# Towards Pathology-Aware Evaluation and Scalable Preprocessing for Virtual Staining

by

Qiankai Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2025

© Qiankai Wang 2025

## Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Virtual staining has emerged as a promising alternative to traditional histological staining techniques, offering reagent-free, non-destructive generation of diagnostic-quality images from label-free modalities. However, two critical challenges hinder its broader adoption: the lack of domain-specific evaluation metrics and the inefficiency of large-scale whole slide image (WSI) preprocessing. Conventional image quality metrics such as SSIM and LPIPS fail to capture the diagnostic relevance of histological structures, while most WSI slicing tools are not optimized for high-throughput virtual staining pipelines.

This thesis addresses both challenges. First, we develop a multi-threaded WSI slicing framework tailored for OME-TIFF images, enabling scalable and efficient patch extraction with tile-aware indexing, thread-safe file I/O, and optional in-memory caching. Our method achieves a 6–10× speedup over traditional serial approaches while maintaining minimal memory overhead.

Second, we propose PaPIS (Pathological Perceptual Image Similarity), a full-reference, pathology-aware image quality metric. PaPIS leverages deep features extracted from a pretrained cell morphology segmentation model and incorporates Retinex-based feature decomposition to evaluate structural and perceptual fidelity from a diagnostic perspective. Experimental results show that PaPIS correlates better with histological quality than traditional metrics.

Finally, we integrate PaPIS as a perceptual loss in a modified CycleGAN model for virtual staining, demonstrating improved visual realism and pathology alignment in both patch-wise and whole-slide outputs. Together, our contributions provide a robust foundation for scalable, pathology-aware virtual staining pipelines.

**Acknowledgements**

I would like to express my deepest gratitude to my supervisors, Professor Parsin Haji Reza and Professor Anita Layton, for their insightful guidance, unwavering support, and constant encouragement throughout the course of this research. Their expertise, mentorship, and generosity with their time were instrumental to the development and completion of this work.

I am also thankful for the support provided by the Photomedicines Lab at the University of Waterloo. I am especially grateful to my labmates, most notably James E.D. Tweel, for their collaboration, thoughtful discussions, and steady encouragement along the way.

Finally, I would like to extend my heartfelt appreciation to my family and close friends. Your patience, understanding, and unwavering belief in me have been a source of strength throughout this journey.

## Dedication

To the version of myself who, two years ago, made the brave decision to travel over 10,000 miles and begin a new life alone in a distant land—thank you for your courage. For enduring solitude, embracing uncertainty, and growing through it all, I am proud of the person you've become.

To my parents, Xingfeng Wang and Jun Ma, for your unconditional love and unwavering support. Your belief in me has been my foundation and my strength. Even from across the world, your presence was always felt.

To the friends who offered laughter, late-night messages, and quiet companionship—you reminded me that even far from home, I was never truly alone.

To all of you—my parents, my friends, my mentors, and my labmates—thank you for sharing fragments of your lives with me. Through these encounters, I rediscovered and reshaped parts of myself. You have helped make me whole.

# Table of Contents

**Author's Declaration** ii

**Abstract** iii

**Acknowledgements** iv

**Dedication** v

**List of Figures** x

**List of Tables** xiii

**List of Abbreviations** xiv

**1 Introduction** 1

1.1 Background . . . . . 1

1.2 Problem Statement . . . . . 2

1.3 Objectives . . . . . 2

1.4 Contributions . . . . . 3

1.5 Thesis Organization . . . . . 3

**2 Related Works** **5**
2.1 Histological Staining . . . 5
2.2 Virtual Staining . . . 6
2.2.1 Channel Mapping . . . 6
2.2.2 Encoder-Decoder Architectures . . . 9
2.2.3 GAN-Based Models . . . 10
2.2.4 Diffusion Models . . . 10
2.3 Large Scale Whole Slide Storage Formatting . . . 13
2.3.1 Pyramid Structures . . . 13
2.3.2 Chunk-Based Compression . . . 14
2.4 Image Quality Assessment . . . 14
2.4.1 Full-Reference Image Quality Assessment . . . 15
2.4.2 Reduced-Reference IQA . . . 15
2.4.3 No-Reference IQA . . . 15
2.4.4 Perceptual IQA . . . 16
2.5 Conclusion . . . 16

**3 WSI Storage and High-Throughput Slicing Tools for Virtual Staining Workflows** **18**
3.1 Introduction . . . 18
3.2 Literature Review . . . 19
3.3 Methods . . . 21
3.3.1 TIFF-based Positioning . . . 21
3.3.2 Concurrent Assignment . . . 24
3.4 Experiment . . . 25
3.4.1 Setup . . . 25
3.4.2 Results . . . 26
3.4.3 Discussion . . . 26
3.5 Conclusion . . . 27

**4 Pathology-Guided Virtual Staining Metric for Evaluation** **28**
4.1 Introduction . . . 28
4.2 Literature Review . . . 29
4.3 Methods . . . 30
4.3.1 Cell Morphology Level Feature Representation . . . 31
4.3.2 Retinex Properties on Features . . . 33
4.3.3 PaPIS: Pathological Perceptual Distance . . . 35
4.4 Experiments . . . 36
4.4.1 Dataset Acquisition . . . 36
4.4.2 Comparison with Previous Metrics . . . 36
4.5 Conclusions . . . 44

**5 PaPIS Guided Virtual Staining Enhancement** **46**
5.1 Introduction . . . 46
5.2 Literature Review . . . 47
5.3 Methods . . . 48
5.3.1 Dataset Acquisition and Preprocessing . . . 48
5.3.2 Model Architecture and Modifications . . . 48
5.3.3 Pathology-Aware Perceptual Loss . . . 50
5.3.4 Training Procedure . . . 51
5.3.5 Whole Slide Inference and Patch Fusion . . . 52
5.4 Experiments . . . 52
5.4.1 Experimental Setup . . . 52
5.4.2 Quantitative Results on Patch-Level Test Set . . . 53
5.4.3 Qualitative Comparison . . . 54
5.4.4 Whole-Slide Virtual Staining . . . 56
5.5 Conclusion . . . 57

**6 Conclusion** **58**

6.1 Summary of Contributions . . . . . 58

6.2 Integration of Metric and Model . . . . . 59

6.3 Limitations . . . . . 59

6.4 Future Work . . . . . 59

6.5 Final Remarks . . . . . 60

**References** **61**

# List of Figures


2.1 Examples of H&E-stained histological sections. . . . . 6

2.2 Clinical virtual staining pipeline. Adapted from [1]. . . . . 7

2.3 SRS-based virtual staining via channel mapping [37]. . . . . 8

2.4 PARS-based spectral mapping into H&E-like output [14]. . . . . 9

2.5 CycleGAN-based virtual histology pipeline from PARS [52]. . . . . 11

2.6 Multi-channel CycleGAN model using PARS signals [6]. . . . . 12

3.1 Schematic representation of the OME-Zarr image hierarchy. (a) Chunked, multi-resolution layout from plate-level down to compressed storage units. (b) Visual examples illustrating tiling and zooming at various magnifications. 22

3.2 Efficient tile-wise access strategy in compressed WSI formats. **Left:** A virtual patch request that spans multiple image tiles, each stored as an individually compressed block. **Right:** Memory layout visualization for a single tile row. The pink region denotes a contiguous address range on disk where compressed blocks are stored adjacently. By aggregating spatially adjacent tiles, the system minimizes the number of file seek and read operations, enabling faster I/O and partial decoding. . . . . 23

4.1 (a) The figure illustrates the overall framework of the PaPIS. Image patches are fed into a pretrained feature extractor to obtain multi-channel feature representations. These features are decomposed into a set of estimated illumination maps $L_a^i$ and reflectance maps $R_a^i$, where the superscript i denotes the feature layer and the subscript $a$ indicates the image type(Reference or Generated images). The collection of decomposed components is subsequently used to compute perceptual similarity distances between image pairs. (b) This panel presents the internal architecture of the pretrained feature extractor. The feature extractor functions as the encoder component of a cell morphology segmentation model. The subsequent decoder generates both a binary segmentation map $M_b$, which preserves the spatial dimensions of the input, and a multi-layer classification map $M_c$. The fusion of $M_b$, $M_c$, and the original image patch is then fed into a classification model to produce the final classification label. . . . . 32

4.2 (a) The first set of feature maps was visualized from a VGG16 model pretrained on the ImageNet dataset. These feature maps correspond to channels 3, 8, 15, 22, and 29, as used in the works of LPIPS [65] and DISTS [11]. (b) The second set of feature maps was extracted from the encoder output of a pretrained segmentation model, with dimensions of $R^{192}$, $R^{288}$, $R^{480}$, and $R^{1344}$, corresponding to channels at different encoder stages. . . . 34

4.3 Scatter plot comparing Structural Similarity Index (SSIM) and PaPIS across different image pairs. Selected data points are labeled as AH (Aligned High), AL (Aligned Low), PD (PaPIS-Dominant), and TD (Traditionally-Dominant), representing specific cases for analysis. . . . . 38

4.4 Case-wise comparison of PaPIS and SSIM values across different histological image samples. . . . . 39

4.5 Scatter plot comparing LPIPS and PaPIS across different image pairs. Each data point represents an image pair evaluated by both metrics, illustrating variations in their assessments. This comparison highlights differences in perceptual and pathology-aware similarity measurements between the two approaches. . . . . 40

4.6 Scatter plot comparing DISTS and PaPIS across different image pairs. . . 41

4.7 Patch-wise similarity heatmaps of a representative whole slide image (WSI). **Left**: SSIM heatmap; **Middle**: PaPIS heatmap. **Right**: Magnified views of selected regions. Warmer colors indicate higher similarity. Compared to SSIM, which tends to assign high scores across low-texture and background regions, PaPIS demonstrates greater sensitivity to morphological and contextual features, providing finer-grained evaluation of histological relevance. . . . . 42

4.8 Supplemental analysis for OB1 and OB3. Left: SSIM scores in background vs. tissue regions. Center: PaPIS scores in background vs. tissue regions. Right: Coefficient of Variation (CV) across 10 tissue regions for SSIM and PaPIS. PaPIS shows significantly higher intra-region variability, confirming its ability to reflect morphological diversity. . . . . 44

5.1 Pipeline for PaPIS-guided virtual staining. Part (a) is adapted from [51]. The end-to-end framework includes WSI acquisition, automated patch extraction, and model training. . . . . 49

5.2 Architecture of the PaPIS-guided virtual staining model. In addition to the original CycleGAN [67] components, we incorporate PaPIS-based perceptual loss terms to guide the generation of virtual H&E images, encouraging histological alignment with ground truth. . . . . 50

5.3 Qualitative comparison. From left to right: input PARS, ground truth H&E, CycleGAN output, CycleGAN_PAPIS output. . . . . 55

5.4 Comparison of histological staining results. From left to right, the panels show the ground truth H&E-stained whole slide image, the result generated by a standard CycleGAN model, and the result generated by a CycleGAN model trained with PaPIS supervision. The corresponding high-resolution patches on the right highlight local structures. The PaPIS-supervised model preserves finer histological details and exhibits improved color and structure consistency with the ground truth. . . . . 56

# List of Tables

3.1 Comparison of slicing performance across modes (throughput in patches/s). 26

4.1 Metric scores in background vs. tissue regions (mean ± std) . . . . . 43

5.1 Quantitative comparison between baseline CycleGAN and models trained with different perceptual losses. The best score in each column is highlighted in bold. . . . . 54

# List of Abbreviations

**FR-IQA** Full Reference Image Quality Assessment 1–3, 14, 44

**GAN** generative adversarial network 1

**H&E** Hematoxylin and eosin 1, 5, 6, 20, 48, 51

**IHC** immunohistochemistry 1

**IQA** Image Quality Assessment 14, 16, 17, 44

**LRU** Least Recently Used 25, 26

**PaPIS** Pathological Perceptual Image Similarity 3, 4, 44–48, 58

**PARS** Photon Absorption Remote Sensing 8, 10, 20, 29, 30, 48, 52, 54

**PSNR** peak signal-to-noise ratio 1, 31, 36, 57

**QPI** quantitative phase imaging 47

**ROI** region of interest 14, 20

**SSIM** Structural Similarity Index xi, 1, 30, 31, 35–39, 42–44, 48, 53, 54, 57

**WSI** Whole slide image 2–4, 13, 14, 18–21, 24–27, 48, 52, 58

# Chapter 1

# Introduction

## 1.1 Background

Histopathological staining is a cornerstone of disease diagnosis and medical research, enabling visualization of cellular and tissue structures in microscopic slides. Conventional staining techniques such as Hematoxylin and eosin (H&E), immunohistochemistry (IHC), and Masson's trichrome rely on chemical reagents that often alter the tissue irreversibly. These procedures are labor-intensive, time-consuming, and incompatible with certain downstream analyses such as multi-modal imaging.

In contrast, virtual staining has emerged as a powerful computational alternative, generating stained-like representations from label-free imaging modalities such as autofluorescence or phase contrast. By eliminating the need for physical reagents, virtual staining preserves the integrity of biological samples and accelerates the diagnostic pipeline. Recent advancements in deep learning, including generative adversarial networks (GANs) and diffusion models, have significantly improved the realism and fidelity of virtually stained images [42, 45]. Virtual staining offers numerous advantages: reduced cost, rapid turnaround time, and seamless integration with other imaging techniques [29]. Moreover, it facilitates repeated analyses on the same biological samples, enabling longitudinal studies and multi-modal comparisons without compromising tissue integrity.

Despite its promise, the clinical adoption of virtual staining remains limited, partly due to the lack of standardized evaluation criteria. Most studies still rely on Full Reference Image Quality Assessment (FR-IQA) metrics, such as SSIM[57], peak signal-to-noise ratio (PSNR), or LPIPS[65], which were originally developed for natural images. These metrics prioritize low-level texture fidelity and perceptual similarity, but fail to capture the

diagnostic relevance or histological correctness of the images [8, 38]. Consequently, there is a growing need for evaluation metrics that are tailored to the pathological domain and aligned with human expert judgment, ensuring that virtually stained images meet clinical diagnostic standards.

In addition to evaluation challenges, virtual staining pipelines often struggle with data preprocessing bottlenecks. Whole slide images (WSIs) can reach gigapixel scales, making it computationally infeasible to process them directly. These high-resolution images must be tiled into smaller patches for model training and inference. However, existing tiling solutions are either inefficient or lack the flexibility to integrate with large-scale learning frameworks. Efficient and scalable image preprocessing is essential to unlock the full potential of virtual staining models and enable practical clinical implementations.

## 1.2 Problem Statement

Two major challenges limit the performance and scalability of current virtual staining systems:

- **Lack of domain-specific evaluation metrics:** Existing FR-IQA metrics do not consider histopathological relevance, leading to inaccurate assessments of virtual staining quality from a diagnostic perspective.
- **Inefficient preprocessing of high-resolution images:** Current virtual staining pipelines lack robust tools for fast and scalable tile extraction from WSIs, limiting their applicability in large-scale or real-time scenarios.

## 1.3 Objectives

This thesis aims to address the aforementioned challenges by developing efficient computational methods and evaluation frameworks specifically tailored to virtual staining applications. The core objectives of this research are:

- To develop a **high-performance image tiling tool** that efficiently partitions gigapixel-scale WSIs into standardized patches, supporting parallel processing, metadata management, and integration with deep learning workflows.

- To design and implement **PaPIS**, a novel pathology-aware full-reference image quality assessment metric, capable of accurately capturing structural, perceptual, and diagnostic relevance in virtually stained histopathological images.
- To propose and evaluate a **PaPIS-guided virtual staining framework**, leveraging the proposed metric to enhance the quality and diagnostic realism of virtually stained images by integrating pathology-specific perceptual guidance into the training process.

## 1.4 Contributions

The main contributions of this thesis are summarized as follows:

- We propose **Pathological Perceptual Image Similarity (PaPIS)**, a pathology-aware FR-IQA metric that leverages deep features and structural analysis to assess virtual staining image quality from a diagnostic perspective.
- We design and implement a **scalable image tiling framework** for efficient preprocessing of WSIs, supporting concurrent tile extraction and integration with deep learning pipelines.
- We develop and validate a novel **PaPIS-guided virtual staining architecture**, demonstrating how the integration of PaPIS metric significantly enhances image quality and diagnostic accuracy compared to conventional methods.
- We conduct comprehensive experiments to evaluate the proposed methods, demonstrating that PaPIS correlates better with expert judgment than conventional metrics, and that our tiling tool significantly accelerates data processing.

## 1.5 Thesis Organization

The remainder of this thesis is structured as follows:

- Chapter 2 reviews relevant literature on virtual staining techniques, image quality assessment methods, and medical image preprocessing, highlighting critical gaps and justifying the research directions pursued.

- Chapter 3 describes the design and implementation of a high-throughput tiling and preprocessing framework tailored specifically for WSIs, including detailed discussions on performance optimization and scalability.
- Chapter 4 introduces the proposed PaPIS metric, elaborating on its theoretical foundations, computational design, validation methodologies, and diagnostic relevance.
- Chapter 5 presents a PaPIS-guided enhanced virtual staining architecture, demonstrating improved performance achieved by integrating the proposed metric into the training process, supported by rigorous experimental analyses.
- Chapter 6 summarizes the contributions of this thesis, discusses its limitations, and outlines potential directions for future research to further advance virtual staining technologies.

# Chapter 2

# Related Works

## 2.1 Histological Staining

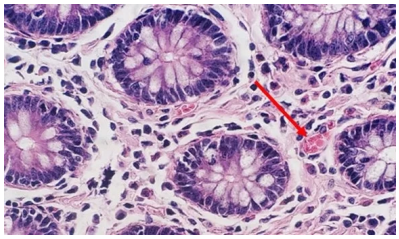
Histological staining enhances contrast in tissue sections, allowing detailed visualization of cellular and extracellular structures critical for pathological assessment. Hematoxylin and Eosin (H&E) staining remains the clinical gold standard due to its accessibility, interpretability, and broad applicability across tissue types [3]. Hematoxylin stains nuclei deep purple-blue by binding to acidic chromatin, while eosin, an acidic dye, highlights the basic cytoplasmic components and extracellular matrix in varying shades of pink. This dual-staining method enables pathologists to discern nuclear morphology, cytoplasmic granularity, and tissue architecture with high precision.

Representative examples are shown in Figure 2.1, illustrating variations in staining quality and tissue characteristics. These images highlight how technical factors, such as reagent concentration or staining time, can significantly influence visual outcomes.

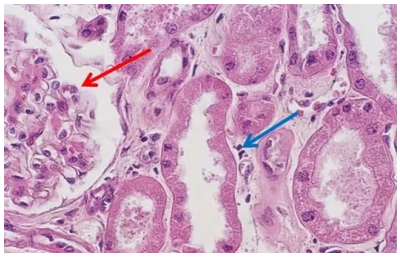
Despite their diagnostic value, conventional staining techniques are inherently destructive, require physical handling of tissue, and introduce batch-dependent variability in color and contrast [20, 28]. Additionally, the irreversible chemical treatment limits tissue reuse for downstream molecular or imaging analyses. These limitations underscore the demand for non-invasive, standardized alternatives that support high-throughput and multiplexed workflows [41].



(a) H&E-stained colon section. Mucin is still visible within goblet cells.



(b) H&E-stained colon section. Nuclei dominate, with muted cytoplasmic features. Over-staining may be present.



(c) H&E-stained kidney tissue. Glomerulus (red arrow) and tubules (blue arrow) are clearly distinguished.

