TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

AI-driven image-to-image translation for histopathology

Umair Akhtar Hasan Khan

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1949 | MEDICA – ODONTOLOGICA | TURKU 2026

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

AI-DRIVEN IMAGE-TO-IMAGE TRANSLATION FOR HISTOPATHOLOGY

Umair Akhtar Hasan Khan

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1949 | MEDICA – ODONTOLOGICA | TURKU 2026

University of Turku

Faculty of Medicine
Institute of Biomedicine
Pathology
Doctoral Programme on Drug Research and Diagnostics (DRD)

Supervised by

Associate Professor Pekka Ruusuvuori
University of Turku
Institute of Biomedicine
Turku, Finland

Adjunct Professor Leena Latonen
University of Eastern Finland
Institute of Biomedicine
Kuopio, Finland

Reviewed by

Professor Tapio Seppänen
University of Oulu
Faculty of Information Technology and Electrical Engineering
Oulu, Finland

Academy Research Fellow Lassi Paavolainen
University of Helsinki
Institute for Molecular Medicine Finland
Helsinki, Finland

Opponent

Professor Nataša Sladoje
Uppsala University
Department of Information Technology
Uppsala, Sweden

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

ISBN 978-952-02-0542-3 (PRINT)
ISBN 978-952-02-0543-0 (PDF)
ISSN 0355-9483 (PRINT)
ISSN 2343-3213 (ONLINE)
Painosalama, Turku, Finland 2026

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Pathology
UMAIR AKHTAR HASAN KHAN: AI-driven image-to-image translation for histopathology
Doctoral Dissertation, 136 pp.
Doctoral Programme on Drug Research and Diagnostics (DRD)
February 2026

ABSTRACT

Histological assessment of tissue specimens is the benchmark for disease diagnosis. Chemical staining of the tissue specimen is the process that enables this assessment by effectively applying distinct colors to different tissue components. A stained tissue reveals a plethora of information that is almost invisible to a human eye in an unstained tissue. However, chemical staining is an age-old technique which has some critical drawbacks such as inconsistent appearance of the stained tissue, waste generation, resource and tissue consumption, and being labor intensive. Although the protocols have advanced over time, the fundamentals have remained largely unchanged. A paradigm shift to the digital medium has the potential to alleviate these problems.

Artificial Intelligence (AI) has been very successful in propelling a range of different domains including biomedical imaging. The major effort in the field has been directed toward disease diagnosis of late. However, these methods can also be utilized for the optimization of various histological processes including virtual staining of unstained tissue images and stain normalization of tissue images with nonuniform color appearance. It is primarily achieved by using AI-based image-to-image translation methods that are capable of transforming images from one domain to another. The studies presented in this work mainly revolve around developing AI-based image-to-image translation methods and comparing existing ones to optimize different histological processes. It also focuses on fine tuning tissue processing and imaging to produce optimal tissue image data for training the virtual staining methods. This thesis highlights the potential of AI-based image-to-image translation methods for virtually recreating the stained tissue images using only unstained images as input. Furthermore, it showcases the challenges of handling variation in inter-laboratory staining appearance and effectiveness of different methods on stain normalization tasks. This work demonstrates the transformative potential of generative AI and its vitality for optimizing histological workflow by enabling faster, leaner, more efficient and sustainable sample processing pipeline, and by improving the consistency of staining appearance. It will also serve as an improved foundation for AI-based diagnostic systems.

KEYWORDS: generative artificial intelligence, image-to-image translation, virtual staining, stain normalization

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Biolääketieteen laitos
Patologia
UMAIR AKHTAR HASAN KHAN: Tekoälypohjainen kuvasta-kuvaan muuntaminen histopatologiassa
Väitöskirja, 136 s.
Lääketutkimuksen ja diagnostiikan tohtoriohjelma (DRD)
Helmikuu 2026

TIIVISTELMÄ

Kudosnäytteiden histologinen arviointi on monien sairauksien diagnosoinnin perusta. Kudosnäytteen kemiallinen värjäys mahdollistaa tämän arvioinnin värjäämällä kudoksen eri osat. Värjätty kudos paljastaa runsaasti informaatiota, joka on lähes näkymätöntä paljaalle silmälle värjäämättömässä kudoksessa. Kemiallinen värjäys on kuitenkin vanha menetelmä, jolla on merkittäviä haittapuolia, kuten kudoksen epätasainen värjäytyminen, jätteen synty, resurssien ja kudoksen kulutus sekä työläys. Vaikka menetelmät ovat kehittyneet ajan myötä, niiden perusperiaatteet ovat pysyneet suurelta osin muuttumattomina. Siirtyminen digitaaliseen ympäristöön voi mahdollistaa näiden ongelmien lieventämisen.

Tekoäly on saavuttanut merkittävää menestystä monilla eri aloilla, mukaan lukien biolääketieteellinen kuvantaminen. Viimeaikainen painopiste alalla on ollut sairauksien diagnosoinnissa. Tekoälyyn perustuvia menetelmiä voidaan kuitenkin hyödyntää myös erilaisten histologisten prosessien, kuten värittömien kudoskuvien virtuaalivärjäyksen ja epätasaisen värjäyksen omaavien kudoskuvien värin normalisoinnin optimoinnissa. Tämä saavutetaan pääasiassa tekoälyyn perustuvilla kuvasta-kuvaan -muuntomenetelmillä, jotka kykenevät muuntamaan kuvia yhdestä esitystyylistä toiseen. Tämä työ keskittyy pääosin näiden menetelmien kehittämiseen ja olemassa olevien menetelmien vertailuun histologisten prosessien optimoimiseksi. Lisäksi työ tarkastelee kudosprosessoinnin ja kuvantamisen hienosäätöä optimaalisien kudoskuvien tuottamiseksi virtuaalivärjäysmenetelmien koulutukseen. Tämä väitöskirja tuo esiin tekoälypohjaisten kuvasta-kuvaan -muuntomenetelmien mahdollisuudet luoda värjättyjä kudoskuvia virtuaalisesti pelkästään värjäämättömien kuvien perusteella. Lisäksi työ havainnollistaa laboratoriokohtaisten värjäyserojen käsittelyn haasteita ja eri menetelmien tehokkuutta värjäyksen normalisoinnissa. Tämä tutkimus osoittaa generatiivisen tekoälyn muutosvoiman ja sen merkityksen histologisen työnkulun optimoinnissa – mahdollistaen nopeamman, kevyemmän, tehokkaamman ja ympäristölle kestävämmän näytekäsittelyprosessin sekä parantaen värjäyksen yhdenmukaisuutta. Tämä tarjoaa myös paremman lähtökohdan tekoälypohjaisille diagnostiikkajärjestelmille.

AVAINSANAT: generatiivinen tekoäly, kuvasta-kuvaan-muunnos, virtuaalivärjäys, värin normalisointi

Table of Contents

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
CT	Computed Tomography
EMA	European Medicines Agency
EQA	External Quality Assurance
FDA	Food and Drugs Administration
FID	Frachet Inception Distance
FFPE	Formalin-Fixed Paraffin-Embedded
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
H&E	Hematoxylin and Eosin
HOG	Histogram of Oriented Gradients
LBP	Local Binary Patterns
MAE	Mean Absolute Error
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
OD	Optical Density
PCC	Pearson Correlation Coefficient
PET	Positron Emission Tomography
PSNR	Peak Signal-to-Noise Ratio
QPI	Quantitative Phase Imaging
SIFT	Scale-Invariant Feature Transform
SNMF	Sparse Non-negative Matrix Factorization
SSIM	Structural Similarity Index Measure
SVD	Singular Value Decomposition
SVM	Support Vector Machine
ViT	Vision Transformers
WSI	Whole Slide Image
UGATIT	Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Koivukoski, S., Khan, U., Ruusuvuori, P., & Latonen, L. Unstained tissue imaging and virtual hematoxylin and eosin staining of histologic whole slide images. Laboratory Investigation, 2023; 103(5), 100070. https://doi.org/10.1016/j.labinv.2023.100070

II Khan, U., Koivukoski, S., Valkonen, M., Latonen, L., & Ruusuvuori, P. The effect of neural network architecture on virtual H&E staining: Systematic assessment of histological feasibility. Patterns, 2023; 4(5). https://doi.org/10.1016/j.patter.2023.100725

III Khan, U., Härkönen J., Friman M., Latonen L., Kuopio, T., Ruusuvuori, P. Staining normalization in histopathology: Multi-center dataset and method benchmarking. *arXiv,* 2025. https://doi.org/10.48550/arXiv.2506.19106

1 Introduction

Pathology is a branch of medical science concerned with the study of the nature and causes of different diseases. Histopathology is a sub-branch of pathology that studies the manifestation of those diseases particularly in tissue. (Hoda, 2020) Modern microscopy has enabled us to tap into a plethora of information packed in tissue specimens and understand how the structural and morphological changes correlate with a disease and progression thereof. Analyzing a tissue specimen's granularities at 20x, 40x, and occasionally at 100x magnification, significantly leaves other methodologies behind in terms of the sheer exactness, detail, and quantity of the information it provides.

Like many other medical specialties, histopathology has also been going through digital transformation for approximately the last three decades. Major strides towards digitalization started appearing in the 1990s with the release of the first commercial tissue slide scanner (circa 1994), designed by James Bacus (Pantanowitz, et al., 2018). The initiatives involving the capture, storage, management, and analysis of tissue images in the digital medium gave birth to the field of digital pathology. The realization of digital pathology is imperative not only because of its inherent benefits such as enablement of efficient workflow and standardization, but also because it serves as an essential and inevitable milestone for maximizing the true potential of histological assessment, i.e., using artificial intelligence-based computational methods for tissue analysis.

The last decade was marked by an exceptional rise of AI (Shao, et al., 2020), primarily because of the synchronous growth of both hardware and AI algorithms. Graphics processing unit (GPU), once restricted to the realm of video games, witnessed large scale adoption for AI model training. On the algorithmic front, Alexnet (Krizhevsky, et al., 2012), an image classification model, introduced the concept of deep learning (LeCun, et al., 2015). Consequently, the evolution of deep learning had a momentous impact on both computer vision (Chai, et al., 2021) and natural language processing (Vaswani, et al., 2017; Devlin, et al., 2019). During this period, in biomedical imaging, the utilization of AI-based computational methods for analyzing ultra-high resolution gigapixel tissue images became increasingly prevalent for both quantitative and qualitative analyses, leading to the emergence of another subspeciality called computational pathology.

Computational pathology has been an active area of research for over a decade, and it has direct applications in diagnostics, prognostics, and precision medicine (Cui & Zhang, 2021; Fuchs & Buhmann, 2011; Hosseini, et al., 2024). In recent past, the

advancements in computational pathology have extended beyond the confines of research and made their way to FDA-approved and CE-marked clinical applications (Aiforia, 2013; Paige, 2013). The diagnostic precision and the time saved by AI-assisted diagnostic tools is the promised value that will not only offset the cost of digitalization, but more importantly, positively impact patient survival outcomes. Diagnostic delays have been reported to significantly affect patient survival, as evidenced by a 10% reduction in survival rates linked with delays in the initiation of cancer treatment (Hanna, et al., 2020).

Qualitative and quantitative analyses, aiding disease diagnosis, have been the nexus of most of the research in computational pathology. Workflow optimization and standardization, although imperative, remains less explored areas in comparison. Tissue processing, staining and imaging are the histopathology workflow steps that suffer from lack of standardization. Variability in these steps could have a significant impact on the appearance of stained tissue specimens and their digital images. This not only poses a challenge for the experts analyzing the tissue either optically or virtually (Salvi, et al., 2023), but also becomes a gordian knot in developing robust AI-based diagnostic tools (Breen, et al., 2024). Following digitization, both optimization and standardization problems can be solved using computational methods. In computer vision, image-to-image translation encompasses a group of computational methods that transforms images from a source domain to a target domain. Applications for image-to-image translation in histopathology include, but are not restricted to, virtual staining of unstained tissue images (virtual staining) (Latonen, et al., 2024), normalization of varying staining appearance caused by different labs, protocols, and imaging equipment (stain normalization) (Breen, et al., 2024), and transforming images of tissue stained with a particular stain to another stain (stain-to-stain transformation) (Bai, et al., 2023) (**Figure 1**).

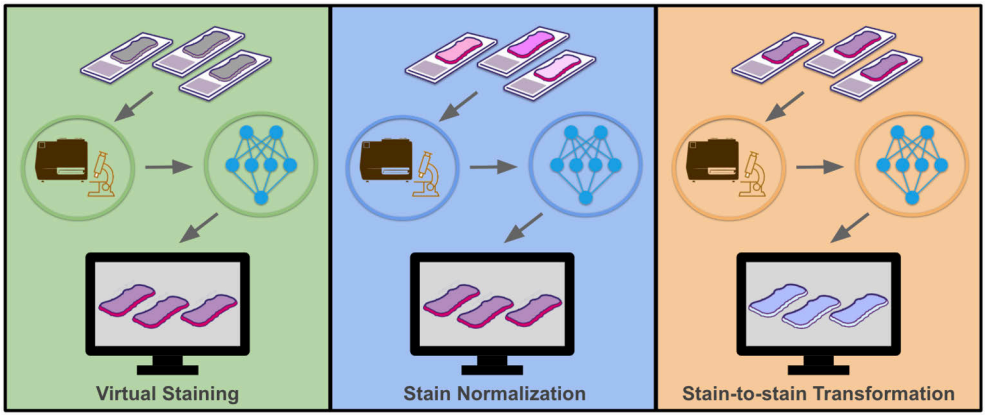

Figure 1. Applications of image-to-image translation in histopathology.

Virtual staining and stain normalization are the focal points of this thesis work. The work presented here explores the prerequisites for virtual staining such as tissue processing and imaging protocols and their impact on resulting images. In continuation, various image-to-image translation methods were assessed for virtual staining. Further investigation was conducted to understand the impact of varying architectural complexities of the chosen methodology, on virtual staining outcomes. Finally, a unique multi-lab tissue image dataset was studied in the context of stain normalization, wherein, both traditional and deep learning methods were used for normalizing the staining variation stemming from different labs.

2 Review of the Literature

2.1 Histopathology

Although the field of histopathology surfaced and evolved mostly during the last two centuries (Titford, 2006), one of the first recorded tissue analyses dates back to the 17th century when a Dutch textile merchant and a self-taught microbiologist and microscopist, Antonie van Leeuwenhoek, used his single-lens microscope to observe a variety of different specimens including his own skin, blood, and saliva (Cole, 1949). A human or animal tissue is inherently delicate, and after extraction, as it is, does not reveal much discernible information. Therefore, one of the most critical aspects of histopathology is the handling and processing of the tissue. Once the tissue is treated and sectioned, it is ready for staining which is one of the most significant steps in the workflow, as it reveals all the complex and elaborate tissue structures and most importantly, the distinctive morphology of both healthy and diseased cells. Interestingly, the first recorded use of histological staining is also attributed to Leeuwenhoek, when he used saffron to dye the tissue making the microscopic structures appear more pronounced revealing details never witnessed before (Schulte, 1991). In essence, the evolution and coalescence of microscopy, pathology, tissue processing, and staining techniques set the foundation stone of the field of histopathology.

2.1.1 Histological Assessment

In medical science, to diagnose chronic diseases such as cancer, the preliminary approach used is non-invasive imaging of the potentially diseased body part. Computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound and X-rays are some of the examples of non-invasive imaging techniques. These diagnostic methods are invaluable and provide inimitable details of the human body. However, microscopic assessment of histologically stained tissue specimens, typically extracted through biopsies and surgical excision, remains the gold standard for cancer diagnosis. (Brown, et al., 2012; Zheng, et al., 2022; Chong, et al., 2021) The level of granularity provided by histological analysis is unparalleled. Direct visualization of tissue morphology and cellular characteristics enables more accurate assessment of the disease. It not only attests the presence of a disease such as cancer, but also provides details to distinguish between different types, subtypes, and grade of cancers with high confidence. This allows for a

comprehensive evaluation that helps to devise appropriate treatment plans and predict patient survival outcomes (Das, et al., 2020; Leong & Zhuang, 2011). It also serves as a validation tool for the preliminary line of examination. For instance, MR or CT scans may indicate the presence of a tumor, but it is the histological analysis that provides confirmation and detailed information about the nature and characteristics of the tumor. In addition to cancer, histological examination of the stained tissue specimens is occasionally used to diagnose other medical conditions such as inflammatory, degenerative, metabolic, genetic, endocrine, hematological, and connective tissue diseases as well (Sternberg, et al., 2004; Kumar, et al., 2013).

2.1.2 Tissue Processing

A freshly extracted clinical or preclinical tissue specimen is quite delicate in nature and is required to undergo a set of treatments for structural and morphological preservation by removal of water and lipid enabling tissue embedding and sectioning. (Bancroft & Gamble, 2008) The principal goal of tissue processing is to prepare it for optimal staining and imaging to further facilitate tissue analysis and disease diagnosis.

Typically, the first step in tissue processing is fixation wherein the tissue block is treated with a fixative solution to prevent it from decaying due to autolysis or putrefaction. (Kiernan, 2015) The process is required to strike a delicate balance between halting selective chemical processes and leaving the tissue components with limited chemical reactivity for specific staining. (Carson & Cappellano, 2015) Formaldehyde-based fixation, is the most commonly used method due to its ability to cross-link proteins and nucleic acids, preserving cellular architecture and antigenicity (Thavarajah, et al., 2012). To remove water, the tissue is then dehydrated through a series of alcohol washes with gradual increase in alcohol concentration in each iteration. This facilitates the embedding media penetration and prevents distortions and shrinkage. (Kiernan, 2015) Tissue dehydration is followed by embedding. In this step, a structural support media such as paraffin wax or acrylic resin is infused in the tissue block, this hardens the tissue to facilitate the execution of the next step, i.e., sectioning. (Pecio & Piprek, 2019) The embedded tissue block is sliced into razor thin tissue sections using a device called microtome, the sectioned tissue is then mounted, typically through warm water as in intermediary medium, onto transparent glass slides. (Bancroft & Gamble, 2008) For typical histological analysis, a section thickness of 4-5 micrometers is preferred. (Alturkistani, et al., 2016) The tissue sections are then ready for staining which is explained in the next section. While formalin-fixed paraffin-embedded (FFPE) processing remains the predominant method in histopathology, some applications such as intraoperative diagnosis use frozen sections and cryosectioning as well which involves instant freezing of tissue specimen and the use of cryostat, a refrigerated microtome, for tissue slicing (Son, et al., 2013; Jaafar, 2006). The work done in this thesis used FFPE processed tissues.

The steps explained above constitute the general tissue processing framework; however, the protocols and execution differ from laboratory to laboratory. Tissues

processing comprises a set of intricate steps, during their execution, determinants such as excessive or insufficient fixation duration, selection of embedding media and temperature, as well as the thickness of tissue sections, may non-trivially contribute to the heterogeneity of the staining results (Troiano, et al., 2009; Chlipala, et al., 2021).

2.1.3 Histological Staining

The thinly sliced tissue section, mounted on the glass slide, is almost transparent and doesn't reveal much meaningful information, therefore, to visually accentuate the constituents of a tissue specimen, it must be stained. In histopathology, there are a variety of stains and stain combinations that can be used to dye a tissue section based on the underlying diagnostic target or examination requirements. (Bancroft & Gamble, 2008) However, Hematoxylin and Eosin (H&E) remains the most commonly used stain combination and widely considered a benchmark in medical diagnosis. (Rosai, 2007) The work presented in this thesis also focuses on H&E-stained tissue samples. Examples of other stains include Immunohistochemistry (IHC) stain family and special stains such as Periodic Acid Schiff (PAS), Mason's Trichrome, Giesma, Reticulin, and Aldehyde Fuchsin.

H&E stain combination has stood the test of time since it was first used in the late 19th century by Wissowzsky (1876) and Busch (1877) (Wittekind, 2003). Hematoxylin induces a purplish-blue chromaticity in cell nuclei, whereas eosin imparts a pink tint to the extracellular matrix and cytoplasm. Different structures within the tissue exhibit a spectrum of abovementioned colors, manifesting through varying amalgamations of these pigments. (Bancroft & Stevens, 1982; Chan, 2014) This helps to emphasize the visual characteristics of a tissue specimen and makes it easier for a pathologist to distinguish different cellular components including organelles such as nuclei cytoplasm, and other components such as the extracellular matrix. Furthermore, the collective chromatic pattern emerging from the stained tissue provides a macro view of the spatial distribution of cells and the tissue structure itself. (Wittekind, 2003) Consequently, pattern recognition is possible for both human experts and computer-aided diagnostic systems to assist those experts.

Chemical staining of a tissue specimen, although invaluable and foundation stone of histological analysis, does have some noticeable drawback as well. It is a resource-intensive and laborious process that is prone to introducing technical artifacts in the stained tissue. Even subtle variation in the staining protocol could influence the appearance of the stained tissue. The heterogeneity introduced, does not only pose a diagnostic challenge for human experts, but also hinders the ability of AI-based tools to generalize across an ever-increasing variety of staining appearances effectively (Van der Laak, et al., 2021; Soliman, et al., 2024). The staining process is also inherently tissue-consuming because typically a tissue section stained with a particular dye is not restrained with another dye, the adjacent section is used if a different staining is required for a different underlying target. Although removing the initial stain is possible in some cases, it is not part of the

standard technique, requires significant efforts, involves the possible risk of damaging the tissue, and some stains are permanent in nature.

2.1.4 Tissue Imaging

Traditionally, microscopes have been the equipment of choice for histological analysis of tissue specimens. However, with the advent of digital pathology tissue analysis has become more convenient because ultra-high resolution scanned tissue images can be more easily analyzed in the digital medium.

Whole slide image (WSI) scanning is the process of digitizing the tissue mounted glass slides meant to facilitate pathology workflows, education, research and digital archiving. (Zarella, et al., 2019). In digital slide scanners, WSI scanning is typically done in small sections either in the form of a tile or a line. The small sections, once imaged, are digitally merged to form the full-resolution WSI (Indu, et al., 2016). Factors influencing the scanning process, depending on the scanner vendor, could also be manually configured. WSI can be imaged at different magnification levels, such as 10x, 20x, and 40x. Modern scanners, through a low magnification scan are capable of automatically detecting tissue boundaries, which in turn, significantly increases scanning pace (Zarella, et al., 2019) (**Figure 2**). The scanned image contains a huge amount of data, typically the size of a whole slide image is in gigabytes. Depending on the vendor, the imaging software also provides compression options: lossy compression such as JP2000 or lossless compression such as LZW.

