

---

# 3SGAN: Semisupervised and Multitask GAN for Stain Normalization and Nuclei Segmentation of Histopathological Images

---

[Yifan Chen](#) , Zhiruo Yang , Guoqing Wu , [Qisheng Tang](#) , [Kay Ka-Wai Li](#) , [Ho-Keung Ng](#) , Zhifeng Shi , [Jinhua Yu](#) , [Guohui Zhou](#) \*

Posted Date: 1 December 2025

doi: 10.20944/preprints202512.0031.v1

Keywords: nuclei segmentation; stain normalization; semisupervised learning; multitask learning; generative adversarial networks; digital pathology; glioma



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# 3SGAN: Semisupervised and Multitask GAN for Stain Normalization and Nuclei Segmentation of Histopathological Images

Yifan Chen <sup>1,†</sup>, Zhiruo Yang <sup>1,†</sup>, Guoqing Wu <sup>1</sup>, Qisheng Tang <sup>2</sup>, Kay Ka-Wai Li <sup>3</sup>, Ho-Keung Ng <sup>3</sup>, Zhifeng Shi <sup>2</sup>, Jinhua Yu <sup>1</sup> and Guohui Zhou <sup>1,\*</sup>

<sup>1</sup> College of Biomedical Engineering, Fudan University, Shanghai, 200438, China

<sup>2</sup> Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai, 200040, China

<sup>3</sup> Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China SAR

\* Correspondence: zhoug@fudan.edu.cn

† These authors contributed equally to this work.

## Simple Summary

Analyzing cancer cells from microscope images is challenging due to inconsistent staining colors and the time-consuming nature of manually labeling cells. We developed 3SGAN, an artificial intelligence system that simultaneously standardizes staining colors and identifies cell nuclei using only 5% manually labeled data. Demonstrating superior performance on a large dataset of 1,408 Whole-Slide Images (WSIs) from two medical institutions—encompassing 101 distinct staining styles—our method significantly outperformed existing approaches in both tasks. This technology could accelerate cancer diagnosis and reduce the workload of pathologists while maintaining high accuracy with minimal labeled data.

## Abstract

**Background/Objectives:** Variations in staining styles—arising from differences in tissue preparation, scanners, and laboratory protocols—severely compromise the robustness of automated cell segmentation algorithms in digital pathology. Moreover, manual nucleus annotation is extremely labor-intensive, leading to a scarcity of large-scale, fully annotated datasets for supervised nucleus segmentation. This study proposes a novel framework that simultaneously mitigates staining variability and achieves high-accuracy nucleus segmentation using only minimal annotations. **Methods:** We present 3SGAN, a multitask dual-branch generative adversarial network (GAN) that jointly performs stain normalization and nucleus segmentation in a semi-supervised manner. The framework adopts a teacher-student paradigm: a lightweight teacher model (AttCycle) equipped with attention gates generates reliable pseudo-labels, while a high-capacity student model (TransCycle) leveraging a hybrid CNN-Transformer architecture further refines performance. 3SGAN was trained and evaluated on a large dataset of 1,408 Whole-Slide Images (WSIs) from two medical institutions, encompassing 101 distinct staining styles, with nucleus-level annotations required for only 5% of the data. **Results:** 3SGAN significantly outperformed state-of-the-art methods, achieving substantial improvements in stain normalization quality (RMSE, PSNR, SSIM) and marked gains in nucleus segmentation performance (F1 score, mean IoU, and AJI). External validation on independent MoNuSeg and PanNuke datasets, as well as on previously untested tumor-rich non-ROI regions from our in-house WSIs, confirmed strong generalizability with excellent stain normalization and top-tier segmentation accuracy across diverse staining protocols, tissue types, and pathological patterns. **Conclusions:** The proposed 3SGAN framework demonstrates that high-performance nucleus segmentation and stain normalization can be achieved with minimal annotation requirements, offering a practical and scalable solution for digital pathology applications across diverse clinical settings and staining protocols.