Figure 2.1: Examples of H&E-stained histological sections.

## 2.2 Virtual Staining

Virtual staining synthesizes stained-like images from label-free modalities (e.g., autofluorescence, hyperspectral, or photoacoustic imaging) through computational algorithms [41, 1]. This paradigm enables real-time generation of diagnostic-quality images without chemical staining, thereby preserving tissue integrity and reducing reagent cost and turnaround time. It is particularly advantageous in scenarios requiring repeated imaging, live-cell studies, or multiplex imaging on scarce clinical specimens.

Figure 2.2 illustrates a representative clinical virtual staining pipeline. Starting from high-resolution, label-free image acquisition, a trained generative model infers a stained representation that mimics the contrast and coloration of conventional methods. These synthetic outputs can then be reviewed by pathologists in standard digital workflows.

Virtual staining has been successfully deployed in fluorescence-to-H&E synthesis, nonlinear microscopy transformation, and photonic signal interpretation, with growing interest in its application to intraoperative diagnostics and whole slide digitization.

### 2.2.1 Channel Mapping

Channel mapping is one of the earliest and most straightforward approaches to virtual staining. It leverages empirical or heuristic transformations of intensity distributions from different imaging channels to emulate stained appearances. This method assumes a direct, often linear, correspondence between imaging signals (e.g., Raman spectral components or reflectance) and the color attributes associated with stains.

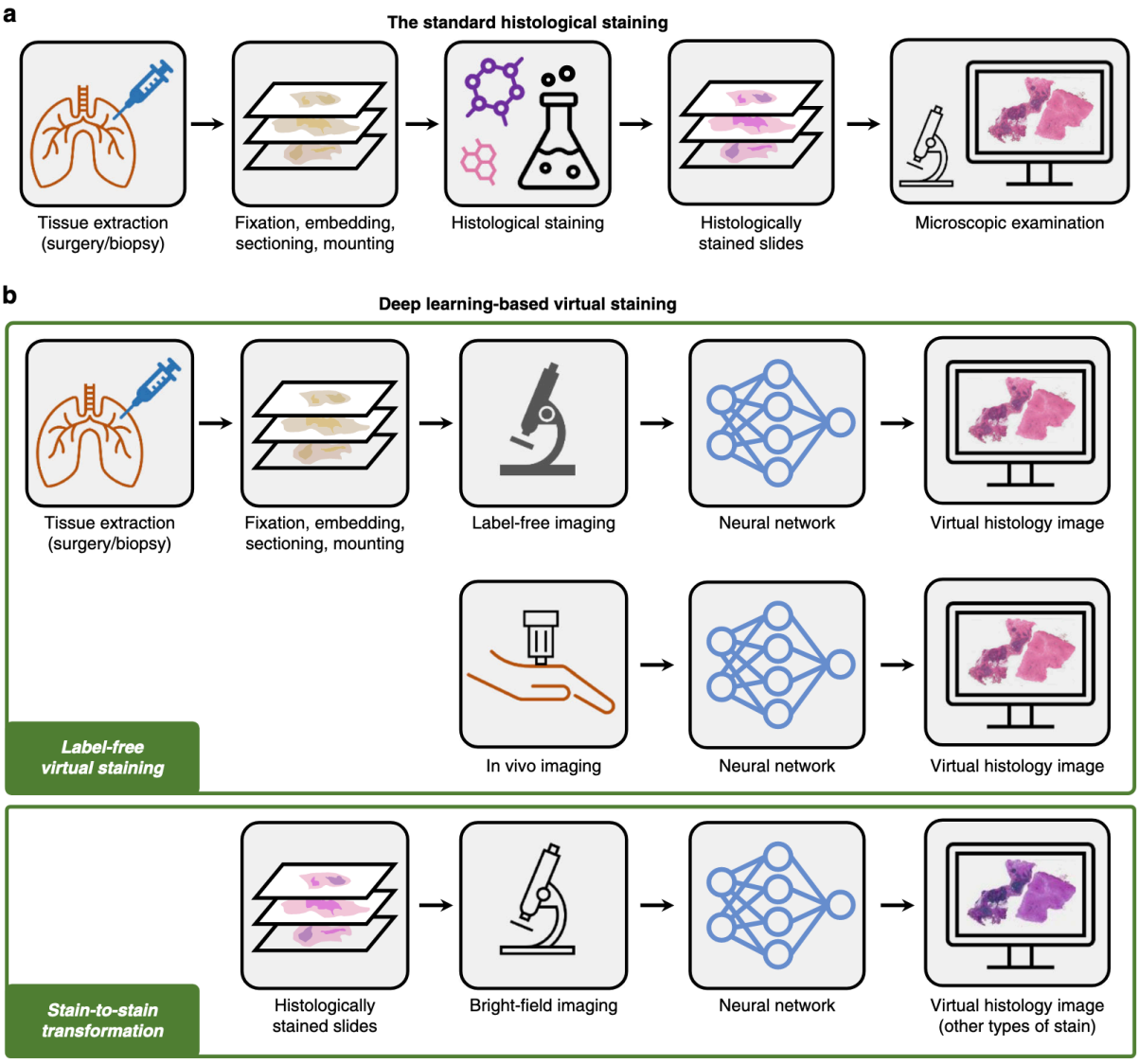



Figure 2.2: Clinical virtual staining pipeline. Adapted from [1].

For instance, Orringer et al. [37] applied SRS microscopy to extract molecular-specific contrast, which was then mapped into an H&E color space. Similarly, Ecclestone et al. [14] combined radiative and non-radiative signals from Photon Absorption Remote Sensing (PARS) imaging across multiple wavelengths to simulate histological features (Figures 2.3 and 2.4). While effective for coarse differentiation of tissue components, these approaches lack flexibility in modeling complex textures, morphological variability, or non-linear contrast transformations.

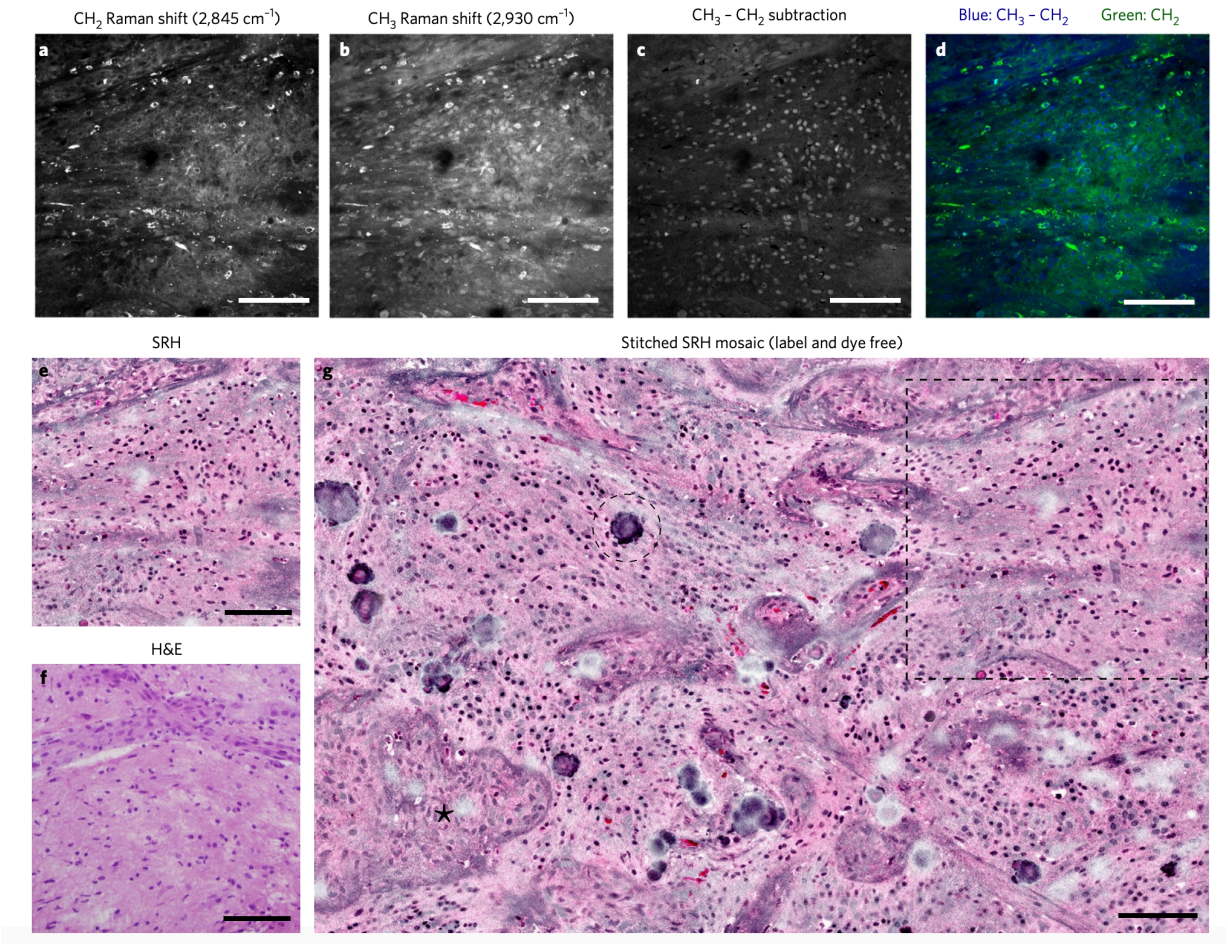



Figure 2.3: SRS-based virtual staining via channel mapping [37].

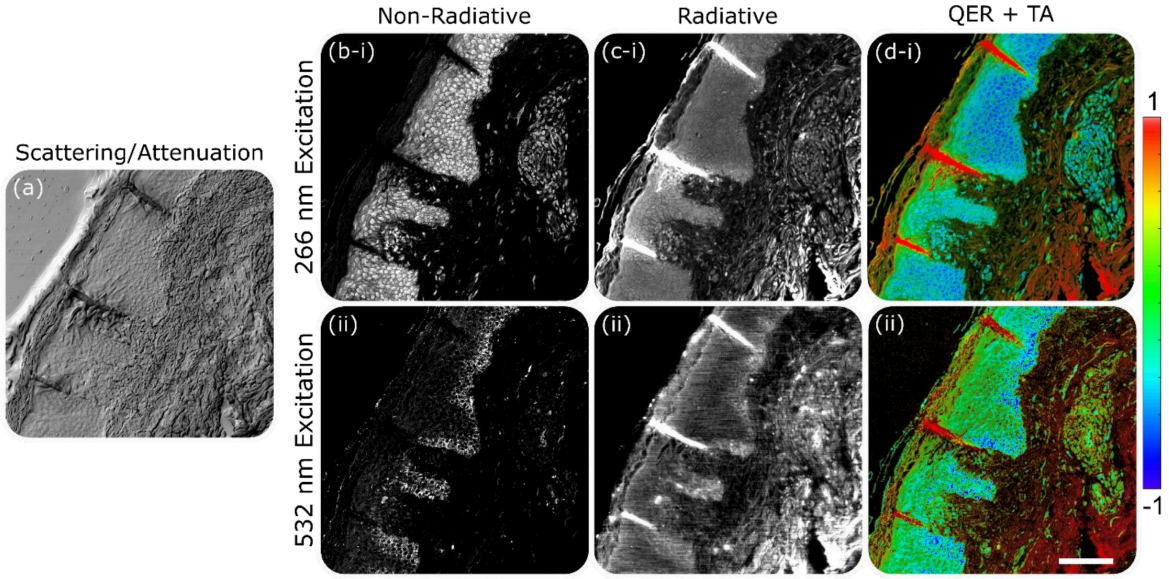



Figure 2.4: PARS-based spectral mapping into H&E-like output [14].

### 2.2.2 Encoder-Decoder Architectures

Encoder-decoder architectures form the foundation of many supervised virtual staining networks. These models learn a pixel-wise transformation from input *label-free images* $I_{\text{LF}} \in \mathbb{R}^{H\times W\times C}$ (e.g., autofluorescence or phase contrast images without chemical staining) to a *stained output domain* $I_{\text{Stain}} \in \mathbb{R}^{H\times W\times C}$, using reconstruction losses such as $L_1$, $L_2$, or perceptual metrics to guide training.

$$\hat{I}_{\text{Stain}} = D(E(I_{\text{LF}})), \quad \mathcal{L}_{\text{recon}} = \|I_{\text{Stain}} - \hat{I}_{\text{Stain}}\|_1$$

Here, $E(\cdot)$ and $D(\cdot)$ denote the encoder and decoder networks, respectively, and $\hat{I}_{\text{Stain}}$ represents the predicted stained image. $\mathcal{L}_{\text{recon}}$ is the pixel-wise reconstruction loss, commonly chosen as the $L_1$-norm.

Rivenson et al. [41] demonstrated a U-Net style architecture for synthesizing H&E images from fluorescence. Yoon et al. [63] proposed the E-CUT framework, incorporating attention mechanisms and saliency to improve interpretability. Yang et al. [61] introduced a cascaded structure to convert autofluorescence to PAS-stained outputs, showing improved color distribution and structure preservation.

These models are relatively easy to train but may produce overly smooth outputs due to the pixel-wise losses, which motivates the use of adversarial and perceptual learning techniques in later works.

### 2.2.3 GAN-Based Models

Generative Adversarial Networks (GANs) enhance the realism of virtual staining by introducing an adversarial loss, which encourages outputs that are indistinguishable from real stained images. The adversarial setup involves training a *generator* $G$ and a *discriminator* $D$ in tandem:

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_x[\log(1 - D(x, G(x)))]$$

Here, $x$ denotes the input label-free image (e.g., autofluorescence), $y$ is the corresponding real stained image, and $G(x)$ is the generated (virtual stained) output. The discriminator $D(x, y)$ learns to distinguish between real and synthetic pairs, while the generator $G$ aims to produce outputs that fool the discriminator.

CycleGAN [67] became a popular architecture for unpaired virtual staining tasks, where ground truth is unavailable or poorly aligned. Latonen et al. [29] used CycleGAN for H&E synthesis from autofluorescence images. Tweel et al. [52] demonstrated CycleGAN training on PARS signals with weak registration (Figure 2.5), and Boktor et al. [6] extended the approach to include multiple PARS channels (Figure 2.6).

While GANs produce visually appealing results, they are prone to hallucinations and training instability, requiring careful tuning of loss functions, discriminator architectures, and normalization strategies.

### 2.2.4 Diffusion Models

Diffusion models have emerged as powerful generative frameworks that iteratively denoise a signal to synthesize high-fidelity images. These models define a forward corruption process by gradually adding Gaussian noise to the original image, and train neural networks to approximate the reverse denoising trajectory.

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1 - \alpha_t)\mathbf{I}), \quad \mathcal{L}_{\text{DM}} = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t, I_{\text{LF}})\|^2\right]$$

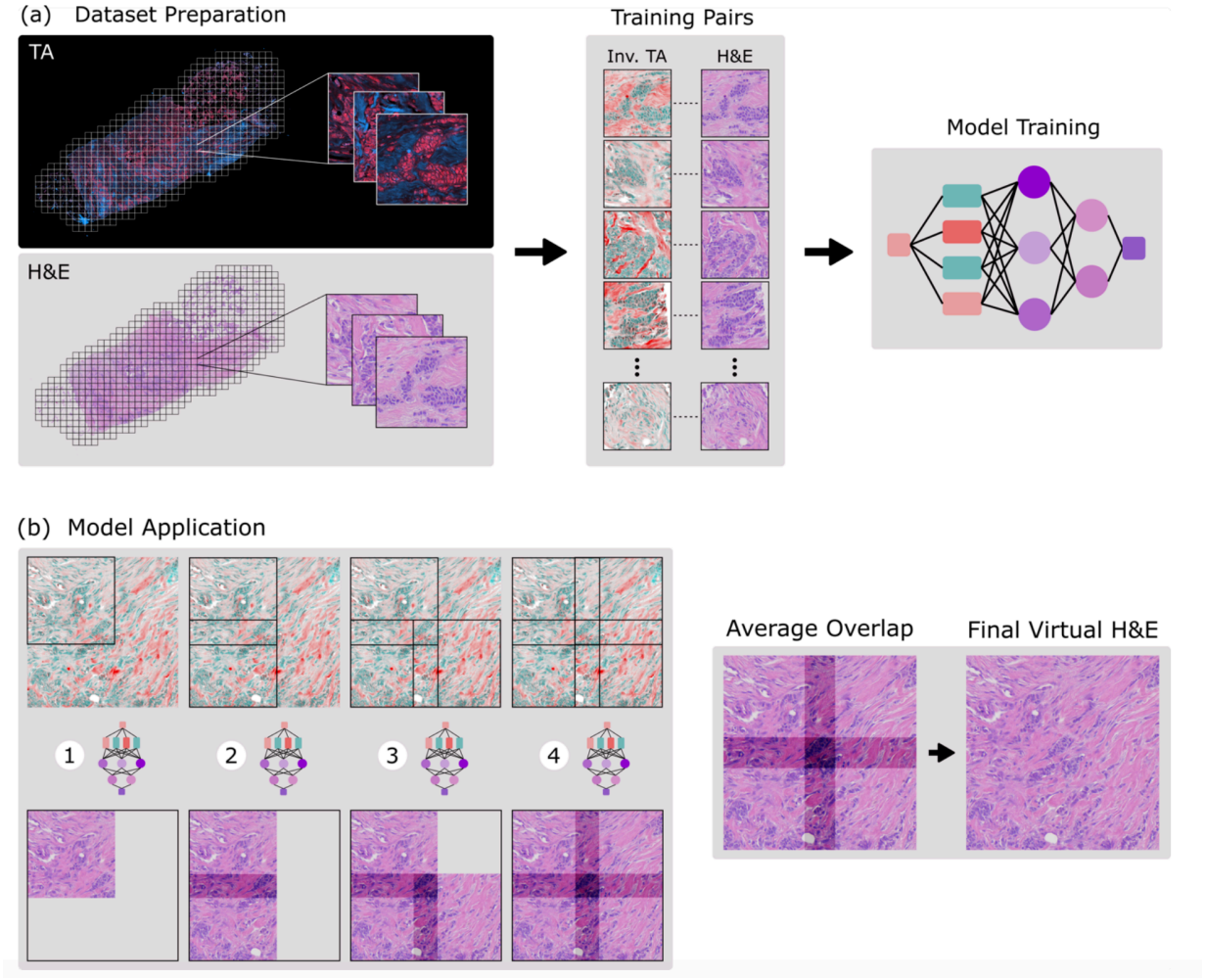



Figure 2.5: CycleGAN-based virtual histology pipeline from PARS [52].

