comparison across different training data scales

Scale	E-UNSB		CycleGAN		CUT	
	FID	KID ($\times10^{-3}$)	FID	KID ($\times10^{-3}$)	FID	KID ($\times10^{-3}$)
Base	75.27	18.1	79.11	21.6	81.02	26.6
Mid	80.97	21.9	87.38	28.9	89.67	28.0
Tiny	87.70	28.6	93.81	31.8	91.21	30.4

3.5 Exploratory study: IHC-like prediction

We further conducted an exploratory qualitative and distribution-level quantitative study of IHC-like marker prediction conditioned on H&E morphology. As shown in Figs. 9 and 10, E-UNSB generated patterns that were qualitatively consistent with the reference images at the regional level, including endothelial-like patterns for CD146 and nuclear-like staining for NKX3.1. These results are intended to be hypothesis-generating and require rigorous validation, particularly given the imperfect correspondence between the input H&E and reference IHC images.

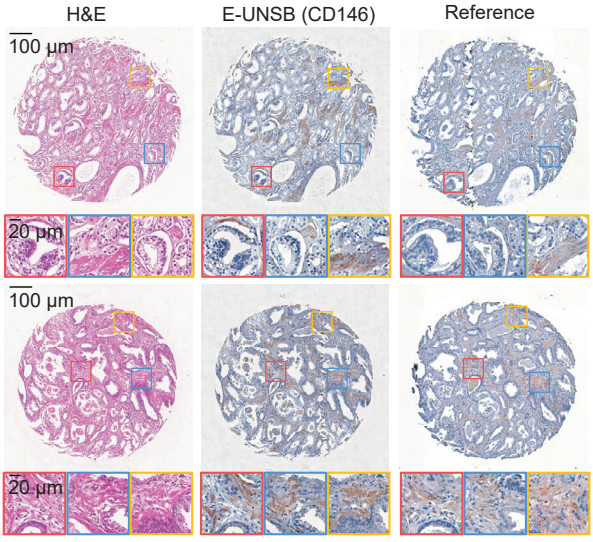


Fig. 9 Experimental results of H&E-to-CD146-like staining translation. E-UNSB produced endothelial-like staining patterns at the regional level despite minor misregistration

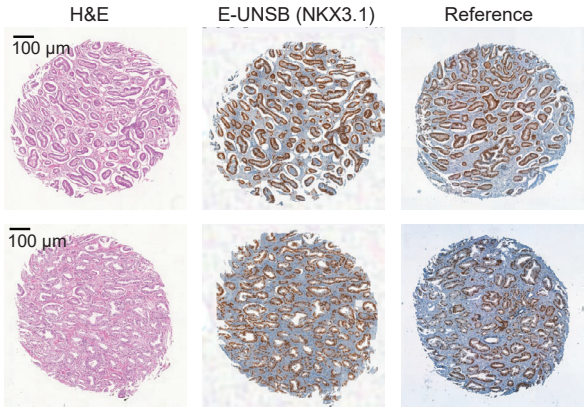


Fig. 10 Experimental results of H&E-to-NKX3.1-like staining translation. E-UNSB produced nuclear-like staining patterns that were qualitatively consistent with the reference at the regional level

Table 6 Distribution-level quantitative evaluation of E-UNSB for H&E-to-IHC-like prediction

Marker	FID ↓	KID ($\times10^{-3}$) ↓
CD146	35.57 ± 1.78	12.2 ± 0.5
NKX3.1	29.22 ± 1.91	10.4 ± 0.7

Lower FID and KID indicate better distributional similarity to reference images. Values are reported as mean ± standard deviation across three evaluation subsets

To complement the qualitative assessment, we evaluated the distributional similarity between the generated and real IHC images using FID and KID. As summarized in Table 6, E-UNSB achieved an FID of 35.57 ± 1.78 and a KID of $(12.2\pm0.5)\times10^{-3}$ for CD146. For NKX3.1, E-UNSB achieved an FID of 29.22 ± 1.91 and a KID of $(10.4\pm0.7)\times10^{-3}$. These results indicate distribution-level similarity between the generated IHC-like images and the corresponding real IHC images under the present evaluation protocol.

