Virtual staining: Unpaired image-to-image translation using Cycle-Consistent Generative Adversarial Networks

Table of Contents

Terms and Definitions	
Histology	The study of the microscopic structure of tissues
H&E	Haematoxylin and Eosin stain
IHC	Immunohistochemistry/Immunohistochemical stain
Virtual staining	Computation technique of digital simulation of stains on tissue images
BCI	Breast Cancer Immunohistochemistry dataset/original paper
KID	Kernel Inception Distance (metric)
LPIPS	Learned Perceptual Image Patch Similarity (metric)
cGAN	Conditional Generative Adversarial Network
CycleGAN	Cycle-Consistent Generative Adversarial Network
SR-GAN	Simulated-Raman Generative Adversarial Network
UTOM	Unsupervised Content-preserving Transformation for Optical Microscopy

Abstract

Traditional histological staining methods are known to be tissue-distorting, costly and time-consuming task. And yet, with increase in demand for cheaper, faster and equitable healthcare access, the utilisation of deep learning as an alternative in medical imaging presents a promising approach that overcomes the said challenges. In this project, we implemented a Cycle GAN model to learn the staining task between on 6968 stained and unstained images sourced from the Breast Cancer Immunohistochemical (BCI) dataset (Liu et al., 2022). The trained model was then used to stain 977 unstained images. 20 hand-picked generated images were qualitatively evaluated by experts through a survey who reported that they appeared decent with the correct colours being translated however the resolution was low. The images had a Kernel Inception Score (KID) of 0.20, a decreased cell count by 96.34% and increased cell size by 55.43%. The expert evaluation showed that there was a 50% chance of mistaking the generated images for real images. There was a high stain fidelity and colour realism, but with low diagnostic confidence and nuclear detailing. This showed that there is still promising progress and room to improve the use of deep learning for histology staining in modern healthcare.

Introduction

Biological tissues and cells are colourless and appear transparent to the human eye even after extreme magnification (Alturkistani et al., 2016). To make them visible, staining has traditionally been the preferred approach. Histologists make use of various chemicals like reagents and dyes that allow for the colorization and thus the visualisation of these tissues for easy examination at the microscopic level (Alturkistani et al., 2016).

The goal for histological staining is to highlight different aspects of tissue samples for downstream analysis and diagnostics (Alturkistani et al., 2016). However, this approach can lead to the degradation, distortion or even destruction of certain tissues which poses a technical challenge (Javaeed et al., 2021). Moreover, as the stains evolve to become more specific and complex, the costs associated with the chemical reagents start to increase exponentially (Javaeed et al., 2021). The quality of staining is further exeberated by proper tissue processing and technician skills, which when done poorly, can yield catastrophic results (Alturkistani et al., 2016). All these problems are challenges in the histopathology field and yet there is a lack of a non-distortive and scalable method that can replicate and improve the ideal results of most traditional stains (Deng et al., 2019). In resource-rich settings, most of these issues can be easily solved by using automated pipelines and machinery, which are not easily accessible in low-resource settings, like sub-Saharan Africa (Javaeed et al., 2021). As such, there is need for alternative approaches to implement staining that are affordable, skill independent, accessible, and easy to implement. These approaches should be able to at least replicate traditional methods, with comparable accuracies, and provision for improved performance and are scalable.

As such, virtual staining provides an alternative approach to traditional staining. It refers to simulating different stains or tissue properties digitally. Staining virtually creates faster workflows, easy digital storage, it is non-distortive and therefore, tissue architecture is preserved, and multiple stains can be simulated on one tissue sample (Latonen et al., 2024). This is done using digital techniques like deep learning. Specifically, it involves the use of Convolutional Neural Networks (CNNs) to automatically capture, extract and model complex spatial relationships and patterns from detailed, high resolution imaging data directly from raw pixel values (van der Laak et al., 2021). Generative modelling, implemented through CNNs, provides an approach to virtual staining. New non-existent information can be generated through these models. They learn a distribution within a dataset and learn to create a new data, based on patterns learned from a dataset they have previously seen.

Background

Modern-day staining techniques

Most of the staining techniques developed in the late 19th century still form the basis of the staining techniques used today (Javaeed et al., 2021). The Haematoxylin and Eosin (H&E) is a popular modern-day stain, and it is known to be economical, quick and reliable for diagnosis (Alturkistani et al., 2016). This is an acid-base stain where Haematoxylin will bind to, and dye acidic cellular structures blue (like the nucleus) and Eosin will bind to, and dye basic cellular structures (like the cytoplasm) pink/red (Javaeed et al., 2021). Nearly all tissues today will be stained with H&E for structural analysis (Alturkistani et al., 2016). In the 20th century, as biological/chemical knowledge and the complexity of microscopes increased, more advanced and specialized stains were developed (Javaeed et al., 2021). Simple stains like H&E struggle to differentiate between similar/similarly charged tissues like muscle, fibrosis and collagen (Dibal et al., 2022).