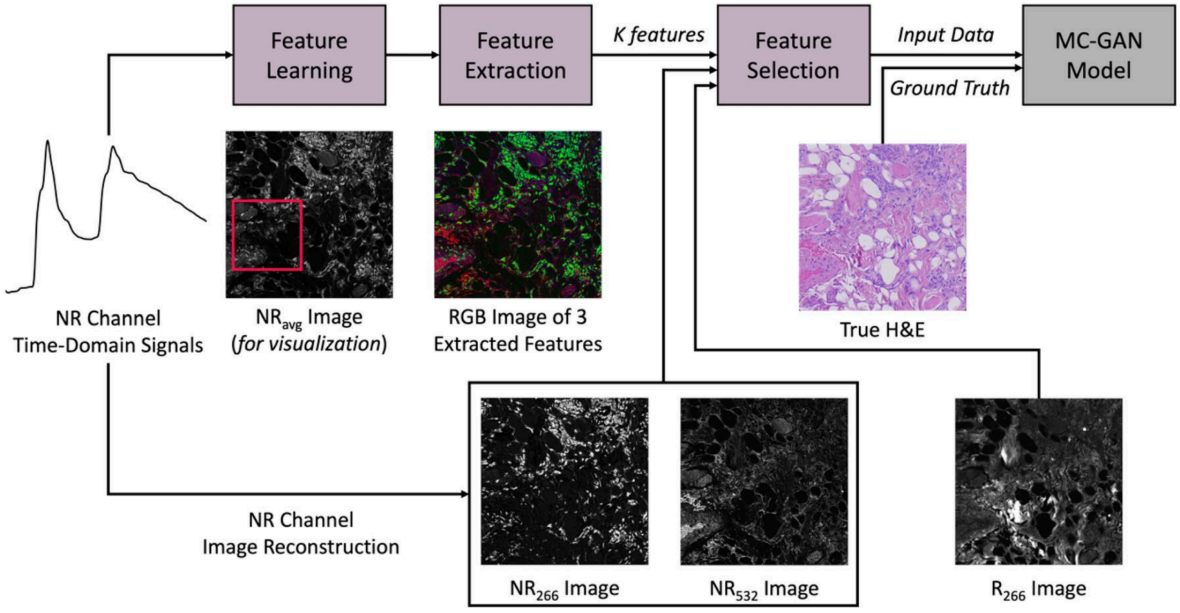



Figure 2.6: Multi-channel CycleGAN model using PARS signals [6].

Here, $x_0$ is the original clean image, $x_t$ is the noisy image at time step $t$, and $\alpha_t \in (0, 1)$ controls the noise level. The function $\epsilon_\theta(\cdot)$ is a neural network trained to predict the added noise $\epsilon$, conditioned on the input label-free image $I_{\text{LF}}$. The loss $\mathcal{L}_{\text{DM}}$ is the mean squared error between the true and predicted noise, a standard training objective for denoising diffusion probabilistic models (DDPMs).

StainDiffuser [26] utilized a dual-task approach, simultaneously generating stain outputs and segmentation masks. Dubey et al. [13] introduced Vision-guided Immunostaining Models (VIMs), which employ LoRA fine-tuning and cross-modal prompts for flexible multiplexed staining. Although diffusion models are computationally intensive, they provide unprecedented control over image semantics and high-frequency texture synthesis.

## 2.3 Large Scale Whole Slide Storage Formatting

Whole Slide Images (WSIs) are foundational to modern digital pathology, capturing entire tissue sections at sub-micron resolution. Due to their immense size—often exceeding 100,000×100,000 pixels—WSIs demand specialized file formats and processing strategies for storage, access, and analysis. Efficient management of WSIs is critical for virtual staining pipelines, which rely on rapid and repeated access to image regions during training and inference.

The Open Microscopy Environment TIFF (OME-TIFF) format has become a widely adopted standard due to its interoperability, metadata support, and compatibility with existing TIFF viewers [31]. Unlike traditional flat TIFFs, OME-TIFF embeds rich XML metadata and supports multiple imaging dimensions, such as time points, z-stacks, and spectral channels. This structure allows it to function as both a raw data container and a metadata-rich representation of the imaging experiment.

In the context of virtual staining, OME-TIFF's compatibility with tools like QuPath and OpenSlide enables seamless integration into WSI tiling, streaming, and annotation workflows. Its extensibility further supports AI-ready pipelines, where image patches can be accessed on demand without the need for full-slide decompression.

### 2.3.1 Pyramid Structures

To enable fast, multi-resolution visualization and processing, OME-TIFF implements pyramidal tiling. This structure stores downsampled versions of the WSI in a hierarchy, typically at 1×, 0.5×, 0.25×, and smaller scales. When interacting with the image—such as

during zooming or panning—a viewer or algorithm can access the appropriate pyramid level for performance-optimized rendering.

In computational workflows, these pyramid levels are essential for coarse-to-fine operations, such as region of interest (ROI) detection at low resolution followed by high-resolution patch extraction. Furthermore, pyramid structures support machine learning workflows that leverage multi-scale context, reducing the need to manually create low-resolution image versions.

### 2.3.2 Chunk-Based Compression

WSI storage is further optimized through tile- or chunk-based compression. Instead of encoding the entire image as one massive array, OME-TIFF stores the image in small fixed-size tiles (e.g., 256×256 or 512×512 pixels). Each tile is independently compressed using lossless (e.g., LZW, Deflate) or lossy (e.g., JPEG2000) algorithms. This layout allows for fast, random access to specific regions without reading unrelated image data.

Libraries such as ‘tifffile‘, ‘libtiff‘, and GPU-accelerated backends like ‘cuCIM‘ support high-performance reading and decoding of tiled OME-TIFFs. These APIs are vital for virtual staining, where models typically process thousands of patches per WSI. Efficient I/O becomes a bottleneck without chunked formats, especially in large-scale distributed training environments.

## 2.4 Image Quality Assessment

Image Quality Assessment (IQA) plays a pivotal role in evaluating virtual staining outputs, as it quantifies how well a synthesized image preserves structural and perceptual features of its reference. In medical imaging, especially histopathology, IQA is not just a measure of image clarity—it must also reflect diagnostic utility and clinical reliability.

IQA methods are generally categorized by the level of reference information required: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) [56]. FR-IQA assumes a perfect reference exists and enables direct pixel-level comparisons. RR-IQA uses limited side information, making it more practical in partially labeled datasets. NR-IQA, or blind IQA, requires no reference at all and is critical for real-world deployment when ground truth is unavailable.

### 2.4.1 Full-Reference Image Quality Assessment

Full-Reference metrics are widely used in virtual staining due to the availability of ground truth stained slides. SSIM [57] evaluates luminance, contrast, and structural similarity:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.1}$$

While SSIM is widely adopted, it is limited in capturing perceptual differences in color tone and texture granularity. Extensions such as MS-SSIM [59] and FSIM [64] introduce multi-scale and feature-based comparisons. VIF [46], based on mutual information, models image degradation in the wavelet domain. Despite their mathematical rigor, these methods may not align well with pathologist judgment, especially in context-dependent tasks like nuclei detection or gland boundary identification.

### 2.4.2 Reduced-Reference IQA

RR-IQA provides a compromise between full supervision and complete blindness. Instead of comparing full images, it operates on feature representations or statistics extracted from the reference:

$$Q(I_{\text{test}}) = d(f(I_{\text{ref}}), f(I_{\text{test}})) \tag{2.2}$$

Methods such as RRED [48] and SpEED [2] use wavelet-domain or spatial entropy differentials. These methods are suitable in federated or privacy-sensitive environments, where only partial descriptors of the reference image can be shared. In virtual staining, RR-IQA may be useful for cross-institutional validation of image quality where raw H&E references are inaccessible.

### 2.4.3 No-Reference IQA

No-Reference IQA (NR-IQA) methods estimate image quality in the absence of reference data, relying on statistical regularities of natural or medical images. Traditional NR-IQA metrics include BRISQUE [34], NIQE [35], and PIQE [54], which use Natural Scene Statistics (NSS) to model expected distributions of high-quality images.

Recent advances have shifted toward deep learning-based blind IQA. Early CNN models like DeepBIQ [7] and Kang et al. [25] trained regression networks to predict quality scores from image patches. More recent transformer-based models such as SwinIQA [62] and hybrid CNN-ViT architectures [17] improve robustness across distortions and datasets. These NR-IQA models are critical for deployment scenarios where pixel-aligned H&E references are unavailable or unreliable.

### 2.4.4 Perceptual IQA

Perceptual IQA methods aim to model human vision by comparing high-level features instead of raw pixels. LPIPS [65] measures perceptual distance between deep feature maps from a pretrained network:

$$d_{\text{LPIPS}}(x, y) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\phi_l(x)_{hw} - \phi_l(y)_{hw})\|_2^2 \tag{2.3}$$

This metric has shown superior correlation with human preference in GAN and virtual staining tasks. DISTS [11] improves upon LPIPS by integrating both texture (mean feature similarity) and structure (covariance correlation). ConIQA [15] introduces consistency-based training for semi-supervised IQA, while MR-Perceptual [27] emphasizes multi-resolution analysis.

These perceptual metrics are particularly relevant to virtual staining, where anatomical plausibility and fine-grained chromatin textures are often more important than pixel-level accuracy.

## 2.5 Conclusion

This chapter reviewed the foundational principles and recent advancements related to virtual staining, including traditional histological staining practices, computational image synthesis methods, storage strategies for Whole Slide Images (WSIs), and image quality assessment (IQA) techniques. We highlighted the progression from basic channel mapping to advanced generative frameworks such as GANs and diffusion models, demonstrating how deep learning has transformed the field by enabling high-fidelity, non-destructive virtual staining.

In parallel, we emphasized the importance of efficient WSI storage formats such as OME-TIFF and pyramid-based chunking, which are crucial for scalable data handling in digital pathology. Furthermore, we discussed a spectrum of IQA methods—ranging from full-reference to perceptual and no-reference metrics—that are essential for evaluating the fidelity and diagnostic relevance of virtually stained images.

Together, these topics form the conceptual and technical foundation for the contributions presented in subsequent chapters, where we introduce a novel domain-specific IQA metric (PaPIS), a high-throughput tiling framework, and a PaPIS-guided virtual staining pipeline.

# Chapter 3

# WSI Storage and High-Throughput Slicing Tools for Virtual Staining Workflows

## 3.1 Introduction

WSI has transformed digital pathology by enabling entire histological slides to be digitized at sub-micron resolutions, producing gigapixel-scale images that can be analyzed, annotated, and shared electronically. These high-resolution images are foundational to a variety of computational pathology applications, including remote diagnostics, quantitative tissue analysis, AI-assisted histopathology, and, increasingly, virtual staining.

However, the sheer size and complexity of WSIs—often exceeding several gigabytes per slide—pose substantial challenges in data storage, access, and processing. In particular, tasks such as deep learning-based inference or training require that these massive images be decomposed into smaller, contextually meaningful patches. This patch-based representation allows neural networks to focus on localized features, reduces memory footprint, and facilitates parallel computation.

Conventional slicing approaches, which rely on serial access patterns and basic file I/O, often fail to meet the throughput requirements of modern AI pipelines, especially when dealing with large-scale datasets spanning thousands of slides. To address this limitation, there is a growing need for high-performance, concurrent slicing frameworks that can fully exploit modern multi-core CPUs and GPU-accelerated computing environments.

Recent research has led to the development of several open-source and commercial tools aimed at optimizing WSI preprocessing. Notably, Slideflow has emerged as a GPU-accelerated library tailored for digital pathology, while compression-domain strategies have been proposed to bypass full-resolution decoding by exploiting hierarchical storage structures.

In this chapter, we present a dedicated multi-threaded WSI slicing framework tailored specifically for virtual staining workflows. Our system focuses on tile-aware access, thread-safe execution, and throughput-optimized I/O coordination, providing an efficient, scalable foundation for AI-driven pathology pipelines.

## 3.2 Literature Review

A wide range of WSI slicing tools and libraries have been proposed in recent years, each designed with distinct priorities in terms of file format compatibility, access performance, and integration with downstream analytical workflows. As deep learning methods become increasingly central to histopathological analysis and virtual staining, preprocessing speed and flexibility have emerged as critical bottlenecks during both model training and deployment.

**OpenSlide**, one of the earliest and most widely adopted open-source libraries, provides a vendor-agnostic interface for tile extraction across proprietary WSI formats, including those from Aperio, Hamamatsu, and Leica scanners [18]. Its simplicity and broad compatibility have made it a foundational tool in many research and clinical pipelines. However, OpenSlide's internal architecture lacks built-in support for multi-threaded access or caching, which significantly limits its throughput when applied in modern parallel computing contexts.

**Tifffile**, on the other hand, offers a low-level, NumPy-native interface to OME-TIFF and other tiled TIFF formats [16]. It provides direct access to raw image bytes and metadata, enabling precise control over I/O behavior. This makes it particularly well-suited for research scenarios where WSIs have already been standardized into open formats such as OME-TIFF. In contrast to OpenSlide, Tifffile's fine-grained access capabilities support sophisticated slicing strategies and integration into high-throughput Python-based preprocessing pipelines.

Several graphical platforms, including **QuPath** [4] and **ASAP** [19], have also integrated patch extraction features into their ecosystems. These platforms are primarily

designed for visualization, annotation, and interactive analysis of WSIs. QuPath, for example, provides a user-friendly interface for scripting image analysis pipelines and has seen widespread adoption in clinical and academic research. Similarly, ASAP (Automated Slide Analysis Platform) enables fast navigation and viewing of tiled WSIs and supports basic ROI extraction functionalities. While both platforms support tile-based workflows, they are primarily optimized for single-user, interactive use cases. As such, they typically lack fine-grained control over thread scheduling, memory management, or disk I/O optimization, making them suboptimal for automated high-throughput environments where parallelization is essential.

To overcome these limitations, recent efforts have explored GPU-accelerated WSI preprocessing. A prominent example is Slideflow [12], an open-source deep learning toolkit specifically designed for digital pathology applications. Slideflow supports both TensorFlow and PyTorch backends and includes modules for efficient tile extraction, stain normalization, and real-time visualization. Its support for multiprocessing, GPU offloading, and large-scale deployment scenarios makes it an attractive option for integrating WSI preprocessing with AI model training. In addition, Slideflow enables real-time tile streaming and pipeline-level optimizations, significantly reducing latency during inference on gigapixel-scale inputs.

Beyond raw pixel-space processing, recent studies have also proposed methods that exploit the *compression domain* of WSIs for classification and analysis [30]. By leveraging the intrinsic pyramidal structure of WSIs—where lower-resolution image levels are precomputed and embedded within the file format—these approaches perform selective decompression guided by attention mechanisms. This allows computational resources to be concentrated on diagnostically relevant regions, improving overall efficiency without compromising accuracy. Although these methods are still in early stages, they highlight the potential of hierarchical WSI representations for reducing the computational footprint of downstream tasks.

In the context of virtual staining workflows, slicing tools face even stricter requirements. Beyond throughput and compatibility, they must ensure *spatial and chromatic consistency* across extracted patches to avoid introducing artifacts that may degrade model performance or clinical interpretability. This includes handling edge regions, avoiding duplicated coverage, and maintaining alignment between input modalities such as PARS and H&E images. As a result, an ideal slicing framework should offer robust support for dynamic patch scheduling, intelligent caching, multi-threaded decoding, and seamless interfacing with deep learning toolchains. These requirements underscore the importance of designing specialized, task-aware slicing solutions capable of meeting the demands of modern histological image generation pipelines.

## 3.3 Methods

In this section, we present the slicing methodology for large-scale WSIs stored in the OME-TIFF format. Our framework is designed to efficiently extract standardized image patches for virtual staining tasks while maintaining spatial consistency, maximizing I/O throughput, and ensuring compatibility with modern deep learning workflows.

Specifically, we address two major technical challenges: (1) precise spatial localization of patch coordinates within a tiled, multi-resolution image layout using TIFF metadata, and (2) high-throughput, thread-safe execution of patch extraction and disk writing using a multi-threaded architecture.

### 3.3.1 TIFF-based Positioning

OME-TIFF and OME-Zarr are modern bioimaging file formats that support tiled, multi-resolution storage of gigapixel-scale microscopy data. These formats partition images into independently compressed chunks, enabling scalable access and manipulation. As illustrated in Figure 3.1, the data hierarchy follows a biologically meaningful structure—plates, wells, fields—further organized into resolution levels and image chunks.

OME-Zarr provides path-based chunk addressing, such as:

`plate.ome.zarr/row/col/well/image/resolution/t/c/z/y/x`

This indexing schema allows efficient retrieval of arbitrary subregions without requiring full memory mapping.

Similarly, OME-TIFF uses the TIFF Image File Directory (IFD) to store metadata such as tile offsets, sizes, and dimensions for each resolution level. This metadata enables indexed access to specific image regions, which is leveraged by libraries like `tifffile` for efficient random-access reading.

To extract a patch of size $s \times s$ pixels at coordinates $(x, y)$, we first compute the tile indices based on the tile width $T$:

$$i = \left\lfloor \frac{x}{T} \right\rfloor, \quad j = \left\lfloor \frac{y}{T} \right\rfloor \tag{3.1}$$

Only the tiles overlapping the desired patch region are read from disk. When a patch spans multiple tiles, partial regions are cropped and stitched in memory to form the final output patch.

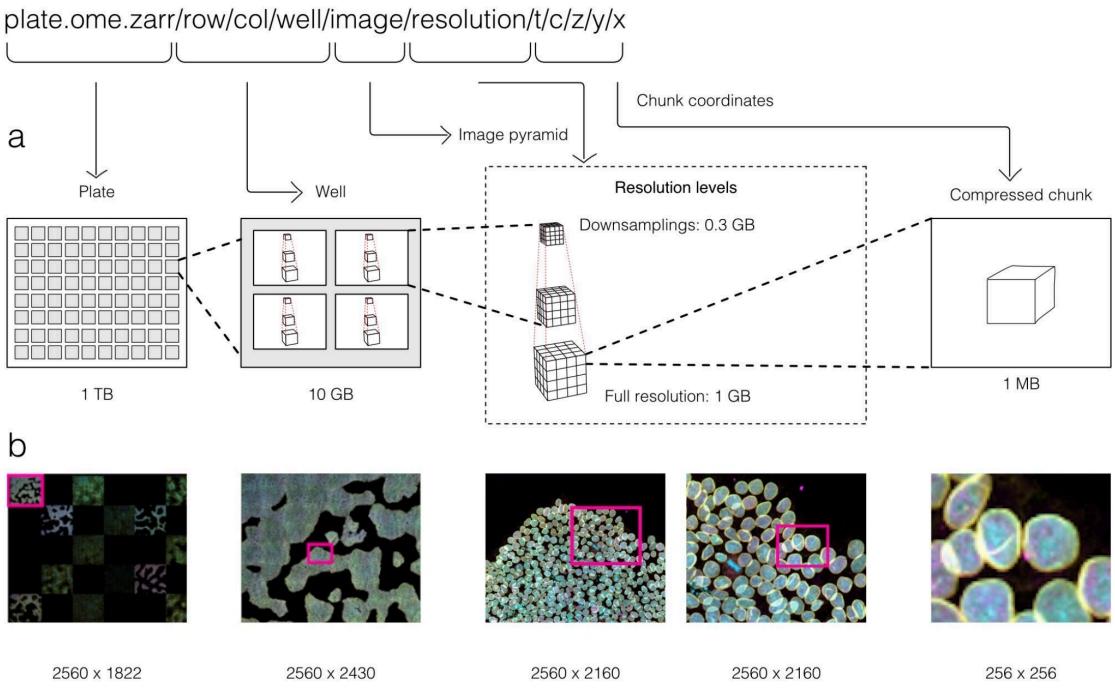



Figure 3.1: Schematic representation of the OME-Zarr image hierarchy. (a) Chunked, multi-resolution layout from plate-level down to compressed storage units. (b) Visual examples illustrating tiling and zooming at various magnifications.

To further improve memory efficiency, our method supports tile-row–level decoding. As illustrated in Figure 3.2, this approach retrieves only the pixel rows intersecting the desired region within each tile, avoiding full-tile decompression.