4 Discussion

In this study, we adapted the UNSB framework to unsupervised virtual staining and developed E-UNSB by incorporating pathology-oriented structural constraints. Unlike generic image translation, virtual staining requires changes in image appearance while preserving diagnostically relevant morphology. To address this requirement, we integrated contrastive regularization, saliency consistency, and global feature consistency into the progressive Schrödinger bridge transport process to promote local structural correspondence, emphasize salient tissue regions, and reduce large-scale structural distortion.

Under the evaluated datasets and experimental protocol, E-UNSB achieved lower FID and KID values and closer morphometric agreement than the selected GAN-based methods and the diffusion-based SynDiff baseline. In the blinded reader study, E-UNSB received higher ratings than CycleGAN and Santa. The bridge step analysis further showed that increasing the number of transport steps improved distributional metrics up to a moderate trajectory depth, supporting the use of progressive transport as a potentially effective strategy for cross-modality image translation with structural preservation.

The exploratory H&E-to-IHC-like experiments suggest that E-UNSB can capture statistical associations between tissue morphology and marker-related staining patterns. However, tissue morphology does not uniquely determine molecular expression. Therefore, the generated IHC-like images should be interpreted as exploratory visual estimates rather than substitutes for experimentally acquired IHC.

Several limitations remain. The iterative bridge process is computationally more expensive than single-step GAN-based methods, and the current validation covers a limited range of datasets, tissue types, imaging conditions, and staining targets. The method may also produce generation artifacts or hallucinated structures, particularly in regions with abrupt or substantial intensity variations in the AF input. In addition, the intensity-based saliency constraint may be sensitive to differences in image acquisition and normalization across devices. More extensive multi-center and cross-device validation is therefore required before broader clinical applicability can be assessed.

As with other generative virtual staining methods, visually plausible outputs may contain spurious nuclei, altered tissue boundaries, missing cellular components, or false-positive and false-negative staining patterns. Accordingly, virtual staining results should be interpreted cautiously and should not be regarded as direct replacements for conventional histological or molecular assays. Future work should include systematic failure case analysis, uncertainty estimation, assessment of rare-structure preservation, and cell- or lesion-level validation using accurately registered reference images when available.

5 Conclusions

E-UNSB extends UNSB to pathology-oriented virtual staining by combining progressive transport with local and global structural regularization. Across the evaluated datasets and protocols, E-UNSB improved distributional similarity, morphometric consistency, and pathologist-perceived staining quality relative to the evaluated baselines, while warranting further validation of robustness, generalizability, and failure modes.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62501030), the Beijing Natural Science Foundation (No. 4254070), and the Beijing Nova Program (No. 202604841259).

Author contributions

Jianqiao YE and Xingye CHEN designed the research. Jianqiao YE and Jiahong SUN processed the data. Jianqiao YE and Xinyang LI performed the experiments. Jianqiao YE and Puti YAN developed the algorithms. Jianqiao YE drafted the paper. Xingye CHEN and Jiandong WANG revised and finalized the paper.

Compliance with ethics guidelines

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Chinese PLA General Hospital (protocol code S2024-711-01 and date of approval: November 28, 2024).

Conflict of interest

Puti YAN is a junior editorial board member of *ENGINEERING Information Technology & Electronic Engineering*, and he was not involved with the peer review process of this paper. All the authors declare that they have no conflict of interest.

Data availability

The datasets analyzed during this study are publicly available in the Zenodo repository. The UTOM dataset can be accessed at https://doi.org/10.5281/zenodo.8376199, the EMPaCT TMA dataset can be accessed at https://doi.org/10.5281/zenodo.10066853, and the breast cancer dataset used for blinded pathological evaluation can be accessed at https://doi.org/10.5281/zenodo.17646116. The code used for data processing, model training, and analysis in this study is available at https://github.com/Cambridgebit/E-UNSB.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used DeepSeek to improve language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Abraham T, Shaw A, O'Connor D, et al., 2020. Slide-free MUSE microscopy to H&E histology modality conversion via unpaired image-to-image translation GAN models. https://arxiv.org/abs/2008.08579

Abraham T, Costa PC, Serafini CE, et al., 2023. Label- and slide-free tissue histology using 3D epi-mode quantitative phase imaging and virtual hematoxylin and eosin staining. *Optica*, 10(12):1605-1618. https://doi.org/10.1364/OPTICA.502859