A highly specialized and advanced stain, and the stain focus of my project, would be the Immunohistochemical stain. Immunohistochemistry (IHC), Immunofluorescence (IF) and Immunocytochemistry (ICC) all describe the use of antibodies to see the visual details about where proteins are abundant or localized within the tissue sample which is a key development in diagnostic pathology (Javaeed et al., 2021). Once enzyme-linked antibody detection systems were established, researchers noticed that when these enzymes were exposed to a chromogenic substrate like DAB or AEC, the enzymes produced a coloured precipitate which was visible under a standard light microscope (Deng et al., 2019). This marked the beginning of immunohistochemistry and highlighted how the specificity of immunology can be combined with the visual clarity of histology (Javaeed et al., 2021).

Early Computational Pathology methods and Deep learning-based Virtual Staining

Computational pathology emerged in the 1960s and the field was commonly known as “Quantitative histopathology” or “Computer-assisted microscopy”. The main goal was to automate tedious manual tasks like cell counting and grading as well as quantifying tissue morphology (cell size, shape density etc) (Bengtsson et al., 2017). Even then, the research they conducted using these algorithms showed a clear correlation between the features extracted from the cells/tissues and the patient outcomes (Bengtsson et al., 2017).

Convolutional neural networks (CNNs) are deep-learning networks that can perform pattern recognition tasks from data such as images, audio and text. There are multiple variations of CNNs with U-Net being one of the most important and widely adopted deep learning architecture in biomedicine (Behrad & Saniee Abadeh, 2022). The invention of U-Net further led to the design of different variants that addressed the limitations of basic U-net models or previous non-DL methods, and these were issues relating to complexity, processing time, noise reduction, low-level feature extraction and image quality preservation (Komosar et al., 2024). Examples of this would be 3D U-Nets (regular U-Nets can only use 2D images) and Residual U-Nets which use skip connections from one layer to a different and deeper layer. For image translation and image generation methods like virtual staining, U-Nets are often used as a backbone for the generator model architectures making it a crucial part of virtual staining (Meng et al., 2021).

Regular GANs (Generative adversarial networks) work by generating an image from random noise. These types of DL architectures have two neural networks built into them; a generator that generates an image, and a discriminator that judges whether the generated image is real or fake (Isola et al., 2017). Isola et al introduced a new type of a supervised deep learning framework called Pix2Pix, which is based on a Conditional Generative Adversarial Network (cGAN) (Latonen et al., 2024). This model preforms paired image-to-image translation which makes it conditional because the output/generated image is dependent on the input (Li et al., 2024). The Pix2pix generator uses a ResU-Net architecture for its discriminator, and this often led to high quality, realistic generated images with better scores than if adversarial losses were used alone (Isola et al., 2017).

Compared to supervised models, the unsupervised nature of CycleGANs makes it applicable in broader areas especially when paired examples simply do not exist (Zhu et al., 2017). This is particularly useful in virtual staining where obtaining paired images from the same stained section is chemically impossible (Bai et al., 2023). Even if that was an option, multi-stage image registration processes are needed to align the image pairs down to the pixel level. This is challenging and laborious which defeats the ultimate purpose of virtual staining (Bai et al., 2023).

	Dates	Principle	Methods / Mechanisms	The problem solved	Advantages	Disadvantages / Limitations
Traditional Histological Staining	1600s–1900s	Chemical interaction of dyes with cellular components (acid–base, affinity-based staining).	Tissue fixation → embedding → sectioning → staining (e.g., H&E, trichrome, PAS). Dyes bind to specific cellular structures (acidophilic/basophilic).	Allowed visualization of cellular/tissue morphology and differentiation of tissue components under light microscopy.	Highly interpretable, gold-standard, high pathologist familiarity, reliable	Dye fading, destructive staining (cannot reuse samples, time consuming, intensive, limited multiplexing and requires specialized infrastructure and skilled technicians
Early Computational Pathology (Pre-ML/AI/DL)	1980s–2010	Digital image analysis using rule-based or classical statistical methods.	Colour deconvolution, thresholding, morphology filters, texture and colour feature extraction; traditional classifiers (k-NN, SVM, PCA, logistic regression).	Automated quantification and analysis of stained slides; reduced observer bias; enabled digitization.	Introduced quantitative metrics, allowed digital storage and sharing and reduced some workload	Dependent on handcrafted features, constrained by computational power, needed laborious and explicit programming and could not simulate staining.
U-Net and Variants (DL-based segmentation / mapping)	~2015 onward	Fully convolutional encoder–decoder with skip connections for pixel-wise prediction.	End-to-end CNN trained on paired data; maps unstained → stained appearance using L1/L2 (regression) loss; sometimes extended with attention or residual blocks.	Learned spatial mapping and contextual features automatically, reducing need for manual feature engineering.	Strong structure preservation, good for segmentation and pixel mapping; simple to train and has an interpretable feature hierarchy.	Produces smooth/blurry outputs needs paired and aligned data; cannot generalize to unpaired datasets, prone to overfitting
Pix2pix (Paired/Conditional GANs)	2016	Conditional adversarial learning — learn mapping from input → output domain using paired examples.	Generator (U-Net-like) + PatchGAN discriminator; adversarial + L1 loss; trained on paired datasets (e.g., unstained ↔ stained).	Introduced texture realism; overcame blurriness of U-Nets by using adversarial training; enhanced stain-like appearance.	High-fidelity, sharp, realistic virtual stains, preserves structure because paired data ensures alignment.	Requires perfectly paired and aligned data down to the pixel level (hard to obtain in histology)
CycleGAN (Unpaired GAN)	2017	Unpaired domain translation with cycle-consistency constraint.	Two generators (A→B, B→A) + two discriminators; uses adversarial + cycle-consistency + identity losses to enforce reversibility.	Eliminated need for paired data — enabled virtual staining when only separate sets of unstained and stained images exist.	Can be trained with large, unpaired datasets, produces realistic stain styles, flexible and scalable.	Can distort fine structures, less accuracy than pix2pix, often produces low resolution images, longer training times.