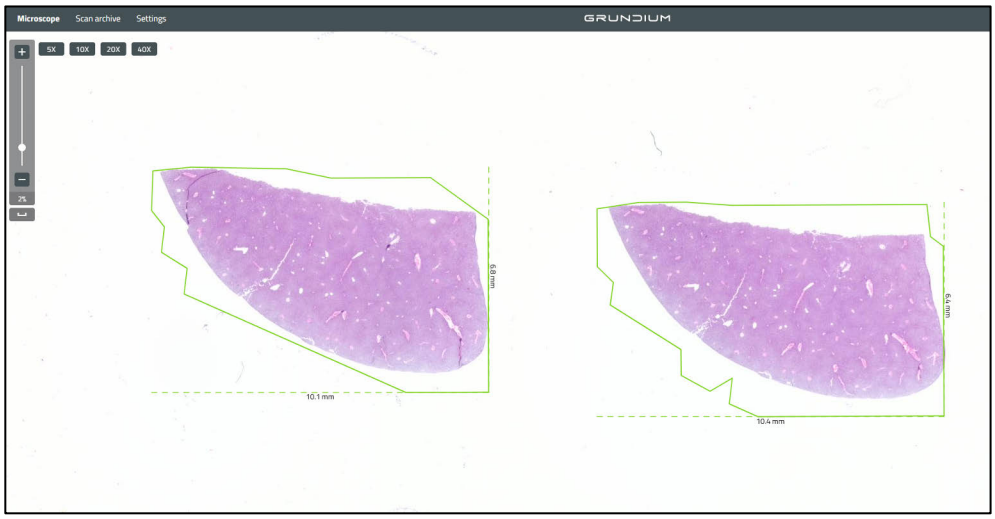

Figure 2. Real-time automated detection of kidney tissue using software developed by Grundium, a company that manufactures compact digital pathology scanners.

An aspect of this research work that makes it ambitious is the use of label-free imaging, i.e., scanning the unstained tissue image. It is a technique which, by

definition, attempts to capture the content of the biological specimen as unaltered and as close to its natural form as possible without physically or chemically affecting it in unintended ways. (Shaked, et al., 2023) There are a range of different imaging modalities capable of performing label-free imaging. (Shaked, et al., 2023) However, brightfield imaging has been the modality of choice in this study because it has high throughput, it is cost-effective and is virtually available in every laboratory and hospital. Although brightfield scanners are capable of scanning unstained tissue, they are generally designed to image stained tissue. Imaging unstained tissue also becomes challenging because it lacks contrast. As a solution, removing the protective coverslip significantly improves the imaging process. Another problem that plagues the imaging step is the lack of standardization in imaging techniques of different digital slide scanners. Same tissue sections scanned with different scanners tend to differ in appearance and this poses similar challenges as the variation in staining or tissue processing. (Dunn, et al., 2024)

The digital medium has an immense potential to tackle the challenges that arise during tissue processing, staining, and imaging. (Madabhushi & Lee, 2016) With advancement in data storage and compute technologies, an unprecedented amount of tissue image data can be stored very efficiently and processed with compute power never experienced before. This makes tissue analysis a quintessential AI application area that can be explored using computational methods requiring troves of data and colossal compute power. (Campanella, et al., 2019)

2.2 Digital & Computational Pathology

Digitalization is inherently a disruptive phenomenon which typically spreads as a result of its advantages significantly outweighing the drawbacks. For instance, digitalization of photography made it cost-effective, highly configurable and editable, portable, easy to access, and most importantly environment friendly. Similarly, digital pathology, a domain that enables the digitization of tissue slides using digital slide scanners, provides efficient image storage, management, analysis, and collaboration solutions. (Pantanowitz, et al., 2011) The concept of whole slide images is the same as analyzing tissue sections with high magnification using a traditional microscope, but digitally, without requiring the physical slide. This is also an effective way of preserving the staining information because the appearance of the physical specimen fades with time. (Vicory, et al., 2015)

At the heart of digital pathology is the slide scanner. For the last two decades, companies such as Philips, Olympus, Hamamatsu Photonics, 3DHistech, and Leica Biosystems have been at the forefront of developing digital slide scanners. Just like microscopes, digital slides scanners can also be categorized as brightfield, fluorescent, and multispectral (Zarella, et al., 2019). Brightfield scanner-based imaging remains the most commonly used and cost-effective. Scanners can also be vertically categorized based on throughput, i.e., in terms of their slide stacking capacity which can range from a single slide to as high as a few hundred slides at a time. (Indu, et al., 2016) **Figure 3** shows examples of different digital slide scanners.

Slide scanning time varies from 30 seconds to several minutes depending on factors such as the size of the tissue and magnification level. (Zarella, et al., 2019)

The value proposition of digital pathology has been largely clear and evident. It can be used in diagnostics, education, and research. It standardizes workflow, enhances collaboration, reduces errors, improves tissue analysis, and consequently, improves patient care as well. Although highly publicized, the cost-effectiveness aspect of digital pathology has been enveloped by ambivalence. What has made it questionable is the upfront transition cost, and this predicament has been one of the biggest impediments to the fast adoption of digital pathology. (Griffin & Treanor, 2017) Early on, diagnostic validation also posed a noticeable challenge primarily because of the prolonged infancy of the domain, the cyclical slow adoption problem, and absence of a robust validation framework. (Parwani, et al., 2014; Pantanowitz, et al., 2013) However, as major medical institutes and laboratories from different parts of the world became early adopters of digital pathology, they significantly helped the advancement of the field through conducting extensive clinical validation studies and outlining validation procedures and protocols. (Jahn, et al., 2020; Evans, et al., 2017; Williams, et al., 2019) These players are also reaping the benefits of digitalization in different ways, one of which is the aforementioned high throughput scanning that is invaluable for establishments with high patient influx (Zarella & Rivera Alvarez, 2022).

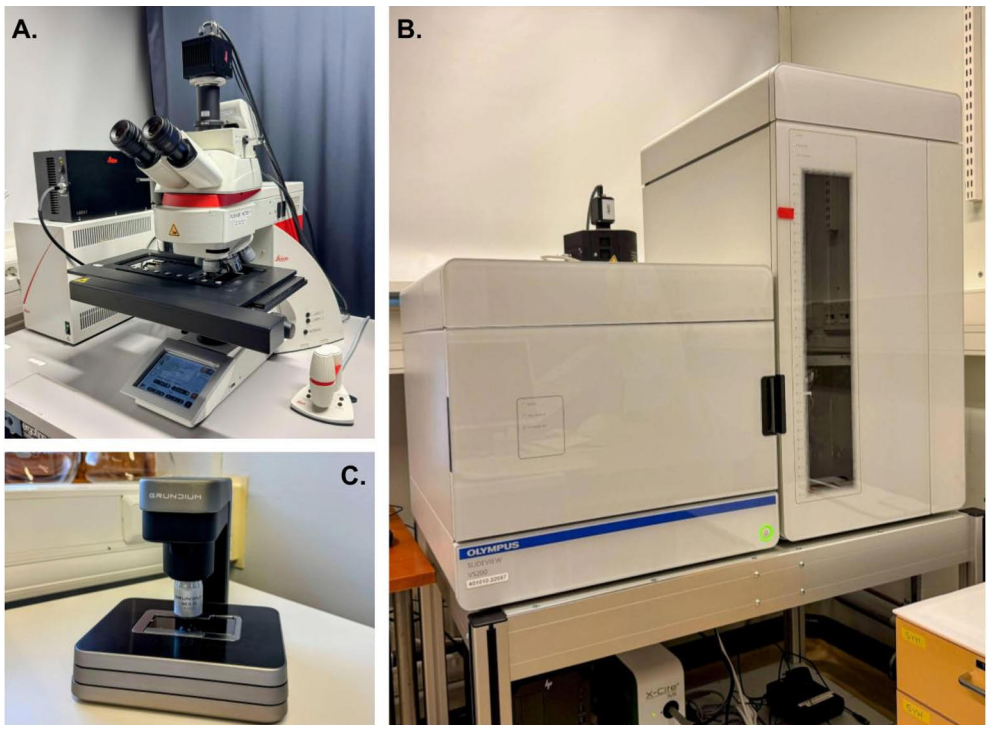

Figure 3. Digital pathology scanners range from small portable scanners to high-throughput ones with high slide loading capacity. **A**) Leica DM6 B **B**) Olympus Slideview VS200 **C**) Grundium Ocus 40

Regulatory agencies such as Food and Drugs Administration (FDA) and European Medicines Agency (EMA) are playing a significant role in developing the regulatory framework towards equipment and software standardization in digital pathology. (Kearney, et al., 2021) Intellisite by Philips is the first comprehensive digital pathology solution to receive FDA-clearance, it primarily consists of an ultra-fast slide scanner, an image management system, a display, and a set of advanced software tools. (Patel, et al., 2021) In the last few decades, digital pathology has come a long way, for instance, scanning time for WSI at 40x has decreased from roughly 1 hour to less than a minute (Patel, et al., 2021). Even with an initial reluctance in adoption of digital pathology, attributed mainly to the daunting implementation costs, early adopters (medical institutes and laboratories) have also been successful in generating invaluable WSI datasets. A substantial number of both proprietary and public datasets are available containing hundreds to millions of WSIs (Bulten, et al., 2022; Bandi, et al., 2018; Vorontsov, et al., 2023; Stettler, 2022). This abundance of WSI data is key to something more powerful that can help to significantly boost the digitalization traction, i.e., the utilization of AI for tissue image analysis.

Computational pathology, a term that ushered in the 2010s, refers to a domain dedicated to solving pathology-related problems through computational methods, with a primary focus on analysis and interpretation of whole slide image and other relevant digitized data. (Abels, et al., 2019) The methodologies are primarily categorized into two approaches. The first approach involves traditional machine learning methods such as linear regression (Montgomery, et al., 2021), support vector machine (SVM) (Cortes & Vapnik, 1995), random forest (Breiman, 2001), and clustering (MacQueen, 1967). They rely mainly on manually engineered features extracted from the images using various image processing algorithms such as scale invariant feature transformation (SIFT) (Lowe, 1999), local binary patterns (LBP) (Ojala, et al., 1996), and histogram of oriented gradients (HOG) (Dalal & Triggs, 2005). The second approach leverages deep learning which eliminates the need for manually engineered features and instead learns to identify relevant features from the data by itself. In computer vision, deep learning has been largely dominated by convolutional neural networks (CNNs) (LeCun, et al., 1998; LeCun, et al., 2015), however, lately vision transformers (ViT) (Dosovitskiy, et al., 2020), have shown a great potential and emerged as a worthy competitor (Deininger, et al., 2022). Generative methods such as generative adversarial network (GAN) (Goodfellow, et al., 2014) and diffusion models (Sohl-Dickstein, et al., 2015; Ho, et al. 2020) also fall in the deep learning category; they are very effective for image synthesis.

The early work in computational pathology was based on traditional machine learning and feature engineering. (Gurcan, et al., 2009; Tabesh, et al., 2007; Doyle, et al. 2008) However, after the landmark computer vision study by Krizhevsky, A. et al. in 2012, the momentum of the domain increasingly started shifting towards deep learning. Following the deep learning revolution, initial investigations primarily focused on highlighting the diagnostic prowess of deep convolutional neural networks on tasks such as cancer classification and grading (Cireşan, et al., 2013; Litjens, et al., 2016; Wang, et al., 2016). These foundational studies paved the

way for subsequent research to scale these methodologies and conduct clinical validation with significantly larger cohorts (Esteva, et al., 2017; Campanella, et al., 2019; Ström, et al., 2020). In prognostics, deep learning has been applied to tissue images to study patient survival, potential prognostic markers and tumor micro-environment (Lu, et al., 2018; Wang, et al., 2020; Schmauch, et al., 2020). As deep learning became more prevalent and compute resources more accessible, a trend of online challenges with open dataset started to emerge. Similarly, in computational pathology online challenges focusing on key problem areas have worked as a catalyst in advancing the field by providing public access to valuable WSI datasets (Bejnordi, et al., 2017; Veta, et al., 2019; Bulten, et al., 2022).

While diagnostic and prognostic applications are the mainstay in computational pathology because of their numerous direct and indirect benefits, there's also value in investigating workflow optimization and enhancement. Artifact detection and removal, stain normalization, virtual histopathology staining and focus quality assessment are some of the examples of workflow optimization that, although investigated by some studies, remain relatively unexplored areas. Deep learning demonstrates significant potential for enhancing these areas as well. This study primarily focuses on virtual histopathology staining and stain normalization, which are discussed in detail in section 2.4 and 2.5 respectively.

2.3 Image-to-image Translation

Image-to-image translation is a technique in computer vision wherein a method learns a transformation from a source domain to a target domain, mostly, although not strictly, by preserving selective source structural characteristics and transforming stylistic characteristics to match that of the target domain. (Pang, et al., 2021) It can also be understood broadly as an image transformation technique that can accentuate the distinctive features and/or suppress unwanted elements of the input image, thereby producing an output that provides more valuable and meaningful information. The technique is used in a range of applications in different domain including remote sensing (satellite images to maps (Ingale, et al., 2021), land cover classification (Yang, et al., 2021)), medical imaging (MR to CT synthesis (Han, 2017), image denoising (Yang, et al., 2018), semantic segmentation (Guo, et al., 2019)), media (photo colorization (Cheng, et al., 2015), image deblurring (Zhang, et al., 2022), style transfer (Liu, et al., 2019), sketches to photorealistic images (Koley, et al., 2023)), agriculture (crop health monitoring (Sahu, et al., 2023) , field segmentation (Mei, et al., 2022)), autonomous driving (semantic segmentation (Papadeas, et al., 2021), depth estimation (Khan, et al., 2022)), and urban planning (spatial analysis (Sun, et al., 2023), infrastructure mapping (Mattyus, et al., 2015)).

Image-to-image translation has been dominated by deep learning-based methods more recently; however, the technique itself is much older than deep learning and has traces of its evolution across the last few decades (Pérez, et al., 2023; Ashikhmin, 2001; De Bonet, 1997). Initially, translation methods were based on manually crafted features, pixel transformations or rigid mathematical frameworks meant for a specific type of

translation which also required domain knowledge. Due to the underlying rigidity and specificity, the methods weren't suitable for complex and precise transformations. Deep learning catapulted the development of robust, precise, scalable and more generic image-to-image translation methods without the requirement of underlying domain knowledge. These models are categorized as the so-called generative models (Xu, et al., 2015; Oussidi & Elhassouny, 2018) that are capable of generating data as opposed to only classifying it (Pang, et al., 2021). Models such as U-Net (Ronneberger, et al., 2015) and fully convolutional network (Long, et al., 2015) used for medical and natural image segmentation, respectively, gained prominence as first-of-kind deep learning based image-to-image translation methods. Their architectural approach also serves as the backbone of more sophisticated image-to-image translation methods such as variants of GANs (Goodfellow, et al., 2014) and diffusion models (Sohl-Dickstein, et al., 2015; Ho, et al. 2020). GAN represents a family of methods primarily associated with image synthesis and translation. GAN models are the central focus of this thesis and will be discussed in detail in the next section.

2.3.1 Generative Adversarial Network (GAN)

GAN is essentially an approach that leverages adversarial training for image synthesis. Typically, a GAN consists of two entities, a generator model G responsible for capturing the underlying data distribution to generate synthetic images from the same distribution and a discriminator model responsible for classifying the images as synthetic (generator-produced) or real (from the training set). Both models are trained in unison until an equilibrium state is reached where the realism of the synthetic images cannot be improved further, and the discriminator cannot distinguish synthetic images from the real ones. (Goodfellow, et al., 2014) This is the fundamental training protocol of GAN models which becomes more complex as the objective changes from image synthesis (from latent space) to different types of image-to-image translation.

There are different types of GAN-based image-to-image translation methods, they follow the same learning paradigms commonly found in machine learning such as supervised, unsupervised, semi-supervised, few-shots, and one-shot learning. However, supervised and unsupervised learning methods remain the most used. Furthermore, they can also be categorized domain-wise as two-domain and multi-domain image-to-image translation models. (Pang, et al., 2021)

2.4 Virtual Histopathology Staining

Virtual histopathology staining of unstained tissue can best be explained by first drawing parallels to conventional staining and analog photography. In analog photography, film negatives are developed into colored images in a darkroom through chemical processing. Similarly, in conventional histopathology staining, tissue specimens are processed and stained in a laboratory using chemical stains to reveal detailed microscopic structures. In both scenarios, the input undergoes

chemical processing to uncover critical information. Photography transitioned to the digital medium almost completely in the first decade of the twenty first century. Although the field of histopathology has been gradually embracing digitalization, the fundamentals of the staining process have largely remained unchanged. Virtual staining aims to revolutionize the process by digitally staining unstained tissue images (Bai, et al., 2023). **Figure 4** shows a visual representation of both conventional and AI-driven virtual histopathology workflows. Virtual staining also refers to transforming images of tissues stained with one type of dye into images that appear as if stained with another dye, however, that is beyond the scope of this thesis. While staining is mostly considered means to an end, improving it through virtualization holds great potential and merits significant attention.

Figure 4. High-level visual representation of histopathology workflow with both conventional and AI-driven virtual staining paths. After tissue extraction and processing, the unstained tissue sections are directly imaged and virtually stained in the AI-driven path, whereas in the conventional path, the unstained tissue sections are first stained and then imaged before they are ready for pathologist's evaluation.

Virtual staining has several benefits. For instance, biopsy results typically take 1-2 days, however, in complicated cases that require thorough analysis, second opinion, multiple stains or other advanced tests the results could take 2 to 3 weeks (Waiting for Your Biopsy or Cytology Test Results, 2023). With virtual staining, multiple stains can be applied to the tissue images within minutes which can potentially reduce the biopsy turnaround time. Given the thin and delicate nature of the tissue sections, they are susceptible to degradation and can be damaged with

extensive processing. Virtual staining as a solution minimizes the repeated use of physical slides after initial processing. If tissue slides are damaged during staining, they can be reconstructed or restained using the original unstained tissue image through virtual staining which reduces the risk of sample loss.

An overlooked drawback of conventional staining is its substantial water consumption. It is estimated that, globally, these processes collectively consume over one million liters of water every year. (Bai, et al., 2023) The process also produces significant chemical waste, and some special stains contain chemicals that are known to be harmful to health and the environment. For example, some special stains containing heavy metals such as arsenic and mercury are carcinogenic and genotoxic, however, they are often key to studying specific anatomical target[1]. (Bordoloi, et al., 2017) Virtual staining can help to reduce the need for using hazardous stains and, in turn, reduce water consumption as well making the process more sustainable. Furthermore, some special stains, such as silver nitrate-based stains, are quite expensive (Kumar & Kiernan, 2010); virtual staining is a scalable technology that has the potential to significantly reduce cost by providing digital alternatives to expensive stains.

It is worth mentioning here that virtual staining is not the only methodology aiming to improve the histopathology workflow. In the past couple of decades, various attempts have been made to modify the current histopathology workflow, mainly by using a combination of imaging devices that provide non-standard contrast and computation methods to accentuate the final result or parts thereof. Some of these methods ambitiously aim to streamline and simplify the assessment process by providing a solution for tumor detection and demarcation during surgical operations, consequently, eliminating the need for conventional histopathology steps such as fixation and chemical staining. This idea is dubbed as slide-free histopathology. (Rivenson, et al., 2020)

Most of the virtual staining methods are based on GANs. The body of work on virtual staining of unstained tissue images covers various (organs) tissue specimens, stain types, and label-free input image modalities, which has substantially enriched the application area of this family of methods. It can be roughly divided into studies that primarily concentrated on establishing the proof of concept by focusing on the quantitative evaluation of virtually stained images and studies that moved the needle to bolster the validation process by assessing the fidelity of virtually stained images against chemically stained ones on downstream tasks such as tumor detection or cancer grade classification.

One of the first studies to use GAN-based (Isola, et al., 2017) virtual staining was by Bayramoglu et al. in which they attempted to digitally produce H&E-stained lung tissue images through hyperspectral unstained input images (Bayramoglu, et al., 2017). Rivenson et al. developed a conditional GAN-based virtual staining method called PhaseStain. The method was trained to digitally reproduce the H&E-

[1] One example is the Golgi-Cox stain, which is essential for studying neuronal morphology and contains mercuric chloride.

, Jones'-, and Masson's trichrome-stained variants from the quantitative phase input images of human skin, kidney, and liver tissue, respectively. (Rivenson, et al., 2019) Borhani et al. used a deep learning approach based on pixel-to-pixel and area-to-pixel mapping and a combination of two-photon excitation fluorescence and fluorescence lifetime imaging as the model input to virtually stain preclinical samples with H&E. (Borhani, et al., 2019) Zhang et al. used pix2pix, a variant of conditional GAN, with brightfield label-free tissue images of murine carotid artery as the model input to synthesize its hematoxylin and eosin-, picrosirius red-, and Verhoeff van Gieson-stained counterparts. (Zhang, et al., 2022)

To establish clinical relevance of virtual staining methods, it is imperative to validate the resulting images on downstream tasks, to that end, Rana et al. proposed a conditional GAN-based virtual H&E staining method for the staining and destaining of prostate (core biopsy) tissue images. The resulting virtually stained images were compared against the original chemically stained images on tumor segmentation task. (Rana, et al., 2020) Picon et al. used a Siamese cycled network that learns a direct and inverse transformation between label-free autofluorescence and H&E-stained tissue images of colon, breast, and lung. They further validated virtually stained colon samples for colorectal cancer prediction against the chemically stained ground truth tissue images. (Picon, et al., 2021) Some studies even presented novel approaches such as multiplex staining, stain-in-stain feature, and virtual staining-enabled real-time fertility assessment. (Zhang, et al., 2020; Nygate, et al., 2020) Although stain-to-stain transformation is beyond the scope of this thesis, it is another similar application area that has been disrupted by deep learning and GAN-based methods. (Gadermayr, et al., 2018; de Haan, et al., 2021; Chen, et al., 2021; Lin, et al., 2022) While GAN-based methods have traditionally dominated the virtual staining space, recent studies have begun to demonstrate the effectiveness of diffusion models for the task (Kataria, et al., 2024; Zhang, et al., 2024a; Zhang, et al., 2024b)

2.5 Stain Normalization

Normalization is the process of transforming measured data from varying scales to a common scale to ensure a comparable and consistent understanding of the data. In histopathology, the process of adjusting the appearance of tissue images stained with the same dye to minimize heterogeneity of color, regardless of the source thereof, is called stain normalization. (Hoque, et al., 2024) There are various factors that contribute to the non-uniform appearance of the stained tissue images, including variations in staining protocols, stain manufacturers, imaging hardware, and post-processing techniques (Weinstein, et al., 2009; Ji, et al., 2025). The lack of standardization poses a non-trivial challenge to large-scale use of AI-based diagnostic tools as it hinders the generalization capability of algorithms powering such tools. Stain normalization has been an effective approach to curb the lack of standardization in the appearance of stained tissue images. It not only helps AI-based

diagnostic tools generalize better on downstream tasks but also allow pathologists and researchers to analyze tissue samples accurately and reliably.

Stain normalization methods are mainly divided into traditional and deep learning-based methods. Traditional methods are based on customized mathematical frameworks that perform transformations in different color spaces such as RGB and the $l\alpha\beta$ color space (ICC, 2004), whereas deep learning-based methods mostly use different forms of GAN-based image-to-image translation to perform stain normalization. Typically, stain normalization is performed between a source or multiple source tissue datasets and a target dataset. Based on representativeness, first, a reference tissue sample is selected from the target dataset then the tissue samples in source dataset(s) are normalized with respect to the reference samples. In the case of deep learning methods, it is possible to select multiple reference tissue samples in order to capture all possible appearance variations of the target dataset. Although traditional methods can also learn from multiple reference samples in theory, their design restricts them from learning in a scalable way. A general stain normalization workflow can be seen in **Figure 5**.