Anyi CL, You HZ, Li HK, et al., 2026. Label-free mapping of subcellular dynamics using wide-field interferometric scattering microscopy and spectral exponent analysis. *PhotoniX Life*, 1:11. https://doi.org/10.3724/PXLIFE.2025-0012

Bai BJ, Wang HD, Li YZ, et al., 2022. Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning. *BME Front*, 2022:9786242. https://doi.org/10.34133/2022/9786242

Bai BJ, Yang XL, Li YZ, et al., 2023. Deep learning-enabled virtual histological staining of biological samples. *Light Sci Appl*, 12(1):57. https://doi.org/10.1038/s41377-023-01104-7

Bankhead P, Loughrey MB, Fernández JA, et al., 2017. QuPath: open source software for digital pathology image analysis. *Sci Rep*, 7(1):16878. https://doi.org/10.1038/s41598-017-17204-5

Belaldavar C, Angadi PV, Mudenagudi U, 2024. QuPath for automated analysis of digital images of oral epithelial dysplasia. *J Oral Maxillofac Pathol*, 28(3):381-386. https://doi.org/10.4103/jomfp.jomfp_499_23

Bińkowski M, Sutherland DJ, Arbel M, et al., 2018. Demystifying MMD GANs. https://arxiv.org/abs/1801.01401

Cao R, Nelson SD, Davis S, et al., 2023. Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nat Biomed Eng*, 7(2):124-134. https://doi.org/10.1038/s41551-022-00940-z

de Haan K, Zhang YJ, Zuckerman JE, et al., 2021. Deep learning-based transformation of H&E stained tissues into special stains. *Nat Commun*, 12(1):4884. https://doi.org/10.1038/s41467-021-25221-2

Feldman AT, Wolfe D, 2014. Tissue processing and hematoxylin and eosin staining. In: Day CE (Ed.), Histopathology: Methods in Molecular Biology. Humana Press, New York, p.31-43. https://doi.org/10.1007/978-1-4939-1050-2_3

Fu H, Gong MM, Wang CH, et al., 2019. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.2422-2431. https://doi.org/10.1109/CVPR.2019.00253

Guo XL, Tao YY, Zhang YZ, et al., 2026. CycleGAN-TMDE: an image dehazing model using cycle generative adversarial network with transmission map and depth estimation. *Chin J Electron*, 35(1):377-391. https://doi.org/10.23919/cje.2024.00.141

Heusel M, Ramsauer H, Unterthiner T, et al., 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Proc 31st Int Conf on Neural Information Processing Systems, p.6629-6640.

Huang JH, Kou TD, He TY, et al., 2025. Advances in deep learning-based virtual staining of pathological tissues. *Laser Optoelectron Prog*, 62(14):1400001 (in Chinese). https://doi.org/10.3788/LOP242293

Huang X, Liu MY, Belongie S, et al., 2018. Multimodal unsupervised image-to-image translation. Proc 15th European Conf on Computer Vision, p.179-196. https://doi.org/10.1007/978-3-030-01219-9_11

Isola P, Zhu JY, Zhou TH, et al., 2017. Image-to-image translation with conditional adversarial networks. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.5967-5976. https://doi.org/10.1109/CVPR.2017.632

Karkampouna S, Kruithof-de Julio M, 2023. Dataset EMPaCT TMA. https://doi.org/10.5281/zenodo.10066853

Kim B, Kwon G, Kim K, et al., 2024. Unpaired image-to-image translation via neural Schrödinger bridge. Proc 12th Int Conf on Learning Representations.