Table 1: An summarised overview of all histological staining techniques starting from traditional wet lab practice to modern deep learning-based virtual staining.All information in this table is sourced from the following papers (alphabetical order): Bamforth et al, 1958;Bai et al., 2023; Javaeed et al., 2021;Latonen et al., 2024;Zhu et al., 2017.

Applications in Biomedical imaging

Regular CycleGANs for unpaired/unsupervised translations are rare in biomedicine due to their limitations. Generic CycleGANs often struggle with complex bioimaging tasks like virtual staining. To enhance stain fidelity and training stability, generic models are modified with specific additions or constraints to address its limitations (Liu et al., 2024).

An example of this is the design and implementation of the Stimulated Raman CycleGAN (SRC-GAN). This modified CycleGAN uses a semi-supervised learning approach where the training dataset goes from paired images to completely unpaired images encouraging the model towards the correct mappings (Liu et al., 2024). The SRC-GAN translates Stimulated Raman Scattering (SRS) images of fresh brain tissue into Formalin-Fixed and Paraffin Embedded H&E quality stains (Liu et al., 2024). This model provides real-time intraoperative diagnosis and can differentiate different subtypes of brain tumours within minutes (Liu et al., 2024).

A true example of unsupervised (unpaired) GAN approach is UTOM. UTOM, a CycleGAN variant, is used to transform label-free autofluorescence human colorectal tissue images into standard H&E-stained images (Li et al., 2021). UTOM is also applied in image restoration and virtual labelling. It can transform low-SNR (signal-to-noise ratio) fluorescence images, widefield images and phase-contrast images into High-SNR, super resolution images or images with specific fluorescence labels (Li et al., 2021).

Research Aim and Objectives

Study Aim

To translate unpaired and unstained tissue images from unstained to an immunohistochemically stained one.

Specific Objectives:

(i) To explore the use of Generative Adversarial Networks for image generation and translation.

(ii) To implement a cycle consistency Generative Adversarial Network for unpaired image to image translation.

(iii) To evaluate the trained Cycle Consistency Generative Adversarial Network on the generated images and evaluating the translated images.

Methodology

To meet our study objectives, we implemented the following tasks for each of the objectives.

Objective	Tasks performed
To explore the use of Generative Adversarial Networks for image generation and translation	• Literature review
To implement a cycle consistency Generative Adversarial Network for unpaired image to image translation	• Dataset acquisition • Data preprocessing • Model Design and Development • Model Training
To evaluate the trained Cycle Consistency Generative Adversarial Network on the generated images	• Testing the model on a separate dataset using KID, LPIPS, and morphology metrics scores and evaluating the translated images.

Dataset Description and Preprocessing

Dataset Preprocessing

We downloaded the H&E and the IHC dataset that were both stained datasets from an open source (Liu et al., 2022). To create an unstained dataset, we converted the H&E dataset to grayscale to have an unstained dataset and discarded the RGB equivalent. This created an unpaired set of H&E grayscale images and IHC images for RGB. We then shuffled and split the data randomly into training and testing using 80% and 20% split for training and testing. We resized the images from 1024*1024 to 256*256, did random horizontal flipping and normalisation.

Data Distribution Preprocessing		
Dataset	**Original Structure**	**Final Structure**
H&E	RGB - Stained	Gray – unstained
IHC	RGB - Stained	RGB - Stained
Data Distribution Postprocessing		
Dataset	**Train**	**Testing**
H&E	3948	977
IHC	3948	977

Table 2: The dataset distribution before and after processing

Technology Stack and Hyper parameters

We trained the cycle consistency GAN for unpaired image translation. We used the following technologies to implement our task, and we trained the model using the following hyperparameters.