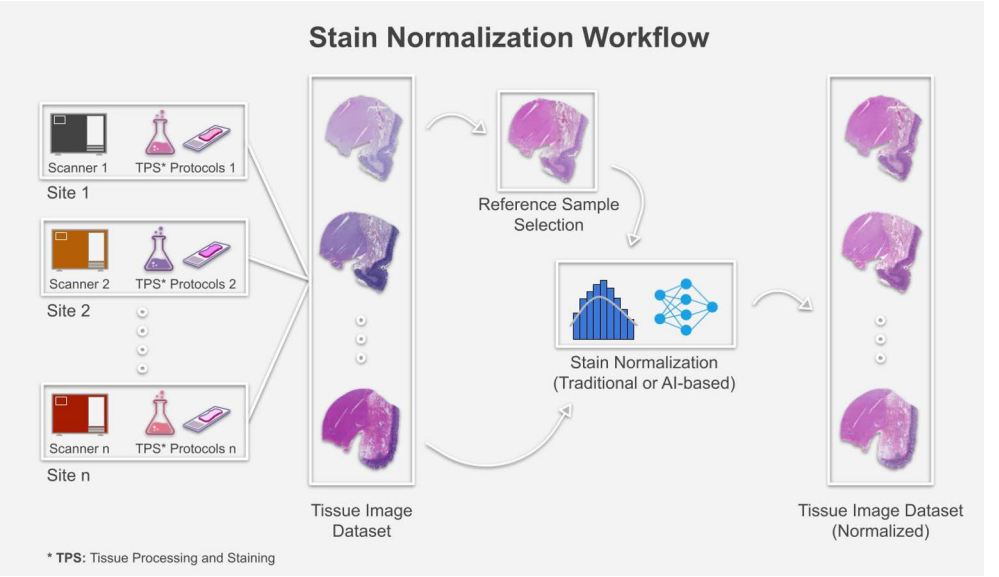

Figure 5. High-level stain normalization workflow that shows tissue image dataset collected from different sites and the various factors that contribute to the heterogeneity of stained tissue, followed by reference sample selection and stain normalization process using a traditional or AI-based method, and finally, a normalized dataset with a uniform appearance of the stained tissue images.

Color quantification and normalization of stained tissue images started in the early 2000s (Ruifrok & Johnston, 2001; Reinhard, et al., 2001). Since then, several different methods have been proposed for stain normalization - initially the aforementioned traditional methods and after the advent of deep learning mostly

GAN-based methods. Some of the most commonly used traditional stain normalization methods, which have been studied in this thesis work, are Reinhard (Reinhard, et al., 2001), Macenko (Macenko, et al., 2009), Vahadane (Vahadane, et al., 2016), and the color histogram matching (Gonzalez & Woods, 2008). GAN-based stain normalization methods, however, have gained more prominence in the recent past. Similar to virtual staining, these methods are primarily driven by CycleGAN (Zhu, et al., 2017) or conditional GAN (Isola, et al., 2017) variants. de Bel et al. proposed a CycleGAN-based stain normalization method using multi-center tissue image dataset to improve the downstream segmentation of renal histopathology. (de Bel, et al., 2018) In a subsequent study, they modified the CycleGAN generator by introducing residual learning and showed that the modified CycleGAN-based (Zhu, et al., 2017) stain normalization outperforms its predecessor method on morphological segmentation of colon and kidney tissues samples. (de Bel, et al., 2021) Similarly, conditional GAN (Isola, et al., 2017) a supervised-learning method, has also been used for the task of stain normalization using grayscale and RGB versions of the same tissue images as training pairs because of their inherent alignment (Shaban, et al., 2019; Salehi & Chalechale, 2020). Cong et al. adapt conditional GAN to develop a method dubbed as color adaptive GAN (CAGAN) that uses a combination of supervised and unsupervised learning for stain normalization. (Cong, et al., 2022) Several other studies have proposed GAN-based stain normalization methods with different architectural, loss function, and training approach modifications (Cho, et al., 2017; Liang, et al., 2020; BenTaieb & Hamarneh, 2017; Altini, et al., 2023; Cai, et al., 2019; Mahapatra, et al., 2020) which further underscores the potential of GAN-based image-to-image translation in histopathology.

3 Aims

The main objectives of this thesis are:

1. Explore optimal tissue processing and brightfield imaging protocols for the development of virtual histopathology staining algorithms.
2. Explore the effects of network architecture of image-to-image translation models on the quality virtual histopathology staining.
3. Benchmark various image-to-image translation methods for histological stain normalization using multi-center data.

4 Materials and Methods

4.1 Data Collection

4.1.1 Virtual Staining Data

In Publication **I**, we studied the optimal tissue and imaging protocols for virtual staining. We prepared adjacent tissue sections with varying specifications both in terms of processing protocols and conditions prior to imaging. In this study, CycleGAN (Zhu, et al., 2017) was chosen for virtual staining because of its ability to handle unaligned images. Tissue sections from mouse anterior prostate, liver, kidney, and testis were analyzed in this study. Two different tissue WSI datasets were prepared. For the first dataset, anterior prostate lobe tissues were fixed in PAXgene tissue fixative (PreAnalytiX GmbH, Hombrechtikon, Switzerland), then sectioned into samples of five different thicknesses: 3, 5, 8, 12, and 20µm. Adjacent sections of each thickness were processed using four distinct methods: (1) no further processing after attachment to slides (unprocessed), (2) paraffin removal followed by rehydration (deparaffinized), (3) paraffin removal, rehydration, dehydration, and coverslip mounting (coverslipped), and (4) standard H&E staining (H&E reference), which encompasses all aforementioned steps (**Figure 6**). The second dataset consisted of liver, kidney, and testis tissues which were formalin-fixed and then cut into 5µm thick sections. These tissue sections were imaged once before staining (deparaffinized and without coverslip) and once after being H&E stained to ensure content correspondence. Imaging was performed using Thunder Imager 3D Tissue slide scanner (Leica Microsystems) equipped with DMC2900 camera (Leica Microsystems) and HC PL APO 40x/0.95 DRY and HC PL FLUOTAR 20x/0.5 DRY objectives (Leica Microsystems). WSIs were stitched together with LAS X (ver 3.7.4.; Leica Microsystems). The pixel size of the acquired WSIs was 0.353µm for 40×, and 0.671µm for 20×, and the images were stored as tiff files.

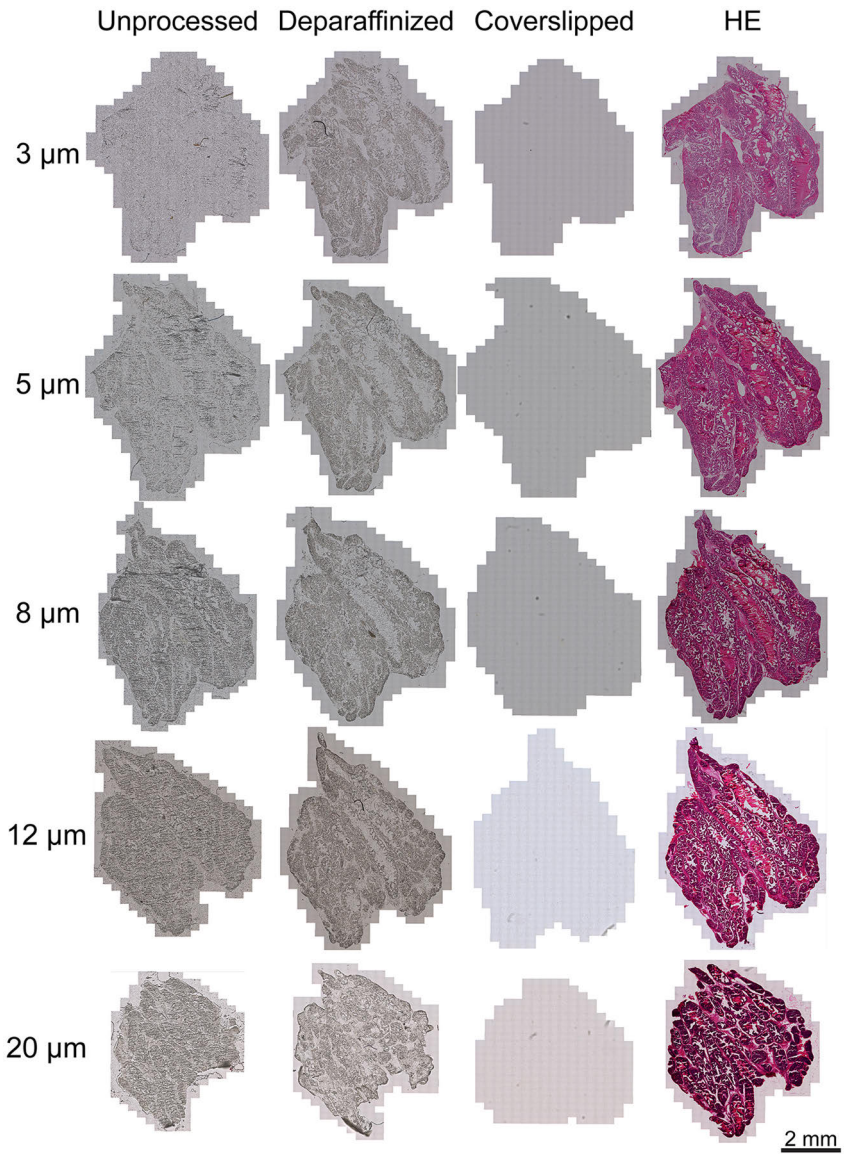

Figure 6. Tissue content of WSIs imaged for each section-processing condition. WSIs obtained using brightfield microscopy for traditionally, chemically stained H&E sections and unstained tissue sections processed differently. The unprocessed and deparaffinized sections provide a well-contrasted signal. The tissue imaged through mounting medium and coverslip (coverslipped) exhibits a practically transparent overall appearance to the human eye, accompanied by the appearance of a grid pattern according to the slide scanner fields of imaging. H&E-stained images are for reference.

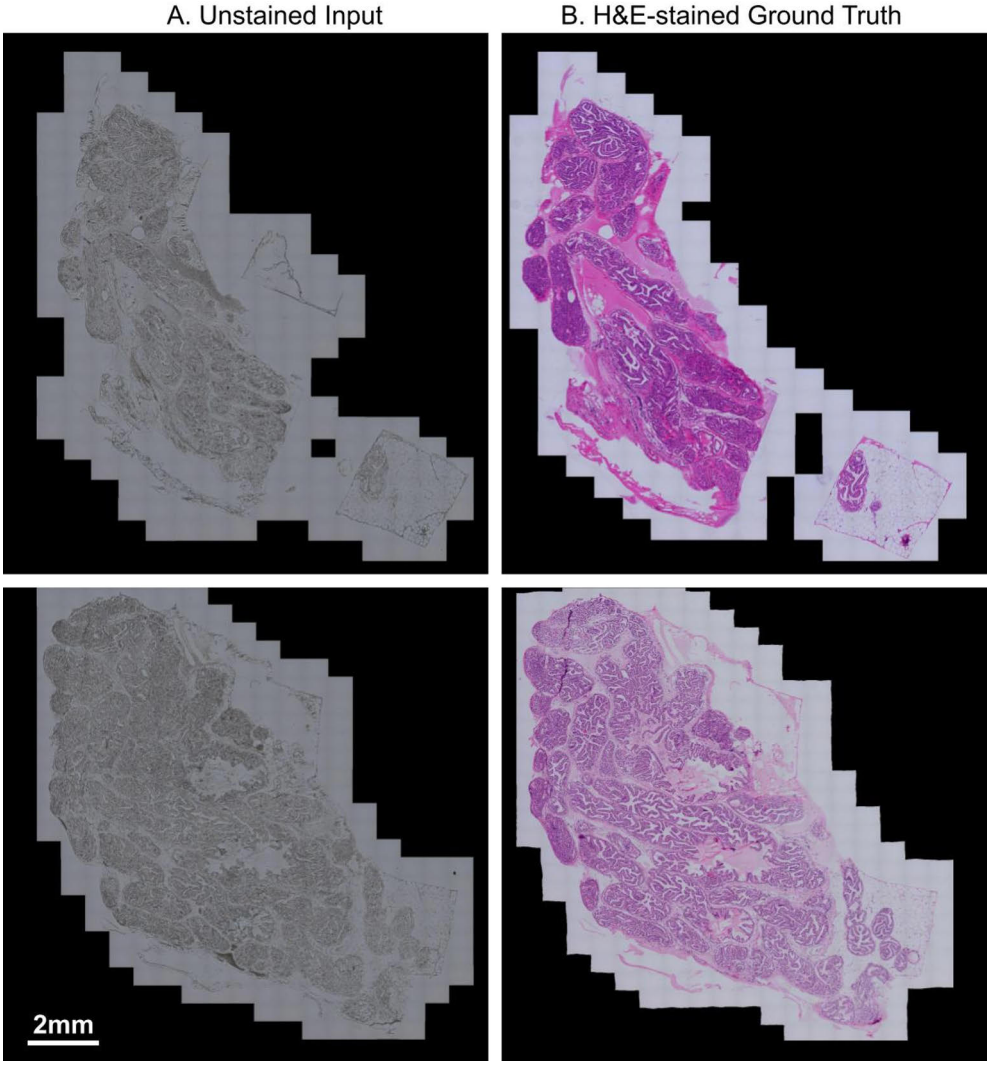

Figure 7. Samples images of unstained and H&E-stained tissue sections. First, unstained tissue sections were imaged without coverslip followed by H&E staining and imaging of the same tissue sections.

In Publication **II**, we studied the impact of neural network architecture on the quality of virtual staining. We used pix2pix (Isola, et al., 2017) and tweaked it to create two more variants, in addition to the baseline model, to compare their performance with each other on the task of virtual staining. An important aspect of this study was the use of brightfield imaging for tissue scanning which is a widely used modality and easily available in laboratories all around the world. Again, two different datasets were prepared in this study. The first dataset consisted of prostate tissues (**Figure 7**) fixed in PAXgeneTM, and the second one consisted of kidney,

liver, spleen, testis, epididymis, and seminal vesicle tissues fixed in formalin. Based on the learning of the Publication **I**, the section thickness was set to 5μm. The first dataset contained 81 WSI pairs of unstained and H&E-stained histological sections of anterior prostate tissue, and one WSI pair of unstained and H&E-stained tissues per organ for the second dataset. Tissue imaging was done using Thunder Imager 3D Tissue slide scanner (Leica Microsystems, Wetzlar, Germany) equipped with DMC2900 camera and HC PL APO 40x/0.95 DRY objective with a pixel resolution of 0.353 μm.

4.1.2 Stain Normalization Data

In Publication III, a unique tissue image dataset, encompassing 66 different laboratories, was collected to understand the breadth of variation caused by differences in tissue processing and staining protocols. The H&E-stained tissue image dataset was obtained as part of an external quality assessment (EQA) initiative coordinated by Labquality, a firm specializing in external quality assessment programs for clinical laboratories located in Helsinki, Finland. These slides featured a tissue microarray section comprising three 6mm punch biopsies, each extracted from normal human skin, kidney, and colon tissue specimens. These tissue samples were obtained from anonymized, formalin-fixed, and paraffin-embedded histological specimens from a reference pathology laboratory. During the EQA round, slides with 3-micron unstained tissue sections were dispatched to EQA participant laboratories. These laboratories were instructed to apply their routine H&E staining methodology, typically used in their daily diagnostic practices. **Figure 8** shows the variation in staining appearance across the EQA participant laboratories. In total, 66 laboratories from 11 different countries participated in the evaluation process. Subsequently, these slides were digitized using a Hamamatsu Photonics NanoZoomer-XR slide scanner, employing a 20× objective lens, with a scanning resolution of 0.46μm per pixel. For stain normalization experiments, the whole slide images (WSI) were resampled to 10×.

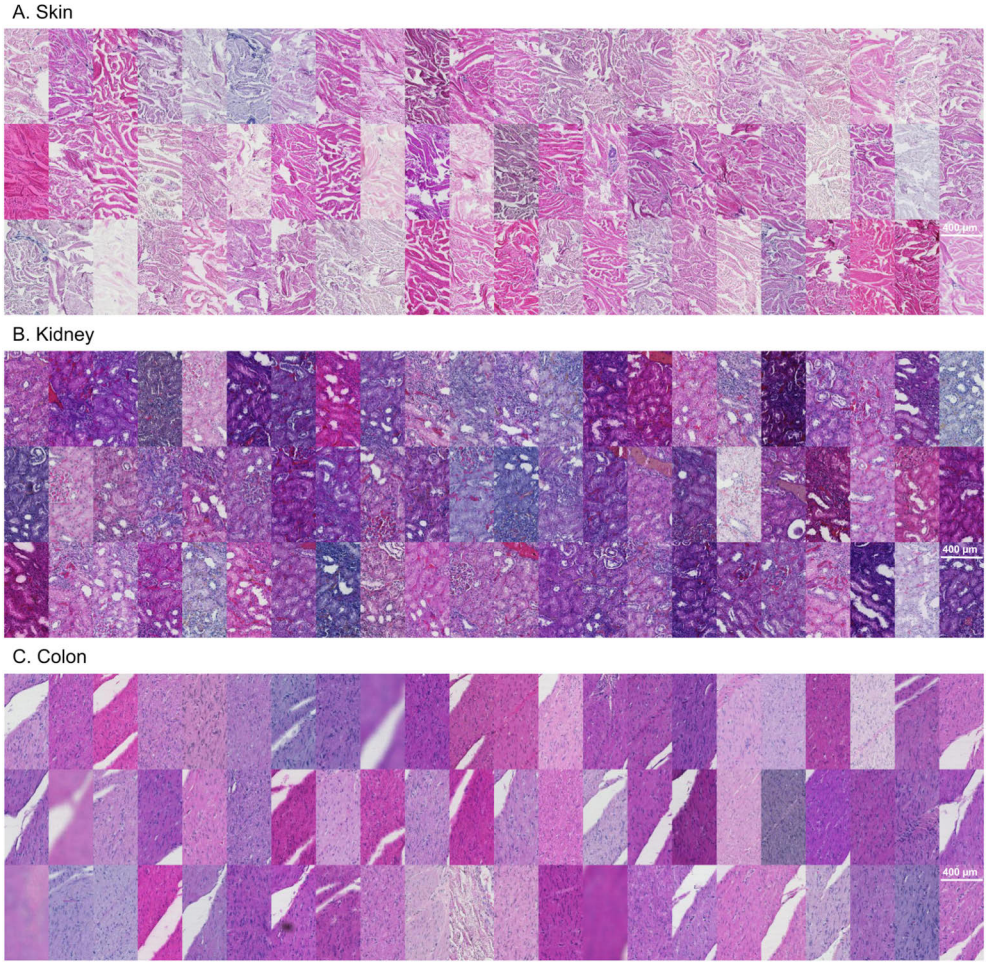

Figure 8. A collection of tissue patches illustrating variation in appearance. **A**) Patches extracted from skin tissue sections. **B**) Patches extracted from kidney tissue sections. **C**) Patches extracted from colon tissue sections.

4.2 Computational Methods

In this thesis, for virtual staining (Publication **I** and **II**) GAN-based methods were used, whereas for stain normalization (Publication **III**), GAN-based and traditional stain normalization methods were used.

4.2.1 GAN-based Image-to-image Translation

A range of different GAN-based methods were tested, first, on the virtual staining task and then on stain normalization. This thesis work is primarily based on variants of CycleGAN (Zhu, et al., 2017), an unsupervised image-to-image translation

method and pix2pix (Isola, et al., 2017), a supervised image-to-image translation method. Other GAN-based methods were also explored; they are mentioned in section 0.

4.2.1.1 CycleGAN

CycleGAN is an unsupervised image-to-image translation method that employs the cyclical training approach A-to-B-to-A. It consists of two generators G_1 and G_2, responsible for A-to-B and B-to-A domain transformations, respectively, and two discriminators D_1 and D_2 that guide quality of transformed images (B_{gen} and A_{gen}), produced by their respective generators (**Figure 9**). The transformed or synthesized images that deviate from the original target domain are penalized more by the discriminators, in turn, the generators are nudged to learn a more accurate image mapping; this forms the foundation of adversarial training. Further, the transformed images B_{gen} and A_{gen} are fed to the opposite generators, which produces A_{rep} and $B_{rep,}$ that are then compared with the original input images A_{org} and B_{org} in a cyclic manner using the cycle-consistency loss, which aims to minimize the difference between *org* and *rep* images (**Figure 9**). CycleGAN doesn't require registered image pairs for training.

Two different generator architectures were tested for CycleGAN. A ResNet-based (He, et al., 2016) generator and a U-Net-based (Ronneberger, et al., 2015) generator. In Publication I, ResNet-based generators were used, whereas in Publication **II** and **III**, both approaches were used. Furthermore, in Publication **II** a CycleGAN-based semi-supervised approach was also explored. The idea was to feed a small percentage of aligned or registered image pairs and use a direct pixel-loss function like mean absolute error (MAE) or mean squared error (MSE) instead of cycle-consistency loss for the registered batches. This approach was tested with different percentages and batch order of supervised data; however, it failed to improve the quality of virtually staining.

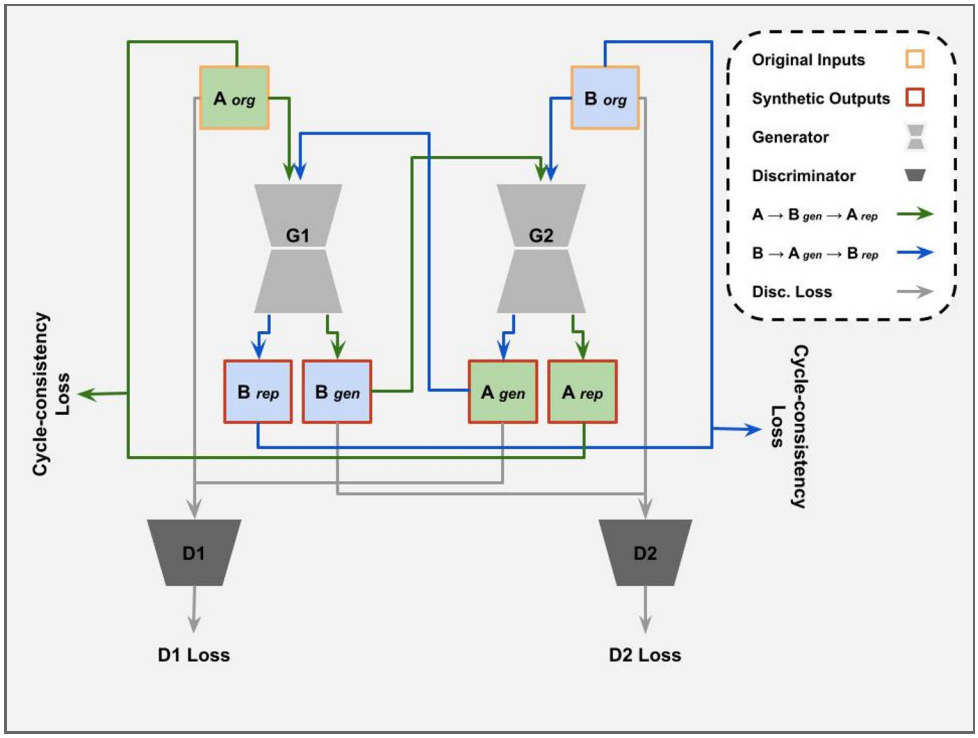

Figure 9. High-level architectural representation of the CycleGAN model is depicted in the figure. The network comprises two generators, G_1 and G_2, along with two discriminators, D_1 and D_2. A_{org} and B_{org} denote the original input images from the two domains. A_{gen} and B_{gen} represent the synthetically generated output images from the respective original counterparts. A_{rep} and B_{rep} denote the repeat synthetically generated images using previously generated B_{gen} and A_{gen}, respectively. The discriminator losses, D_1 loss and D_2 loss, are computed using the original input and the generated output. The cycle consistency loss is computed using the original input and the repeat generated output.