**Keywords:** nuclei segmentation; stain normalization; semisupervised learning; multitask learning; generative adversarial networks; digital pathology; glioma

---

## 1. Introduction

Histopathology based on hematoxylin and eosin (H&E) staining remains the cornerstone and gold standard for most clinical diagnoses. However, achieving cellular-level analysis of pathology images presents significant challenges. Variations in staining techniques lead to inconsistencies in staining styles, complicating automated analysis. Additionally, extracting quantitative measurements of cell morphology and density from histopathology images is hindered by the vast amount of data requiring processing. Moreover, pixel-level annotation at the cellular level is time-consuming and labor-intensive. These challenges underscore the need for efficient and robust methods for cellular-level intelligent analysis.

Cellular-level intelligent diagnosis depends on precise assessment of cell proportions and spatial composition. However, variations in histochemical staining, particularly H&E, introduce significant obstacles for automated computational pathology. H&E staining, widely used in histopathology, highlights tissue and cellular structures by imparting distinct color contrasts. Yet, differences in staining protocols often result in color variations, known as staining differences, which, while negligible for visual diagnosis by pathologists, impair automated tasks such as cell segmentation and classification [1]. Consequently, the importance of staining style normalization has gained increasing recognition in recent years [2–4].

The scarcity of pixel-level annotations further complicates automated nuclei segmentation in pathology images, hindering progress in cellular-level analysis. In digital pathology, fully supervised deep learning models such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) dominate nuclei segmentation [5] and classification tasks [6,7]. Among CNN-based approaches, U-Net and its variants are widely adopted for nuclei segmentation [8,9], while ViT-based architectures and their derivatives have recently gained prominence due to their strong performance on histopathology benchmarks [10]. However, these models require extensive pixel-level annotations to achieve robust segmentation and reliable classification. ViTs, with their larger network capacity and more complex architecture, typically demand even more annotated data than CNNs. Given the labor-intensive nature of acquiring pixel-level labels, fully supervised methods face significant constraints in practice, making semi-supervised learning frameworks a more practical solution for nuclei segmentation.

To address staining variations, which manifest as differences in staining intensity and color, we propose a multitask network structure based on a cycle-consistency Generative Adversarial Network (GAN) [11]. This network enhances the generators of the cycle-consistency GAN by adopting a dual-decoder structure with a hard parameter-sharing strategy. By leveraging shared features extracted from a common encoder, this approach facilitates accurate staining normalization and nuclei segmentation with improved quality.

To overcome the limitations in nuclei segmentation accuracy due to insufficient pixel-level annotations, we introduce a semi-supervised teacher-student framework. The teacher model generates high-quality pseudo-labeled datasets from limited training data, including pseudo-labeled nucleus segmentation and staining normalization outputs. The student model, trained on augmented datasets, further refines segmentation and normalization performance. To capitalize on the stronger modeling capabilities of larger models, the framework employs a low-capacity teacher model and a high-capacity student model.

By integrating the semi-supervised framework with the multitask network architecture, we develop 3SGAN. The teacher model, AttCycle, incorporates dual-decoder generators with attention gates [9] within the multitask network, chosen for its efficacy in learning from small datasets. The student model, TransCycle, integrates a hybrid CNN-Transformer encoder into the dual-decoder

generator, leveraging its capacity for large datasets. The selection of teacher and student models is guided by their network parameter capacities, reflecting their suitability for datasets of varying sizes.

The main contributions of this study are:

- A novel multitask network architecture that simultaneously performs staining normalization and nuclei segmentation, with the two tasks designed to mutually enhance performance, improving overall segmentation accuracy.
- A semi-supervised teacher-student paradigm that mitigates performance degradation in nuclei segmentation due to scarce pixel-level annotations, leveraging unlabeled data to boost accuracy.
- The integration of the multitask architecture and semi-supervised paradigm into 3SGAN, achieving superior staining normalization and nuclei segmentation performance with minimal annotated data on a large dataset of 1,408 WSIs from two medical institutions, encompassing 101 staining styles and only 5% nucleus-level annotations.