Technology Stack	**Version**
Python	3.10.1
PyTorch	2.0
Cuda	11.8.0_520.61.05
ILIFU GPU	NVIDIA A100 40GB
Hyper parameters	**Value**
Batch size	1
Learning Rate	1 x10 $^{-5}$
Cycle consistency weight (λ)	10
Identity Loss	0
Epochs	200

Table 3: The technology stack and hyperparameters used to train the CycleGAN model

Analysis of Results

We utilised two approaches to analyse our results. Quantitatively, we calculated the Morphology metric, the Kernel inception distance (KID) and the Learned Perceptual Image Patch Similarity (LPIPS) metric. The morphology metric is the only metric used in this project that does not calculate based on pre-trained models. This method measures and quantifies structural/morphological histology features such as nucleus area, circularity, cell density, size and shape and the preservation of tissue boundaries and overall shape (Helgadottir et al., 2020). The KID, an unbiased estimator, was used to measure the statistical similarity between the feature distributions of real and fake images (Lin et al., 2025). Lastly, the LPIPS score was calculated to measure the perceptual distance between two images by comparing their feature maps (Lin et al., 2025).

Qualitatively, we conducted an online survey using google forms (https://forms.gle/ZLnodcoaPnUHeDyN7) to capture the perceptions of the generated images among pathologists that have domain understanding on stained breast cancer tissues. We reached out to 10 pathologists with a response rate of 30%. To eliminate bias among the pathologists, they evaluated 10 images that were both real and fake. This survey had closed ended questions, to determine which image was real, which was fake or whether they could not tell. They were also asked to rate the generated images based on aspects like stain fidelity, contrast, diagnostic confidence and morphological fidelity.

Results

Objective 1 - To explore the use of Generative Adversarial Networks for image generation and translation

Through extensive literature review, we discovered that Generative Adversarial Networks (GANs) are used for many tasks, including text to image, image to image, and noise to image tasks, as well as image super resolution (Dash et al., 2024). For image generation tasks they make use of a DCGAN. For this GAN, the generator receives a random noise and tries generates a new image (Dash et al., 2024). We also discovered the use of GANs for image-to-image translation, that includes both paired and unpaired image data (Lin et al., 2025; Bai et al., 2023).

GAN models involve adversarial learning between two networks: a generator and discriminator. The training characteristic of GANs involve a competitive "game" between the two models (Irvine et al., 2024). The generator (G) produces a fake image, the discriminator (G) critically evaluates the image and labels it either real or fake, then the generator uses that judgment to generate a "better" fake images. The discriminator receives the generated image, the actual image, and the source image, determines whether the generated image is real or fake, and provides feedback to the generator, to improve its performance.

There is a constant back and forth between the two networks and eventually the generator learns to generate realistic images (Isola et al., 2017). This framework is called adversarial training, and it follows the objective function below

$$L_{GAN}(G, D, X, Y) = \underbrace{E_{y \sim p_{data}(y)}[\log D(y)]}_{\text{Discriminator loss on real data}} + \underbrace{E_{x \sim p_{data}(x)}[\log(1 - D(G(x)))]}_{\text{Discriminator loss on fake data}}$$

Figure 1: The adversarial loss equation. Image sourced from Dash et. al, 2024

Mathematically, the adversarial loss encourages the generator to minimize the discriminator's ability to classify generated images as fake (Figure 1). This is often expressed as a binary cross-entropy or least-squares loss between the discriminator's predictions and the true scores (real = 1, fake = 0). In the first term, $Ey\sim_{pdata}(y)$ represents the distribution of all the real images in the target domain. Because these are real images, the discriminator wants to minimize this value so that when Dy=1, log (1) ≈0. In the second term, $Ex\sim_{pdata}(x)$ represents the distribution of the noise or generated images. Here the

discriminator aims to maximise the value so that when G(x)=0, log(0) = -infinity. If the generated images (G(x)) gets a score for real (1), it would mean that $Ex\sim_{pdata}(x)(\log(1-1)) = -\infty$ which is what it is trying to avoid therefore the discriminator must criticize as much as possible and get as close to a probability of 0 as possible for all fake data (Dash et al., 2024; Isola et al., 2017)

For paired image translation, they often make use of cGANs like Pix2Pix (Isola et al., 2017). The generator receives an image from one domain and learns to translate it to another domain. The discriminator judges the generated image by comparing it to its image pair. This model falls under supervised deep learning because the generated images are dependent on the source image (Irvine et al., 2024; Behrad & Saniee Abadeh, 2022).