4.2.1.2 Pix2pix

Pix2pix or a conditional GAN (Isola, et al., 2017) is a type of supervised image-to-image translation method. It consists of a generator and a discriminator model (**Figure 10**). As opposed to a standard GAN that maps a latent vector to an image in a target domain, a conditional GAN is provided with a condition in the form of a label or any structured data. In case of pix2pix, the network is trained to learn a mapping between images of a source domain to the target domain and source images are model input during inference. In an adversarial manner, the discriminator is trained to penalize synthesized images that deviate from the target domain in appearance, this guides the generator to transform images to match target domain distribution as closely as possible. Unlike CycleGAN, pix2pix requires image pairs to be registered prior to training. In addition to adversarial loss, pix2pix also employs an additional L1 loss term for the generator network by leveraging aligned data, thus

producing better-quality translated images as compared to unsupervised learning methods.

In Publication **I**, a baseline variant of pix2pix was used that consists of a U-Net-based generator (Ronneberger, et al., 2015) and a PatchGAN-based discriminator (Ronneberger, et al., 2015) (**Figure 11**). In Publication **II**, pix2pix was modified to construct two more variants (**Figure 12**). In the first variant, convolution layers were doubled at each level of both the encoder and decoder of the generator. The same modification in the discriminator did not yield positive results as the discriminator began to dominate early on disturbing the training equilibrium which forced the generator to produce suboptimal images. In the second variant, convolution layers were replaced with a more sophisticated unit called dense convolutions inspired by dense U-Net (Dong, et al., 2019) a network designed for the segmentation of small objects in satellite imaging (**Figure 12**).

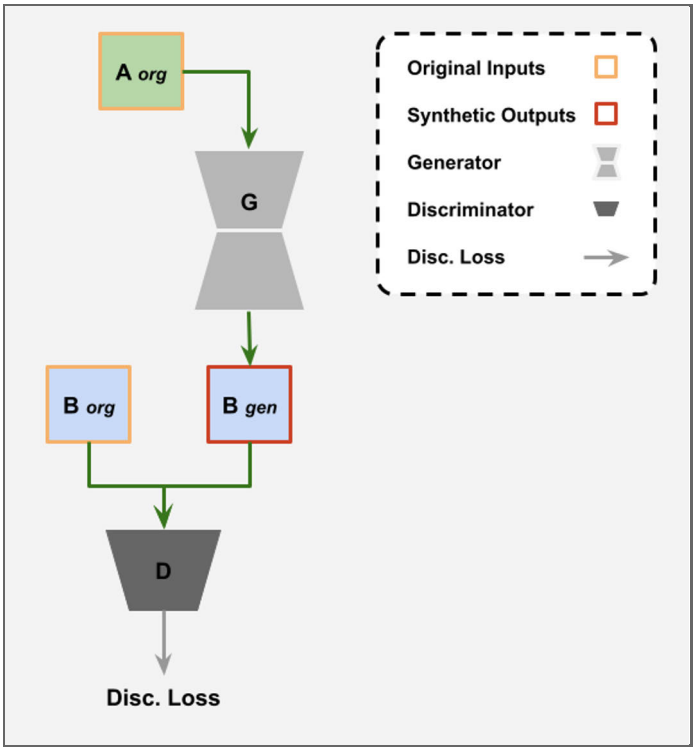

Figure 10. Schematic representation of the pix2pix network. The network has a generator G and a discriminator D. A_{org}, B_{org}, and B_{gen} represent the original input image, the registered ground truth image, and the synthetically generated output, respectively. The network uses a combination of GAN loss and L1 loss for the generator output.

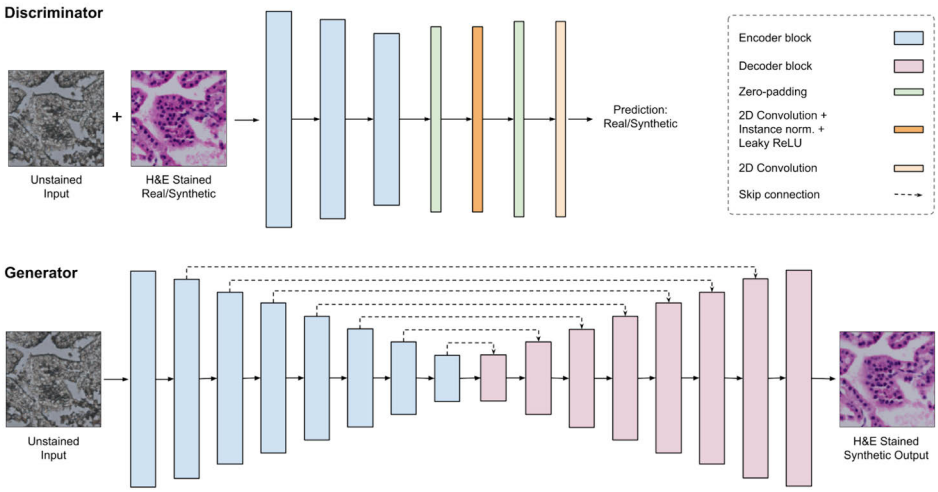

Figure 11. Pix2pix model architecture. The PatchGAN discriminator comprises three encoding blocks, the first one without instance normalization, followed by a zero-padding layer, a 2D convolution layer, instance normalization, a leaky rectified linear unit activation, another zero-padding layer, and finally a 2D convolution layer. The U-net generator consists of eight encoding and eight decoding blocks. Each encoder block further consists of a 2D convolution layer with a stride size of two, followed by instance normalization and leaky rectified linear unit activation. Each decoder block consists of a 2D transposed convolution, with a stride size of two, followed by instance normalization, and leaky rectified linear unit activation. Skip connections were used in the generator, which means that the output of each encoding block is concatenated with the output of the corresponding decoding block excluding the first encoding and last decoding block.

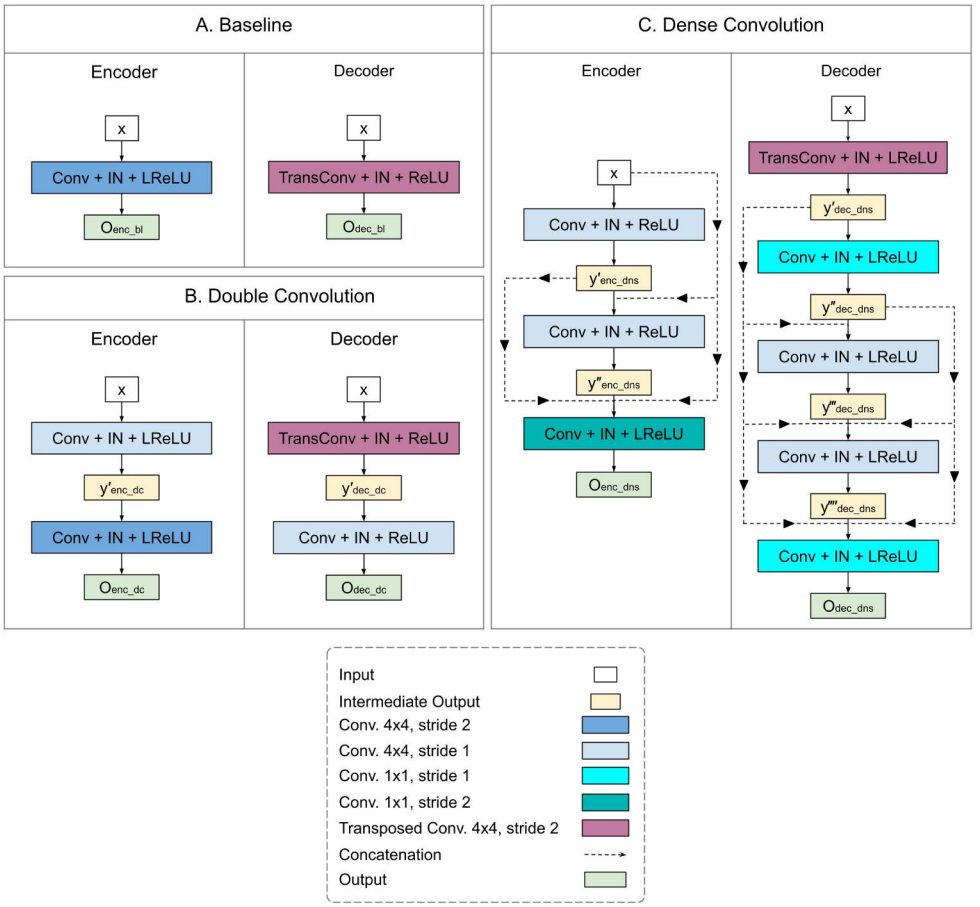

Figure 12. **A**) Baseline: the reference pix2pix implementation with single convolution for each encoder and decoder block. **B**) Double convolution: additional convolutional layer for each encoder and decoder block. **C**) Dense convolution: a more complex unit inspired by DenseU-net for each encoder and decoder block.

4.2.1.3 Other GAN-based Methods

In Publication **II**, other GAN-based methods were also explored. This exploration was driven by the motivation to search for a method that could help to reduce or possibly eliminate the load and potential source of errors pertaining to preprocessing steps. The key factors taken into consideration were the possibility of bypassing image registration and minimizing the quantity of data required for training.

Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation (UGATIT) (Kim, et al., 2019) a GAN-based unsupervised image-to-image translation method was explored. It is a CycleGAN-inspired method that incorporates the concepts of attention feature maps to focus on important regions in the source and target image and hybrid

normalization layers with learnable parameters derived from the attention feature maps. Meta translation GAN (MT-GAN) (Lin, et al., 2020), a few-shots learning method was also used. It is also an unsupervised image-to-image translation method that presents a solution to tackle the model disposability, i.e., a model once trained on a specific set of data or domains cannot be easily adapted to new domains. TuiGAN (Lin, et al., 2020), a one-shot learning method was ambitiously put to test for the task of virtual staining. The method uses a single source-target image pair and attempts to learn a mapping in a coarse-to-fine manner where the generated image is constructed by first rendering high-level features and then iteratively adding fine-grained details. Unfortunately, all of the aforementioned attempts resulted in suboptimal virtual staining of unstained tissue images.

A novel approach, constructing the generator of a pix2pix model by combining U-Net (Ronneberger, et al., 2015) and vision transformer (Dosovitskiy, et al., 2020), dubbed U-Net Transformer Hybrid GAN (UTHGAN), was also employed for the virtual staining task. The approach was inspired by TransUNet, a vision transformer- and U-Net-based method developed for medical image segmentation (Chen, et al., 2021). The approach showed potential initially (**Figure 13**), however, transformers are notoriously time- and resource-consuming to train. Therefore, it was deemed appropriate to proceed with more established image-to-image translation methods. The work on UTHGAN is not a part of any of the publications and has not been independently published yet.

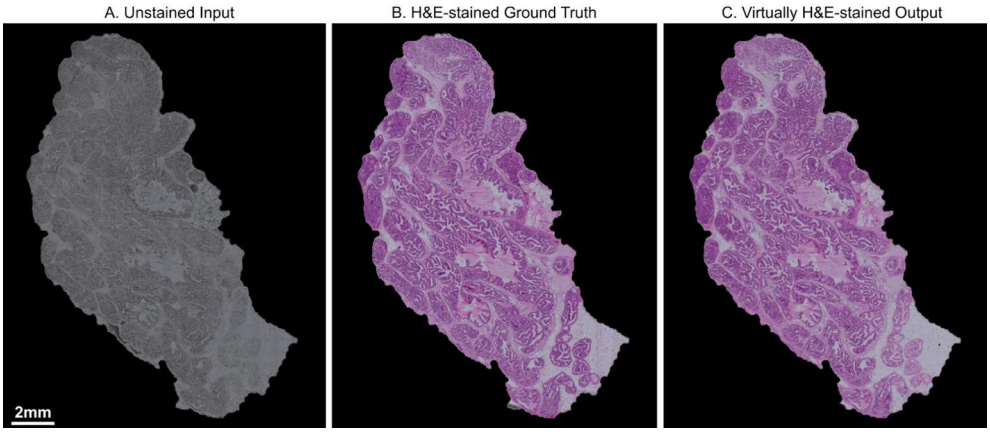

Figure 13. Virtual staining using UTHGAN. **A**) Unstained tissue image as the model input. **B**) H&E-stained ground truth image from the test set. **C**) Virtually H&E-stained image generated by UTHGAN.

4.2.2 Traditional Stain Normalization Methods

Typically, traditional normalization is applied at the tile level, so is reference sample selection. The drawback of this approach is that the infinitesimally small amount of tissue content variation preserved in a small tile as compared to the whole tissue section. Therefore, in Publication **III**, we decided to use the whole tissue image as the reference

sample and also applied stain normalization at WSI-level to alleviate tiling artifacts that usually appear when normalization is applied at tile-level and the tiles are stitched as a post processing step, especially in the case of traditional stain normalization methods. Four different traditional stain normalization methods were used in Publication **III**.

4.2.2.1 Histogram Matching

Histogram matching is an image processing technique that adjusts pixel values in an image to alter its color distribution, represented by its histogram, to align it with the color distribution of a reference image (Gonzalez & Woods, 2008). By aligning the color distribution of the source image with the reference image, this process aims to achieve color consistency between the two images, facilitating effective image analysis and visualization under varying imaging conditions. Although histogram matching is a generic color normalization technique not specifically designed for histopathology images, it has been notably applied in this domain. Given the experimental setup wherein samples stained in different laboratories originate from the same tissue block and thus share highly similar morphology, histogram matching is well-suited method for this type of stain normalization.

4.2.2.2 Macenko

In Macenko stain normalization, distinct Hematoxylin and Eosin (H&E) vectors are computed using singular value decomposition (SVD) in the optical density (OD) space, which is a logarithmic transformation of the RGB color space. The OD space facilitates easier manipulation of stains by enabling linear coupling and decoupling of stain components. By scaling the stain vectors of the source image to match that of a reference image, the Macenko method standardizes staining variations, ensuring consistent color representation across images (Macenko, et al., 2009).

4.2.2.3 Reinhard

Reinhard normalization (Reinhard, et al., 2001), like histogram matching, is a generic image normalization technique not specifically tailored for histopathology images. In this method, images are first converted to the $l\alpha\beta$ color space (ICC, 2004) then statistical standardization is applied to the source images (subtracting the mean value and dividing by the standard deviation). Subsequently, the images are linearly scaled using the standard deviation and mean of the reference image, and this process performed separately for each channel. This normalization is effective because, unlike RGB, the $l\alpha\beta$ color space exhibits low correlation between color channels.

4.2.2.4 Vahadane

Another stain normalization technique employs color deconvolution in the OD space, but instead of SVD, it utilizes sparse non-negative matrix factorization (SNMF), which

has proven more effective. This method optimizes the deconvolution process over successive iterations to generate two matrices: one representing color appearance and the other representing stain density. The source stain density is scaled, similar to Macenko's approach, and then combined with the color appearance of the reference image to reconstruct a normalized source image. (Vahadane, et al., 2016)

4.2.3 Evaluation

We performed both quantitative and qualitative evaluation of the results. In Publication **I**, Frachet Inception Distance (FID) (Heusel, et al., 2017) was used to compute the similarity between the reference H&E-stained and virtually H&E-stained tissue image. FID is a distance metric commonly used to compare high-level abstracted features from both virtual and real domains. In Publication **II**, three different quantitative metrics were used, i.e., structural similarity index Measure (SSIM) (Zhang, et al., 2011), peak signal-to-noise ratio (PSNR) (Hore & Ziou, 2010), and Pearson's correlation coefficient (PCC). Additionally, the virtually stained WSIs were compared against chemically stained H&E for nuclei-level quantification using Hover-Net (Graham, et al., 2019). The nuclei count for the variants of pix2pix were compared against ground truth using Pearson's correlation coefficient, whereas nuclei segmentation was evaluated using F-score, precision, and recall.

In Publication **III**, three different types of quantitative evaluations were performed. First, to assess the performance of the method on color/style transfer, the normalized images were transformed to the $l\alpha\beta$ color space (ICC, 2004) and then the normalized histogram of each channel was compared with the reference image histograms using intersection, Pearson correlation coefficient (PCC), Euclidean distance, and Jensen-Shannon (JS) divergence. Then, InceptionV3-extracted (Szegedy, et al., 2016), high-level features of the normalized and reference images were compared using FID score (Heusel, et al., 2017). Lastly, SSIM was used to ensure the structural integrity of the normalized images with respect to the original input images.

While quantitative evaluation is an essential measure for assessing and comparing the overall performance of different methods, it can take us only so far when more granular and nuanced inspection is required which is imperative for analyzing biological specimens. Therefore, in all three publications, a thorough qualitative evaluation was also performed by domain experts which revealed critical aspects that weren't evident in quantitative evaluation alone, providing deeper insights into the efficacy and limitations of the methods used in the studies.

4.2.4 Pre-processing

Whole slide image (WSI) preprocessing has been a crucial and significant part of all three publications. Pre-processing includes image registration, background removal, and tiling.

4.2.4.1 Registration

Image registration is the process of aligning images typically from different modalities such that the structural content of the image pairs has correspondence at both micro and macro levels. The image that is transformed in the image pair is called the moving image and the other is fixed or static image. The fixed image serves as the reference for the alignment process. Rigid and elastic registration techniques are commonly used to align WSI pairs. Rigid registration is a technique that aligns images using translation, rotation, and uniform scaling without deforming their shape. Elastic registration is a method that aligns images by allowing nonlinear deformations to account for complex distortions and morphological changes.

Image registration was a key step in Publication **II** as it primarily uses a supervised method which relies on aligned data for training. A three-phase registration procedure was applied to the whole slide image pairs. Initially, a subset of WSI pairs underwent rigid registration. This subset was employed to train a baseline model designed to generate intermediate H&E-like outputs for all the unstained WSIs. Although the rigid alignment of the training WSI pairs resulted in suboptimal virtual staining quality, the intermediate output was crucial for the subsequent phase due to its inherent alignment with the input unstained WSI. Subsequently, the ground truth H&E WSIs were elastically registered to the intermediate H&E-like WSIs, using them as surrogates for the unstained WSIs. Due to the similar appearances, alignment between the initially rigidly registered unstained and ground truth H&E WSI pairs was refined.

4.2.4.2 Background Removal

In addition to high-resolution tissue images, a WSI also contains a background that is usually not uniform and occasionally also contains processing debris and unwanted artifacts. The background can be removed using a binary mask. Mask generation consists of the following steps: grayscale conversion of the WSI, thresholding to remove majority of the background, dilation, binary hole filling, and then erosion in the same order to get a refined mask. Background removal was applied to all the tissue images in all three publications.

4.2.4.3 Tiling

Deep learning models are generally trained on images that are orders of magnitude smaller in dimensions compared to the high-resolution WSIs. Therefore, the binary masks generated in the previous step are used to split WSIs into much smaller patches called tiles, the masks guided the tiling process to exclude tiles containing only background or undesired content like debris or artifacts.

All the GAN-based methods in this thesis work were trained using tiled tissue image data. In Publication **II**, we tested different tile dimensions: 256x256 pixels, 512x512 pixels, 1024x1024 pixels and observed that 512x512 pixels tile dimension was optimal for enough morphological or structural context and a reasonable batch

size of 16 distributed over four GPUs during training. For CycleGAN (Zhu, et al., 2017) and Pix2pix (Isola, et al., 2017) experiments in all three publications, the tile size was increased to 2048x2048 during inference because of the models being fully convolutional, this has been reported to generate more consistent colors (de Bel, et al., 2018). As the inverse of the tiling process, the virtually stained tiles generated during inference were merged to recreate the WSI, presenting the unified tissue section as a whole. Inspired by an approach proposed by de Bal et al., overlapping tiles were merged using a weighted merging strategy to reduce the tiling effect (de Bel, et al., 2018).

4.2.5 Heuristic Methods

In Publication **III**, semi-heuristic approaches were used for reference sample selection and training set formation for CycleGAN experiments.

4.2.5.1 Reference Sample Selection

Ideally, a reference sample should comprehensively represent the morphological variations present in the tissue image dataset and exhibit staining colors preferred by the expert. Since the tissue sections used in Publication **III**, were prepared from the same block of tissue extracted through punch biopsy, therefore, morphologically the dataset was quite consistent. Therefore, we opted for a computational approach that mainly relied on the color balance of the tissue images. The method used the mean intensity ratio of the red and blue channels to select the reference sample. For each tissue type, the sample with a red-to-blue ratio closest to value one was designated as the reference sample. This selection criterion is based on the characteristic color expression of H&E-stained tissue samples, in which shades of blue and red are prominent.

4.2.5.2 Clustering

In Publication **III**, each set (colon, kidney, skin) of tissue samples was sourced from a different laboratory, making it difficult to distribute the data into test and training sets. Therefore, we adopted a clustering-based approach to find out if there exists a natural grouping in the dataset. First, the tissue images were transformed to the $l\alpha\beta$ color space (ICC, 2004) to enable the use of more independent and distinguishable representation in each channel. Normalized histograms of each channel were used as features. Then, using principal component analysis, the first two components were used to plot the samples in the Cartesian plane (**Figure 14A**). The same components were used to perform k-means clustering (MacQueen, 1967) of the tissue images. Within-cluster sum of squares was used to find the optimal number of clusters. To ensure uniformity, for each organ tissue 8 clusters were chosen and the sample closest to the cluster mean was chosen as the representative sample of that cluster; examples of tissue clusters are shown in **Figure 14** (B and C). Altogether 8 representative samples were chosen that were used in CycleGAN trainings as the source samples.