The paper is organized as follows. Section 2 reviews related work. Section 3 details the proposed method and dataset. Section 4 presents experimental results and comparisons with state-of-the-art methods. Section 5 discusses the proposed method's effectiveness in addressing the two challenges. Finally, Section 6 concludes with a summary of findings and directions for future research.

## 2. Related Work

### 2.1. Stain Normalization of Histopathological Images

Stain normalization is a critical preprocessing step in histopathological image analysis, addressing variability in staining styles caused by differences in staining protocols, scanner configurations, and tissue reactivity. Existing methods for stain normalization can be broadly categorized into color-matching techniques and style transfer generative models.

Color-matching methods align the color histograms of reference and target images. For instance, Reinhard et al. [12] proposed a global histogram matching approach by converting RGB images to LAB color space and aligning their histograms. However, histogram-based methods often introduce visual artifacts. To mitigate this, stain separation techniques have been developed [13,14]. Khan et al. [14] employed a color-based classifier to estimate the stain matrix, followed by nonlinear color normalization. Such transformations, however, risk altering the image's biological structure. To address this, Vahadane et al. [15] introduced a stain separation method using Sparse Non-negative Matrix Factorization (SNMF) and structure-preserving color normalization (SPCN), which better preserves tissue morphology. Several other SNMF-based approaches have been proposed for stain separation and color unmixing in histopathology [16], demonstrating quantitative benefits for color consistency and downstream analysis. In our work, SNMF-derived stain features are used to characterize stain-style variability and support data-driven style clustering, whereas stain normalization and nuclei segmentation are handled jointly by the proposed 3SGAN framework.

Style transfer generative models, particularly those based on Generative Adversarial Networks (GANs) [17], treat stain normalization as a style transfer problem. Various GAN structures have been explored in medical imaging, including conditional GANs for paired translation [18], cycle-consistent GANs for unpaired translation [19], multi-domain StarGAN that can normalize to multiple target stains with one model [20], and StyleGAN-based methods offering superior high-resolution style control [21]. In histopathology, BenTaieb and Hamarneh [22] combined a stain generation network with an auxiliary task-specific discriminator. However, excessive focus on global appearance may overlook biologically critical regions. Recent works therefore leverage specialized variants such as StainGAN [23] and Colour Adaptive GAN [24], which build upon CycleGAN and incorporate ideas from StarGAN and StyleGAN families to achieve both high visual fidelity and better preservation of tissue structure.

In recent years, diffusion models have emerged as a powerful alternative for stain normalization. Jeong et al. [25] proposed a method using score-based diffusion models, which involves separating H&E stains via SNMF and applying patch-wise normalization with an overlapped moving window

strategy to avoid grid artifacts in Whole-Slide Images (WSIs). This approach ensures high-fidelity normalization while preserving tissue structures. Similarly, Yang et al. [26] introduced SAStainDiff, a self-supervised method based on denoising diffusion probabilistic models. SAStainDiff transforms the normalization task into a pixel-wise color-to-color mapping problem, leveraging stain augmentation and a rescheduled sampling strategy to improve inference speed without compromising performance. These diffusion-based methods demonstrate superior generalization and the ability to handle diverse staining styles, making them promising for robust stain normalization in histopathology.

Despite these advancements, existing unpaired image-to-image translation methods, including diffusion-based approaches, have not fully exploited the synergistic relationship between stain normalization and downstream tasks like nuclei segmentation. Moreover, they often fail to adequately address the challenge of limited annotations, which is critical for accurate segmentation in digital pathology.

## 2.2. Nuclei Segmentation

Nuclei segmentation in histopathology images is a challenging task, complicated by staining variability, overlapping nuclei, and limited pixel-level annotations. Fully supervised deep learning models, such as CNNs, have been widely employed but often struggle to effectively integrate multi-resolution information, resulting in suboptimal pixel-level segmentation outcomes.