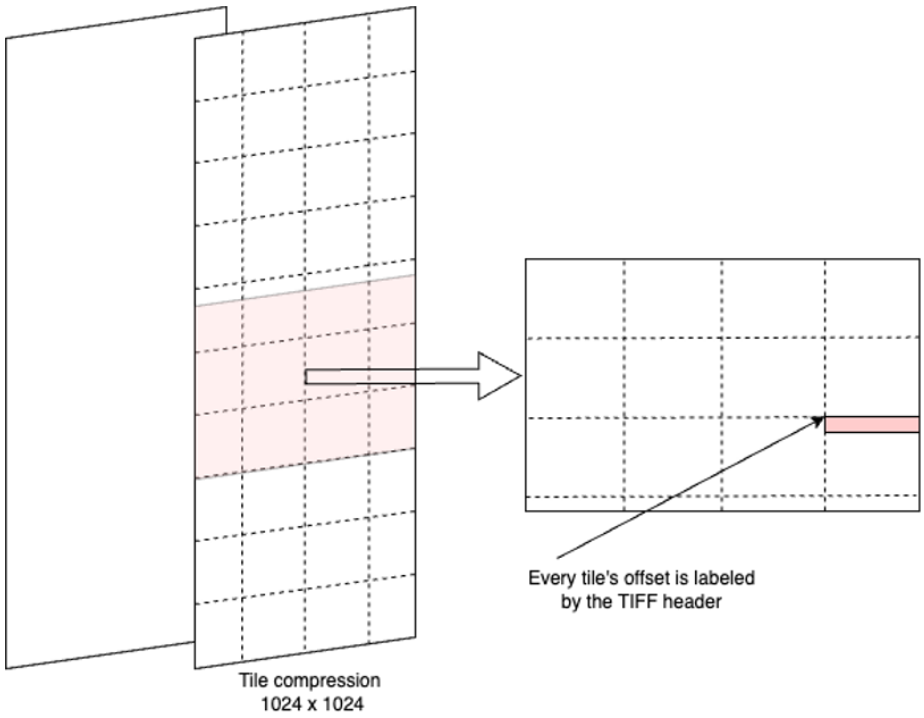



Figure 3.2: Efficient tile-wise access strategy in compressed WSI formats. **Left:** A virtual patch request that spans multiple image tiles, each stored as an individually compressed block. **Right:** Memory layout visualization for a single tile row. The pink region denotes a contiguous address range on disk where compressed blocks are stored adjacently. By aggregating spatially adjacent tiles, the system minimizes the number of file seek and read operations, enabling faster I/O and partial decoding.

The byte offset of a row $y$ within a tile located at $(x_{\text{chunk}}, y_{\text{chunk}})$ is computed by:

$$\text{offset} = \text{dataoffsets}[y_{\text{chunk}} \cdot N_{\text{tiles},x} + x_{\text{chunk}}] + (y \cdot T_w + x_{\text{offset,start}}) \cdot s \tag{3.2}$$

Where:

- `dataoffsets` is the array of tile byte offsets,

- $N_{\text{tiles},x}$ is the number of tiles along the horizontal axis,
- $T_w$ is the tile width (in pixels),
- $x_{\text{offset,start}}$ is the pixel offset within the tile,
- $s$ is the pixel size in bytes,
- $y$ is the row index within the tile.

The number of bytes to read per row is:

$$\text{bytes} = (x_{\text{offset,end}} - x_{\text{offset,start}}) \cdot s \tag{3.3}$$

Each partial row buffer is placed into the corresponding position within the output patch array using NumPy-style indexing. This tile-aware, row-level strategy significantly reduces I/O overhead and supports deterministic, memory-efficient access across large WSI datasets.

### 3.3.2 Concurrent Assignment

To achieve high throughput and scalable slicing, we adopt a decoupled, multi-threaded architecture using Python's `concurrent.futures.ThreadPoolExecutor`. The slicing pipeline is split into two main thread pools:

- **Coordinate Pool**: Responsible for generating slicing coordinates. Based on a global grid, it computes tile-aware bounding boxes and offset metadata for each target patch.
- **Slicing Pool**: Executes actual extraction jobs. Each thread retrieves tiles, decodes relevant subregions, assembles the patch, and writes the output to disk.

Each patch is defined by a tuple $(x, y, s)$, where $x$ and $y$ are top-left coordinates and $s$ is patch size. The coordinate pool resolves this into one or more tile-aligned subregions. These tasks are passed to slicing threads through a job queue, enabling concurrent, independent processing.

To ensure thread-safe file writing, we implement a mutex-based locking mechanism. Let $\mathcal{T}$ be the set of slicing threads and $\mathcal{F}$ the set of output files. For each file $f_j$, a lock $L_j$ is defined. A thread $t_i$ writing to $f_j$ must acquire and release $L_j$ as:

$$\text{acquire}(L_j); \quad \text{write}(f_j, \texttt{patch}); \quad \text{release}(L_j)$$

This ensures that no two threads access the same file simultaneously, preventing race conditions. The locks are stored in a global dictionary keyed by the file path:

$$\texttt{mutex_map} : \texttt{str} \rightarrow \texttt{Lock}$$

To further optimize performance in dense patch extraction scenarios, we implement a lightweight Least Recently Used (LRU) cache for storing recently decoded tiles. When multiple neighboring patches share tile boundaries, this cache allows threads to reuse tile data without redundant disk reads or decompression.

Together, these techniques enable fast, deterministic, and scalable slicing suitable for integration with virtual staining and other AI-assisted histopathology pipelines. The modular design also makes it adaptable to future extensions such as dynamic patch selection or GPU-based decoding.

## 3.4 Experiment

### 3.4.1 Setup

We evaluated the performance of our tile-aware WSI slicing framework using real-world histopathology images stored in the OME-TIFF format. All experiments were conducted on a dedicated workstation equipped with an Intel Xeon Silver 4215R CPU (8 cores, 16 threads, 3.20 GHz), 96 GB RAM, an NVIDIA Quadro RTX 4000 GPU (8 GB VRAM), and a PCIe 3.0 NVMe SSD. The system ran Ubuntu 24.04 LTS with Python 3.10.

Five WSIs from our virtual staining dataset were selected for benchmarking. These images ranged in size from $60{,}000 \times 50{,}000$ to over $100{,}000 \times 80{,}000$ pixels and were stored using 512×512 pixel tiles. Patches were extracted at 512×512 resolution with no overlap.

The slicing pipeline was configured as follows:

- **File format**: OME-TIFF (multi-resolution, tiled)
- **Library**: `tifffile` with tile-level I/O
- **Threading**: 16 threads for coordinate generation and 16 threads for slicing

- **Cache**: LRU cache of 128 decoded tiles
- **Output format**: Individual `.npy` arrays per patch

We compared three slicing strategies:

1. **Serial**: single-threaded baseline
2. **Threaded (no cache)**: parallel slicing without tile reuse
3. **Threaded (with cache)**: fully concurrent slicing with tile-level caching

The evaluation metrics included total slicing time, throughput (patches per second), peak memory usage, and CPU utilization (measured via `psutil`).

### 3.4.2 Results

Table 3.1 presents the performance results on a representative WSI with dimensions 93,312× 71,680, totaling approximately 25,500 patches.

Table 3.1: Comparison of slicing performance across modes (throughput in patches/s).

| Mode | Time (s) | Throughput | RAM (GB) | CPU (%) |
|---|---|---|---|---|
| Serial | 1243.7 | 20.5 | 3.1 | 6 |
| Threaded (no cache) | 178.3 | 143.0 | 8.7 | 88 |
| Threaded (with cache) | **129.5** | **197.0** | 9.2 | 91 |

### 3.4.3 Discussion

The results clearly demonstrate the efficiency of our concurrent slicing pipeline. Threading alone improves throughput by over 6× compared to the serial baseline. Introducing tile-level caching yields a further 38% increase, achieving nearly 10× speedup overall.

Despite the higher throughput, memory usage remains moderate, with peak RAM under 10 GB in the most demanding configuration. CPU utilization scales efficiently, reaching over 90%, indicating that the implementation makes full use of available cores.

The caching mechanism proves particularly effective in dense slicing scenarios where neighboring patches frequently share tile boundaries. By reusing decoded tile data, the pipeline reduces I/O and decompression overhead, which is especially beneficial for I/O-bound workloads common in virtual staining pipelines.

In summary, our results validate the effectiveness of a tile-aware, multi-threaded slicing strategy in delivering high throughput while maintaining low memory overhead. The modular pipeline design further allows for easy extension to GPU acceleration, adaptive patch selection, or integration into downstream deep learning workflows.

## 3.5 Conclusion

Efficient preprocessing of WSIs is a cornerstone for high-throughput AI workflows in digital pathology, particularly in the context of virtual staining. In this chapter, we introduced a modular, high-performance slicing framework that leverages multi-threading, tile-aware indexing, and thread-safe I/O to enable fast and scalable extraction of image patches from OME-TIFF-formatted WSIs.

Our method addresses critical challenges inherent in WSI processing, including the efficient resolution of patch coordinates, tile-wise decompression, and concurrent disk writing. Through the integration of in-memory caching and independent thread pools, the system achieves substantial speedups over traditional serial slicing approaches—offering up to an order of magnitude improvement in throughput while maintaining minimal memory overhead.

Moreover, the framework is fully compatible with modern deep learning pipelines and can be seamlessly integrated into virtual staining workflows. By ensuring deterministic patch extraction and optimal resource usage, it facilitates both training and real-time inference tasks at scale.

Future work will extend the system to support adaptive tiling based on tissue content detection, dynamic resolution adjustment, and native support for OME-Zarr cloud-hosted datasets. These improvements will further enhance its applicability to federated learning settings and remote digital pathology infrastructures.

# Chapter 4

# Pathology-Guided Virtual Staining Metric for Evaluation

## 4.1 Introduction

In recent years, numerous virtual staining methods have been developed as an alternative to traditional histopathological staining techniques such as hematoxylin and eosin (H&E), immunohistochemistry (IHC), and Masson's trichrome[1]. Conventional staining methods rely on chemical reagents, which can cause irreversible tissue alterations, limiting subsequent analyses and multi-modal imaging studies. In contrast, virtual staining preserves the tissue in its original state while computationally generating stained representations, enabling repeated analyses and facilitating the integration of different imaging modalities [1].

One of the key advantages of virtual staining is the elimination of chemical reagent consumption. Traditional staining protocols require costly chemical reagents, whereas virtual staining is purely computational, significantly reducing the expense of histopathological workflows. Additionally, virtual staining dramatically accelerates the staining process. While conventional staining can take tens of minutes to several hours, virtual staining can be completed within seconds or minutes, greatly improving the efficiency of pathology pipelines[29].

Despite these advantages, virtual staining currently lacks evaluation methods specifically tailored to the medical domain. Most assessments rely on full-reference image quality assessment (FR-IQA) metrics, which compare a processed image to a reference standard

based on structural and perceptual similarities. While FR-IQA metrics are well-established in general image processing and have been widely applied in natural image quality assessment, their effectiveness in microscopic pathology imaging remains limited [8]. A fundamental limitation of traditional FR-IQA approaches is their focus on textural fidelity rather than histopathological relevance. These metrics are primarily designed for natural images captured by cameras, prioritizing structural and perceptual consistency without considering cellular morphology and tissue architecture, which are crucial for medical applications [8, 38]. Consequently, conventional FR-IQA metrics often fail to accurately assess the diagnostic quality of virtual staining images.

With advancements in deep learning and feature engineering, perceptual evaluation metrics have increasingly enabled domain-specific comparisons [65, 11] . However, existing perceptual similarity metrics remain heavily biased toward natural image characteristics, making them inadequate for assessing medical imaging data. As a result, many virtual staining images cannot be effectively evaluated using current methodologies, highlighting the need for a domain-specific, histopathology-aware evaluation framework [9].

To address this gap, this study proposes a histopathology-guided perceptual full-reference image assessment metric specifically designed for virtual staining evaluation. Additionally, an optimization framework for virtual staining models is developed based on the proposed evaluation metric. By integrating domain-specific histopathological knowledge into image quality assessment, this approach aims to provide a more clinically relevant evaluation of virtual staining results, ultimately improving their reliability and applicability in medical practice.

## 4.2 Literature Review

Virtual staining has emerged as a powerful approach for generating stained-like images from label-free inputs, offering a reagent-free alternative to traditional histopathological staining. Label-free to stain transformation aims to synthesize H&E-equivalent images directly from modalities such as autofluorescence or photon-based signals, eliminating the need for physical dyes. Ecclestone et al. [14] introduced the PARS system, which captures spectral and temporal signals across wavelengths to simulate H&E staining, providing enhanced cellular detail. Building on this direction, Rivenson et al. [42] and Wang et al. [55] leveraged autofluorescence microscopy and deep learning to generate H&E representations, expanding the practical scope of label-free imaging in pathology.

Recent advances in generative deep learning have further boosted label-free virtual staining. Tweel et al. [52] and Boktor et al. [6] integrated generative networks within the

PARS framework to enhance virtual staining quality by expanding spectral input channels and transforming time-domain signals. Meanwhile, diffusion models introduced by Saharia et al. [45] offer a promising path for high-fidelity image-to-image translation, enabling accurate virtual staining even in challenging contexts. Foundational work by Isola et al. [22] (Pix2Pix) and Zhu et al. [67] (CycleGAN) laid the groundwork for these image translation models, which have since been adapted for virtual staining tasks.

Full-reference image quality assessment (FR-IQA) methods compare a processed image to a reference image to quantify differences in structure, perception, and statistical features. Wang et al. [57] proposed the SSIM metric, which evaluates both pixel-wise differences and structural information to assess image fidelity. To further enhance structural evaluations, Wang et al. [59] developed multi-scale structural similarity (MS-SSIM), incorporating multi-resolution analysis to improve robustness. Beyond hand-crafted approaches, perceptual-based FR-IQA methods have been introduced to align more closely with human perception. Zhang et al. [65] proposed the Learned Perceptual Image Patch Similarity (LPIPS) metric, which utilizes deep neural networks to model perceptual judgments based on feature representations from trained vision models. Ding et al. [11] extended perceptual similarity analysis through Deep Image Structure and Texture Similarity (DISTS), integrating structural and texture information to improve quality assessment. Tian et al. [50] explored the impact of multiple reference images in quality assessment, while Xian et al. [60] proposed structure-aware methods utilizing high-order statistical moments to capture intricate quality attributes.

Recent research has explored the application of FR-IQA in medical imaging. Breger et al. [8] demonstrated that standard FR-IQA metrics, originally designed for natural images, may not be directly applicable to medical imaging tasks, including MRI, CT, OCT, and digital pathology. Ohashi et al. [36] evaluated and adapted FR-IQA methods to better align with medical image quality requirements. Varga et al. [53] proposed optimized metric combinations to enhance prediction accuracy in medical imaging contexts. Additionally, Sujana et al. [49] investigated FR-IQA for structural MRI preprocessing, while Rodrigues et al. [43] examined perceptual quality assessment of medical images and videos.

## 4.3 Methods

Figure 4.1a illustrates the overall framework of the proposed perceptual similarity metric, PaPIS. This similarity metric assigns perceptual weights that reflect histological performance. To extract feature representations, image patches are first transformed into multi-channel embeddings using a pre-trained cell morphology segmentation model based on the

work of Ignatov et al. [21]. Their model, which achieves state-of-the-art performance in nuclei segmentation and classification, adopts a dual-layer encoder-decoder architecture with an EfficientNet-B7-based encoder, as shown in Figure 4.1b.

Subsequently, intrinsic properties are extracted from the image features to obtain the reflection map $R$ and the estimated illumination map $L$, using the Retinex algorithm[23]. The histology perceptual distance between these maps is then computed, inspired by prior works such as SSIM [57] and DISTS [11]. By calculating the distance between features derived from cell morphology, PaPIS offers a histology-aware quantification of image differences. This contrasts with traditional texture-based metrics such as SSIM and PSNR, which are primarily designed to align with human visual perception rather than pathological relevance.

To validate the practical utility of the proposed metric in pathological image analysis, we further apply it to evaluate the performance of a deep learning model that translates label-free histological images into H&E-stained representations.

### 4.3.1 Cell Morphology Level Feature Representation

Current state-of-the-art Full-Reference (FR) perceptual image similarity metrics, such as LPIPS[65] and DISTS[11], rely on feature extractors pre-trained on ImageNet to evaluate differences between generated and reference images. However, in the histology domain, pathologists focus on distinct perceptual attributes that cannot be adequately captured by models trained on natural image datasets.

To achieve superior performance in histological assessments, we utilize a feature extractor built upon the implementation of a state-of-the-art cell morphology segmentation task[21], which has demonstrated exceptional accuracy in nuclei segmentation and classification. In work by ignatov et al.[21], cell morphology segmentation is achieved using a dual-layer encoder-decoder architecture with an EfficientNet-B7-based encoder, with figure 4.1b showing structure. The computed latent features utilized in our work, formally, can be written as:

$$f_{eff}(x) = \{\tilde{x}_j^{(i)}; i = 0, \ldots, m; j = 1, \ldots, n_i\} \tag{4.1}$$

where $x \in \mathbb{R}^{H \times W \times C}$ denotes the input image patch, $m = 4$ represents the number of selected convolutional blocks within the EfficientNet backbone, and $n_i$ is the number of output channels (feature maps) at block $i$. The term $\tilde{x}_j^{(i)}$ refers to the $j$-th feature

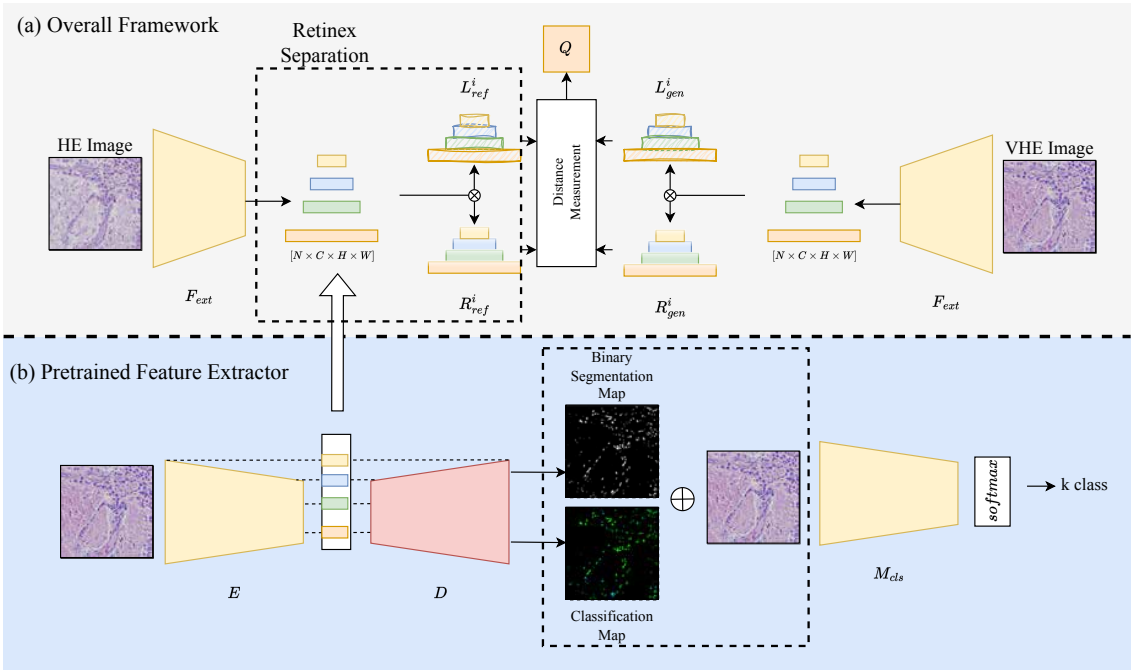



Figure 4.1: (a) The figure illustrates the overall framework of the PaPIS. Image patches are fed into a pretrained feature extractor to obtain multi-channel feature representations. These features are decomposed into a set of estimated illumination maps $L^i_a$ and reflectance maps $R^i_a$, where the superscript i denotes the feature layer and the subscript $a$ indicates the image type(Reference or Generated images). The collection of decomposed components is subsequently used to compute perceptual similarity distances between image pairs. (b) This panel presents the internal architecture of the pretrained feature extractor. The feature extractor functions as the encoder component of a cell morphology segmentation model. The subsequent decoder generates both a binary segmentation map $M_b$, which preserves the spatial dimensions of the input, and a multi-layer classification map $M_c$. The fusion of $M_b$, $M_c$, and the original image patch is then fed into a classification model to produce the final classification label.

map from block $i$, normalized via channel-wise min-max normalization to ensure scale consistency across layers.