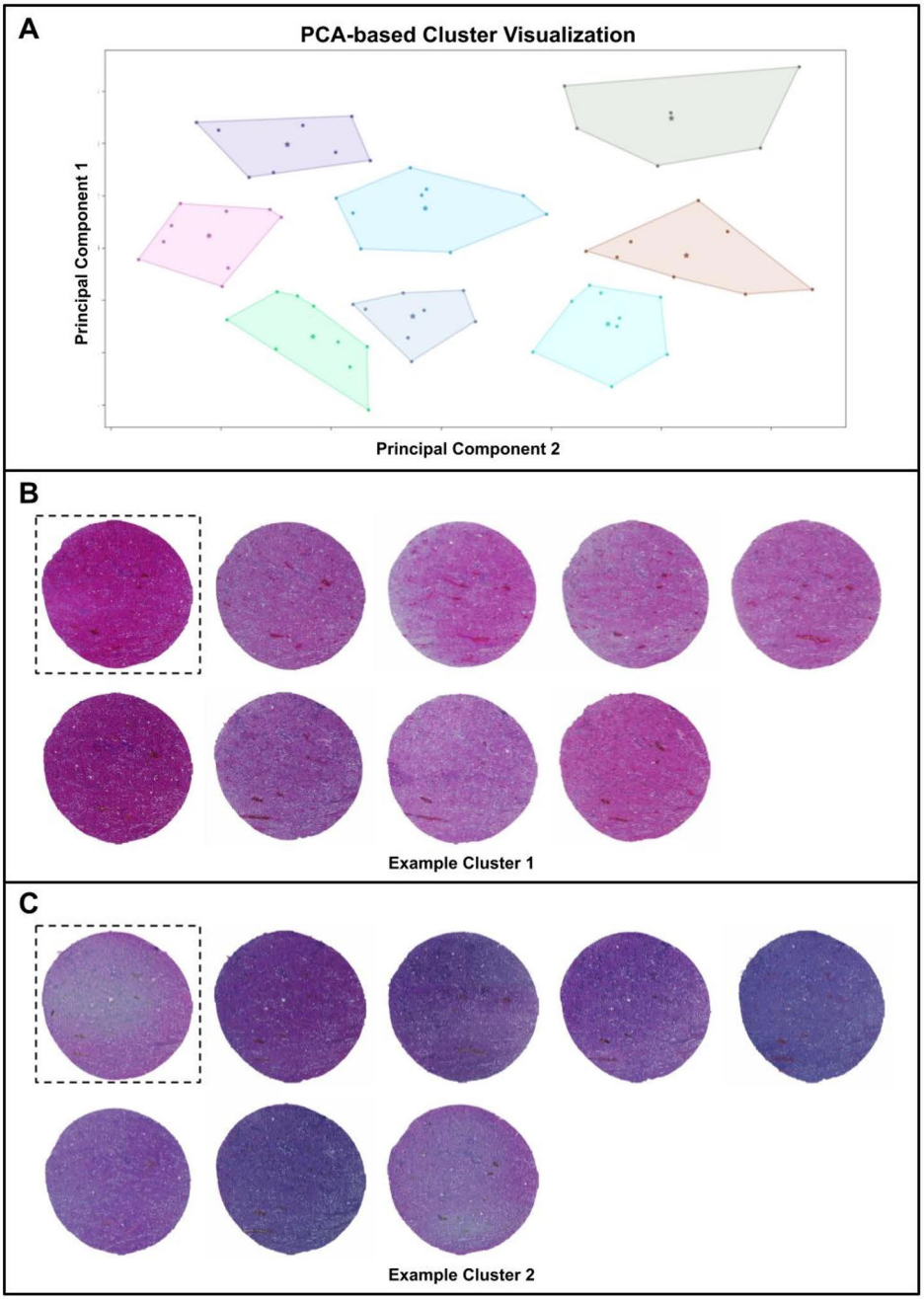

Figure 14. **A**) PCA-based K-means clustering of kidney tissues (normalized histograms of $l\alpha\beta$ channels) visualized using principal components 1 and 2 **B**) Example of a kidney tissue cluster with red-dominant staining manifestation. **C**) Example of a kidney tissue cluster with blue-dominant staining manifestation. Cluster representatives shown within the dashed box.

4.2.6 Computing and Software

All the computing experiments were conducted in the high-performance computing (HPC) environment provided by Finnish IT Center for Science (CSC). Each compute node is equipped with four NVIDIA Volta V100 GPUs, most of the trainings were parallelized over all four GPUs in the node. The implementation was containerized using a virtualization tool called Singularity and then deployed in the HPC environment. The code base for all the publications was implemented in the Python programming language. For Publication **II** & **III**, the dataset and code has been made publicly available. All the deep learning models were implemented using Tensorflow (Abadi, et al., 2016) with Keras (Chollet, 2021) interface. We used tiff file format for handling WSIs, and read/write operations were performed using tifffile and pytiff libraries. WSI registration was performed using a publicly available tool called wsireg (https://github.com/NHPatterson/wsireg).

5 Results

5.1 Optimal Tissue Processing and Imaging Protocols for Virtual Staining

In Publication **I**, we studied the impact of tissue preparation and imaging on virtual staining. In virtual staining, imaging unstained tissue sections is a pivotal step, however, it poses some challenges. When the same tissue section is required to be stained, removing the coverslip to apply staining could potentially damage the tissue. Secondly, unlike stained tissue, unstained tissue sections lack contrast, initial imaging showed that applying the coverslip further diminishes the contrast, resulting in a barely perceptible tissue image. To understand these challenges better, we decided to image both versions of unstained tissue, with and without coverslip. Additionally, we also imaged the untreated, paraffinized tissue on glass slide, since it has proven to be beneficial in various techniques required for follow up analyses (Pilling & Gardner, 2016) . Overall, we imaged H&E-stained tissue along with three variants of unstained tissue: with coverslip (coverslipped), without coverslip (deparaffinized), and in paraffin (unprocessed), which also represent the sequence of adjacent sections from a tissue block. Section thickness is another important aspect of tissue preparation as it pertains to stain absorption, therefore, we prepared the above-mentioned sequence in five different thicknesses: 3, 5, 8, 12, and 20μm. All tissue images with the exception of coverslipped exhibited reasonable contrast. The coverslipped unstained tissue images lacked contrast and appeared near transparent to the human eye. Additionally, these images exhibited grid-like artifacts through the whole slide image (WSI), likely introduced by the scanner during the post-processing tile stitching step.

5.1.1 Unsupervised Virtual Staining

The primary tissue dataset in Publication **I** was prepared using adjacent tissue sections which did not have pixel-to-pixel content correspondence between the unstained and stained tissue pairs, therefore, CycleGAN (Zhu et al., 2017), a very commonly used unsupervised image-to-image translation method was chosen for initial experimentation. Each unstained variant (unprocessed, deparaffinized, and coverslipped) was paired with H&E-stained tissue. The WSIs were halved making

sure both halves had enough representative morphology, the bottom half was used for training and the top for inference. The same procedure was repeated for each of the five section thicknesses. We employed FID scores (Heusel, et al., 2017) for quantitative evaluation, widely used to measure the similarity between synthetic and real images, particularly in domains where aligned images are not available. A lower FID score indicates greater similarity in content and appearance between the sets of images. Although FID operates at a high level of abstraction which may not be an ideal metric to assess virtual staining, because of the inapplicability of a more precise metric in this scenario, it was used as a proxy to determine the optimal unstained imaging specifications by measuring the resemblance between virtually stained and H&E-stained ground truth images. Quantitative evaluation revealed a trend of increasing FID scores with increased tissue section thickness (**Figure 15**). Furthermore, FID scores indicated that virtual staining results deviated from the ground truth in 12µm samples the most, whereas coverslipped unstained samples showed the high variance. Quantitative analysis indicates that deparaffinized samples, with a section thickness in the range of 3-5µm, are likely the optimal choice for virtual staining. **Figure 16**A shows an example of the top-half input WSIs, their respective virtually H&E-stained results and the H&E-stained reference.

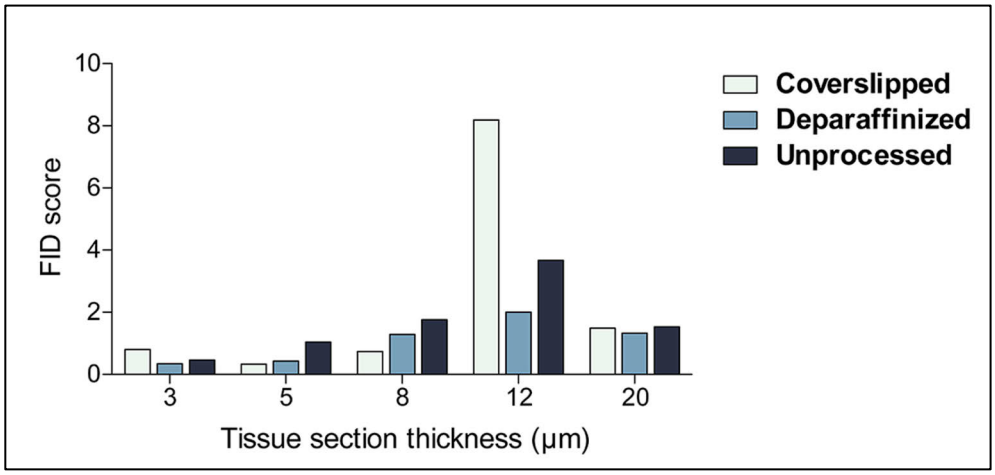

Figure 15. Quantitative evaluation of CycleGAN model virtual staining performance with FID scores for each tissue section thickness and processing method.

The qualitative evaluation of virtually stained sections, compared with reference H&E sections also indicated thinner sections performing better which aligns well with the FID-based quantitative analysis. Deparaffinized samples, with the lowest FID scores, produced the best results in terms of tissue morphology and hematoxylin and eosin localization. Unprocessed samples also performed well, but the residual paraffin created artifacts such as small creases or folds, especially in thicker sections.

Coverslipped sections performed poorly, exhibiting pronounced vignetting in the virtually stained image which is most likely caused by the grid-like artifact in the low-contrast input images. Higher magnification analysis showed that 5µm sections performed better than the rest, in general, and deparaffinized tissues showed the most accurate morphology and nuclei count, consistent with WSI-level results and FID scores (**Figure 15**).

Morphologically, epithelial tissue was interpreted well by the virtual staining algorithm at higher magnifications (**Figure 16C**), but distinguishing cells was particularly challenging in unprocessed tissues, in 5µm and 8µm sections. Coverslipped and deparaffinized tissues showed some cell distinction, though shapes were not always entirely accurate. All virtual staining results lacked inflammatory and basal cells, whereas the morphology of blood vessels was generally well reproduced, except for endothelial cell nuclei. In the stroma, 3µm and 5µm deparaffinized tissues reflect the most accurate results for nuclei and cytoplasm. Adipose tissue was best reproduced in 3µm deparaffinized sections, despite challenges in reproducing small, irregular adipocyte nuclei.

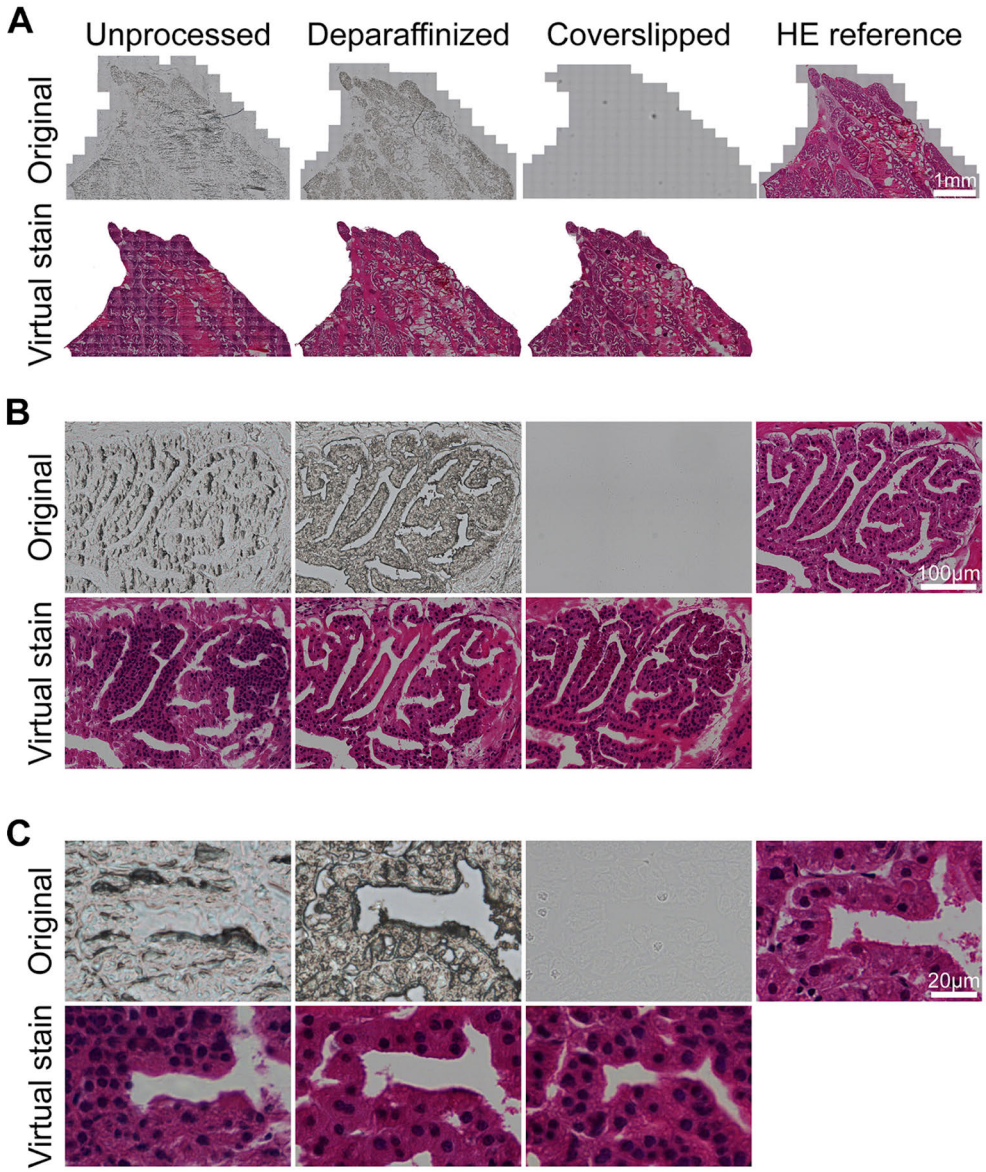

Figure 16. Histologic performance of unsupervised virtual H&E staining with images of differentially processed tissue sections. Comparative views of the original unstained images and virtual staining and H&E reference images. An example of section thickness 5 µm is shown. **A**) A WSI-level view at low magnification. **B**) A view of a prostate gland structure. **C**) A high-magnification view of prostate epithelial cells.

5.1.2 Supervised Virtual Staining with Optimized Protocols

The experimentation with unpaired data from adjacent tissue sections provided a reasonable idea for optimal tissue processing and imaging protocols. However, this

evaluation left room to assess the performance of virtual staining using the same tissue section imaged before and after chemical H&E staining. This approach would provide a direct ground truth with pixel-to-pixel correspondence between the WSI pairs, allowing for more precise comparison and validation. Using the learnings from the unsupervised virtual staining experiments, same tissue sections were imaged before (deparaffinized and without coverslip) and after H&E staining to generate alignable pairs of unstained and H&E-stained WSIs. This was followed by an image registration step to align the images ensuring correspondence between the two modalities. We used pix2pix (Isola, et al., 2017), a commonly used supervised image-to-image translation method that leverages aligned image pairs for learning. The quality of virtual staining improved significantly with the supervised learning method. The virtually stained tissue images had high morphological resemblance with the H&E-stained ground truth on a macro level. High magnification analysis further revealed a significant improvement in the representation of basal cells in the epithelium, an area where the unsupervised method particularly struggled (**Figure 17**).

To evaluate the comprehensiveness of supervised virtual staining, we utilized a multi-organ tissue dataset consisting of liver, kidney, and testis tissues, representing various morphologies. The quality of virtual staining in multi-organ tissues was comparable to that observed in prostate tissue virtual staining. The general morphology was well reproduced, and major structural components of liver, kidney, and testis can be accurately distinguished in the virtually stained tissue (**Figure 17B**). However, the method faced some challenges in representing certain less prominent cellular details particularly in testis tissue. Furthermore, in addition to 40× magnification, virtual staining was also tested on WSI imaged at 20× magnification, and the results were on par with 40×. Computationally upscaling unstained input images from 20× to 40× resulted in subpar virtual staining.

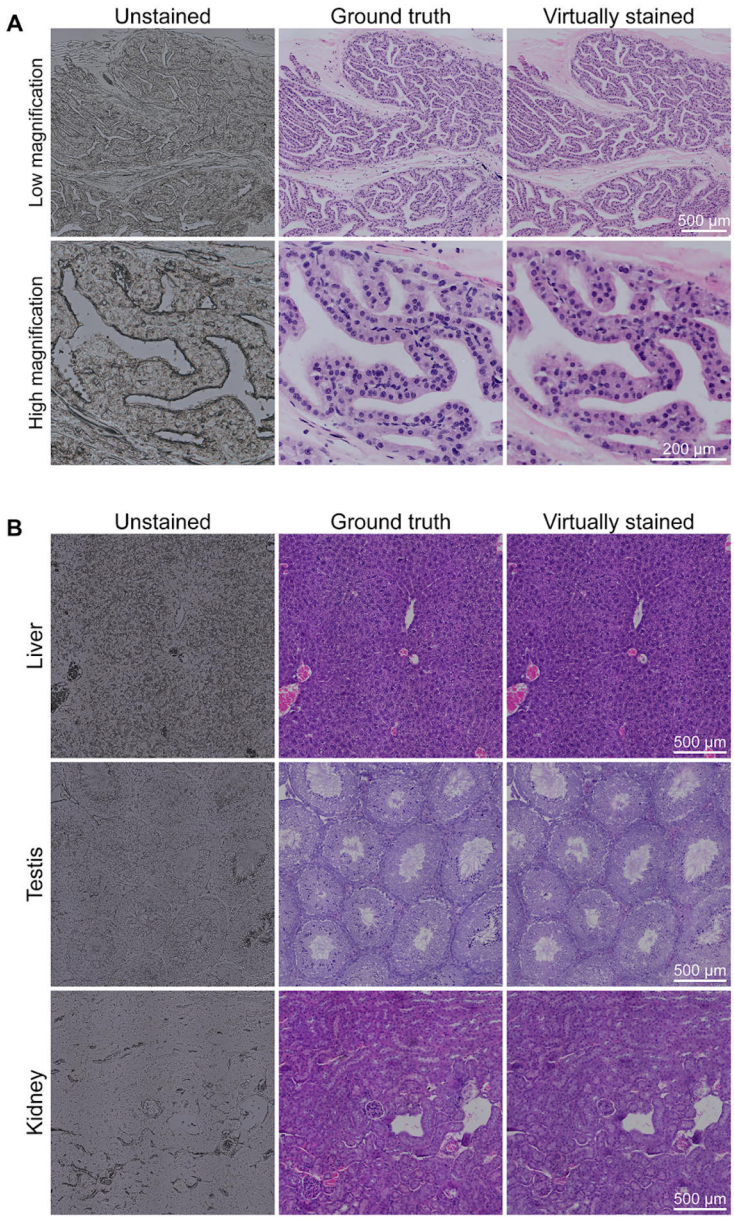

Figure 17. **A**) Histologic performance of supervised virtual H&E staining of prostate gland structures. Comparative views of the original unstained images and virtual staining and H&E reference images of prostate gland structure. An overview (top row) and a close-up (bottom row). Basal cell nuclei are indicated using arrowheads. **B**) Histologic performance of supervised virtual H&E staining of different organ tissues. Comparative views of original unstained images and virtual staining and H&E reference images of histology of liver (top row), testis (middle row), and kidney (bottom row).

5.2 Virtual Staining with Increased Neural Network Architecture Complexity

After identifying the optimal tissue imaging and processing protocols in Publication **I**, we studied the effect of neural network architecture complexity on virtual staining quality in Publication **II**. In addition to CycleGAN (Zhu, et al., 2017), an unsupervised learning method, we also tested methods from other learning paradigms such as semi-supervised, multi-shot, and one-shot learning, before selecting pix2pix (Isola, et al., 2017), a supervised learning method for further exploration. Although methods from other learning paradigms are useful in various domains, they lack the translation accuracy required to digitally reconstruct a detail-rich biomedical imaging modality. We used murine prostate samples, 81 pairs of tissue images, scanned before and after H&E staining for initial experimentation. The WSI pairs were registered and tiled before training. We used the original implementation of pix2pix as the baseline version and modified the generators using double convolutions and dense convolutions (explained in 4.2.1.2, **Figure 12**) to create two more variants, respectively. The three variants of the pix2pix model, namely, baseline, double convolution, and dense convolution, were trained on registered tissue images using the same training parameters and experimental conditions. At a macro level, all three variants of pix2pix demonstrated comparable performance, effectively replicating the macroscopic tissue morphology similar to that achieved with chemical H&E staining (**Figure 18**).

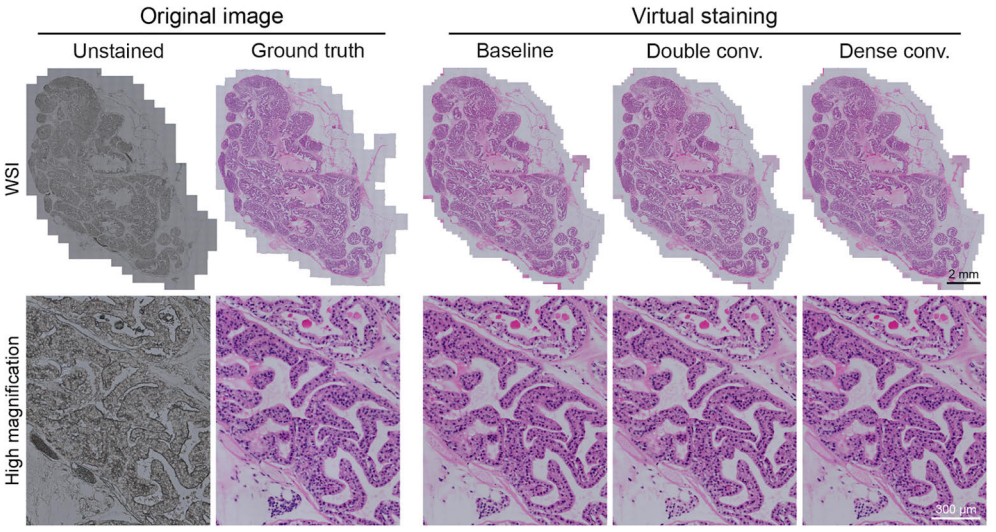

Figure 18. Macro- and micro-level views showing the performance of virtual staining. Left to right columns: original images of unstained and H&E-stained tissue, virtually stained tissue with baseline, double convolution, dense convolution architectures.