Long et al. [27] proposed Fully Convolutional Networks (FCNs), which surpass traditional CNNs in segmentation tasks by leveraging multi-scale information. However, FCNs exhibit inconsistencies in information fusion, leading to reduced sensitivity to fine details. Encoder-decoder architectures, such as U-Net [8], mitigate this issue by incorporating skip connections to better capture contextual details. Nevertheless, the localized nature of convolutional layers restricts U-Net's capacity to model long-range dependencies [28]. To address this limitation, TransUNet [28] integrates transformers into an encoder-decoder framework, enabling global context modeling. Despite its potential, TransUNet's high computational complexity and dependence on large annotated datasets render it less feasible for scenarios characterized by staining variability and limited labeled data.

To overcome these challenges, weakly supervised and semi-supervised approaches have gained traction. X-Net [29], a weakly supervised model combining master and slave U-Nets, employs advanced strategies to enhance segmentation accuracy with fewer annotations. Similarly, Meta-MTL [30] utilizes optimization-based meta-learning and multi-task learning to reduce reliance on extensive labeled data. However, these methods may not consistently deliver robust performance across diverse staining styles.

More recently, foundation models such as the Segment Anything Model (SAM) [31] have been adapted for nuclei segmentation in histopathology. Domain-specific large-scale models, such as MedSAM [32], have demonstrated superior zero-shot segmentation performance in medical imaging contexts. Building on this, pathology-specific segmentation models have emerged. Griebel et al. [33] introduced PathoSAM, a vision foundation model trained on diverse histopathology datasets, achieving state-of-the-art performance in both automated and interactive nucleus instance segmentation.

Despite the impressive zero-shot segmentation capabilities of large-scale models, their general-purpose design limits the scope for prompt engineering optimization, hindering their ability to fully exploit the potential of private datasets. In particular, significant variations in staining styles can lead to unstable segmentation performance. Furthermore, the substantial GPU requirements and prolonged inference times of these models present considerable barriers to their practical implementation in clinical settings.

## 2.3. Multitask Strategy

Multitask learning has shown promise in improving image style transfer by integrating complementary tasks. Zhang et al. [34] combined image-to-image translation with segmentation,

using independent segmentors to enhance synthetic image quality. However, the quality and quantity of synthetic data can destabilize training [34]. Ren et al. [35] proposed a segmentation-renormalization approach to mitigate inter-scanner heterogeneity while preserving anatomical structures across imaging modalities. Similarly, Tomczak et al. [36] employed auxiliary classification and segmentation networks to improve the quality of reconstructed stained images. Jin et al. [37] introduced an asymmetric semi-supervised generative adversarial network (ASSGAN) for breast ultrasound segmentation, leveraging dual generators to mutually supervise segmentation tasks and enhance boundary delineation, though challenges remain in balancing labeled and unlabeled data.

Although these approaches highlight the value of auxiliary tasks in improving image-to-image translation, their effectiveness in segmentation tasks with limited annotations remains constrained. Moreover, incorporating auxiliary tasks elevates training complexity, underscoring the need for efficient network architectures and optimized training strategies.

#### 2.4. Semisupervised Learning

Semisupervised learning offers a robust solution for medical image processing when labeled data is scarce, often by generating pseudo-labeled data to augment training. Yalniz et al. [38] developed a teacher-student pipeline that improved ImageNet performance using up to 1 billion unlabeled images. Shaw et al. [39] demonstrated that semi-supervised teacher-student models can rival fully supervised methods. Earlier, Lee [40] explored a teacher-student paradigm with a high-capacity teacher and a low-capacity student to optimize inference efficiency.