For unpaired image translation, we discovered that the model learns the mappings of input domain X to target domain Y without the need for aligned image pairs (Zhu et al., 2017). This model achieves this by having two generators and two discriminators where instead of only translating X→Y, the second set of networks translates Y→X (Zhu et al., 2017). Discriminators D(x) and D(y) compare the generated images to the distribution of real images in in the rest of the dataset for it to calculate its judgment (Zhu et al., 2017). The model supervises its learning at the level of image sets, meaning that it only requires images belonging to domain *X* and a different collection belonging to domain *Y* to perform its task (Hu, 2024). CycleGANs have a special loss calculation called the Cycle Consistency Loss, this loss encourages the generators to learn the reverse mappings of the image which ensure that if the generated image was to be converted back to its original domain, it must look exactly like the original image (Hu, 2024).

$$L_{cyc}(G_{XY}, G_{YX}) = \mathbb{E}_{x\sim p_{data}(x)}[||G_{YX}(G_{XY}(x)) - x||_1] + \mathbb{E}_{y\sim p_{data}(y)}[||G_{XY}(G_{YX}(y)) - y||_1]$$

Figure 2: The cycle consistency loss equation. Image sourced from Dash et. al, 2024

$G_{YX}(G_{XY}(x))$ represent the forward cycle where X is being translated into Y, $G_{XY}(G_{YX}(y))$ represents the backwards pass where Y is translated back to X. "||…||1" represents the L1 norm calculation. Because there is no ground truth example to compare the generated image to, the discriminator makes it judgment by comparing it to the distribution of the target domain within the dataset (Dash et al., 2024).

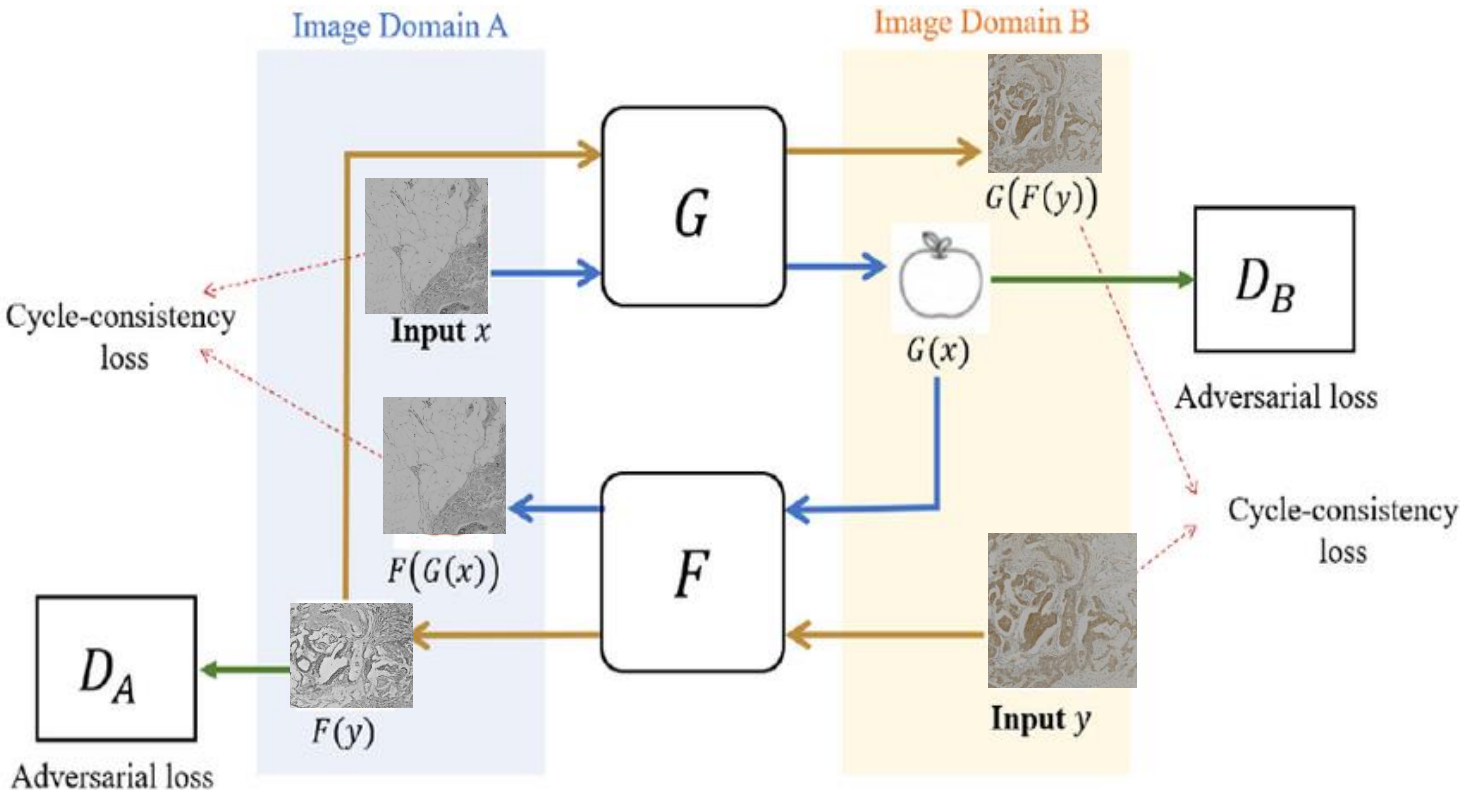


Figure 3: A diagram of the CycleGAN training process. Domain A represents unstained tissue images and Domain B represents stained tissue images. This diagram is an edited open-source image from Research-gate (Ji et. al, 2020)

Objective 2 **-** To implement a cycle consistency Generative Adversarial Network for unpaired image to image translation.