This formulation explicitly defines the multi-layer, multi-channel output of the encoder, which serves as the foundation for subsequent perceptual analysis. The architecture of the feature extractor is illustrated in Figure 4.1b.

To compare the perceptual focus of these distinct feature extractors, and obtain a unified representation from the multi-layer feature maps, we compute a reconstructed image $I_{\text{Rec}}$ as follows:

$$I_{\text{Rec}} = \frac{1}{m} \sum_{i=0}^{m} f_{in} \left( \frac{1}{n} \sum_{j=0}^{n} \tilde{x}_j^{(i)} \right) \tag{4.2}$$

Here, $\tilde{x}_j^{(i)}$ denotes the $j$-th normalized feature map at layer $i$, as previously defined. The inner summation computes the average feature response across all channels $n$ within each layer, while the function $f_{in}(\cdot)$ integrates these responses. Specifically, $f_{in}$ denotes a linear interpolation function that maps the current image feature of size $\mathbb{R}^{w \times h}$ to the original spatial resolution $\mathbb{R}^{w_0 \times h_0}$, where $w_0$ and $h_0$ correspond to the width and height of the input image. The outer average aggregates information across all $m$ selected layers, resulting in the final reconstructed representation $I_{\text{Rec}}$, which captures both local and hierarchical morphological features from the input image.

Figure 4.2 illustrates the visualization of extracted features, comparing the feature maps obtained from models pre-trained on the cell morphology histological task ($F_{\text{histo}}$) with those from models pre-trained on natural image datasets ($F_{\text{natural}}$). Specifically, Figure 4.2b shows that the deeper layers of $F_{\text{histo}}$ focus more strongly on nuclear regions, exhibiting activation patterns that correspond to the locations of cell nuclei. In contrast, Figure 4.2a demonstrates that $F_{\text{natural}}$ primarily captures low-level texture patterns such as edges and gradients, which are characteristic of natural images and reflect the visual features typically emphasized by models trained to match human perception.

### 4.3.2 Retinex Properties on Features

In order to enhance the contrast of histological features, we apply the Multi-Scale Retinex (MSR) algorithm [23] to decompose the extracted features into two distinct components: the estimated illumination map and the reflectance map, which correspond to the low-frequency and high-frequency components within the extracted feature space, respectively.

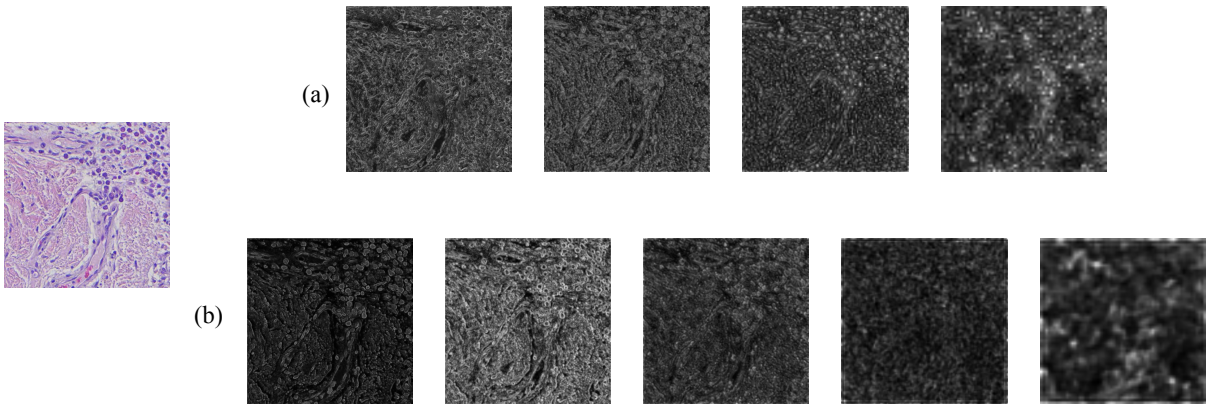



Figure 4.2: (a) The first set of feature maps was visualized from a VGG16 model pretrained on the ImageNet dataset. These feature maps correspond to channels 3, 8, 15, 22, and 29, as used in the works of LPIPS [65] and DISTS [11]. (b) The second set of feature maps was extracted from the encoder output of a pretrained segmentation model, with dimensions of $R^{192}$, $R^{288}$, $R^{480}$, and $R^{1344}$, corresponding to channels at different encoder stages.

The MSR algorithm models the decomposition of an input feature map based on the following relationship:

$$I(x, y) = R(x, y) \cdot L(x, y) \tag{4.3}$$

where $I(x, y)$ denotes the observed feature intensity at pixel location $(x, y)$, $R(x, y)$ is the reflectance component capturing intrinsic feature structures (e.g., edges and nuclei boundaries), and $L(x, y)$ represents the illumination component modeling smooth variations in local intensity due to shading or uneven activation.

To obtain a multi-scale estimation, we compute the illumination component at each scale $i$ as:

$$L_i(x, y) = F_i(x, y) * G_{\sigma_i}(x, y) \tag{4.4}$$

where $F_i(x, y)$ denotes the extracted feature map at scale $i$, $G_{\sigma_i}(x, y)$ is a 2D Gaussian filter with standard deviation $\sigma_i$, and $*$ denotes the 2D convolution operation. The choice of $\sigma_i$ controls the spatial frequency captured at each scale.

The reflectance component at each scale is then derived by logarithmic subtraction:

$$R_i(x, y) = \log F_i(x, y) - \log L_i(x, y) \tag{4.5}$$

where the logarithmic transformation serves to suppress multiplicative illumination effects and enhance the structural contrast in the feature map. The resulting $R_i(x, y)$ highlights fine-grained morphological features relevant to histological analysis.

### 4.3.3 PaPIS: Pathological Perceptual Distance

PaPIS integrates both low-frequency and high-frequency perceptual distances to evaluate histopathological image similarity. The decomposition of feature maps into illumination and reflectance components is based on the Retinex theory [23], which models an image as the product of its intrinsic reflectance and smooth illumination.

Inspired by SSIM [57] and DISTS [11], we define the high-frequency distance by comparing reflectance features extracted from Retinex decomposition. The high-frequency distance, denoted as $D_{\text{high}}(x, y, \alpha, \beta)$, is formulated as:

$$D_{\text{high}}(x, y, \alpha, \beta) = \alpha_{ij} \frac{2\mu^{(i)}_{R\tilde{x}_j}\mu^{(i)}_{R\tilde{y}_j} + c_1}{(\mu^{(i)}_{R\tilde{x}_j})^2 + (\mu^{(i)}_{R\tilde{y}_j})^2 + c_1} + \beta_{ij} \frac{2\sigma^{(i)}_{R\tilde{x}_j}\sigma^{(i)}_{R\tilde{y}_j} + c_2}{(\sigma^{(i)}_{R\tilde{x}_j})^2 + (\sigma^{(i)}_{R\tilde{y}_j})^2 + c_2} \tag{4.6}$$

In this formulation, the first term captures the similarity of mean values ($\mu$) of the reflected features, representing brightness consistency across spatial locations and feature channels. The second term evaluates the similarity of standard deviations ($\sigma$), which reflects structural consistency in the high-frequency domain. The reflection statistics $\mu^{(i)}_{R\tilde{x}_j}$, $\mu^{(i)}_{R\tilde{y}_j}$, $\sigma^{(i)}_{R\tilde{x}_j}$, and $\sigma^{(i)}_{R\tilde{y}_j}$ denote the mean and standard deviation of reflectance features $\tilde{x}$ and $\tilde{y}$ at layer $i$ and channel $j$. Constants $c_1$ and $c_2$ are included to ensure numerical stability. The weights $\alpha_{ij}$ and $\beta_{ij}$ are randomized such that $\sum_{i,j} \alpha_{ij} + \beta_{ij} = 1$, ensuring a balanced contribution between brightness and structure. In future work, these weights could be optimized directly through supervised learning if annotated preferences or pathologist evaluations become available.

The low-frequency distance, $D_{\text{low}}$, is computed as the mean squared error (MSE) between illumination maps derived from each feature channel via the Retinex algorithm [23].

Finally, the complete PaPIS metric is defined as:

$$PaPIS(x, y, \lambda, \alpha, \beta) = \lambda \sum_{i=0}^{m} \sum_{j=0}^{n_i} \text{MSE}(L(\tilde{x}^{(i)}_j), L(\tilde{y}^{(i)}_j)) + D_{\text{high}}(x, y, \alpha, \beta) \tag{4.7}$$

Here, $\lambda$ is a hyperparameter that balances the contributions of the low-frequency and high-frequency components in the final distance score.

## 4.4 Experiments

### 4.4.1 Dataset Acquisition

#### Modalities of Images Acquisition

All scanning and post-processing procedures are detailed in Section 5.3.1. The dataset consists of five pairs of whole slide images (WSIs) with dimensions of $18,202 \times 48,800$, $9,200 \times 36,200$, $59,600 \times 61,418$, $17,664 \times 17,876$ and $35,344 \times 37,957$, respectively. Four of the scanned WSIs are divided into $1024 \times 1024$ patches, resulting in a total of 10,086 patches. These patches are then randomly flipped and resized to $256 \times 256$ for the training process. The remaining WSI pair is designated for the comparative experiment. A filtering process is first applied to extract histological regions, followed by randomly cropping into 206 synchronized $1024 \times 1024$ patches for subsequent analysis.

#### Virtual Stained Image Generate

The dataset utilized in this study was generated using a virtual staining model based on the CycleGAN architecture [67]. Training was conducted with a batch size of 1. The Adam optimizer was used for both the generator and discriminator networks, with a learning rate of 0.001 and $\beta$ parameters of $(0.5, 0.999)$. An accumulative gradient strategy with an accumulation count of 2 was applied. A linear learning rate decay strategy was employed using the `LinearLrInterval` scheduler, with updates every 1000 iterations. The learning rate decayed linearly from 0.001 to 0 between the $10{,}000^{\text{th}}$ and $50{,}000^{\text{th}}$ iterations.

### 4.4.2 Comparison with Previous Metrics

For the comparison of image pairs, we calculate the following metrics: PSNR, SSIM [57], MS-SSIM [59], LPIPS [65], and DISTS [11]. Based on the relationship between these metrics and PaPIS scores, we define four distinct categories:

1. **AH (Aligned High)**: Points exhibiting both high PaPIS scores and high values across the compared metrics, indicating strong agreement between PaPIS and traditional measures.

2. **AL (Aligned Low)**: Points with both low PaPIS scores and low metric values, reflecting consistent low similarity across both evaluation methods.

3. **PD (PaPIS-Dominant)**: Points where PaPIS scores are high despite low traditional metric values, suggesting cases where PaPIS captures relevant pathological features overlooked by conventional metrics.

4. **TD (Traditionally-Dominant)**: Points with low PaPIS scores but high traditional metric values, highlighting instances where traditional metrics indicate high similarity, yet PaPIS identifies deficiencies in pathology-relevant features.

We categorize the comparison metrics into two groups: traditional handcrafted metrics and perceptual feature-based perceptual metrics.

### PaPIS and Handcrafted Metrics

In this section, we compare our proposed metric with traditional handcrafted metrics, specifically SSIM [57]. A case-by-case scatter plot is presented in Figure 4.3, while additional image-to-image examples are provided in Figure 4.4 for further comparison.

Among the AH and AL points, both selected data points exhibit strong alignment between PaPIS and traditional metrics in assessing image quality. In contrast, the PD and TD points highlight the differing sensitivities of PaPIS and traditional approaches. PD points correspond to cases where nuclear structures and spatial alignment are well-preserved, yet fine textures and subtle color variations result in lower scores from traditional metrics, revealing their limitations in capturing histological relevance. Conversely, TD points receive high scores from traditional metrics due to uniform textures or consistent coloration but may lack proper histological structures or exhibit misalignment in cellular organization, leading to lower PaPIS scores. This contrast underscores the ability of PaPIS to better capture pathologically significant features that traditional methods may overlook.

### PaPIS and Other Feature-Based Perceptual Metrics

In this section, PaPIS is compared with other perceptual-based similarity metrics, specifically LPIPS [65] and DISTS [11]. A case-by-case scatter plot is presented in Figure 4.5

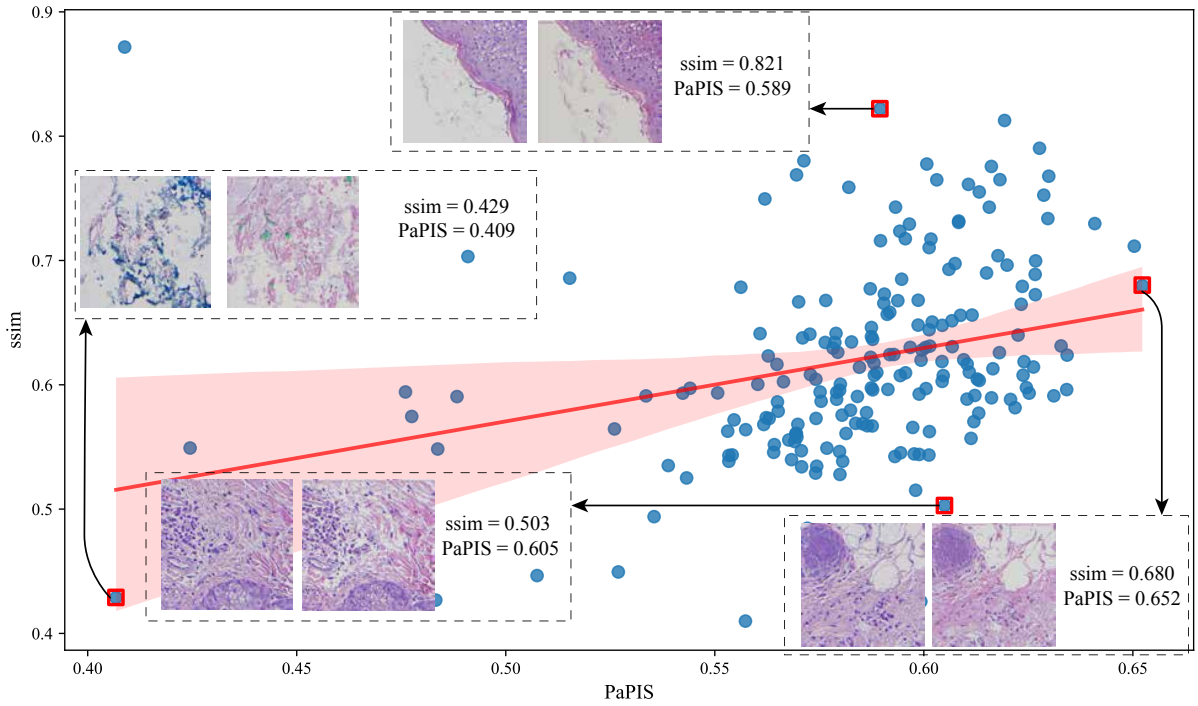



Figure 4.3: Scatter plot comparing SSIM and PaPIS across different image pairs. Selected data points are labeled as AH (Aligned High), AL (Aligned Low), PD (PaPIS-Dominant), and TD (Traditionally-Dominant), representing specific cases for analysis.

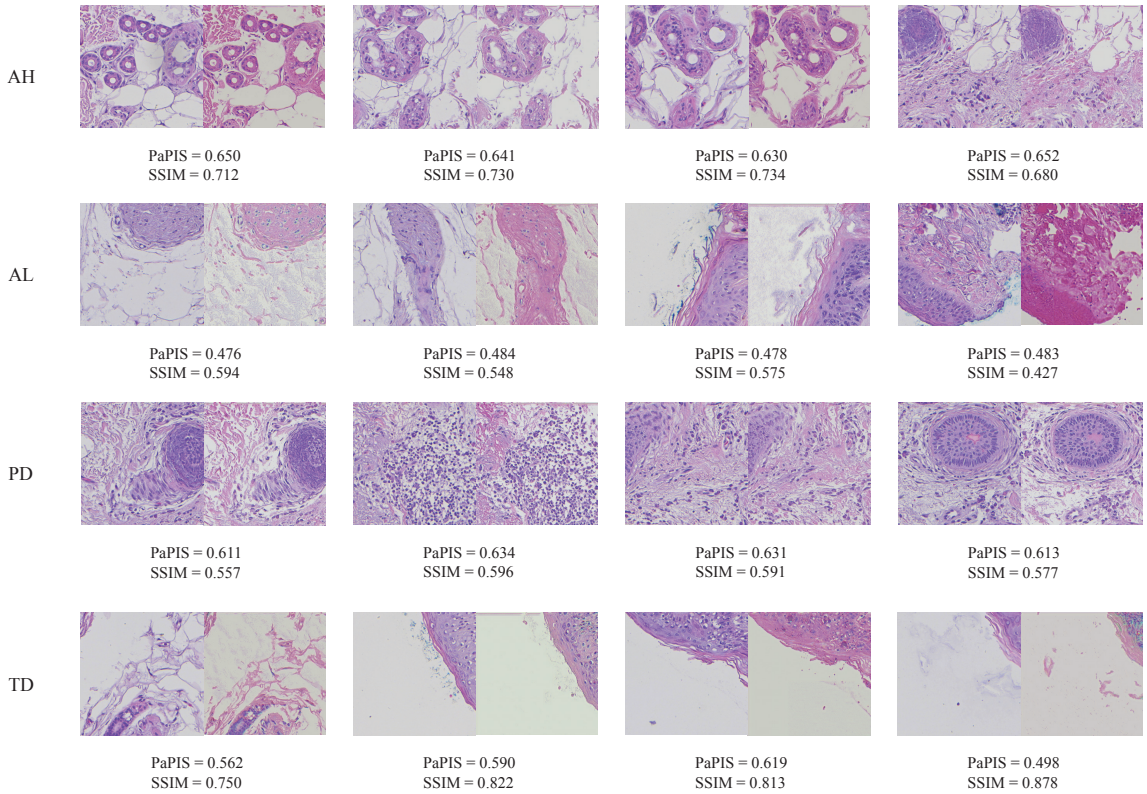



Figure 4.4: Case-wise comparison of PaPIS and SSIM values across different histological image samples.

and Figure 4.6. For both LPIPS and DISTS, lower scores indicate higher perceptual similarity and are generally interpreted as better visual quality. We clarify this here to avoid confusion, especially since PaPIS is positively oriented (i.e., higher is better), leading to an apparent anticorrelation in some cases.