5.2.1 Quantitative and Qualitative Evaluation of Virtual Staining

Next, we employed SSIM, PSNR, and PCC to quantitatively assess the performance of the models. The baseline model achieved a mean PSNR of 22.609, a mean SSIM of 0.725, and a mean PCC of 0.903. The second variant, a pix2pix model incorporating double convolution encoder-decoder blocks, was trained to assess whether it enhances the quality of virtual staining compared to the benchmark model. The mean PSNR and SSIM did not improve, but PCC improved slightly; the values were 22.214, 0.720, and 0.904, respectively. Finally, the third variant, a pix2pix model with dense convolution encoder-decoder blocks inspired by DenseU-net, was trained to investigate the impact of a more advanced approach to increasing network capacity on the quality of virtual staining. The dense convolution variant outperformed the previous two with a mean PSNR of 22.865, a mean SSIM of 0.746, and a mean PCC of 0.916. Quantitative evaluation indicated that increasing the network capacity in a more sophisticated way positively impacts the virtual staining quality. Only tiles containing tissue content were used in quantitative evaluation.

Quantitative evaluation was followed by histological evaluation. All three variants were observed to have produced the general morphology and major structural components of the H&E-stained prostate tissue with considerable accuracy. Components with less structural complexity such as adipose or secretions were represented accurately across the board, whereas the more complex ones were relatively more challenging. At a more granular level, it was observed that the round nuclei in epithelial tissues had accurate shapes, sizes, and proximity (**Figure 19**). Most round and strongly colored nuclei in other tissues were also reproduced accurately, but accuracy decreased for elongated or irregular-shaped nuclei, or if they were very pale. This is likely due to the infrequent occurrence of such nuclei. Eosinophilic areas and color tones in all tissues, especially in muscle, nerve, and connective tissue, were reproduced accurately (**Figure 19**). GAN-based methods are known for creating false details called hallucinations, which were found in baseline network images, particularly around clusters of red and white blood cells. However, hallucinations reduced significantly with increased network capacity in the dense convolution experiments. Some minor artifacts were also observed such as fine patterning and occasional miscoloring. While some patterns were observed across all three variants of pix2pix with varying intensity, they did not compromise histological interpretability.

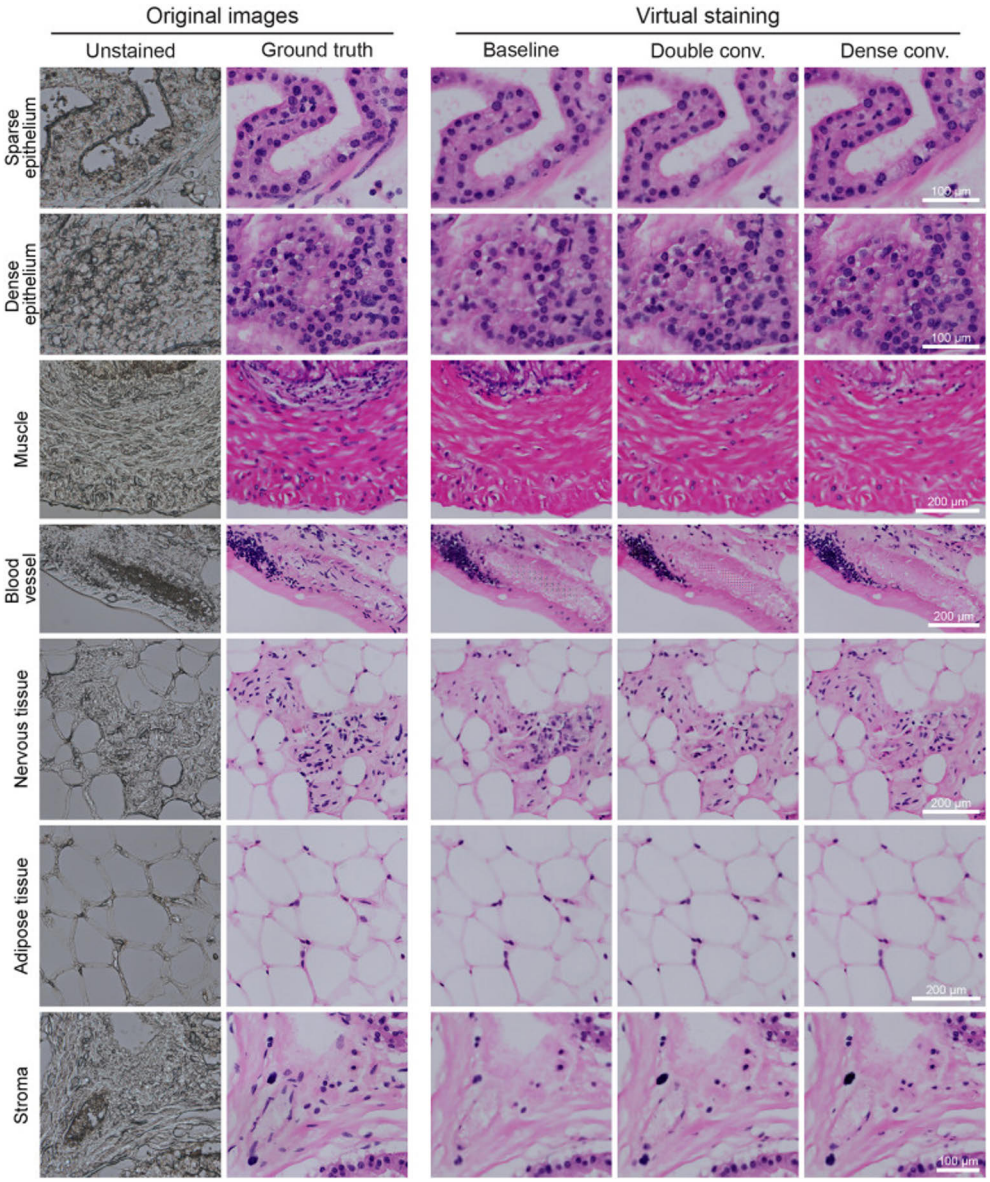

Figure 19. Representative images of different virtually H&E-stained tissue types and the corresponding unstained tissue and ground truth.

5.2.2 Nucleus-level Statistics of Virtually Staining

In addition to quantitative and histological evaluation, we also performed downstream analysis by comparing the similarity between H&E and virtual staining at a single-cell level by analyzing nucleus detection statistics. The Hover-Net (Graham, et al., 2019) model was employed to segment nuclei in tiles from both

H&E-stained reference images and their virtually stained counterparts. We extracted 135 image tiles, representing significant tissue types (epithelial, stromal, adipose, muscle), from H&E ground truth and virtually H&E-stained samples and evaluated nucleus segmentation accuracy using F-score, precision, and recall as evaluation metrics.

Figure 20A shows example tiles with nucleus segmentation overlaid for H&E, baseline, double convolution, and dense convolution methods. The correlation between nucleus counts in H&E and virtually stained tiles demonstrates strong correspondence, with PCC values between 0.912 and 0.952 (**Figure 20B**). Although the nucleus count is similar across the three variants, the F-score (**Figure 20C**) indicates suboptimal object-level correspondence, with nuclei locations not fully aligning. This also indicates the possibility of granular registration errors resulting in misalignment of nuclei location between the unstained input and the H&E ground truth. Among the variants, dense convolution outperforms others in F-score due to its higher precision. These results confirm the accuracy of virtual staining in reproducing nuclei, especially in epithelial areas, while highlighting suboptimal performance in elongated stromal nuclei, likely due to the abundance of epithelial nuclei and the relative scarcity of elongated stromal nuclei in the tissue image dataset.

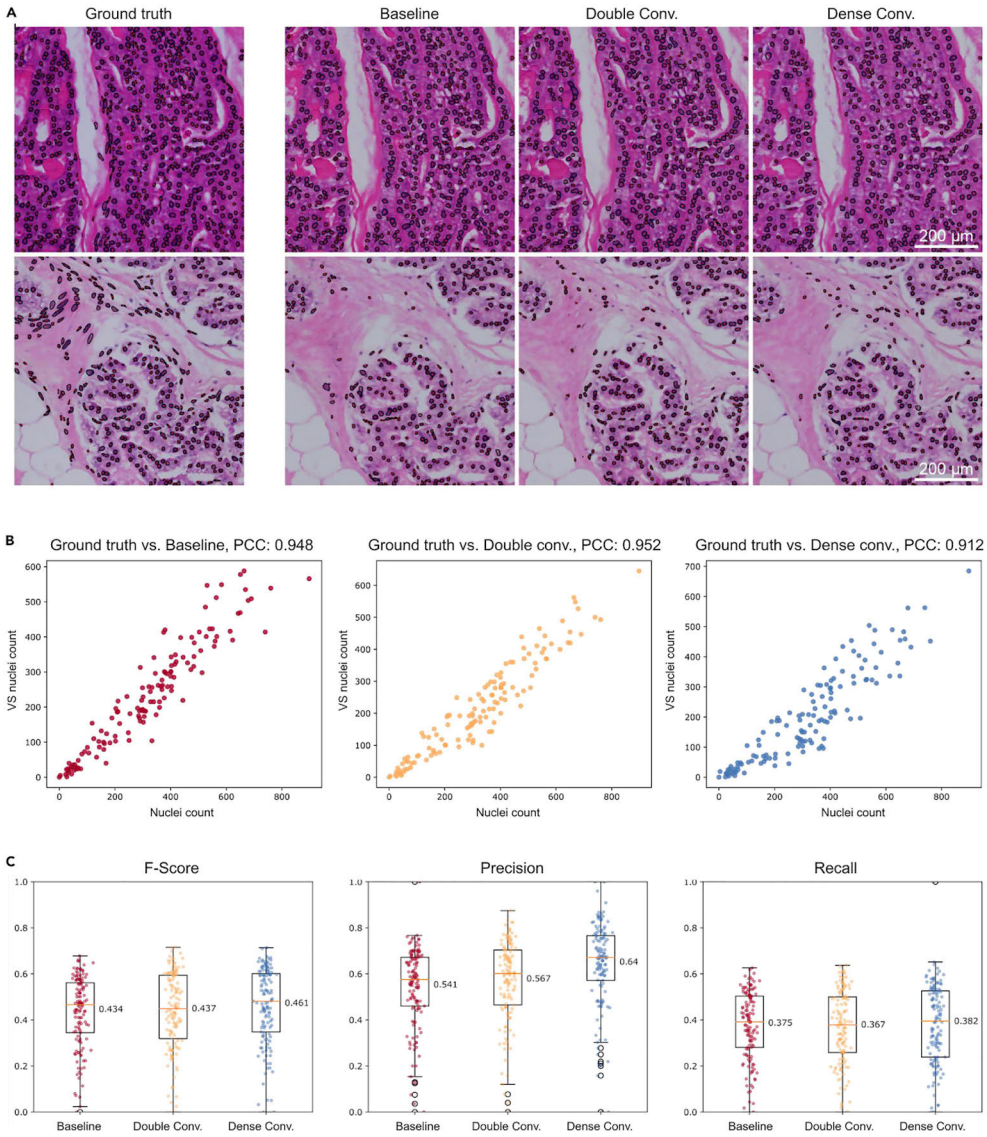

Figure 20. The performance of Hover-Net nucleus segmentation for virtually stained images was quantitatively evaluated by using segmentation obtained for the H&E staining as the ground truth. **A**) Examples of nuclei segmentation masks based on H&E-stained tissue, and virtually stained tissue with baseline, double convolution, and dense convolution architecture. **B**) Nuclei count correlations between ground truth H&E staining and pix2pix variant virtual stainings. Each data point corresponds to the number of nuclei detected in a tile. **C**) Tile-level averages of nucleus detection F-score, precision, and recall for virtual stainings against the segmentation result obtained for the H&E-stained ground truth. Boxplots show the mean, first and third quartiles, and 1.5x interquartile range as whiskers.

5.3 Stain Normalization Methods Comparison

In Publication **III**, we investigated the variation in tissue appearance resulting from different laboratory staining protocols and evaluated traditional and deep learning-based stain normalization methods using a tissue image dataset collected from 66 laboratories. The dataset consisted of skin, kidney and colon tissue sections. As illustrated in **Figure 8**, the tissue sections exhibited substantial variability which made it a well-suited dataset for this study. We compared four traditional and four deep learning-based methods. Traditional methods included histogram matching, Macenko, Reinhard, and Vahadane normalization, whereas deep learning-based methods were two variants of CycleGAN and pix2pix each. The reference sample was selected based on the red-blue color ratio criterion mentioned in section 4.2.5.1.

5.3.1 Quantitative Evaluation

Quantitative evaluation focused on measuring two key attributes, the degree of similarity between the normalized images and the reference image in terms of color or cosmetics, and the extent to which original tissue morphology and structural components are preserved during the stain normalization process. To assess color similarity, we used the normalized histogram of the tissues images in $l\alpha\beta$ color space (ICC, 2004) to compute intersection, Pearson correlation coefficient, Euclidean distance, and Jensen-Shannon (JS) divergence between the reference image and the rest of the images before and after stain normalization. Subsequently, FID (Heusel, et al., 2017) was computed using abstract features extracted from InceptionV4 (Szegedy, et al., 2016) to gauge a high-level similarity before and after stain normalization. Finally, SSIM was employed to evaluate the preservation of structural components after normalization.

In skin tissue images, histogram matching significantly improved performance across all metrics measuring color/cosmetic correspondence (**Figure 21A**), it achieved the highest mean scores in intersection, PCC, Euclidean distance, and JS divergence against the original images when compared with other methods. It also showed the best results in abstract feature-based similarity with the lowest Fréchet Inception Distance (FID) score. However, Vahadane normalization preserved structural similarity the best, as indicated by the highest mean SSIM score, although it underperformed in other metrics. Overall, all methods effectively preserved tissue structure, with SSIM scores consistently above 0.92.

In kidney tissue images, histogram matching again led in all key metrics, performing exceptionally well likely due to the morphological uniformity observed in this particular tissue type (**Figure 21B**). Histogram matching also achieved the lowest FID score, indicating superior high-level feature concordance. Vahadane normalization, similar to the skin results, best preserved structural similarity with the highest mean SSIM score. All methods demonstrated strong performance in maintaining tissue structure, as reflected by mean SSIM scores above 0.945.

In colon tissue images, despite the higher intra-tissue morphological heterogeneity, histogram matching outperformed other methods in all color/cosmetics similarity measuring metrics, including intersection, PCC, Euclidean distance, and JS divergence. However, in high-level feature similarity, Reinhard and both CycleGAN variants performed better than histogram matching, attaining better FID scores. Vahadane normalization once again led structural similarity preservation, closely followed by Reinhard and histogram matching, all of which achieved mean SSIM scores above 0.95.

Although Vahadane proved to be the most effective method for preserving post-normalization structural similarity in all three tissue types, visual inspection of its results revealed an almost linear overall color shift, rendering the normalization ineffective, and in turn, structural similarity performance irrelevant. This underscores the significance of employing multiple metrics to assess various aspects of image transformation and highlights why detailed histological analysis is indispensable in digital and computational pathology.

5.3.2 Histological Evaluation

We chose colon tissues for histological evaluation because of its heterogeneity as compared to the other two organs tissue. The most prominent features observed in the ground truth tissue images were over-staining, under-staining, and inconsistent staining. Our analysis concentrated on addressing these issues. Furthermore, we assessed the normalized tissue images against the reference sample to determine the effectiveness of the different stain normalization methods used in Publication **III**.

High-quality H&E staining relies on a clear distinction between the blue/purple of hematoxylin and the pink/red of eosin, making hue a crucial factor. Generally, normalization methods introduced a shift toward pink/red, which could reduce the visibility of certain tissue structures. The Vahadane method, in particular, often altered the background color and diminished the hematoxylin contrast. While histogram matching improved hue in over-stained images, it introduced a strong pink/red shift in under-stained ones (**Figure 22**B). Macenko and Reinhard methods were most effective in enhancing under-stained images but struggled with over-staining. Digital artifacts was another issue observed in some of the methods. The Macenko method altered erythrocyte and lamina propria colors to blue, while deep-learning models like CycleGAN (ResNet) and Pix2pix (DenseUnet) generated artifacts in adipose tissue and smooth muscle nuclei, respectively (**Figure 22**C). Tile artifacts were noted with CycleGAN (ResNet) at tissue borders. No artifacts were seen with histogram matching and Reinhard, but neither method consistently improved visual quality across all staining conditions, histogram matching performed better with over-stained images, while Reinhard was more effective with under-stained ones.

We also assessed how well each method normalization results matched the reference tissue image. As mentioned previously, Vahadane performed poorly by completely removing hematoxylin, while Reinhard increased color intensity, altering

contrast and causing a cloudy appearance in overstained samples. Macenko showed variation depending on the original image, particularly in hematoxylin hue and background, with occasional blue artifacts in eosin. Although some images showed a slight pink tint in the background, histogram matching results were the most consistent and closely aligned with the reference image.

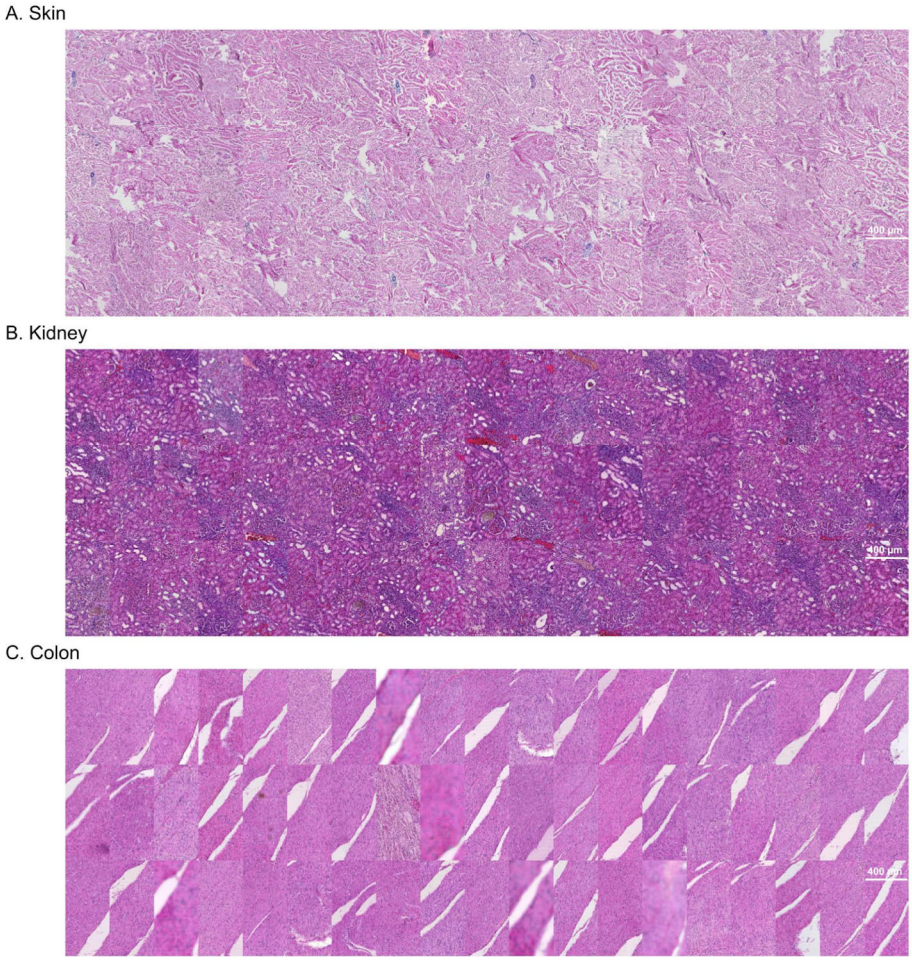

Figure 21. Post-normalization tissue patch collage showing the color uniformity. Tissue images stain normalized using histogram matching. **A**) Image patches extracted from skin tissue sections. **B**) Image patches extracted from kidney tissue sections. **C**) Image patches extracted from colon tissue sections.

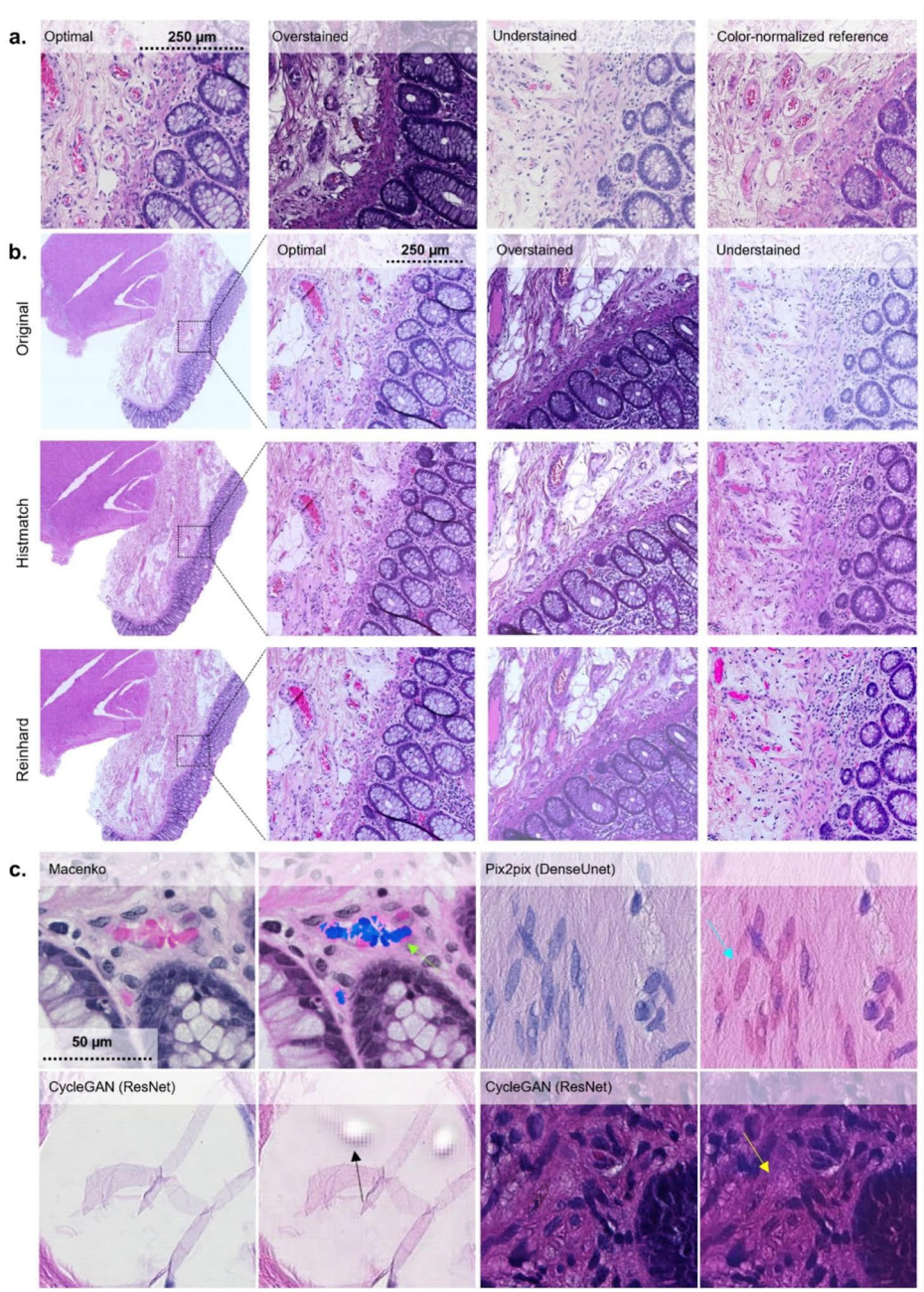

Figure 22. **A**) Representative images of optimal as well as over- and understained H&E-slides. **B**) Comparisons between original, Histogram matching and Reinhard normalized images from colonic mucosa-submucosa border in optimally and suboptimally stained slides. **C**) Artifacts emerging from image processing.