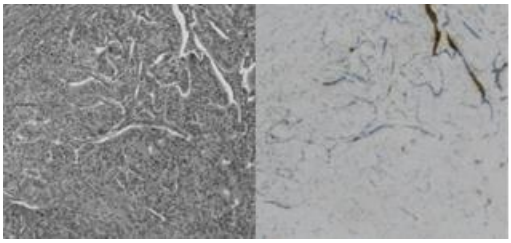
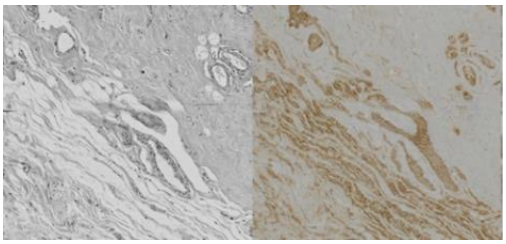
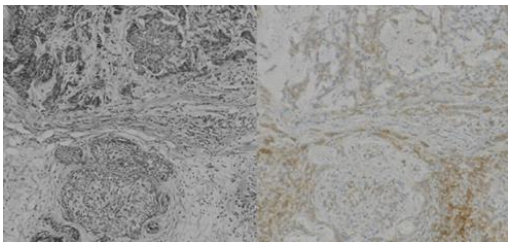
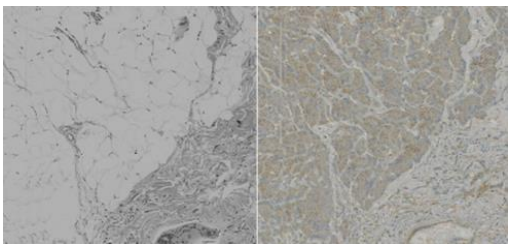
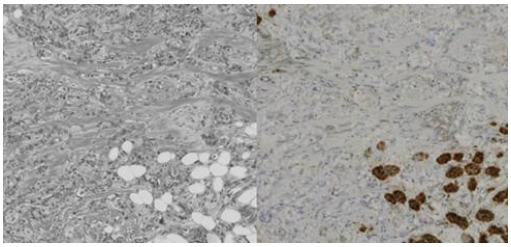
HER2 Expression Level	Translated Image: Unstained → Virtually stained
0	
1	
2	
3	
3	

Table 4:Generated images corresponding to the level of HER2 expression. 0- HER2-negative, 1- low HER2 expression but cancer is considered HER2 Negative, 2- HER2 borderline, confirmatory tests needed and. 3- HER2 positive, high expression meaning patient is a candidate for targeted therapy

Objective 3 - To evaluate the trained Cycle Consistency Generative Adversarial Network on the generated images and evaluating the translated images.

Quantitative Results

	Morphology Metrics			**LPIPS**	**KID**
	Cell Count	**Avg. Cell Size**	**Avg. Circularity**		
Unstained	594643	1,224 px^2	0.45	0.496	N/A
Stained	71,678	13,926 px^2	0.51	0.515	N/A
Generated -Stained	2,625	21,643 px^2	0.45	0.503	0.203

Qualitative Results

From the 10 pathologists who received the google form, 3 of the 10 responded. 2 of the 3 pathologists could not clearly distinguish whether the images provided to them were real or fake.When analysing the quality of virtual staining, the pathologists rated the images highly on stained fidelity and colour realism. However, they did not rate the images highly on nuclear detail clarity and diagnostic confidence.

Discussion

The virtually stained images generally preserved tissue morphology and showed staining patterns consistent with HER2 expression levels. The model accurately reproduced key structures such as epithelial rings around ducts. This can be seen in table 4 where it's evident that the structures were generally preserved after the translation process. This is also supported by the results of the blind survey done by the pathologists. 2 out of 3 were unable to identify which image was real and which image was fake. However, the model occasionally misinterpreted luminal spaces as epithelial tissue, causing abnormal staining and faint or missing nuclei, particularly in dense stromal regions (Table 4, Row 2). A potential reason for this could be that the dense stromal regions of the breast tissue were underrepresented in the dataset leading to poor performance when staining these areas.

Despite these artifacts and some loss of structural context, the generated images displayed biologically plausible patterns and HER2-dependent staining variation. In Table 4, we can see that the model could learn and preserve the varying HER2 expression levels without receiving any explicit guidance from the unstained input images. This highlights its ability to capture subtle but very biologically relevant staining patterns, like protein expression, from unpaired datasets.

The morphological metrics indicate a reduced cell counts and an increased average cell size in the generated images (Table 5). This is likely influenced by the pre-processing process blending cells together into one and the model learning that that as the size of a cell. The circularity of cells remained consistent, indicating that post-translation, circular shapes could still be detected meaning that the individual cell shapes were maintained.