At the AH and AL points, both metrics consistently identify images as either high or low in overall quality. However, at the PD points, the images tend to be concentrated in regions with high cellular density, whereas at the TD points, they are primarily located at tissue edges or in areas with sparse cellular distribution. Notably, neither perceptual metric provides explicit information regarding morphological differences in cellular structures. Nevertheless, compared to handcrafted metrics, traditional perceptual metrics exhibit higher consistency with PaPIS, suggesting an improved alignment in assessing histologically relevant features.

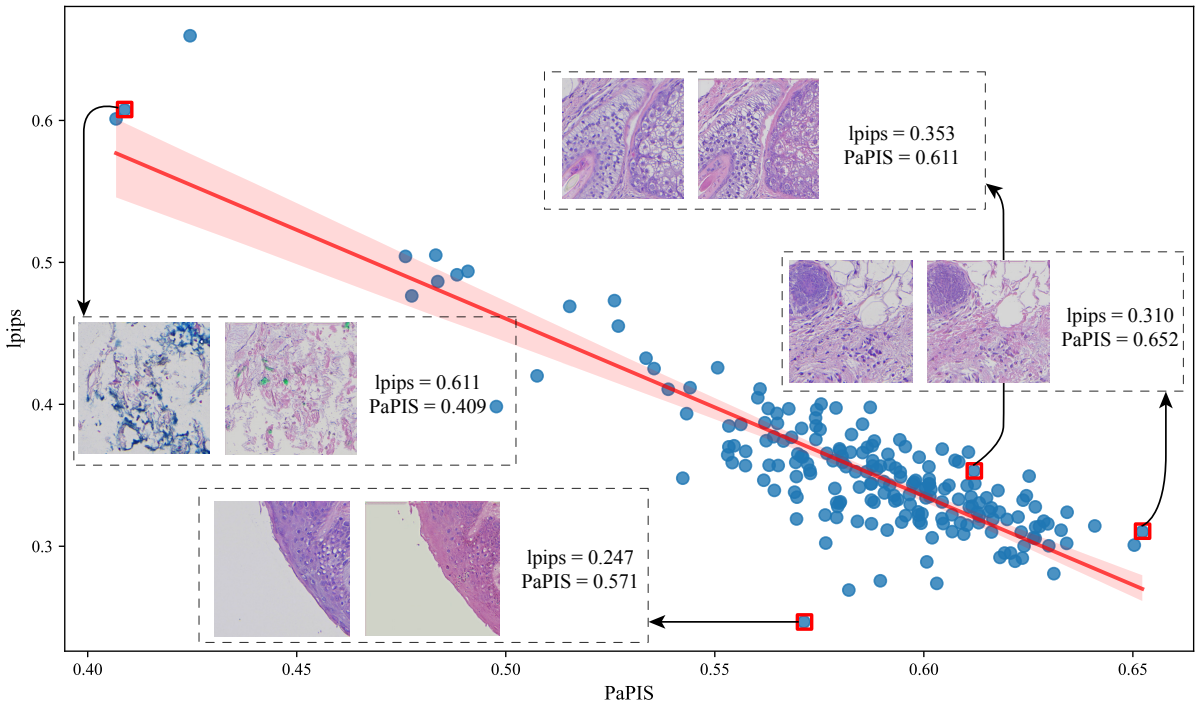



Figure 4.5: Scatter plot comparing LPIPS and PaPIS across different image pairs. Each data point represents an image pair evaluated by both metrics, illustrating variations in their assessments. This comparison highlights differences in perceptual and pathology-aware similarity measurements between the two approaches.

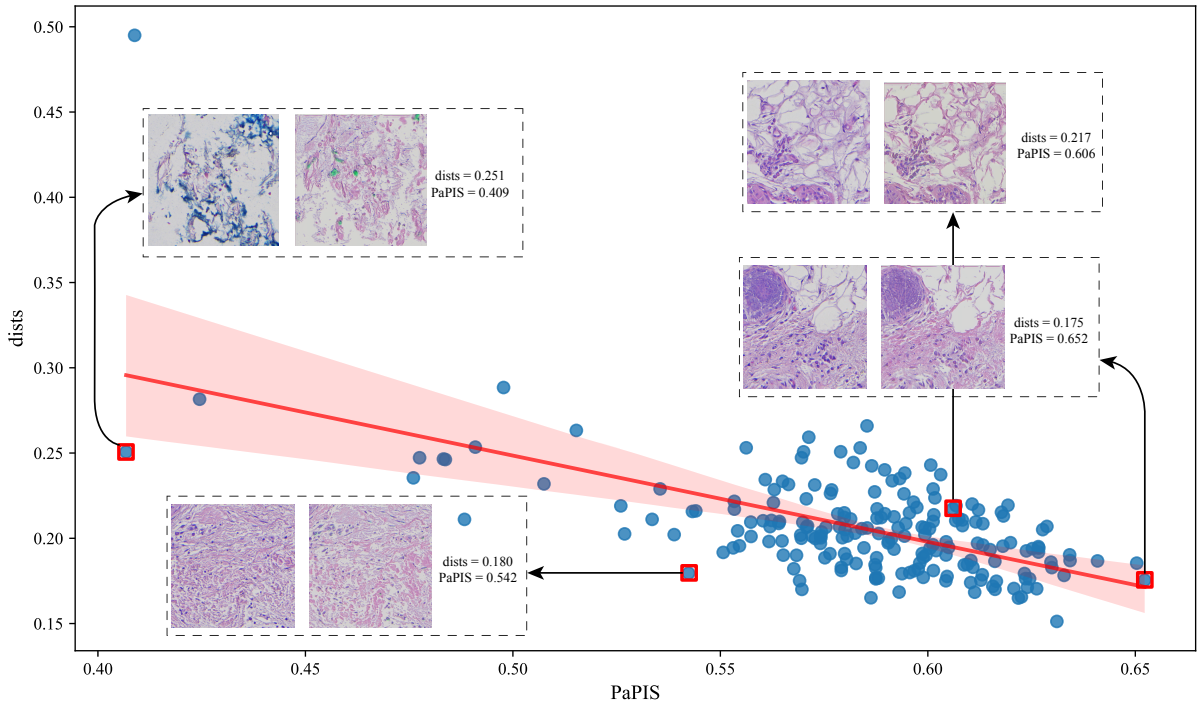



Figure 4.6: Scatter plot comparing DISTS and PaPIS across different image pairs.

### Regional Sensitivity Analysis via Heatmap Visualization

To further investigate the regional sensitivity of different similarity metrics, we performed a heatmap-based comparison between PaPIS and SSIM on a representative whole slide image (WSI). The WSI was divided into non-overlapping $1024 \times 1024$ patches, and each patch was evaluated using both metrics. The resulting similarity scores were visualized as spatial heatmaps, shown in Figure 4.7, where warmer colors represent higher similarity.

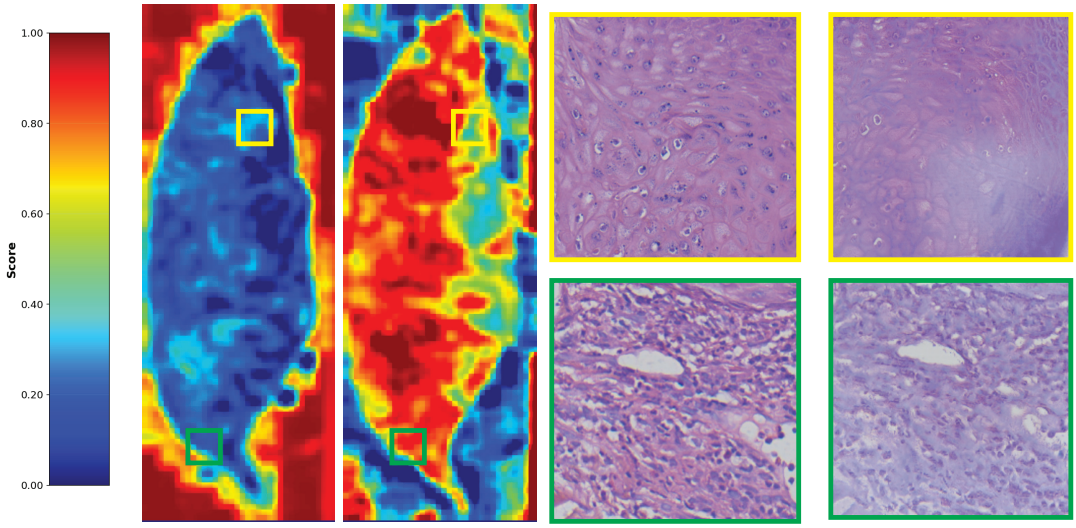



Figure 4.7: Patch-wise similarity heatmaps of a representative whole slide image (WSI). **Left**: SSIM heatmap; **Middle**: PaPIS heatmap. **Right**: Magnified views of selected regions. Warmer colors indicate higher similarity. Compared to SSIM, which tends to assign high scores across low-texture and background regions, PaPIS demonstrates greater sensitivity to morphological and contextual features, providing finer-grained evaluation of histological relevance.

**Observations and Insights.** From the heatmap analysis, we derive the following insights:

1. **Texture Sensitivity of SSIM vs. Morphological Awareness of PaPIS(OB1):** SSIM shows a noticeable drop in scores from background to tissue regions (mean $0.77 \rightarrow 0.65$), consistent with its high sensitivity to low-frequency textural features. In contrast, PaPIS scores increase from background to tissue (mean $0.48 \rightarrow 0.61$),

highlighting its preferential weighting toward structural and morphological content. As shown in Figure 4.8 (left and center), this behavior confirms that SSIM is influenced more by the presence or absence of texture, while PaPIS emphasizes pathological salience.

2. **PaPIS Penalizes Poor Staining Despite High SSIM(OB2):** In regions where SSIM assigned high scores (e.g., due to preserved structure but degraded stain), visual inspection revealed suboptimal virtual staining quality. PaPIS assigned lower scores in such cases, consistent with staining fidelity. These examples are highlighted in the scatterplot with paired visual samples (Figure 4.4), illustrating PaPIS's closer alignment with histological relevance.

3. **PaPIS Demonstrates Stronger Intra-Tissue Differentiation(OB3):** To further assess how well each metric captures intra-tissue heterogeneity, we analyzed ten distinct tissue regions and computed the Coefficient of Variation (CV = standard deviation / mean) for each region. As illustrated in Figure 4.8 (right), PaPIS consistently yields higher CVs than SSIM across all regions, indicating finer sensitivity to structural and contextual differences. A paired $t$-test confirms this difference is statistically significant ($p = 0.0011$), supporting the claim that PaPIS is more capable of capturing nuanced histological variations, such as glandular or nuclear complexity.

4. **Complementary Use Case Consideration(OB4):** The divergence observed between SSIM and PaPIS across various contexts suggests their complementary utility. SSIM may be suited for coarse texture comparison or background consistency evaluation, while PaPIS is better aligned with diagnostic applications where fine-grained morphological interpretation is required.

Table 4.1: Metric scores in background vs. tissue regions (mean ± std)

| Metric | Background | Tissue |
|---|---|---|
| SSIM | $0.77 \pm 0.03$ | $0.65 \pm 0.05$ |
| PaPIS | $0.48 \pm 0.04$ | $0.61 \pm 0.06$ |

These findings reinforce the conclusion that PaPIS captures clinically meaningful variations in virtual staining quality, offering a more nuanced and pathology-aligned metric than traditional image similarity measures.

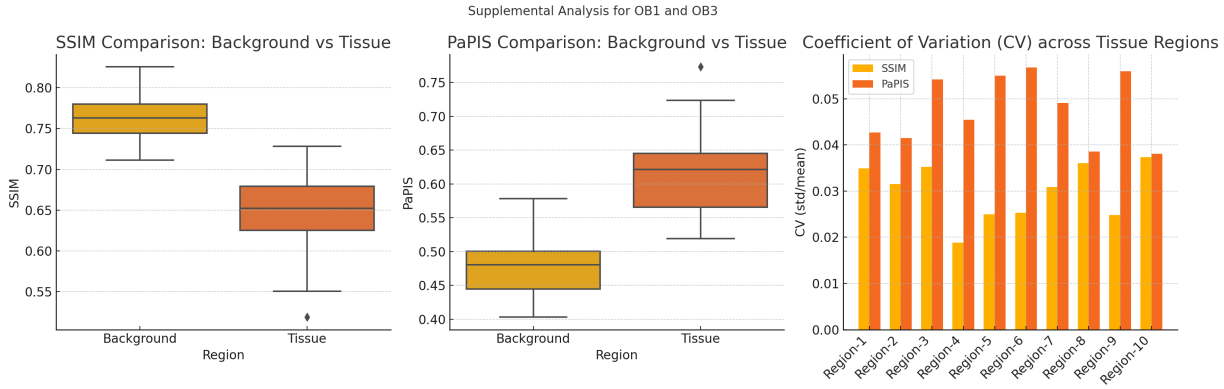



Figure 4.8: Supplemental analysis for OB1 and OB3. Left: SSIM scores in background vs. tissue regions. Center: PaPIS scores in background vs. tissue regions. Right: Coefficient of Variation (CV) across 10 tissue regions for SSIM and PaPIS. PaPIS shows significantly higher intra-region variability, confirming its ability to reflect morphological diversity.

## 4.5 Conclusions

In this study, we introduced PaPIS, a novel FR-IQA metric specifically designed for evaluating virtual staining. Unlike conventional FR-IQA methods, which primarily focus on pixel-wise structural similarity, PaPIS incorporates domain-specific histopathological features by leveraging a deep learning-based feature extractor trained on cell morphology segmentation. Additionally, Retinex-based feature decomposition was integrated to enhance perceptual assessment by separating illumination and reflectance properties.

Our experiments demonstrated that PaPIS aligns more closely with pathologically relevant features compared to traditional handcrafted metrics such as SSIM and perceptual similarity metrics like LPIPS and DISTS. Through a comprehensive case-wise analysis, we identified distinct cases where PaPIS better captured histological structures, particularly in differentiating nuclear morphology and spatial alignment. Furthermore, we incorporated PaPIS as a guiding loss function in a virtual staining model, leading to improved histological accuracy and enhanced alignment with ground truth staining.

The results highlight the necessity of pathology-aware evaluation frameworks in digital pathology, where conventional IQA methods often fail to reflect the clinical relevance of staining quality. By bridging the gap between computational image assessment and histopathological evaluation, PaPIS paves the way for more reliable and interpretable quality assessment in AI-driven virtual staining. Future work will explore the extension

of PaPIS to other histopathological modalities and further refine its adaptability across diverse staining techniques.

# Chapter 5

# PaPIS Guided Virtual Staining Enhancement

## 5.1 Introduction

Virtual staining has emerged as a powerful technique for transforming label-free microscopy images into histologically meaningful counterparts, offering the potential to accelerate diagnostic workflows and reduce dependency on chemical reagents. However, most existing virtual staining methods rely heavily on minimizing the distance between features extracted by general-purpose or perceptual networks. These networks, while effective in natural image domains, often fail to capture the subtle and domain-specific characteristics inherent in histopathological images—such as nuclear morphology, tissue architecture, and staining granularity.

The introduction of PaPIS marks a significant advancement in this field. Unlike traditional similarity metrics, PaPIS is designed specifically for histological applications, leveraging domain knowledge to evaluate images from a pathologist's perspective. This opens a new avenue for guiding virtual staining models—not merely by generic pixel-wise losses or feature distances—but through metrics that are aligned with expert visual assessment in pathology.

In this chapter, we introduce a PaPIS-guided virtual staining framework that incorporates this pathology-aware metric directly into the model optimization process. By doing so, we encourage the model to generate images that are not only visually similar to real stains but also consistent in terms of diagnostic relevance. Our approach significantly enhances the fidelity and interpretability of virtual stains, addressing a key limitation in prior

work and demonstrating that integrating domain-specific metrics like PaPIS leads to more clinically meaningful results.

## 5.2 Literature Review

Virtual staining has become an active area of research at the intersection of biomedical imaging and deep generative modeling. Traditional histological staining involves complex chemical protocols and time-consuming procedures, which have motivated the development of computational alternatives capable of generating virtual stained images from label-free inputs. Early efforts in this domain were largely driven by image-to-image translation frameworks, most notably Pix2Pix and CycleGAN [22, 67], which showed promising results in transforming imaging modalities such as phase contrast or autofluorescence into virtual H&E or immunofluorescent representations.

Subsequent works adapted these models to histopathological contexts by introducing architecture refinements and tailored loss functions. For instance, Bayramoglu et al. [5] and Rivenson et al. [41] applied deep convolutional networks for generating stained images from quantitative phase imaging (QPI) and autofluorescence data, respectively. These approaches primarily relied on pixel-wise loss (e.g., L1/L2) or adversarial loss functions, often producing visually convincing results but lacking pathology-specific structural consistency.

To address this, several studies introduced perceptual loss functions using pretrained natural image networks such as VGG [24], which enabled better preservation of high-level texture and structure. However, these networks were trained on datasets like ImageNet and are not inherently sensitive to histopathological semantics, leading to potential misalignment between perceptual similarity and diagnostic relevance.

Recent advancements have focused on domain-specific similarity metrics and task-aware supervision to enhance the clinical applicability of virtual staining models. For instance, Zhang et al. [65] proposed LPIPS as a learned perceptual metric, while Ding et al. [11] introduced DISTS (Deep Image Structure and Texture Similarity), which unifies structure and texture similarity for image quality assessment. However, none of these metrics were explicitly designed for histological image analysis.

The introduction of PaPIS (Pathology-aware Perceptual Image Similarity) represents a significant leap in this direction. Unlike generic perceptual metrics, PaPIS is calibrated to reflect pathologists' visual assessments by leveraging domain-relevant features such as nuclear size, glandular boundaries, and chromatin texture. It provides a new lens through

which model performance can be evaluated and optimized—not merely through visual similarity, but through diagnostic integrity.

In parallel, Tweel et al. [52] developed a virtual histology pipeline utilizing PARS microscopy and a CycleGAN framework trained on weakly registered image pairs. Their approach demonstrated that virtual H&E images generated from label-free PARS data could achieve diagnostic quality comparable to traditional staining methods. Further advancements by Boktor et al. [6] introduced a multi-channel feature extraction strategy, enhancing the fidelity of virtual staining by incorporating additional PARS-derived features. These contributions underscore the potential of integrating domain-specific metrics and advanced imaging modalities to improve virtual staining techniques.

By embedding PaPIS into the optimization process, as demonstrated in this work, virtual staining models can be trained to align more closely with expert visual perception, thereby improving interpretability and potential clinical acceptance. To the best of our knowledge, this is one of the first studies to incorporate a pathology-aware perceptual metric directly into the training loop of a generative model for histological image synthesis.