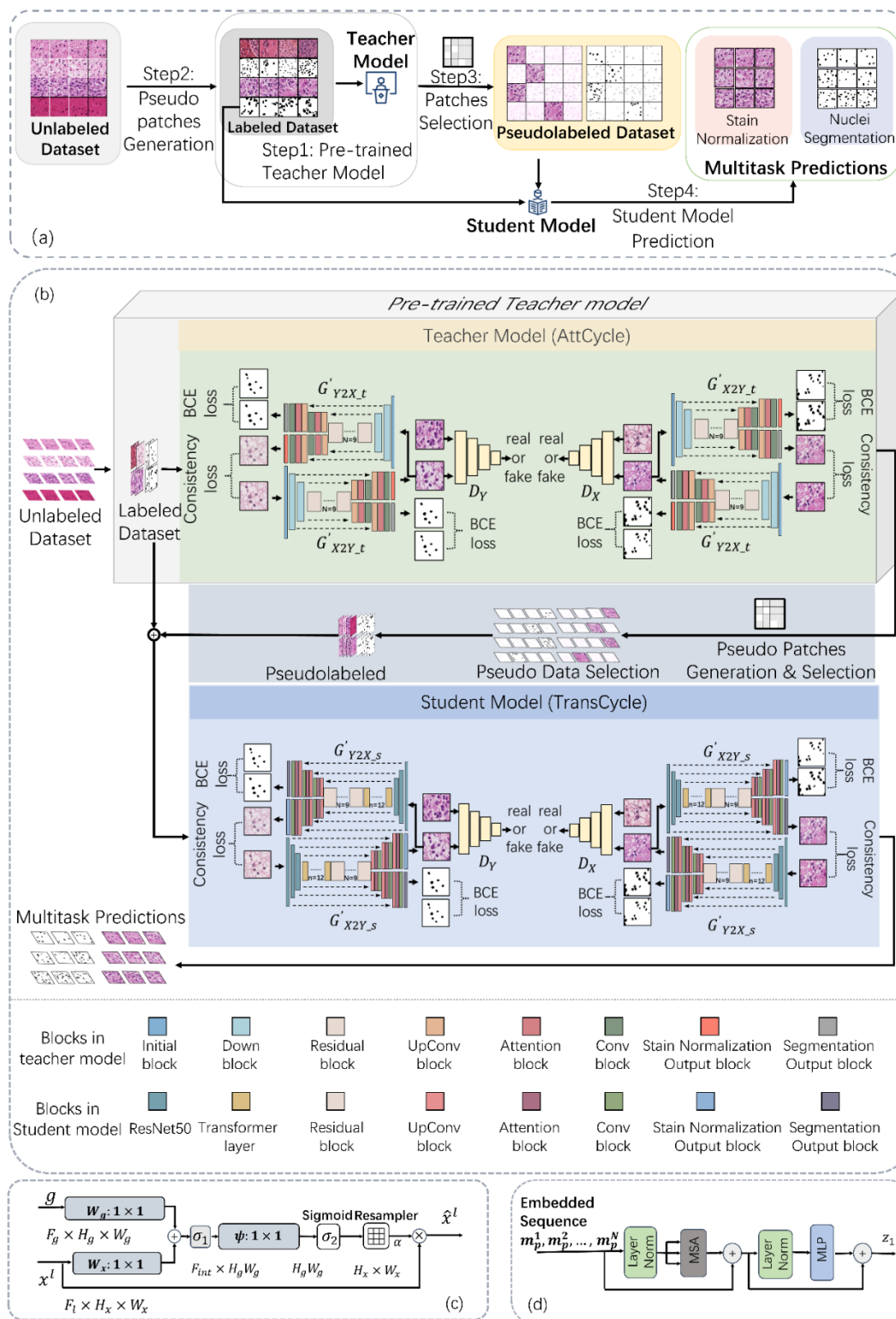
Recent advances include tailored teacher-student frameworks for medical imaging. Zhang et al. [41] introduced a Weakly-Supervised Teacher-Student network (WSTS) for liver tumor segmentation without contrast agents, effectively handling weak labels. Xiao et al. [42] proposed a Dual-Teacher structure combining CNN and transformer teachers to guide a student segmentation model, significantly improving performance with unlabeled data. Lu et al. [43] proposed a framework combining high-uncertainty-aware pseudo-labeling with dual-consistency regularization, using cycle-loss regularization to improve uncertainty estimation and enhance consistency in teacher-student networks. These studies highlight the efficacy of teacher-student frameworks in addressing limited or imperfect annotations in medical imaging.

Despite progress in stain normalization and nuclei segmentation, few studies have tackled both challenges simultaneously, particularly under limited annotations. This paper investigates this gap using a multi-task learning approach, integrating stain normalization and nuclei segmentation to achieve robust performance with minimal labeled data.