Perceptual similarity scores (LPIPS ~0.5 and extremely low KID of 0.02) suggested statistical similarity but also highlighted limitations of using pretrained models on histology images. Sources explain how there is a lack of accurate quantitative metrics for evaluating the outputs of a Cycle-GAN model (Klöckner et al., 2025). The LPIPS and KID scores prove that the generated images can be visually realistic and statistically accurate but still biologically inaccurate.

These scores do not guarantee, nor do they measure if the colour of the stains correspond correctly with the tissue structure or if the morphological structures are truly preserved (Klöckner et al., 2025). This is supported by the results where open spaces (lumens of the milk ducts) were stained as epithelial tissues. This is further supported by the rated evaluation, with results indicating the synthetic images were visually realistic and had high stain fidelity but lower nuclear clarity and diagnostic confidence.

In recent work, researchers aimed to solve this issue. Unsupervised content-preserving transformation for optical microscopy (UTOM) is a CycleGAN framework that introduces saliency constraint to the model which locates specific image content and ensures that the mappings remain similar after transformation (Li et al., 2021). In a study by Li et al (2021), the researchers discovered that UTOM (a modified Cycle-GAN) could generate the correct stain translations and preserve image content compared to a regular CycleGAN which still struggled with achieving both even when an identity loss was included (Li et al., 2021). While cycle consistency ensures that the output of the second domain is similar to the input of the first generator, saliency enforces structural and morphological consistency between the two domains by mapping the edges, textures and intensity maps before and after the translation (Li et al., 2021).

A simple mathematical function was used to calculate the morphology metrics however, using Cellpose, a deep learning-based instance segmentation algorithm, to calculate the morphology metrics could have improved the accuracy and biological relevance of the results. Cellpose is more precise when measuring individual cells and tissue structures therefore providing a better and reliable quantification of morphological features such as cell size, shape, and spatial distribution (Mezei et al., 2024).

Gaps and limitations

We had a series of limitations in this study. There were no unstained images that were readily accessible. To address that, we created a grayscale dataset, which represented the unstained images. This distorted some features about the images, likely explaining the significant differences in the various morphology metrics. The morphology scores further demonstrate the low quality of translation when using the cycle consistency GAN for this task (Helgadottir et al., 2020).

In addition, there are no universally adopted metrics to assess unpaired image translation tasks, therefore making it difficult to benchmark our results in comparison to other studies (Klöckner et al., 2025). Moreover, the identified metrics rely on pretrained models (Inception v3 and ImageNet), which are trained using natural images and not histological images. Therefore, the calculated scores could be inaccurate and divergent from the expected scores (Klöckner et al., 2025).

Recommendations

To potentially improve the effectiveness of the cycle consistency GAN in image translation of unpaired images for histological analysis, we recommend the collection and use of actual unstained images for model training and prediction. We also recommend modifying the architecture to include the addition of a saliency constraint. As mentioned previously, the Cycle-GAN variant, UTOM, showed strict content-preserving stain translations which is crucial for bioimaging tasks where structural integrity is vital (Li et al., 2021). We also recommend the exploration of using transformers or attention mechanisms as an alternative to improve the quality of translation (Behrad & Saniee Abadeh, 2022).

Conclusion

The project successfully achieved its goal by generating visually convincing IHC-stained images. However, visual realism did not always correspond to biological accuracy (Irvine et al., 2024). The CycleGAN model showed difficulties in correctly staining certain tissue regions, particularly the stromal and interlobular areas, where it often confused luminal spaces with epithelial cytoplasm. These regions were also prone to mode collapse during training. While this study highlights the CycleGAN's ability to translate unpaired images and preserve most spatial structures, it reveals challenges in meeting proper staining standards for diagnosis. This indicating a room for improvement in achieving reliable and biologically consistent translations as well as the development of evaluation methods for them (Helgadottir et al., 2020). GAN-based virtual staining is a rapidly advancing domain with future directions including multitranslation virtual staining and 3D staining (Latonen et al., 2024; Irvine et al., 2024). Overall, this project shows how more modifications and additions to the regular CycleGAN model need to be made before any real-world implementations and human experts are crucial in the development and evaluation of any virtual staining tasks.