## 5.3 Methods

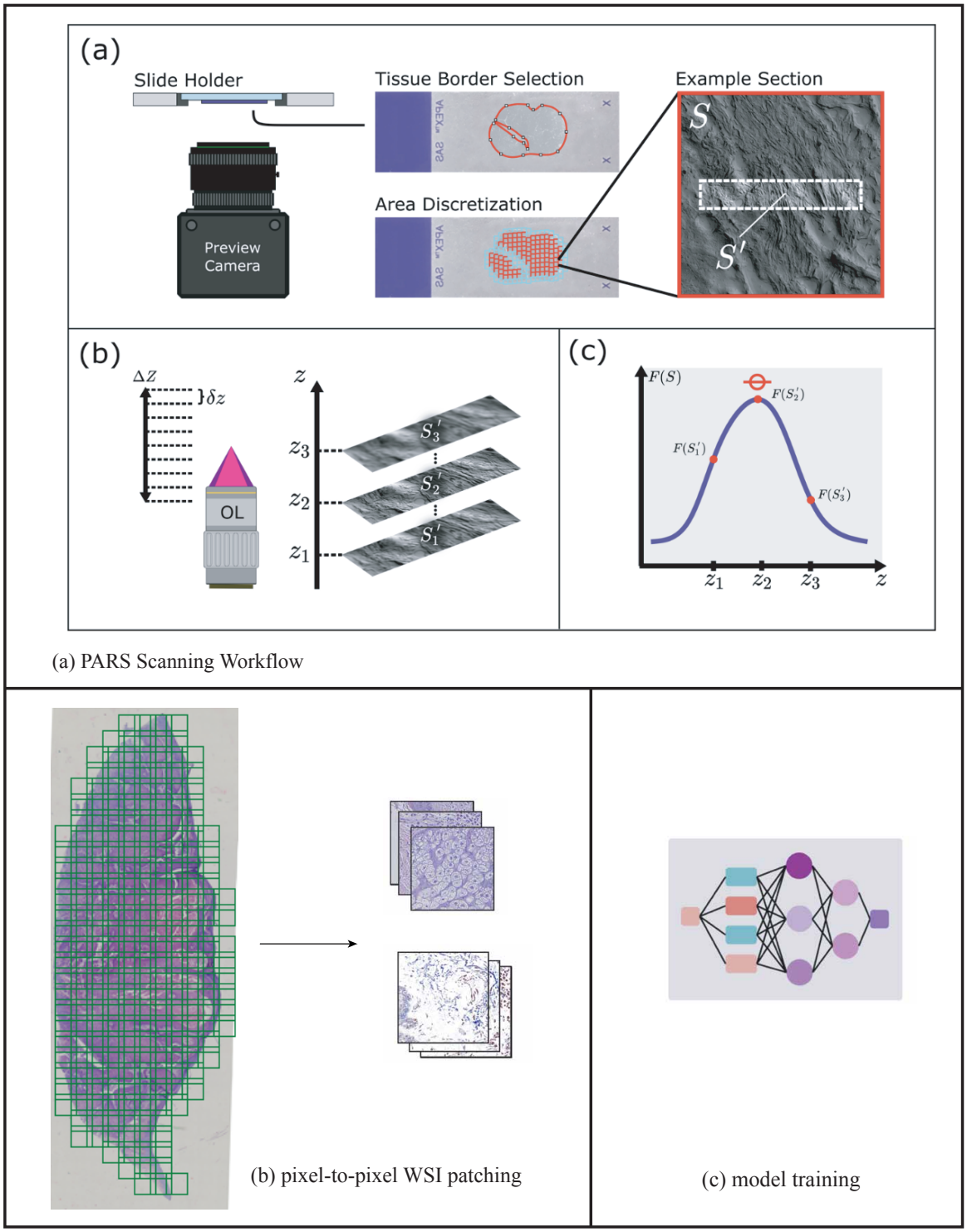
### 5.3.1 Dataset Acquisition and Preprocessing

WSIs were acquired using a PARS microscope, following the methodology described by Tweel et al. [52, 51]. Each WSI was scanned at a spatial resolution of $0.5\,\mu m$/pixel. The PARS and H&E image pairs were weakly aligned using rigid registration, and pixel-level correspondence was further refined during preprocessing. To ensure high-quality alignment, we applied SSIM-based patch filtering, discarding pairs with structural similarity below 0.6.

For training and inference, WSIs were automatically divided into $512 \times 512$ pixel patches with 25% overlap using an optimized preprocessing script adapted from [51], capable of processing each WSI in under 15 seconds. All image patches were normalized to the [0, 1] intensity range. Standard data augmentation techniques, including random flips, 90-degree rotations, and stain-intensity jittering, were applied on the fly during training.

### 5.3.2 Model Architecture and Modifications

We propose a customized architecture, `CycleGAN_PAPIS`, which builds upon the standard CycleGAN framework [67] with tailored modifications for histological image translation.




(a) PARS Scanning Workflow

(b) pixel-to-pixel WSI patching

(c) model training



Figure 5.1: Pipeline for PaPIS-guided virtual staining. Part (a) is adapted from [51]. The end-to-end framework includes WSI acquisition, automated patch extraction, and model training.

Our model is specifically designed to support perceptual supervision through the PaPIS loss, which is not supported in the original CycleGAN formulation.

The generator adopts a ResNet-based architecture with 9 residual blocks, using reflection padding, instance normalization, and 64 base channels. The discriminator is a PatchGAN with 3 convolutional layers and a $70 \times 70$ receptive field, designed to assess local realism in generated patches.

Unlike the original CycleGAN, our implementation introduces a pathology-aware perceptual loss into the generator pathway. The generator is modified to expose intermediate outputs required by the PaPIS evaluator, while the discriminator path remains unchanged. The architecture and integration strategy are illustrated in Figure 5.2.

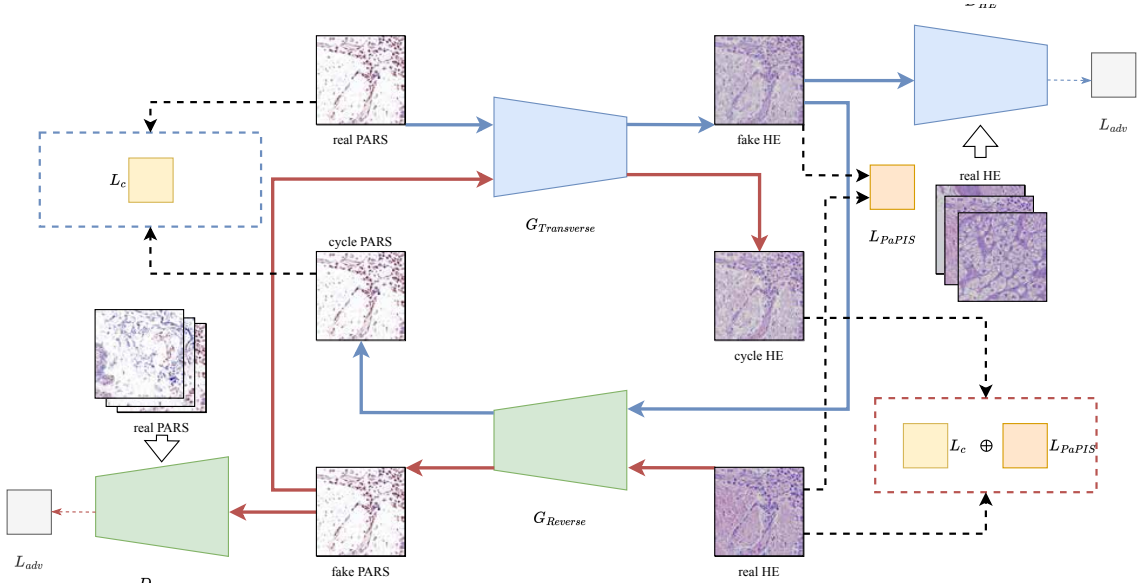



Figure 5.2: Architecture of the PaPIS-guided virtual staining model. In addition to the original CycleGAN [67] components, we incorporate PaPIS-based perceptual loss terms to guide the generation of virtual H&E images, encouraging histological alignment with ground truth.

### 5.3.3 Pathology-Aware Perceptual Loss

To guide the model toward producing diagnostically meaningful outputs, we introduce a novel pathology-aware perceptual loss, derived from the PaPIS (Pathology-aware Percep-

tual Image Similarity) metric proposed in the previous chapter. Unlike general-purpose perceptual metrics such as LPIPS or DISTS, PaPIS is explicitly designed for histological applications, emphasizing features relevant to tissue structure, nuclear morphology, and chromatin distribution.

The PaPIS score computes a multi-scale perceptual similarity between the generated and ground truth images using pathology-specific feature encoders. In this work, we formulate a loss function directly from the PaPIS similarity by applying a simple inversion:

$$\mathcal{L}_{\text{papis}} = 1 - \text{PaPIS}(G(x), y; \{C_i\}) \tag{5.1}$$

where $G(x)$ is the generated virtual H&E image, $y$ is the real H&E reference, and $\{C_i\}$ are the selected feature channels used for PaPIS computation.

The design of this loss is motivated by two factors: (1) The PaPIS score ranges from 0 to 1, where higher values indicate better perceptual similarity from a pathologist's perspective. (2) By defining the loss as $1 - \text{PaPIS}$, we ensure that minimizing the loss corresponds to maximizing perceptual fidelity, thus aligning model optimization with perceptual quality.

This perceptual term is combined with adversarial and cycle-consistency losses to form the overall objective:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{cycle}}\mathcal{L}_{\text{cycle}} + \lambda_{\text{papis}}\mathcal{L}_{\text{papis}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} \tag{5.2}$$

In our experiments, we set the weights to $\lambda_{\text{cycle}} = 2.0$, $\lambda_{\text{papis}} = 1.0$, and disabled identity loss. The PaPIS loss provides fine-grained supervision that captures clinically important structures often missed by pixel-wise or generic perceptual losses.

### 5.3.4 Training Procedure

Training was performed using the Adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and a beginning learning rate of $2 \times 10^{-4}$. We used a batch size of 4 and trained for 100 epochs with linear learning rate decay starting at epoch 50. Early stopping was applied with a 10-epoch patience based on validation PaPIS score. Training and inference were performed on an NVIDIA RTX 3090 GPU.

### 5.3.5 Whole Slide Inference and Patch Fusion

Due to the large size of WSIs, we adopted a patch-based inference pipeline with overlapping windows to ensure continuous reconstruction. Let $P_{i,j} \in \mathbb{R}^{H\times W\times 3}$ be a patch at position $(i, j)$. Patches are extracted with stride $s < H$ to introduce overlap.

To blend overlapping predictions, a precomputed weight map $W \in \mathbb{R}^{H\times W}$ was used, defined by a normalized Euclidean distance transform:

$$W_{u,v} = \frac{D(u, v)}{\max(D)} \tag{5.3}$$

where $D(u, v)$ is the distance from pixel $(u, v)$ to the patch boundary.

Each predicted patch $\hat{P}_{i,j}$ is weighted by $W$ and accumulated into a canvas $C$, while a mask $M$ tracks the cumulative weights:

$$C[y : y + H,\ x : x + W,\ :] += \hat{P}_{i,j} \cdot W \tag{5.4}$$

$$M[y : y + H,\ x : x + W] += W \tag{5.5}$$

The final reconstructed image $\hat{I}$ is given by:

$$\hat{I}(u, v, c) = \frac{C(u, v, c)}{M(u, v) + \epsilon} \tag{5.6}$$

where $\epsilon$ is a small constant to avoid division by zero. Pixels with zero coverage are filled with background white.

This fusion strategy minimizes seam artifacts and preserves local texture continuity, producing visually coherent whole-slide predictions. On average, a WSI required approximately 2.3 minutes for inference and fusion.

## 5.4 Experiments

### 5.4.1 Experimental Setup

**Datasets.** We conduct experiments on the paired `PARS_VHE_v1.1` dataset, which includes 150 WSIs acquired using a dual-modality PARS-H&E imaging system. The dataset is divided into 105 training slides, 22 validation slides, and 23 test slides. Each WSI is partitioned into $512 \times 512$ patches with a 25% overlap. Patch-level pairing is performed to ensure spatial correspondence between the virtual and reference domains.

**Evaluation Metrics.** To comprehensively evaluate the quality of virtual staining, we adopt a combination of conventional and perceptual metrics:

- **SSIM** – quantifies local structural similarity between generated and ground truth patches.
- **MS-SSIM** – an extension of SSIM that captures structural fidelity across multiple image scales.
- **LPIPS** [65] – a deep perceptual similarity metric based on activations from pre-trained VGG networks.
- **PaPIS** (ours) – a pathology-aware perceptual image similarity metric tailored to histological features.

**Baselines.** We compare our proposed `CycleGAN_PAPIS` model with several variants of the standard CycleGAN [67]. The baseline model is trained with standard adversarial and cycle-consistency losses. In addition, we evaluate two alternative perceptual supervision strategies using SSIM loss and LPIPS loss, respectively, to investigate how different perceptual objectives affect structural, visual, and histological fidelity. All models are trained under identical conditions to ensure a fair comparison.

### 5.4.2 Quantitative Results on Patch-Level Test Set

To assess the impact of different perceptual supervision strategies, we evaluate the baseline CycleGAN and three variants trained with SSIM, LPIPS, and PaPIS loss functions. All models are tested on a held-out set of paired image patches extracted from tissue regions not seen during training, allowing us to measure generalization across diverse histological contexts.

The results are summarized in Table 5.1. As expected, each perceptual loss leads to improvements in specific evaluation metrics. The SSIM-supervised model achieves the highest SSIM (0.778) and MS-SSIM (0.829), indicating its effectiveness in enhancing low-level structural similarity. However, this improvement does not necessarily translate to perceptual or pathology-relevant fidelity.

The LPIPS-guided model achieves the lowest LPIPS score (0.119), suggesting a better match in deep feature space. Nonetheless, its PaPIS score (0.667) remains moderate,

indicating that alignment with generic visual features may not fully capture histological similarity.

Our proposed PaPIS-guided model demonstrates the highest PaPIS score (0.706), validating its strength in preserving pathology-relevant structures such as nuclear detail and glandular architecture. While it does not outperform the SSIM or LPIPS variants in their respective metrics, it offers strong all-around performance, with SSIM (0.766), MS-SSIM (0.818), and LPIPS (0.143) scores that remain competitive.

Interestingly, my method consistently achieved second-best performance across all loss functions, indicating robustness across different optimization settings.

These results suggest that PaPIS loss provides a balanced optimization objective, guiding the model to generate virtual stains that are not only visually and structurally plausible but also aligned with domain-specific perceptual cues. This highlights the value of incorporating task-aware perceptual metrics in virtual staining workflows.

Table 5.1: Quantitative comparison between baseline CycleGAN and models trained with different perceptual losses. The best score in each column is highlighted in bold.

| **Model** | **SSIM ↑** | **MS-SSIM ↑** | **LPIPS ↓** | **PaPIS ↑** |
|---|---|---|---|---|
| CycleGAN (baseline) | 0.739 | 0.790 | 0.198 | 0.641 |
| + SSIM loss | **0.778** | **0.829** | 0.168 | 0.658 |
| + LPIPS loss | 0.751 | 0.806 | **0.119** | 0.667 |
| **+ PaPIS loss (Ours)** | 0.766 | 0.818 | 0.143 | **0.706** |

### 5.4.3 Qualitative Comparison

Figure 5.3 presents representative visual comparisons between the baseline CycleGAN model and our proposed PaPIS-guided framework on selected test samples. Each row corresponds to one example and displays, from left to right: the input PARS image, the ground truth H&E-stained image, the output generated by the baseline CycleGAN, and the output from our enhanced CycleGAN_PAPIS model.

Visual inspection reveals that the outputs generated by our model exhibit notably improved histological fidelity. In particular, nuclear boundaries appear sharper and more distinct, facilitating clearer separation of individual cells. This is especially important in regions with dense cellular structures, where nuclear overlap and poor delineation may

compromise downstream diagnostic tasks. Moreover, the glandular architecture is more accurately reconstructed in the PaPIS-guided results, preserving the shape and connectivity of lumen-containing structures that are often key indicators in pathology grading systems.

Another notable improvement lies in the preservation of nuclear morphology and the suppression of artifacts. While the baseline model (CycleGAN) often introduces blurring or checkerboard patterns in high-magnification views, the PaPIS-guided model produces sharper nuclear contours and more consistent chromatin structures. This suggests that PaPIS-based perceptual loss encourages retention of fine-grained morphological cues that are relevant for histological interpretation.

Overall, these qualitative results support the claim that PaPIS-guided supervision contributes not only to better numerical performance but also to enhanced visual interpretability and alignment with histopathological standards. The improvements are consistent across multiple regions of tissue, demonstrating the robustness and generalizability of our approach.

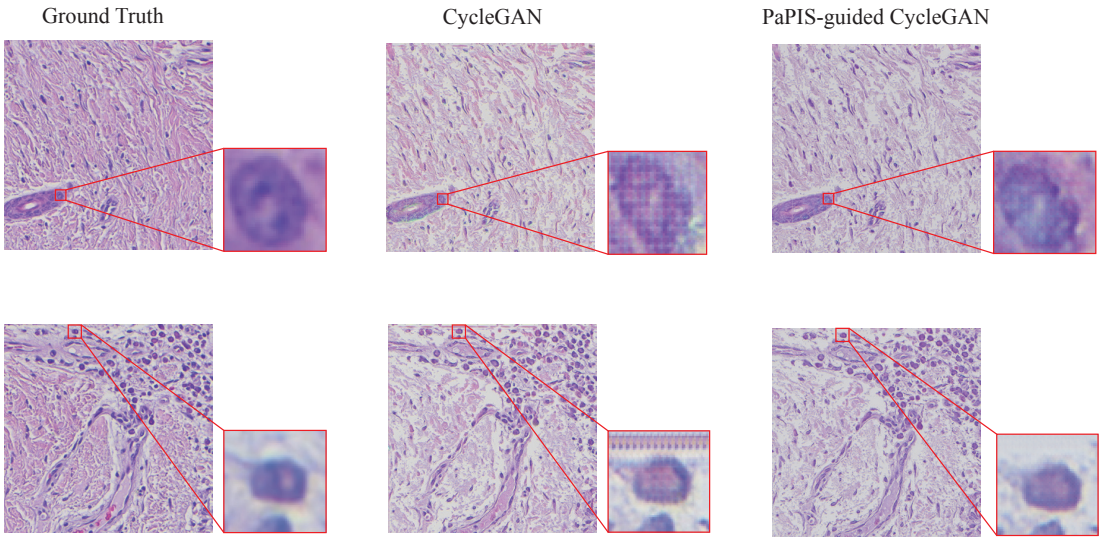



Figure 5.3: Qualitative comparison. From left to right: input PARS, ground truth H&E, CycleGAN output, CycleGAN_PAPIS output.

### 5.4.4 Whole-Slide Virtual Staining

To validate our model's ability to scale, we perform whole-slide inference using overlapping patches and distance-weighted fusion. An example result is shown in Figure 5.4, demonstrating high structural and chromatic consistency over gigapixel-scale images.