### 3. Materials and Methods

#### 3.1. Methodological Overview

To tackle the challenges in histopathology at the cellular level, this study introduces 3SGAN, a novel semisupervised, multitask generative adversarial network. The 3SGAN framework adopts a teacher-student paradigm, with both the teacher model (AttCycle) and student model (TransCycle) built upon a proposed multitask network. This network enhances CycleGAN's generators by incorporating a dual-decoder structure. The teacher and student models in 3SGAN integrate specialized modules to distinguish their functionalities. Figure 1 illustrates the 3SGAN architecture: (a) outlines the data flow, (b) depicts the network structures of AttCycle and TransCycle, and (c) and (d) detail the core modules of the teacher and student models, respectively.



**Figure 1.** The illustration of the proposed 3SGAN. Figure 1(a) shows the data processing flow of the proposed semisupervised teacher-student pipeline. Figure 1(b) shows the data processing flow and model architecture of the 3SGAN. The teacher model (AttCycle, shown in Figure 1(b) teacher model) is trained with the Joint Glioma Labeled Dataset and then infers pseudolabeled data with large quantities of unlabeled data. The student model (TransCycle, shown in Figure 1(b) student model) is trained with a composition of the Joint Glioma Labeled Dataset and pseudolabeled data. The attention gates (AGs) module and transformer module architectures used in AttCycle and TransCycle are illustrated in Figure 1(c) and (d), respectively.

While 3SGAN aligns with the conventional teacher-student paradigm, it introduces key innovations. Both AttCycle and TransCycle leverage the multitask network's dual-decoder structure, enabling simultaneous execution of multiple tasks to improve individual task performance. The teacher model's dual-decoder generator incorporates attention gates to produce high-quality pseudolabeled data for the student model. In contrast, the student model's generator integrates a CNN-Transformer architecture to elevate its performance ceiling. The Transformer's strength in global feature modeling shines when abundant training data is available, allowing the student model to capitalize on the teacher's high-quality pseudolabeled data. This approach mitigates the scarcity of annotated data, enhancing outcomes in nuclei segmentation and staining normalization. By combining CNN's local detail preservation with the Transformer's global feature modeling, the student model optimizes both local and global feature relationships. Figure 1(a) details the semisupervised framework's data processing steps:

- The teacher model (AttCycle) is pretrained using a limited set of labeled samples.
- The pretrained teacher model generates nuclei segmentation and staining normalization results from large volumes of unlabeled data.
- A pseudolabeled dataset selection algorithm (described in Section 3.4) filters the teacher's outputs to create a high-quality pseudolabeled dataset, encompassing pseudolabeled nuclei and staining normalization data.
- The student model (TransCycle) is trained on both the pseudolabeled dataset and the original labeled dataset used for the teacher model.

The teacher model employs attention gates in its generators, while the student model adopts a CNN-Transformer-based encoder architecture, further refining performance.

Figure 1(b) delineates the architectural framework of the teacher model (AttCycle) and the student model (TransCycle) within the 3SGAN system. Consistent with the CycleGAN paradigm, each model comprises two generators and two discriminators. However, unlike the conventional CycleGAN architecture, the generators in 3SGAN adopt a dual-decoder design. Specifically, in the teacher model (AttCycle), the generator  $G'_{X_2Y_t}$  processes an input  $X_t$  and simultaneously produces two outputs: a stain-normalized image  $\tilde{X}_{rs,t}$  achieved by translating image patches from domain  $X$  to domain  $Y$ , and a nuclei segmentation map  $\tilde{X}_{seg,t}$ . This dual-output process is formally represented as:

$$\tilde{X}_{rs,t}, \tilde{X}_{seg,t} = G'_{X_2Y_t}(X_t)$$

Conversely, the generator  $G'_{Y_2X_t}$  performs the inverse style translation. The discriminator  $D_{Y_t}$  evaluates the translated image  $\tilde{X}_{rs,t}$  to ensure its fidelity to the target domain  $Y$ . Given the unpaired nature of the cross-domain data, a cycle-consistency loss is employed to enforce structural preservation, ensuring that the reconstructed image, defined as  $\tilde{X}_{rs,t} = G'_{Y_2X_t}(G'_{X_2Y_t}(X_t))$ , closely approximates the original input  $X_t$ . This approach effectively reduces staining variations in histopathological images, thereby improving the reliability of downstream image analysis. A parallel adversarial mechanism operates within the student model (TransCycle), utilizing the generators  $G'_{X_2Y_s}$  and  $G'_{Y_2X_s}$ , alongside the discriminator  $D_{Y_s}$ .