6 Discussion

Digital and computational advancements in histopathology have come a long way, from digitizing the first tissue slides in early 1990s to the gradual integration of AI-powered fully digital tools into clinical workflows at present. Although recent developments on the computational front have largely revolved around diagnostic and prognostic improvements, they have also shown potential to improve and optimize workflows. The work done in this thesis explores AI-based methods that could help to improve standardization and optimize histopathology workflow; it particularly focuses on the use of generative AI-based image-to-image translation methods to achieve virtual histopathology staining (Bai, et al., 2023) and stain normalization (Hoque, et al., 2024).

6.1 Achievement of Objectives

The first objective of this thesis was to establish optimal tissue processing and imaging protocols for virtual staining. Tissue processing and imaging are intricate steps that could contribute to variation in the appearance of tissue images (Webster & Dunstan, 2014). The unstained tissue inherently lacks contrast; therefore, it is important to gauge the impact of different factors on the resulting unstained tissue images, and consequently, on the virtually stained tissue images they're transformed into. Factors such as section thickness, paraffinization/deparaffinization, and coverslip application if left untested, could have unintended consequences on the end results. For instance, thicker unstained tissue sections may not appear different from thinner sections, but after staining they tend to appear darker than the thinner samples because they absorb more stain. Determining the ideal section thickness is crucial for virtual staining. Similarly, light reflects very differently from a tissue slide with coverslip as compared to one without coverslip. Deparaffinization of a tissue section is a mandatory step before chemical staining is applied, however, if paraffinized sections can somehow be used to digitally stain tissue images it can significantly aid follow-up analyses that rely on paraffinized tissue sections. Establishing optimal protocols is the first step toward ensuring robustness and uniformity of virtual staining results.

In Publication **I**, we iterated through different sets of tissue imaging and processing parameters and observed noticeable variation in the appearance of the unstained tissue image as well as resulting virtually stained tissue images. Initial

experiments showed that 3-5µm deparaffinized unstained tissue sections without coverslip produced virtually H&E-stained tissue images that resembled the adjacent chemically H&E-stained tissue images the most. They also revealed that unsupervised methods such as CycleGAN (Zhu, et al., 2017) can be effective for preliminary and exploratory analysis, and especially when alignable image data is not available, which was the case with the first dataset. However, for more accurate reproducibility of H&E staining digitally, an image dataset consisting of registered pairs of unstained and H&E-stained images of the same tissue is required that could be employed to train supervised image-to-image translation methods such as pix2pix (Isola, et al., 2017). In the subsequent experiments, registered tissue image data (prepared using the optimal protocols identified through previous experiments) was employed to train a pix2pix model, which clearly highlighted the superior quality of the resulting virtually H&E-stained tissue images. In addition to virtually staining prostate tissue, we also showed that the approach is scalable by virtually staining other tissue types such as liver, kidney and testis.

When it comes to tissue processing and imaging there are quite a few moving parts, and it is practically not possible to gauge all possible combinations of all the factors to find the optimal one. The factors involved, have varying degrees of significance, and the ones selected for this study, tend to be consequential for the tissue appearance and its significance in subsequent analyses. Hence, the findings of Publication **I** align well with Objective **I**, establishing that there exists a set of tissue processing and imaging protocols, optimal for virtually staining unstained tissue images, and it further indicated that scalability and accuracy of virtual staining could be improved significantly with guided image translation methodology.

Generative AI and specifically generative adversarial networks have been transformative for image-to-image translation. The very first GAN model (Goodfellow, et al., 2014) was only capable of translating features from the latent space to synthetic samples in the target domain. Over the next few years, many different GAN-based models, capable of all sorts of image-to-image translation, started surfacing (Pang, et al., 2021). In Publication **II**, we continued exploring different learning paradigms for image-to-image translation. In addition to supervised and unsupervised learning, we also experimented with semi-supervised, few-shots (Lin, et al., 2020), and one-shot learning (Lin, et al., 2020) image-to-image translation methods, only to learn that they are not even remotely suitable for the biomedical image translation. The primary focus in Publication **II**, however, was to understand the effect of architectural complexity of an image-to-image translation neural network on the quality of virtually stained tissue images. Pix2pix, an incarnation of conditional GAN (Isola, et al., 2017) and a very well-established image-to-image translation model, was chosen for further experimentation. Additionally, we used registered pairs of the same tissue specimen unstained and H&E-stained to train the models.

Three different variants of pix2pix, namely, baseline, double convolution, and dense convolution were tested. Initial experiments showed that even the baseline variant produced high-fidelity virtual staining images with considerable resemblance

to the H&E-stained ground truth. While double convolution did not show discernible improvements, both qualitative and quantitative analysis indicated that enhancing the generator model architecture through more sophisticated dense convolutions units, inspired by the dense U-Net architecture (Dong, et al., 2019), improved the quality of virtual staining in reproducing the high-level tissue morphology as well as subcellular structures. Although the inference time almost doubled in the case of dense convolution as compared to the baseline and given the fact that it was still under three minutes per WSI with the hardware setup used in the study, it's substantially faster than histochemical staining. Dense convolution also helped to minimize hallucination artifacts prevalent in virtually stained tissue images generated by double convolution and baseline variants.

Although three different metrics were used for quantitative evaluation of virtual staining, namely, SSIM, PSNR, and PCC, histological analysis was a significant part of the evaluation process as it revealed subtle aspects of virtual staining that were not obvious from quantitative evaluation. For instance, typical quantitative metrics treat all morphologies equally, however, some tissue types and structures are more significant histologically as compared to others. Histological analysis provided a more nuanced understanding of the virtual staining such as nuclear fusion being a common problem in all three variants or high occurrence of falsely hyperchromatic nuclei and other artifacts in baseline compared to the other two variants. Additionally, we incorporated nuclei counting, a morphological quantification task, as part of the quantitative evaluation. It showed high performance in nuclei reproducibility, quite close to the H&E-stained ground truth, across the three pix2pix variants. However, the location of the nuclei in virtually stained images lacked precision. This is a potential area of improvement that does not only rely on the image-to-image translation model but also on the quality of the image registration process. In Publication **II**, we concluded that architectural complexity of an image-to-image translation model can in fact improve the quality of virtual staining, answering the question raised in objective II.

In Publication **I** and **II** along with the main objectives, we also established the strengths and weaknesses of unsupervised and supervised learning approaches in the context of image-to-image translation on a high level. For instance, while supervised methods such as pix2pix are capable of producing high-fidelity virtually stained tissue images, they are strictly dependent on pre-requisites like image registration, an image preprocessing step that is memory-intensive and susceptible to subtle inaccuracies that could affect the end result in unintended ways. Another important aspect of these publications was the use of brightfield imaging for unstained tissue images. Despite its frequent availability, only a handful of studies used brightfield in the context of unstained tissue imaging (Zhang, et al., 2022). Most of the other studies employed specialized modalities such as autofluorescence imaging, primarily because of the level of details it is capable of capturing in unstained images (Picon, et al., 2021; Zhang, et al., 2020; Rana, et al., 2020), which consequently produces better quality virtually stained tissue images. However, autofluorescence imaging setup is more complex and infrequently available as compared to brightfield

imaging, which presents a strong case for further investigation into improving virtual staining produced by brightfield-acquired unstained tissue images.

Another significant workflow optimization task is stain normalization in which color heterogeneity of the tissue images is reduced to make the appearance more uniform across tissue sample sets. The tissue specimens are stained with the same stain or stain combinations, the variation in appearance stems from other factors such as subtle changes in laboratory-specific staining protocols, different scanner manufacturers, batch effects, etc. Despite numerous claims regarding the superiority of various stain normalization methods, the task remains inherently challenging because of the ever-changing underlying variables. Studying and gauging all sources of variation at the same time is a highly non-trivial task. A major shortcoming of most of the recent studies is the limited scope, wherein, datasets utilized are often sourced from a handful of sites (de Bel, et al., 2018; de Bel, et al., 2021). Consequently, the generalizability of these methods across diverse datasets and different histopathological conditions remains uncertain. Furthermore, the effectiveness of a particular stain normalization method can vary significantly depending on the specific problem statement and downstream analysis task. Thus, a method proven to be optimal for one application may not necessarily perform well for another.

In Publication **III**, we obtained a unique H&E-stained tissue image dataset from 66 different laboratories in 11 different countries. We attempted to understand and quantify the variation of the stained tissue appearance and compared a set of traditional and GAN-based stain normalization methods. The dataset consisted of three organ tissues, namely, skin, kidney, and colon. The dataset presented in the study showcased a broad spectrum of color variation (**Figure 8**). We employed four commonly used traditional stain normalization methods and two variants of pix2pix and CycleGAN each. Traditional methods included histogram matching, Macenko, Reinhard, and Vahadane normalization. Whereas, GAN-based methods included two variants of pix2pix, one with UNet-based generator, the other with DenseUnet-based generator, and two variants of CycleGAN, one using UNet-based generator and the other using ResNet-based generator.

A prevalent assumption is that GAN or deep learning-based methods would be far more performant as compared to the traditional ones because of the recent advancement, momentum, and adoption of AI across several fields and their panacea-like status. In Publication **III**, it was observed that GAN-based methods produced reasonable results in terms of matching the reference colors and largely maintaining structural integrity, however, they did not succeed in establishing a discernible superiority over traditional methods. Whereas traditional stain normalization methods had mixed results with the exception of histogram matching. While Vahadane normalization completely failed to produce any meaningful normalization results as it colored the tissue image uniformly with a pink hue including the background, histogram matching produced results were the most concordant with the reference sample, observed both quantitatively and qualitatively, for all three tissue types. Some of the factors that could have

contributed to the exceptional performance of histogram matching include the decision to use the whole tissue image as reference sample as opposed to traditionally choosing a small patch. Also, since the tissue samples were extracted from the same tissue block, there appeared to be reasonable morphological uniformity across tissue samples. This made global statistical attributes adjustments, that histogram matching is based on, more effective. In this study, methods-related minutiae may not be the major takeaway, but in fact the high-level findings that despite several studies indicating superior performance of GAN-based stain normalization method, traditional stain normalization methods cannot be written off completely, as they may be more suitable and effective for some specific cases. And, although GAN-based stain normalization methods work quite well in most of the cases, they also have shortcomings such as the unwanted hallucination artifacts.

Moreover, it also shows that as the number of laboratories increases in a study, the variation in the stained tissue appearance is also likely to increase which is reflected by the color spectrum observed in this study (**Figure 8**). Thus, Publication **III**, successfully meets the goal of benchmarking different stain normalization methods, and in addition to that, it also provides the dataset as a freely available public resource which can substantially contribute to a better understanding of the topic and be used for experimentation with other techniques such as stain augmentation. It is also important to highlight the necessity of a more extensive dataset that constitutes several samples from each laboratory of a more effective benchmarking of GAN-based stain normalization methods.

6.2 Limitations

In Publication **I** factors such as choosing the most suitable scanner for tissue imaging would not have been very fruitful given the fact that vendors vary from laboratory to laboratory. Also, switching vendors is a costly affair and may not be feasible for every laboratory, especially only for a specific use case (Lujan, et al., 2021). Procedural factors, such as the ones selected in this study, are easily modifiable and can be applied in any laboratory setting.

Publications **I** and **II** present key findings about virtual staining, however, there were certain limitations that can be addressed in future studies. For instance, batches of H&E-stained tissue images with varying appearance were trained separately; future studies could employ stain normalization (Hoque, et al., 2024) and analyze its effect on virtual staining results. In open imaging systems, where an external source of light can also affect the appearance of unstained tissue images, normalization can also be applied to unstained tissue images and the impact of uniformity on virtual staining results can be further examined in combination with normalized stained images. Another challenging aspect was the lack of metrics to quantify hallucinations which has recently become an active area of research in both computer vision (Tivnan, et al., 2024) and large language models (Luo, et al., 2024).

In publication **III**, even though the WSI dataset offered a broad spectrum of color variation, one sample per laboratory poses a challenge for deep learning methods

because along with quality they also rely on the quantity of data to be able to generalize better. A single tissue sample from each laboratory was proven insufficient for the training strategies employed in GAN-based stain normalization methods, ultimately failing to identify a definitive optimal approach.

6.3 Image-to-image Translation and High-Performance Computing (HPC)

It is also important to note that deep learning methods are not only data-hungry but also resource-hungry. Generative models, especially GAN-based image-to-image translation models, are much bigger and architecturally more complex as compared to a typical classification model, for instance, pix2pix has more than double the number of parameters as compared to inceptionV3 (Szegedy, et al., 2016) and CycleGAN has quadruple the number of parameters since it has two generator-discriminator pairs. Although it is not completely infeasible to train GAN-based image-to-image translation models on GPU-powered local workstations, given the scale of WSI data, it is nonetheless a tedious and slow process. Access to high-performance computing (HPC) environments like Puhti and LUMI, maintained by Finnish IT Center for Science (CSC), has been crucial for tackling such high-dimensional problems. The multi-GPU powered compute nodes have been indispensable for this thesis work as they facilitated the parallelization of image-to-image translation models training without requiring the direct management of the underlying hardware infrastructure. Additionally, handling terabytes of WSI data and memory-intensive tasks such as image registration was significantly more manageable on HPC as compared to executing these operations on local workstations. High computational requirements also raise the question of the feasibility of scaling and commercializing the virtual staining technology for clinical and laboratory use; the answer may lie in the commercial cloud technology that provides a similar scalable infrastructure which could facilitate the continuous development and hosting of the virtual staining application.

6.4 Future Research

This work is a testament to the fact that AI has the potential to not only disrupt the diagnostics applications, but also impactfully improve clinical and laboratory workflows. The AI-driven image-to-image translation approach presented in this thesis to optimize histopathology workflow has a strong potential to end up in future clinical solutions, but a lot of effort is required at every level starting from bright-field imaging enhancements for unstained tissue image acquisition to increasing the type of stains for training the models. Larger, multi-center image datasets, consisting of a broad range of organ tissues, can also make invaluable contribution to validate the accuracy and efficacy of this approach. Although GAN-based methods have been the focus of this thesis, generative AI's unprecedented recent traction is now driving

rapid advancements in the field, with new approaches, such as diffusion (Sohl-Dickstein, et al., 2015; Ho, et al., 2020) and vision transformer (Dosovitskiy, et al., 2020) models, gradually taking center stage. Therefore, these new image-to-image translation methods must be rigorously tested on histopathology applications.

7 Conclusions

This body of research is focused on the applications of AI-based image-to-image translation in the field of histopathology. It explores a variety of different methods and assesses the feasibility to use them in applications such as virtual histopathology staining and stain normalization. The key findings of the thesis work can be summarized as follows.

Tissue processing and imaging protocols are critical factors in virtual staining. Prior to implementing a virtual staining solution, it is essential to optimize these protocols. In Publication **I**, the key findings included determining the optimal section thickness and processing level for virtual staining. Additionally, we noted that unstained tissue sections imaged without coverslip are more suitable for virtual staining, it also reduces the risk of tissue damage during the subsequent staining process if paired data is required for training. Furthermore, we established that while unsupervised image-to-image translation provides some valuable insights, it lacks the precision required for virtual staining. In contrast, supervised image-to-image translation provides a more direct and efficient learning mechanism, allowing for direct comparison with H&E-stained ground truth tissue images.

Building on the initial learnings, in Publication **II**, we used the optimal tissue processing and imaging protocols to prepare a registered unstained and H&E-stained WSI pair dataset for virtual staining. Using that dataset, we established that brightfield imaging, that is readily available in most laboratories, has the potential to be used for virtual staining. Furthermore, we showed that increasing the network complexity in a sophisticated way could help to improve virtual staining quality, especially by curbing hallucination artifacts which is a common drawback of generative models.

Staining color variation is a common histopathology problem and stain normalization, another application of image-to-image translation, can help to solve this problem. Stain normalization is an important image preprocessing step in histopathological analysis. In Publication **III**, using a unique tissue image dataset acquired from 66 different laboratories, we established that stained tissue appearance can vary drastically because of subtle variations in the staining protocols of different laboratories even when other factors, such as scanners, are kept constant. Additionally, we showed that while deep learning-based models reasonably normalized the staining appearance, they were not ideal because of the nature of the dataset and traditional methods such as histogram matching performed well by leveraging the morphological uniformity across tissue images.

Acknowledgements

This doctoral research has been an interesting character-building exercise for me. I had never thought that I had the consistency or perseverance to undertake such a gargantuan task and complete it, yet that very uncertainty spurred me onward. As I am nearing the end of my Ph.D., I take satisfaction in having risen to the challenge. This has been a deeply immersive journey for me. I experienced many different facets of research and academia that I couldn't during my previous studies. This accomplishment, however, would not have been possible without the invaluable support of many people.

I owe a huge thanks to my supervisor, Pekka Ruusuvuori. His guidance and support were so important as I navigated my doctoral studies. I especially appreciated his constant optimism and positivity, which were a perfect antidote to my, at times, self-critical demeanor. His trust in my ability to work independently made a massive difference. That freedom truly allowed me to grow and find my own way in my research. I am also grateful to Leena Latonen for providing the support and much-needed guidance in the application area, a domain, I had very little knowledge about. The tissue image data from her group was sine quo non for the success of this research. I'd also like to take the opportunity to thank Heikki Huttunen who introduced me to Pekka and perhaps without his recommendation this story wouldn't have even begun.

During my research, most of my peers - the members of BIIT group - were in Tampere and I was in Turku, but somehow, we managed to develop a great working relationship. I consider myself very fortunate to have been part of such a talented, dedicated, and supportive group of researchers. I'd especially like to thank Masi Valkonen, Mira Valkonen, and Kaisa Liimatainen for all the engaging discussions and insightful sparring sessions. I deeply cherish the times we were able to spend together in person. I'd also like to thank my co-author Sonja Koivukoski, for her work exudes dedication and rigor, and the tissue samples she produced for our research were nothing short of a work of art. I am also grateful to my other co-authors: Jouni Härkönen, Marjukka Friman, and Teijo Kuopio for their contributions.

I would like to express my deepest appreciation to my parents for their extraordinary love, effort, and energy poured into my upbringing and education. Their trust and belief in me have been a constant source of strength. To my wonderful siblings, thank you for your unwavering support and understanding. Above all, to my beloved wife, Amna, I couldn't have asked for a more understanding and

supportive partner. Your boundless kindness and the belief you have in me, often exceeding my own, has been a constant motivation, especially during moments of self-doubt. I couldn't have done this without your love and support. And a special thank you to my amazing 18-month-old daughter, Dina, for reminding me to take breaks and for providing the best kind of distraction and an unlimited source of energy.

This research was conducted in the Bioimage informatics research group, at University of Turku, Faculty of Medicine between 2021 and 2025. The research and doctoral studies were financially supported by University of Turku, ERAPerMed, and Academy of Finland.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). TensorFlow: A system for Large-Scale machine learning. *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, (pp. 265-283).

Abels, E., Pantanowitz, L., Aeffner, F., Zarella, M. D., Van der Laak, J., Bui, M. M., . . . others. (2019). Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *The Journal of pathology, 249*, 286-294.

Aiforia. (2013). *Aiforia's clinical solutions*. Retrieved April 4, 2024, from AIFORIA: <https://www.aiforia.com/aiforia-clinical-solutions>

Altini, N., Marvulli, T. M., Zito, F. A., Caputo, M., Tommasi, S., Azzariti, A., . . . De Summa, S. (2023). The role of unpaired image-to-image translation for stain color normalization in colorectal cancer histology classification. *Computer Methods and Programs in Biomedicine, 234*, 107511.

Alturkistani, H. A., Tashkandi, F. M., & Mohammedsaleh, Z. M. (2016). Histological stains: a literature review and case study. *Global journal of health science, 8*, 72.

Ashikhmin, M. (2001). Synthesizing natural textures. *Proceedings of the 2001 symposium on Interactive 3D graphics*, (pp. 217-226).

Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., & Ozcan, A. (2023). Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications, 12*, 57.

Bancroft, J. D., & Gamble, M. (2008). *Theory and practice of histological techniques.* Elsevier health sciences.

Bancroft, J. D., & Stevens, A. (1982). The haematoxylins. *Theory and practices of histological technique*, 109-121.

Bandi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., . . . others. (2018). From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE transactions on medical imaging, 38*, 550-560.

Bayramoglu, N., Kaakinen, M., Eklund, L., & Heikkila, J. (2017). Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. *Proceedings of the IEEE international conference on computer vision workshops*, (pp. 64-71).

Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., . . . others. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama, 318*, 2199-2210.

BenTaieb, A., & Hamarneh, G. (2017). Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging, 37*, 792-802.

Bordoloi, B., Jaiswal, R., Siddiqui, S., & Tandon, A. (2017). Health hazards of special stains. *Saudi Journal of Pathology and Microbiology, 2*, 175-178.

Borhani, N., Bower, J, r., Boppart, S. A., & Psaltis, D. (2019). Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy. *Biomedical optics express, 10*, 1339-1350.

Breen, J., Zucker, K., Allen, K., Ravikumar, N., & Orsi, N. M. (2024). Generative Adversarial Networks for Stain Normalisation in Histopathology. In *Applications of Generative AI* (pp. 227–247). Springer.

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32.

Brown, M. V., McDunn, J. E., Gunst, P. R., Smith, E. M., Milburn, M. V., Troyer, D. A., & Lawton, K. A. (2012). Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. *Genome medicine, 4*, 1-12.

Bulten, W., Kartasalo, K., Chen, P.-H. C., Ström, P., Pinckaers, H., Nagpal, K., . . . others. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nature medicine, 28*, 154-163.

Cai, S., Xue, Y., Gao, Q., Du, M., Chen, G., Zhang, H., & Tong, T. (2019). Stain style transfer using transitive adversarial networks. *Machine Learning for Medical Image Reconstruction: Second International Workshop, MLMIR 2019, Held in Conjunction with MICCAI 2019* (pp. 163-172). Springer.

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., . . . Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine, 25*, 1301-1309.

Carson, F. L., & Cappellano, C. H. (2015). Histotechnology. *A Self-Instructional Text*.

Chai, J., Zeng, H., Li, A., & Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications, 6*, 100134.

Chan, J. K. (2014). The wonderful colors of the hematoxylin--eosin stain in diagnostic surgical pathology. *International journal of surgical pathology, 22*, 12-32.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., . . . Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Chen, Z., Yu, W., Wong, I. H., & Wong, T. T. (2021). Deep-learning-assisted microscopy with ultraviolet surface excitation for rapid slide-free histological imaging. *Biomedical Optics Express, 12*, 5920-5938.