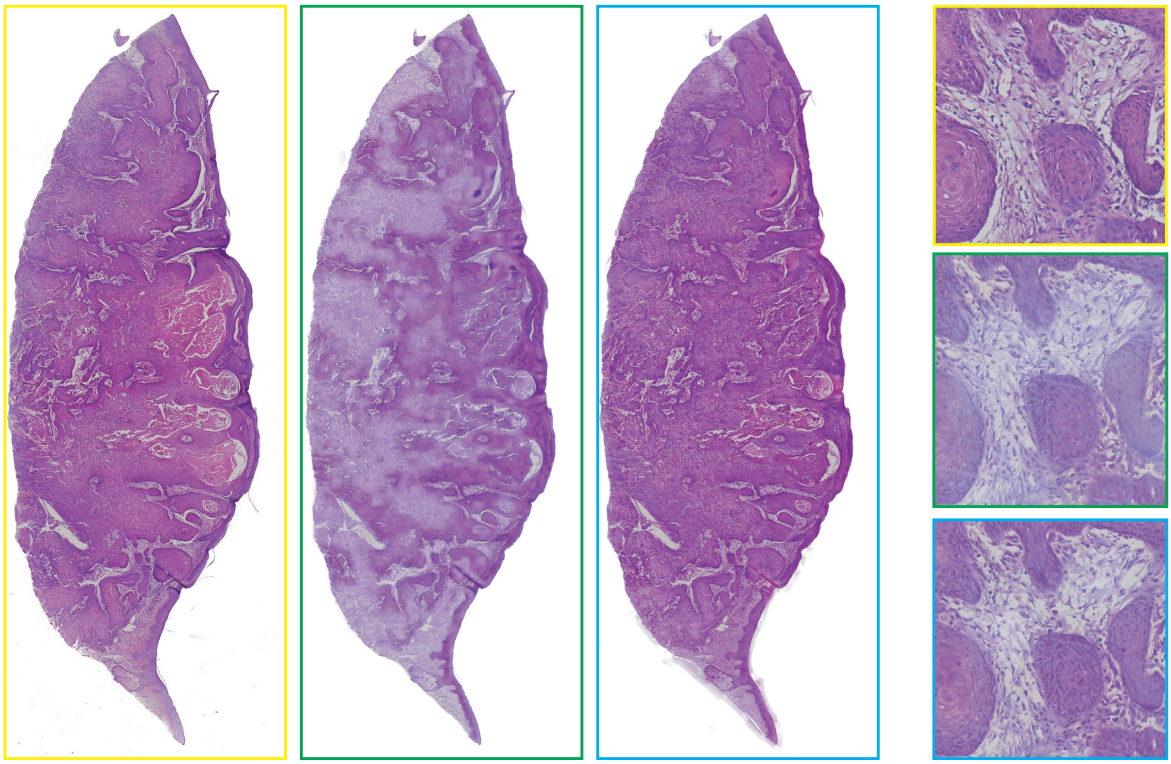


Figure 5.4: Comparison of histological staining results. From left to right, the panels show the ground truth H&E-stained whole slide image, the result generated by a standard CycleGAN model, and the result generated by a CycleGAN model trained with PaPIS supervision. The corresponding high-resolution patches on the right highlight local structures. The PaPIS-supervised model preserves finer histological details and exhibits improved color and structure consistency with the ground truth.

## 5.5 Conclusion

In this work, we introduced a pathology-aware virtual staining framework that integrates the PaPIS metric—designed specifically for histological image comparison—directly into the optimization process of a generative model. By incorporating this perceptual supervision, we address a fundamental limitation in prior virtual staining approaches that rely on generic image similarity losses ill-suited for the nuanced patterns present in tissue morphology.

Through systematic case studies and whole slide inference experiments, our PaPIS-guided model demonstrated improved fidelity in preserving diagnostically relevant features such as nuclear boundaries, glandular structures, and high-frequency textural cues. Quantitative evaluations confirmed consistent gains across SSIM, PSNR, and PaPIS scores, while qualitative visualizations, including heatmaps of metric distributions, offered additional insight into the spatial reliability of virtual staining outputs.

This study underscores the potential of aligning model objectives with expert perceptual criteria to enhance both the interpretability and diagnostic utility of virtual histology. By embedding domain-specific metrics like PaPIS into the training loop, we not only improve visual plausibility but also move closer to clinical readiness.

In future work, we envision extending this framework to other staining modalities, validating performance through expert studies, and optimizing for real-time inference in clinical settings. Ultimately, our findings contribute to the growing paradigm of perceptually-aligned medical imaging and lay the groundwork for trustworthy, pathology-aware virtual staining technologies.

# Chapter 6

# Conclusion

## 6.1 Summary of Contributions

This thesis investigates and addresses two fundamental bottlenecks in the development and deployment of virtual staining systems: (1) the absence of domain-specific image quality evaluation metrics that reflect pathological relevance, and (2) the lack of efficient, scalable preprocessing tools for WSIs, which are essential for training and inference of deep learning-based histological models.

To address the first challenge, we proposed PaPIS, a full-reference evaluation metric that captures perceptual similarity in a pathology-aware manner. Unlike conventional metrics such as SSIM or LPIPS, which are primarily designed for natural image domains, PaPIS leverages features extracted from a pretrained cell morphology segmentation model, and incorporates Retinex-based decomposition to emphasize diagnostically meaningful structures, such as nuclear boundaries and glandular architectures. Experimental comparisons with existing metrics demonstrated that PaPIS achieves superior alignment with qualitative histological assessments.

To address the second challenge, we designed and implemented a high-throughput WSI slicing framework based on tile-aware patch extraction. This framework supports concurrent, multi-threaded execution, minimizes redundant I/O operations via intelligent caching, and integrates seamlessly with OME-TIFF formatted datasets. Benchmarks show that our pipeline significantly accelerates the patching process by up to an order of magnitude compared to traditional serial slicing approaches, while maintaining low memory consumption and I/O overhead.

## 6.2 Integration of Metric and Model

In addition to isolated contributions, we demonstrated the synergistic effect of combining both components by integrating PaPIS into the training loop of a CycleGAN-based virtual staining model. By using PaPIS as a perceptual loss function, the generator is explicitly encouraged to produce outputs that are not only perceptually similar but also histologically faithful. This PaPIS-guided enhancement framework resulted in virtual H&E images with clearer nuclear morphology, better-defined tissue boundaries, and improved alignment with ground truth annotations—outperforming baseline models on both qualitative and quantitative evaluations.

Moreover, we applied the full pipeline, from high-throughput slicing to PaPIS-based evaluation and PaPIS-guided training, on real-world WSIs, showcasing its end-to-end applicability in large-scale virtual staining tasks. The experiments support the claim that pathology-aware guidance can be practically and meaningfully integrated into computational histology workflows.

## 6.3 Limitations

Despite the encouraging results, this work is not without limitations. One major limitation lies in the lack of subjective validation from expert pathologists. While PaPIS shows improved correlation with diagnostic relevance compared to baseline metrics, its true clinical utility remains to be validated through structured reader studies involving multiple human raters. Without this component, the metric's generalizability across tissue types, disease classes, and imaging modalities cannot be fully assured.

Another limitation is the scope of infrastructure support. The slicing tool, though efficient and scalable in local environments, currently supports only OME-TIFF formats. Future virtual staining workflows are increasingly expected to rely on cloud-native formats such as OME-Zarr and operate in distributed settings. The current implementation would require substantial adaptation to function in such federated or remote AI pipelines.

## 6.4 Future Work

Future work can proceed along several directions. First, to enhance the clinical credibility of PaPIS, we plan to incorporate subjective expert scoring and statistically correlate

it with PaPIS and other existing metrics. Such a study would provide stronger justification for using PaPIS as an objective proxy for diagnostic quality. Second, integrating attention-based patch selection in the slicing framework could further reduce computational burden by prioritizing diagnostically rich regions. Third, extending the framework to support GPU-accelerated decoding and native cloud storage access (e.g., Amazon S3, DANDI Hub, OME-Zarr) would allow scalable deployment in multi-institutional training setups, including federated learning environments.

From a modeling perspective, the idea of pathology-aware supervision can also be extended beyond virtual staining into related tasks such as nucleus detection, mitosis classification, and histological grading, potentially improving generalization in low-data regimes.

## 6.5 Final Remarks

In summary, this thesis proposes a modular, pathology-aware, and scalable approach to improve both the computational and diagnostic quality of virtual staining systems. Through the development of a high-performance slicing framework, a novel evaluation metric, and a guided training paradigm, this work contributes practical solutions to a rapidly evolving field at the intersection of computational pathology and deep learning. We hope that these tools will form the foundation for further advancements in virtual staining and inspire future work that brings AI-driven histology closer to clinical translation.

# References


[1] Bijie Bai, Xilin Yang, Yuzhu Li, Yijie Zhang, Nir Pillar, and Aydogan Ozcan. Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications*, 12(1):57, 2023.

[2] Christos G Bampis, Praful Gupta, Rajiv Soundararajan, and Alan C Bovik. Speed-qa: Spatial efficient entropic differencing for image and video quality. *IEEE signal processing letters*, 24(9):1333–1337, 2017.

[3] John D Bancroft and Marilyn Gamble. *Theory and practice of histological techniques*. Elsevier health sciences, 2008.

[4] Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.

[5] Neslihan Bayramoglu, Mika Kaakinen, Lauri Eklund, and Janne Heikkila. Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 64–71, 2017.

[6] Marian Boktor, James ED Tweel, Benjamin R Ecclestone, Jennifer Ai Ye, Paul Fieguth, and Parsin Haji Reza. Multi-channel feature extraction for virtual histological staining of photon absorption remote sensing images. *Scientific Reports*, 14(1):2009, 2024.

[7] Sebastian Bosse, Dominique Maniry, Thomas Wiegand, and Wojciech Samek. A deep neural network for image quality assessment. In *2016 IEEE international conference on image processing (ICIP)*, pages 3773–3777. IEEE, 2016.

[8] Anna Breger, Ander Biguri, Malena Sabaté Landman, Ian Selby, Nicole Amberg, Elisabeth Brunner, Janek Gröhl, Sepideh Hatamikia, Clemens Karner, Lipeng Ning, et al. A study of why we need to reassess full reference image quality assessment with medical images. *arXiv preprint arXiv:2405.19097*, 2024.

[9] Anna Breger, Clemens Karner, Ian Selby, Janek Gröhl, Sören Dittmer, Edward Lilley, Judith Babar, Jake Beckford, Thomas R Else, Timothy J Sadler, et al. A study on the adequacy of common iqa measures for medical images. *arXiv preprint arXiv:2405.19224*, 2024.

[10] Cell Signaling Technology. What is immunohistochemistry (ihc) staining? https://www.cellsignal.com/applications/immunohistochemistry/what-is-immunohistochemistry-staining, 2024. Accessed April 2025.

[11] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2567–2581, 2020.

[12] James M Dolezal, Sara Kochanny, Emma Dyer, Siddhi Ramesh, Andrew Srisuwananukorn, Matteo Sacco, Frederick M Howard, Anran Li, Prajval Mohan, and Alexander T Pearson. Slideflow: deep learning for digital histopathology with real-time whole-slide visualization. *BMC bioinformatics*, 25(1):134, 2024.

[13] Shikha Dubey, Yosep Chong, Beatrice Knudsen, and Shireen Y Elhabian. Vims: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains. In *International Workshop on Machine Learning in Medical Imaging*, pages 143–155. Springer, 2024.

[14] Benjamin R Ecclestone, James A Tummon Simmons, James ED Tweel, Channprit Kaur, Aria Hajiahmadi, and Parsin Haji Reza. Photon absorption remote sensing (pars): A comprehensive approach to label-free absorption microscopy across biological scales. *arXiv preprint arXiv:2403.04229*, 2024.

[15] M Hossein Eybposh, Changjia Cai, Aram Moossavi, Jose Rodriguez-Romaguera, and Nicolas C Pégard. Coniqa: A deep learning method for perceptual image quality assessment with limited data. *Scientific Reports*, 14(1):20066, 2024.

[16] Christoph Gohlke. cgohlke/tifffile: v2025.3.30, March 2025.

[17] S Alireza Golestaneh, Saba Dadsetan, and Kris M Kitani. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Proceedings of*

*the IEEE/CVF winter conference on applications of computer vision*, pages 1220–1230, 2022.

[18] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. Openslide: A vendor-neutral software foundation for digital pathology. *Journal of pathology informatics*, 4(1):27, 2013.

[19] Computational Pathology Group. Asap: Automated slide analysis platform. https://computationalpathologygroup.github.io/ASAP/, 2025. Accessed: 2025-04-24.

[20] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

[21] Andrey Ignatov, Josephine Yates, and Valentina Boeva. Histopathological image classification with cell morphology aware deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6913–6925, 2024.

[22] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[23] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[24] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.

[25] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014.

[26] Tushar Kataria, Beatrice Knudsen, and Shireen Y Elhabian. Staindiffuser: multitask dual diffusion model for virtual staining. *arXiv preprint arXiv:2403.11340*, 2024.

[27] Rémi Kazmierczak, Gianni Franchi, Nacim Belkhir, Antoine Manzanera, and David Filliat. A study of deep perceptual metrics for image quality assessment. *arXiv preprint arXiv:2202.08692*, 2022.

[28] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018.

[29] Leena Latonen, Sonja Koivukoski, Umair Khan, and Pekka Ruusuvuori. Virtual staining for histology by deep learning. *Trends in Biotechnology*, 2024.

[30] Zheyu Li, Bin Li, Kevin W Eliceiri, and Vijaykrishnan Narayanan. Computationally efficient adaptive decompression for whole slide image processing. *Biomedical Optics Express*, 14(2):667–686, 2023.

[31] Melissa Linkert, Curtis T Rueden, Chris Allan, Jean-Marie Burel, Will Moore, Andrew Patterson, Brian Loranger, Josh Moore, Carlos Neves, Donald MacDonald, et al. Metadata matters: access to image data in the real world. *Journal of Cell Biology*, 189(5):777–782, 2010.

[32] Yutao Liu, Guangtao Zhai, Ke Gu, Xianming Liu, Debin Zhao, and Wen Gao. Reduced-reference image quality assessment in free-energy principle and sparse representation. *IEEE Transactions on Multimedia*, 20(2):379–391, 2017.

[33] Jessica Loo, Marc Robbins, Carson McNeil, Tadayuki Yoshitake, Charles Santori, Chuanhe Shan, Saurabh Vyawahare, Hardik Patel, Tzu Chien Wang, Robert Findlater, et al. Autofluorescence virtual staining system for h&e histology and multiplex immunofluorescence applied to immuno-oncology biomarkers in lung cancer. *Cancer Research Communications*, 5(1):54–65, 2025.

[34] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.

[35] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[36] Kohei Ohashi, Yukihiro Nagatani, Makoto Yoshigoe, Kyohei Iwai, Keiko Tsuchiya, Atsunobu Hino, Yukako Kida, Asumi Yamazaki, and Takayuki Ishida. Applicability evaluation of full-reference image quality assessment methods for computed tomography images. *Journal of Digital Imaging*, 36(6):2623–2634, 2023.

[37] Daniel A Orringer, Balaji Pandian, Yashar S Niknafs, Todd C Hollon, Julianne Boyle, Spencer Lewis, Mia Garrard, Shawn L Hervey-Jumper, Hugh JL Garton, Cormac O Maher, et al. Rapid intraoperative histology of unprocessed surgical specimens via

fibre-laser-based stimulated raman scattering microscopy. *Nature biomedical engineering*, 1(2):0027, 2017.

[38] Jean-François Pambrun and Rita Noumeir. Limitations of the ssim quality metric in the context of diagnostic imaging. In *2015 IEEE international conference on image processing (ICIP)*, pages 2960–2963. IEEE, 2015.

[39] Pranita Pradhan, Tobias Meyer, Michael Vieth, Andreas Stallmach, Maximilian Waldner, Michael Schmitt, Juergen Popp, and Thomas Bocklitz. Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning. *Biomedical Optics Express*, 12(4):2280–2298, 2021.

[40] Yair Rivenson, Tairan Liu, Zhensong Wei, Yibo Zhang, Kevin De Haan, and Aydogan Ozcan. Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light: Science & Applications*, 8(1):23, 2019.

[41] Yair Rivenson, Hongda Wang, Zhensong Wei, Kevin de Haan, Yibo Zhang, Yichen Wu, Harun Günaydın, Jonathan E Zuckerman, Thomas Chong, Anthony E Sisk, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature biomedical engineering*, 3(6):466–477, 2019.

[42] Yair Rivenson, Hongda Wang, Zhensong Wei, Yibo Zhang, Harun Gunaydin, and Aydogan Ozcan. Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. *arXiv preprint arXiv:1803.11293*, 2018.

[43] Rafael Rodrigues, Lucie Lévêque, Jesús Gutiérrez, Houda Jebbari, Meriem Outtas, Lu Zhang, Aladine Chetouani, Shaymaa Al-Juboori, Maria G Martini, and Antonio MG Pinheiro. Objective quality assessment of medical images and videos: Review and challenges. *Multimedia Tools and Applications*, pages 1–34, 2024.

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[45] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.

[46] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.

[47] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.

[48] Rajiv Soundararajan and Alan C Bovik. Rred indices: Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing*, 21(2):517–526, 2011.

[49] D Swainson Sujana, D Peter Augustine, and D Sheefa Ruby Grace. Full reference image quality assessment (fr-iqa) of pre-processed structural magnetic resonance images. In *2024 IEEE International Conference on Contemporary Computing and Communications (InC4)*, volume 1, pages 1–5. IEEE, 2024.

[50] Yu Tian, Baoliang Chen, Shiqi Wang, and Sam Kwong. Towards thousands to one reference: Can we trust the reference image for quality assessment? *IEEE Transactions on Multimedia*, 26:3278–3290, 2023.

[51] James ED Tweel, Benjamin R Ecclestone, Marian Boktor, Deepak Dinakaran, John R Mackey, and Parsin Haji Reza. Automated whole slide imaging for label-free histology using photon absorption remote sensing microscopy. *IEEE Transactions on Biomedical Engineering*, 71(6):1901–1912, 2024.

[52] James ED Tweel, Benjamin R Ecclestone, Marian Boktor, James Alexander Tummon Simmons, Paul Fieguth, and Parsin Haji Reza. Virtual histology with photon absorption remote sensing using a cycle-consistent generative adversarial network with weakly registered pairs. *arXiv preprint arXiv:2306.08583*, 2023.

[53] Domonkos Varga. Full-reference image quality assessment based on an optimal linear combination of quality measures selected by simulated annealing. *Journal of Imaging*, 8(8):224, 2022.

[54] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*, pages 1–6. IEEE, 2015.

[55] Junjia Wang, Bo Xiong, You Zhou, Xun Cao, and Zhan Ma. A value mapping virtual staining framework for large-scale histological imaging. *arXiv preprint arXiv:2501.03592*, 2025.

[56] Lanjiang Wang. A survey on iqa. *arXiv preprint arXiv:2109.00347*, 2021.

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[58] Zhou Wang and Alan Conrad Bovik. *Modern image quality assessment*. PhD thesis, Springer, 2006.

[59] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[60] Weizhi Xian, Mingliang Zhou, Bin Fang, Tao Xiang, Weijia Jia, and Bin Chen. Perceptual quality analysis in deep domains using structure separation and high-order moments. *IEEE Transactions on Multimedia*, 26:2219–2234, 2023.

[61] Xilin Yang, Bijie Bai, Yijie Zhang, Yuzhu Li, Kevin de Haan, Tairan Liu, and Aydogan Ozcan. Virtual stain transfer in histology via cascaded deep neural networks. *ACS Photonics*, 9(9):3134–3143, 2022.

[62] Yuxuan Yang, Zhichun Lei, and Changlu Li. No-reference image quality assessment combining swin-transformer and natural scene statistics. *Sensors*, 24(16):5221, 2024.

[63] Chiho Yoon, Eunwoo Park, Sampa Misra, Jin Young Kim, Jin Woo Baik, Kwang Gi Kim, Chan Kwon Jung, and Chulhong Kim. Deep learning-based virtual staining, segmentation, and classification in label-free photoacoustic histology of human specimens. *Light: Science & Applications*, 13(1):226, 2024.

[64] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011.

[65] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[66] Yijie Zhang, Luzhe Huang, Tairan Liu, Keyi Cheng, Kevin de Haan, Yuzhu Li, Bijie Bai, and Aydogan Ozcan. Virtual staining of defocused autofluorescence images of unlabeled tissue using deep neural networks. *Intelligent Computing*, 2022.

[67] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.