Kreiss L, Jiang SW, Li X, et al., 2023. Digital staining in optical microscopy using deep learning - a review. *PhotoniX*, 4(1):34. https://doi.org/10.1186/s43074-023-00113-4

Li X, 2023. UTOM Dataset: Whole-Slide Pathological Images of Various Kinds of Cancers. https://doi.org/10.5281/zenodo.8376199

Li XY, Zhang GX, Qiao H, et al., 2021. Unsupervised content-preserving transformation for optical microscopy. *Light Sci Appl*, 10(1):44. https://doi.org/10.1038/s41377-021-00484-y

Liu ZJ, Chen LC, Cheng HX, et al., 2024. Virtual formalin-fixed and paraffin-embedded staining of fresh brain tissue via stimulated Raman CycleGAN model. *Sci Adv*, 10(13):eadn3426. https://doi.org/10.1126/sciadv.adn3426

Özbey M, Dalmaz O, Dar S, et al., 2023. Unsupervised medical image translation with adversarial diffusion models. *IEEE Trans Med Imag*, 42(12):3524-3539. https://doi.org/10.1109/TMI.2023.3290149

Park T, Efros AA, Zhang R, et al., 2020. Contrastive learning for unpaired image-to-image translation. Proc 16th European Conf on Computer Vision, p.319-345. https://doi.org/10.1007/978-3-030-58545-7_19

Pati P, Karkampouna S, Bonollo F, et al., 2024. Accelerating histopathology workflows with generative AI-based virtually multiplexed tumour profiling. *Nat Mach Intell*, 6(9):1077-1093. https://doi.org/10.1038/s42256-024-00889-5

Qiu FH, Zhang YY, Guo XG, et al., 2024. Weakly supervised virtual immunohistochemistry staining via Schrödinger bridge method. Proc IEEE Int Conf on Bioinformatics and Biomedicine, p.2369-2375. https://doi.org/10.1109/BIBM62325.2024.10822509

Rivenson Y, Wang HD, Wei ZS, et al., 2019. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*, 3(6):466-477. https://doi.org/10.1038/s41551-019-0362-y

Sun JH, Ye JQ, Chen SY, et al., 2026. Artificial intelligence assisted multi-model pathological diagnosis of breast cancer based on multispectral autofluorescence images. *npj Breast Cancer*, 12(1):62. https://doi.org/10.1038/s41523-026-00915-2

Wang JJ, Xiong B, Zhou Y, et al., 2025. VM-GAN: a value mapping framework for virtual staining of optical microscopy images. *Opt Expr*, 33(12):25641-25655. https://doi.org/10.1364/OE.566786

Wang Q, Fernandes S, Williams GOS, et al., 2022. Deep learning-assisted co-registration of full-spectral autofluorescence lifetime microscopic images with H&E-stained histology images. *Commun Biol*, 5(1):1119. https://doi.org/10.1038/s42003-022-04090-5

Wang XY, Yang S, Zhang J, et al., 2022. Transformer-based unsupervised contrastive learning for histopathological image classification. *Med Image Anal*, 81:102559. https://doi.org/10.1016/j.media.2022.102559

Xie SA, Xu YW, Gong MM, et al., 2023. Unpaired image-to-image translation with shortest path regularization. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10177-10187. https://doi.org/10.1109/CVPR52729.2023.00981

Yang XL, Bai BJ, Zhang YJ, et al., 2022. Virtual stain transfer in histology via cascaded deep neural networks. *ACS Photon*, 9(9):3134-3143. https://doi.org/10.1021/acsphotonics.2c00932

Ye JQ, Sun JH, 2025. Deep Learning Virtual Staining Dataset: Breast, Mouse, and Organoids. https://doi.org/10.5281/zenodo.17646116

Zhang RR, E HH, Yuan LF, et al., 2024. FGM-SPCL: open-set recognition network for medical images based on fine-grained data mixture and spatial position constraint loss. *Chin J Electron*, 33(4):1023-1033. https://doi.org/10.23919/cje.2023.00.081

Zhang SW, Yang B, Yang HP, et al., 2024. Potential rapid intraoperative cancer diagnosis using dynamic full-field optical coherence tomography and deep learning: a prospective cohort study in breast cancer patients. *Sci Bull*, 69(11):1748-1756. https://doi.org/10.1016/j.scib.2024.03.061

Zhang YJ, Huang LZ, Pillar N, et al., 2025. Pixel super-resolved virtual staining of label-free tissue using diffusion models. *Nat Commun*, 16(1):5016. https://doi.org/10.1038/s41467-025-60387-z

Zhu JY, Park T, Isola P, et al., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proc IEEE Int Conf on Computer Vision, p.2242-2251. https://doi.org/10.1109/ICCV.2017.244