Additionally, the structural configurations of the attention gate and transformer layer are depicted in Figures 1(c) and Figure 1(d), respectively. The attention gate, originally proposed in Attention U-Net [9], is integrated into the teacher model (AttCycle). This mechanism enhances the network's focus on nucleus-specific features, yielding more accurate segmentation and preserving stylistic integrity during normalization. Meanwhile, the student model (TransCycle) incorporates transformer layers to bolster its global feature extraction capabilities, making it particularly adept for tasks requiring extensive training data.

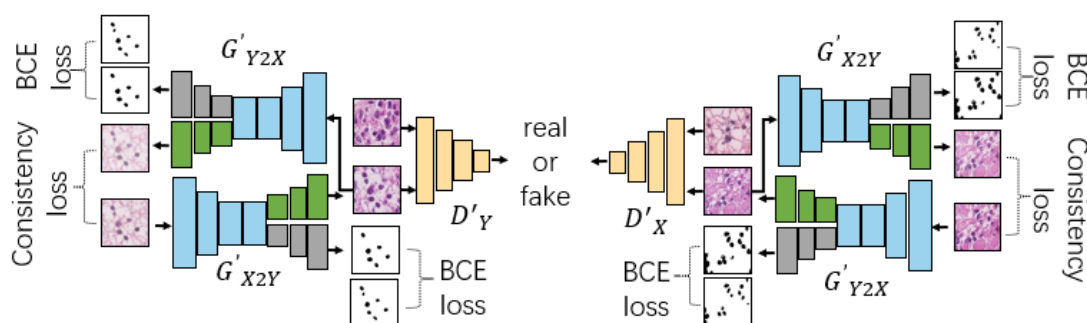
### 3.2. Stain Normalization and Segmentation Multitask Network

In histopathological image analysis, staining normalization and cell nucleus segmentation are mutually influential tasks that enhance each other's performance. Variations in staining across histopathological images can significantly impair the accuracy of computer-aided analysis,

particularly in segmentation tasks. For instance, when processing identical tissue slices with the same deep learning segmentation model, differing staining styles can lead to markedly inconsistent segmentation outcomes. Standardizing the staining style mitigates this issue by reducing noise introduced by staining variability, thereby improving the consistency and precision of segmentation results. Conversely, the segmentation task's emphasis on the biological structure of the nucleus contributes to refining the quality of staining normalization.

To capitalize on this interdependence, we propose a novel multitask network, as illustrated in Figure 2. The network comprises:

- Shared encoders (illustrated in blue)
- Stain normalization branches (depicted in green)
- Segmentation branches (represented in gray)
- PatchGAN-based discriminators (shown in yellow)



**Figure 2.** The schematic diagram of the network model derived from the proposed multi-task strategy to improve the generators in CycleGAN. The blue module represents the shared encoder, the green module denotes the stain normalization branch, the gray module represents the segmentation branch, and the yellow module represents the discriminator using PatchGAN.

This multitask framework enhances both the AttCycle teacher model and the TransCycle student model, as depicted in Figure 1(b).

Our approach leverages a hard parameter sharing strategy to improve the generators within the CycleGAN framework. Specifically, the multitask network employs a shared encoder, enabling the two parallel task-specific decoders to share weights and provide mutual support. This design not only increases the network's efficiency but also facilitates the extraction of features pertinent to both tasks, ultimately leading to superior overall performance.