References

1. Alturkistani, H. A., Tashkandi, F. M., & Mohammedsaleh, Z. M. (2016). Histological Stains: A Literature Review and Case Study. *Global Journal of Health Science*, *8*(3), 72–79. https://doi.org/10.5539/gjhs.v8n3p72
2. Javaeed, A., Qamar, S., Ali, S., Mustafa, M. A. T., Nusrat, A., & Ghauri, S. K. (2021). Histological Stains in the Past, Present, and Future. *Cureus*, *13*(10), e18486. https://doi.org/10.7759/cureus.18486
3. Deng, F., Miller, J., & Swanson, T. (2019). Histology Revolution: From Inefficient, Two-Dimensional, and Low-Resolution Techniques to High-Throughput, Three-Dimensional and High-Resolution Techniques. *EJH (European Journal of Histochemistry)*, *43*(2), 373–. https://doi.org/10.21608/ejh.2019.16977.1168
4. Latonen, L., Koivukoski, S., Khan, U., & Ruusuvuori, P. (2024). Virtual staining for histology by deep learning. *Trends in Biotechnology*, *42*(9), 1177–1191. https://doi.org/10.1016/j.tibtech.2024.02.009
5. van der Laak, J., Litjens, G., & Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. *Nature Medicine*. https://doi.org/10.1038/s41591-021-01343-4
6. Dibal, N. I., Garba, S. H., & Jacks, T. W. (2022). Histological stains and their application in teaching and research. Asian Journal of Health Sciences, 8(2), 43. https://doi.org/10.15419/ajhs.v8i2.514
7. Bamforth, J., & Osborn, G. R. (1958). Diagnosis from cells. *J. clin. Path.*, *11*(6), 473–482. https://doi.org/10.1136/jcp.11.6.473
8. Bengtsson, E., Danielsen, H., Treanor, D., Gurcan, M. N., MacAulay, C., & Molnár, B. (2017). Computer-aided diagnostics in digital pathology. Cytometry Part A, 91A(6), 551–554. https://doi.org/10.1002/cyto.a.23151
9. Behrad, F., & Saniee Abadeh, M. (2022). An overview of deep learning methods for multimodal medical data mining. *Expert Systems With Applications, 200*, 117006. https://doi.org/10.1016/j.eswa.2022.117006
10. Komosar, A., Stefanović, D., & Sladojević, S. (2024). An overview of image processing in biomedicine using U-Net convolutional neural network architecture. *Journal of Computer and Forensic Sciences, 3*(1). https://doi.org/10.5937/jcfs3-48848
11. Meng, X., Li, X., & Wang, X. (2021). A computationally virtual histological staining method to ovarian cancer tissue by deep generative adversarial networks. *Computational and Mathematical Methods in Medicine, 2021*, Article 4244157. https://doi.org/10.1155/2021/4244157

12. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5967–5976). IEEE. https://doi.org/10.1109/CVPR.2017.632
13. Li, Y., Pillar, N., Li, J., Liu, T., Wu, D., Sun, S., Ma, G., de Haan, K., Huang, L., Zhang, Y., Hamidi, S., Urisman, A., Haran, T. K., Wallace, W. D., Zuckerman, J. E., & Ozcan, A. (2024). Virtual histological staining of unlabeled autopsy tissue. *Nature Communications, 15*, Article 444. https://doi.org/10.1038/s41467-024-46077-2
14. Li, X., Zhang, G., Qiao, H., et al. (2021). Unsupervised content-preserving transformation for optical microscopy. *Light: Science & Applications, 10*, Article 44. https://doi.org/10.1038/s41377-021-00484-y
15. Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2242–2251). IEEE. https://doi.org/10.1109/ICCV.2017.244
16. Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., & Ozcan, A. (2023). Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications, 12*, Article 57. https://doi.org/10.1038/s41377-023-01104-7
17. Klöckner, P., Teixeira, J., Montezuma, D., Fraga, J., Horlings, H. M., Cardoso, J. S., & Oliveira, S. P. (2025). H&E to IHC virtual staining methods in breast cancer: an overview and benchmarking. *NPJ Precision Oncology*. https://doi.org/10.1038/s41746-025-01741-9
18. Irvine, S. C., Lucas, C., Bootbool, M., Galli, S., Zeller-Plumhoff, B., & Moosmann, J. P. (2024). Multi-modal image registration and machine learning for the generation of 3D virtual histology of bone implants. In B. Müller & G. Wang (Eds.), *Developments in X-Ray Tomography XV* (page 70). SPIE.
19. Lin, W., Hu, Y., Zhu, R., Wang, B., & Wang, L. (2025). Virtual staining for pathology: Challenges, limitations and perspectives. *Intelligent Oncology*, *1*, 105–119. https://doi.org/10.1016/j.intonc.2025.03.005
20. Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., & Jin, M. (2022.). BCI: Breast cancer immunohistochemical image generation through pyramid Pix2pix [Technical report]. Beijing University of Posts and Telecommunications.
21. Helgadottir, S., Midtvedt, B., Pineda, J., Sabirsh, A., Adiels, C. B., Romeo, S., Midtvedt, D., & Volpe, G. (2020). *Extracting quantitative biological information from bright-field cell images using deep learning*. (Report/Manuscript).

22. Dash, A., Ye, J., & Wang, G. (2024). A Review of Generative Adversarial Networks (GANs) and Its Applications in a Wide Variety of Disciplines: From Medical to Remote Sensing. *IEEE Access, 12*, 1475–1504.
23. Hu, Y. (2024). Impact of hyperparameters on the quality of image translation using CycleGAN. *AIDML, 5*.
24. Mezei, T., Kolcsár, M., Joó, A., & Gurzu, S. (2024). Image analysis in histopathology and cytopathology: From early days to current perspectives. J. Imaging, 10(10), 252. https://doi.org/10.3390/jimaging10100252