Cheng, Z., Yang, Q., & Sheng, B. (2015). Deep colorization. *Proceedings of the IEEE international conference on computer vision*, (pp. 415-423).

Chlipala, E. A., Butters, M., Brous, M., Fortin, J. S., Archuletta, R., Copeland, K., & Bolon, B. (2021). Impact of preanalytical factors during histology processing on section suitability for digital image analysis. *Toxicologic Pathology, 49*, 755-772.

Cho, H., Lim, S., Choi, G., & Min, H. (2017). Neural stain-style transfer learning using GAN for histopathological images. *arXiv preprint arXiv:1710.08543*.

Chollet, F. (2021). *Deep learning with Python.* Simon and Schuster.

Chong, Y., Thakur, N., Lee, J. Y., Hwang, G., Choi, M., Kim, Y., . . . Cho, M. Y. (2021). Diagnosis prediction of tumours of unknown origin using ImmunoGenius, a machine learning-based expert system for immunohistochemistry profile interpretation. *Diagnostic Pathology, 16*, 1-9.

Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2013: 16th International Conference* (pp. 411-418). Springer.

Cole, F. J. (1949). *The Collected Letters of Antoni van Leeuwenhoek.* Nature Publishing Group UK London.

Cong, C., Liu, S., Di Ieva, A., Pagnucco, M., Berkovsky, S., & Song, Y. (2022). Colour adaptive generative networks for stain normalisation of histopathology images. *Medical Image Analysis, 82*, 102580.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*, 273-297.

Cui, M., & Zhang, D. Y. (2021). Artificial intelligence and computational pathology. *Laboratory Investigation*, 412-422.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). 1*, pp. 886-893. IEEE.

Das, A., Nair, M. S., & Peter, S. D. (2020). Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. *Journal of digital imaging, 33*, 1091-1121.

de Bel, T., Bokhorst, J.-M., van der Laak, J., & Litjens, G. (2021). Residual cyclegan for robust domain transformation of histopathological tissue slides. *Medical Image Analysis, 70*, 102004.

de Bel, T., Hermsen, M., Kers, J., van der Laak, J., & Litjens, G. (2018). Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning* (pp. 151-163). PMLR.

De Bonet, J. S. (1997). Multiresolution sampling procedure for analysis and synthesis of texture images. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, (pp. 217-226).

de Haan, K., Zhang, Y., Zuckerman, J. E., Liu, T., Sisk, A. E., Diaz, M. F., . . . others. (2021). Deep learning-based transformation of H&E stained tissues into special stains. *Nature communications, 12*, 1-13.

Deininger, L., Stimpel, B., Yuce, A., Abbasi-Sureshjani, S., Schönenberger, S., Ocampo, P., . . . Gaire, F. (2022). A comparative study between vision transformers and cnns in digital pathology. *arXiv preprint arXiv:2206.00389*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers*, (pp. 4171-4186).

Dong, R., Pan, X., & Li, F. (2019). DenseU-net-based semantic segmentation of small objects in urban remote sensing images. *IEEE access, 7*, 65347--65356.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., . . . others. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008). Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. *2008 5th IEEE international symposium on biomedical imaging: from nano to macro* (pp. 496-499). IEEE.

Dunn, C., Brettle, D., Cockroft, M., Keating, E., Revie, C., & Treanor, D. (2024). Quantitative assessment of H&E staining for pathology: development and clinical evaluation of a novel system. *Diagnostic Pathology, 19*, 42.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*, 115-118.

Evans, A. J., Henricks, W. H., & Pantanowitz, L. (2017). Implementation of whole slide imaging for clinical purposes: issues to consider from the perspective of early adopters. *Archives of pathology & laboratory medicine, 141*, 944-959.

Fuchs, T. J., & Buhmann, J. M. (2011). Computational pathology: challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics, 35*, 515-530.

Gadermayr, M., Appel, V., Klinkhammer, B. M., Boor, P., & Merhof, D. (2018). Which way round? A study on the performance of stain-translation for segmenting arbitrarily dyed histological images. *Medical Image Computing and Computer Assisted Intervention--MICCAI 2018: 21st International Conference* (pp. 165-173). Springer.

Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing.* Prentice Hall.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems, 27*.

Graham, S., Vu, Q. D., Raza, S. E., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis, 58*, 101563.

Griffin, J., & Treanor, D. (2017). Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology, 70*, 134-145.

Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q. (2019). Deep learning-based image segmentation on multimodal medical imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences, 3*, 162-169.

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE reviews in biomedical engineering, 2*, 147-171.

Han, X. (2017). MR-based synthetic CT generation using a deep convolutional neural network method. *Medical physics, 44*, 1408-1419.

Hanna, T. P., King, W. D., Thibodeau, S., Jalink, M., Paulin, G. A., Harvey-Jones, E., . . . Aggarwal, A. (2020). Mortality due to cancer treatment delay: systematic review and meta-analysis. *bmj, 371*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 770-778).

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems, 30*.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems, 33*, 6840-6851.

Hoda, S. A. (2020). *Robbins and cotran pathologic basis of disease.* Oxford University Press US.

Hoque, M. Z., Keskinarkaus, A., Nyberg, P., & Seppänen, T. (2024). Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion, 102*, 101997.

Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. *2010 20th international conference on pattern recognition* (pp. 2366-2369). IEEE.

Hosseini, M. S., Bejnordi, B. E., Trinh, V. Q.-H., Chan, L., Hasan, D., Li, X., . . . others. (2024). Computational pathology: a survey review and the way forward. *Journal of Pathology Informatics*, 100357.

ICC. (2004). *Image technology colour management-Architecture, profile format, and data structure.* Specification ICC. 1: 2004-10 (Profile version 4.2. 0.0, International Color Consortium.

Indu, M., Rathy, R., & Binu, M. (2016). "Slide less pathology": Fairy tale or reality? *Journal of Oral and Maxillofacial Pathology, 20*, 284-288.

Ingale, V., Singh, R., & Patwal, P. (2021). Image to image translation: Generating maps from satellite images. *arXiv preprint arXiv:2105.09253*.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 1125-1134).

Jaafar, H. (2006). Intra-operative frozen section consultation: concepts, applications and limitations. *The Malaysian journal of medical sciences: MJMS, 13*, 4.

Jahn, S. W., Plass, M., & Moinfar, F. (2020). Digital pathology: advantages, limitations and emerging perspectives. *Journal of clinical medicine, 9*, 3697.

Ji, X., Salmon, R., Mulliqi, N., Khan, U., Wang, Y., Blilie, . . . others. (2025). Physical Color Calibration of Digital Pathology Scanners for Robust Artificial Intelligence Assisted Cancer Diagnosis. *Modern Pathology, 38*, 100715.

Kataria, T., Knudsen, B., & Elhabian, S. Y. (2024). StainDiffuser: multitask dual diffusion model for virtual staining. *arXiv preprint arXiv:2403.11340*.

Kearney, S. J., Lowe, A., Lennerz, J. K., Parwani, A., Bui, M. M., Wack, K., . . . Abels, E. (2021). Bridging the gap: the critical role of regulatory affairs and clinical affairs in the total product life cycle of pathology imaging devices and software. *Frontiers in Medicine, 8*, 765385.

Khan, F., Salahuddin, S., & Javidnia, H. (2022). Deep learning-based monocular depth estimation methods—a state-of-the-art review. *Sensors, 20*, 2272.

Kiernan, J. (2015). *Histological and histochemical methods.* Scion publishing ltd.

Kim, J., Kim, M., Kang, H., & Lee, K. (2019). U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830*.

Koley, S., Bhunia, A. K., Sain, A., Chowdhury, P. N., Xiang, T., & Song, Y.-Z. (2023). Picture that sketch: Photorealistic image generation from abstract sketches. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 6850-6861).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*.

Kumar, G. L., & Kiernan, J. A. (2010). *Special Stains and H & E: Education Guide* (Second ed.). Dako North America, Inc.

Kumar, V., Abbas, A. K., & Aster, J. C. (2013). *Robbins basic pathology.* Elsevier Health Sciences.

Latonen, L., Koivukoski, S., Khan, U., & Ruusuvuori, P. (2024). Virtual staining for histology by deep learning. *Trends in Biotechnology*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*, 2278-2324.

Leong, A. S.-Y., & Zhuang, Z. (2011). The changing role of pathology in breast cancer diagnosis and treatment. *Pathobiology, 78*, 99-114.

Liang, H., Plataniotis, K. N., & Li, X. (2020). Stain style transfer of histopathology images via structure-preserved generative learning. *Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI* (pp. 153-162). Springer.

Lin, J., Pang, Y., Xia, Y., Chen, Z., & Luo, J. (2020). Tuigan: Learning versatile image-to-image translation with two unpaired images. *Computer Vision--ECCV 2020: 16th European Conference* (pp. 18-35). Springer.

Lin, J., Wang, Y., Chen, Z., & He, T. (2020). Learning to transfer: unsupervised domain translation via meta-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*, pp. 11507-11514.

Lin, Y., Zeng, B., Wang, Y., Chen, Y., Fang, Z., Zhang, J., . . . Zhang, Y. (2022). Unpaired multi-domain stain transfer for kidney histopathological images. *Proceedings of the AAAI Conference on Artificial Intelligence*, (pp. 1630-1637).

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., . . . Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports, 6*, 26286.

Liu, L., Xi, Z., Ji, R., & Ma, W. (2019). Advanced deep learning techniques for image style transfer: A survey. *Signal Processing: Image Communication, 78*, 465-470.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 3431-3440).

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision. 2*, pp. 1150-1157. IEEE.

Lu, C., Romo-Bucheli, D., Wang, X., Janowczyk, A., Ganesan, S., Gilmore, H., . . . Madabhushi, A. (2018). Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory investigation, 98*, 1438-1448.

Lujan, G., Quigley, J. C., Hartman, D., Parwani, A., Roehmholdt, B., Van Meter, B., . . . Sowards, C. (2021). Dissecting the business case for adoption and implementation of digital pathology: a white paper from the digital pathology association. *Journal of Pathology Informatics, 12*, 17.

Luo, J., Li, T., Wu, D., Jenkin, M., Liu, S., & Dudek, G. (2024). Hallucination detection and hallucination mitigation: An investigation. *arXiv preprint arXiv:2401.08358*.

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., . . . Thomas, N. E. (2009). A method for normalizing histology slides for quantitative analysis. *2009 IEEE international symposium on biomedical imaging: from nano to macro* (pp. 1107-1110). IEEE.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*, pp. 281-297.

Madabhushi, A., & Lee, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis, 33*, 170-175.

Mahapatra, D., Bozorgtabar, B., Thiran, J.-P., & Shao, L. (2020). Structure preserving stain normalization of histopathology images using self supervised semantic guidance. *Medical Image Computing and Computer Assisted Intervention--MICCAI 2020: 23rd International Conference* (pp. 309-319). Springer.

Mattyus, G., Wang, S., Fidler, S., & Urtasun, R. (2015). Enhancing road maps by parsing aerial images around the world. *Proceedings of the IEEE international conference on computer vision*, (pp. 1689-1697).

Mei, W., Wang, H., Fouhey, D., Zhou, W., Hinks, I., Gray, J. M., . . . Jain, M. (2022). Using deep learning and very-high-resolution imagery to map smallholder field boundaries. *Remote Sensing, 14*, 3046.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis.* John Wiley & Sons.

Nygate, Y. N., Levi, M., Mirsky, S. K., Turko, N. A., Rubin, M., Barnea, I., . . . Shaked, N. T. (2020). Holographic virtual staining of individual biological cells. *Proceedings of the National Academy of Sciences, 117*, 9223-9231.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition, 29*, 51-59.

Oussidi, A., & Elhassouny, A. (2018). Deep generative models: Surve. *2018 International conference on intelligent systems and computer vision (ISCV)* (pp. 1-8). IEEE.

Paige. (2013). *Paige Receives First Ever FDA Approval for AI Product in Digital Pathology*. Retrieved April 4, 2024, from Paige: <https://paige.ai/paige-receives-first-ever-fda-approval-for-ai-product-in-digital-pathology/>

Pang, Y., Lin, J., Qin, T., & Chen, Z. (2021). Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia, 24*, 3859-3881.

Pantanowitz, L., Sharma, A., Carter, A. B., Kurc, T., Sussman, A., & Saltz, J. (2018). Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. *Journal of pathology informatics, 9*, 40.

Pantanowitz, L., Sinard, J. H., Henricks, W. H., Fatheree, L. A., Carter, A. B., Contis, L., . . . Parwani, A. V. (2013). Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Archives of Pathology and Laboratory Medicine, 137*, 1710-1722.

Pantanowitz, L., Valenstein, P. N., Evans, A. J., Kaplan, K. J., Pfeifer, J. D., Wilbur, D. C., . . . Colgan, T. J. (2011). Review of the current state of whole slide imaging in pathology. *Journal of pathology informatics, 2*, 36.

Papadeas, I. a., Amanatiadis, A., & Pratikakis, I. (2021). Real-time semantic image segmentation with deep learning for autonomous driving: A survey. *Applied Sciences, 11*, 8802.

Parwani, A. V., Hassell, L., Glassy, E., & Pantanowitz, L. (2014). Regulatory barriers surrounding the use of whole slide imaging in the United States of America. *Journal of pathology informatics, 5*.

Patel, A., Balis, U. G., Cheng, J., Li, Z., Lujan, G., McClintock, D. S., . . . Parwani, A. (2021). Contemporary whole slide imaging devices and their applications within the modern pathology department: a selected hardware review. *Journal of Pathology Informatics, 12*, 50.

Pecio, A., & Piprek, R. P. (2019). Introduction to histological techniques. *The Histology of Fishes* (pp. 1-12). CRC Press.

Pérez, P., Gangnet, M., & Blake, A. (2023). Poisson image editing. *Seminal Graphics Papers: Pushing the Boundarie*, *2*, pp. 577-582.

Picon, A., Medela, A., Sanchez-Peralta, L. F., Cicchi, R., Bilbao, R., Alfieri, D., . . . Saratxaga, C. L. (2021). Autofluorescence image reconstruction and virtual staining for in-vivo optical biopsying. *IEEE Access, 9*, 32081-32093.

Pilling, M., & Gardner, P. (2016). Fundamental developments in infrared spectroscopic imaging for biomedical applications. *Chemical Society reviews, 45*, 1935-1957.

Rana, A., Lowe, A., Lithgow, M., Horback, K., Janovitz, T., Da Silva, A., . . . Shah, P. (2020). Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. *JAMA network open, 3*, e205111-e205111.

Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer graphics and applications, 21*, 34-41.

Rivenson, Y., de Haan, K., Wallace, W. D., & Ozcan, A. (2020). Emerging advances to transform histopathology using virtual staining. *BME frontiers*.

Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., de Haan, K., & Ozcan, A. (2019). PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light: Science & Applications, 8*, 23.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention--MICCAI 2015: 18th international conference* (pp. 234-241). Munich: Springer.

Rosai, J. (2007). Why microscopy will remain a cornerstone of surgical pathology. *Laboratory investigation, 87*, 403-408.

Ruifrok, A. C., & Johnston, D. A. (2001). Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology, 23*, 291-299.

Sahu, P., Chug, A., Singh, A. P., & Singh, D. (2023). Classification of crop leaf diseases using image to image translation with deep-dream. *Multimedia Tools and Applications, 82*, 35585-35619.

Salehi, P., & Chalechale, A. (2020). Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis. *2020 International conference on machine vision and image processing (MVIP)* (pp. 1-7). IEEE.

Salvi, M., Caputo, A., Balmativola, D., Scotto, M., Pennisi, O., Michielli, N., . . . Fraggetta, F. (2023). Impact of Stain Normalization on Pathologist Assessment of Prostate Cancer: A Comparative Study. *Cancers, 15*, 1503.

Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., . . . others. (2020). A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature communications, 11*, 3877.

Schulte, E. K. (1991). Standardization of biological dyes and stains: pitfalls and possibilities. *Histochemistry, 95*, 319-328.

Shaban, M. T., Baur, C., Navab, N., & Albarqouni, S. (2019). Staingan: Stain style transfer for digital histological images. *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*, 953-956.

Shaked, N. T., Boppart, S. A., Wang, L. V., & Popp, J. (2023). Label-free biomedical optical imaging. *Nature photonics, 17*, 1031-1041.

Shao, Z., Shen, Z., Yuan, S., Tang, J., Wang, Y., Wu, L., & Zheng, W. (2020). AI 2000: A decade of artificial intelligence. *Proceedings of the 12th ACM Conference on Web Science*, (pp. 345-354).

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. *International conference on machine learning* (pp. 2256-2265). PMLR.

Soliman, A., Li, Z., & Parwani, A. V. (2024). Artificial intelligence's impact on breast cancer pathology: a literature review. *Diagnostic Pathology, 19*, 1-18.

Son, A. I., Sokolowski, K., & Zhou, R. (2013). Cryosectioning. *Neural Development: Methods and Protocols*, 301-311.

Sternberg, S. S., Mills, S. E., & Carter, D. (2004). *Sternberg's diagnostic surgical pathology* (Vol. 1). Lippincott Williams & Wilkins.

Stettler, M. (2022, July 2). *Histopathology Datasets*. (M. Stettler, Producer) Retrieved May 2024, from github.com: <https://github.com/maduc7/Histopathology-Datasets>

Ström, P., Kartasalo, K., Olsson, H., Solorzano, L., Delahunt, B., Berney, D. M., . . . others. (2020). Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *The Lancet Oncology, 21*, 222-232.

Sun, H., Xu, H., He, H., Wei, Q., Yan, Y., Chen, Z., . . . Li, T. (2023). A Spatial Analysis of Urban Streets under Deep Learning Based on Street View Imagery: Quantifying Perceptual and Elemental Perceptual Relationships. *Sustainability, 15*, 14798.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 2818-2826).

Tabesh, A., Teverovskiy, M., Pang, H.-Y., Kumar, V. P., Verbel, D., Kotsianti, A., & Saidi, O. (2007). Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE transactions on medical imaging, 26*, 1366--1378.

Thavarajah, R., Mudimbaimannar, V. K., Elizabeth, J., Rao, U. K., & Ranganathan, K. (2012). Chemical and physical basics of routine formaldehyde fixation. *Journal of oral and maxillofacial pathology, 16*, 400-405.

Titford, M. (2006). A short history of histopathology technique. *Journal of Histotechnology*, 99-110.

Tivnan, M., Yoon, S., Chen, Z., Li, X., Wu, D., & Li, Q. (2024). Hallucination Index: An Image Quality Metric for Generative Reconstruction Models. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 449-458). Springer.

Troiano, N. W., Ciovacco, W. A., & Kacena, M. A. (2009). The effects of fixation and dehydration on the histological quality of undecalcified murine bone specimens embedded in methylmethacrylate. *Journal of histotechnolog, 32*, 27-31.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., . . . Navab, N. (2016). Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging, 35*, 1962--1971.

Van der Laak, J., Litjens, G., & Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. *Nature medicine, 27*, 775-784.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

Veta, M., Heng, Y. J., Stathonikos, N., Bejnordi, B. E., Beca, F., Wollmann, T., . . . others. (2019). Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. *Medical image analysis, 54*, 111-121.

Vicory, J., Couture, H. D., Thomas, N. E., Borland, D., Marron, J., Woosley, J., & Niethammer, M. (2015). Appearance normalization of histology slides. *Computerized Medical Imaging and Graphics, 43*, 89-98.

Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Liu, S., . . . others. (2023). Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778*.

Waiting for Your Biopsy or Cytology Test Results. (2023, August 1). Retrieved May 2024, from cancer.org: <https://www.cancer.org/cancer/diagnosis-staging/tests/biopsy-and-cytology-tests/waiting-for-your-biopsy-or-cytology-test-results.html>

Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.

Wang, S., Rong, R., Yang, D. M., Fujimoto, J., Yan, S., Cai, L., . . . others. (2020). Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer research, 80*, 2056-2066.

Webster, J., & Dunstan, R. (2014). Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology. *Veterinary pathology, 51*, 211-223.

Weinstein, R. S., Graham, A. R., Richter, L. C., Barker, G. P., Krupinski, E. A., Lopez, A. M., . . . Gilbertson, J. R. (2009). Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human pathology, 40*, 1057-1069.

Williams, B. J., Knowles, C., & Treanor, D. (2019). Maintaining quality diagnosis with digital pathology: a practical guide to ISO 15189 accreditation. *Journal of Clinical Pathology, 72*, 663-668.

Wittekind, D. (2003). Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. Value and limitations of the hematoxylin-eosin stain. *Biotechnic & histochemistry, 78*, 261--270.

Xu, J., Li, H., & Zhou, S. (2015). An overview of deep generative models. *IETE Technical Review, 32*, 131-139.

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., . . . Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE transactions on medical imaging, 37*, 1348--1357.

Yang, Y., Yang, D., Wang, X., Zhang, Z., & Nawaz, Z. (2021). Testing accuracy of land cover classification algorithms in the qilian mountains based on gee cloud platform. *Remote Sensing, 13*, 5064.

Zarella, M. D., & Rivera Alvarez, K. (2022). High-throughput whole-slide scanning to enable large-scale data repository building. *The Journal of Pathology, 257*, 383-390.

Zarella, M. D., Bowman, D., Aeffner, F., Farahani, N., Xthona, A., Absar, S. F., . . . Hartman, D. J. (2019). A practical guide to whole slide imaging: a white paper from the digital pathology association. *Archives of pathology & laboratory medicine, 143*, 222-234.

Zhang, G., Ning, B., Hui, H., Yu, T., Yang, X., Zhang, H., . . . He, W. (2022). Image-to-images translation for multiple virtual histological staining of unlabeled human carotid atherosclerotic tissue. *Molecular Imaging and Biology*, 1-11.

Zhang, K., Ren, W., Luo, W., Lai, W.-S., Stenger, B., Yang, M.-H., & Li, H. (2022). Deep image deblurring: A survey. *International Journal of Computer Vision, 130*, 2103-2130.

Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing, 20*, 2378--2386.

Zhang, Y., de Haan, K., Rivenson, Y., Li, J., Delis, A., & Ozcan, A. (2020). Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light: Science & Applications, 9*, 78.

Zhang, Y., Huang, L., Pillar, N., Li, Y., Chen, H., & Ozcan, A. (2024a). Super-resolved virtual staining of label-free tissue using diffusion models. *arXiv preprint arXiv:2410.20073*.

Zhang, Y., Huang, L., Pillar, N., Li, Y., Migas, L. G., Van de Plas, R., . . . Ozcan, A. (2024b). Virtual Staining of Label-Free Tissue in Imaging Mass Spectrometry. *arXiv preprint arXiv:2411.13120*.

Zheng, Q., Yang, R., Ni, X., Yang, S., Xiong, L., Yan, D., . . . others. (2022). Accurate diagnosis and survival prediction of bladder cancer using deep learning on histological slides. *Cancers, 14*, 5807.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, (pp. 2223-2232).

List of Figures

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2026

ISBN 978-952-02-0542-3 (PRINT)
ISBN 978-952-02-0543-0 (PDF)
ISSN 0355-9483 (PRINT)
ISSN 2343-3213 (ONLINE)