The nuclei segmentation and stain normalization branches are integrated and included as decoders in  $G'_{X2Y}$  and  $G'_{Y2X}$ . Specifically,  $G'_{X2Y}$  denotes the multitask network that translates image patches from domain X to domain Y, while  $G'_{Y2X}$  performs the inverse transformation. The segmentation branch is trained with nuclear labels of the input images, allowing  $G'_{X2Y}$  to produce outputs for both tasks simultaneously: the nuclei segmentation output  $\tilde{X}_{seg}$  and the stain normalization output  $\tilde{X}_{rs}$ . Hence, we have  $\tilde{X}_{seg}, \tilde{X}_{rs} = G'_{X2Y}(X)$ . The segmentation output  $\tilde{X}_{seg}/\tilde{Y}_{seg}$  in  $G'_{X2Y}/G'_{Y2X}$  is obtained by training the segmentation branch with nuclei annotations  $X_{label}/Y_{label}$  paired with input X/Y. For each image x in domain X and each y in domain Y, the objectives can be expressed as follows:

- Segmentation Terms: We utilized a binary cross entropy loss for the segmentation task in our approach. The use of this segmentation loss allowed for the effective weight updates of the segmentation task decoder within the multitask network. Through this approach, we were able to improve the stain normalization results by focusing specifically on the cell nucleus region with the aid of the segmentation loss. This resulted in more accurate segmentation of cell nuclei.

$$L_{Seg}(\tilde{X}_{seg}, X_{label}) = -E_{\tilde{x}_{seg} \sim \tilde{x}_{seg}, x_{label} \sim X_{label}} \left[ x_{label} \log(\tilde{x}_{seg}) + (1 - x_{label}) \log(1 - \tilde{x}_{seg}) \right] \quad (1)$$

Adversarial Terms: We use the least square loss for adversarial training.  $G'_{X2Y}$  and  $D_Y$  are optimized with analogous optimization applied for  $G'_{Y2X}$  and  $D_X$ .  $\tilde{x}_{rs}$  is the staining normalization output of  $G'_{X2Y}$ . Through the use of this loss function, the dual-decoder generator of our multitask network is capable of generating images that closely resemble the target domain images in terms of color style. Meanwhile, the discriminator is more effective in discerning whether the generated images match the color style of the target domain images.

$$\min_{D_Y} L_{GAN}(D_Y) = \frac{1}{2} E_{y \sim Y} [\|1 - D_Y(y)\|_2^2] + \frac{1}{2} E_{\tilde{x}_{rs} \sim \tilde{X}_{rs}} [\|D_Y(\tilde{x}_{rs})\|_2^2] \quad (2)$$

$$\min_{G'_{X2Y}} L_{GAN}(G'_{X2Y}) = \frac{1}{2} E_{\tilde{x}_{rs} \sim \tilde{X}_{rs}} [\|D_Y(\tilde{x}_{rs}) - 1\|_2^2] \quad (3)$$

- Cycle-Consistency Terms: The cycle-consistency loss is used to enforce the consistency of histopathological structures between X and  $\tilde{X}_{restained}$ :

$$L_{Cyc}(G'_{X2Y}, G'_{Y2X}) = E_{x \sim X} [\|G'_{Y2X}(G'_{X2Y}(x)) - x\|_1] + E_{y \sim Y} [\|G'_{X2Y}(G'_{Y2X}(y)) - y\|_1] \quad (4)$$

- Total objective: This total loss function establishes a connection between the segmentation task and the stain normalization task by facilitating the hard parameter sharing of the common encoder during network training. This approach yields enhanced results for both tasks, leading to improved cell nucleus segmentation and stain normalization. The complete objective function of this total loss to minimize is summarized as following.

$$L(G'_{X2Y}, G'_{Y2X}, D_Y, D_X, X_{label}) = [\lambda_{X2Y} L_{GAN}(G'_{X2Y}) + L_{GAN}(D_Y) + L_{GAN}(G'_{Y2X}) + L_{GAN}(D_X)] \\ + \lambda_{Cyc} L_{Cyc}(G'_{X2Y}, G'_{Y2X}) + \lambda_{Seg} L_{Seg}(\tilde{X}_{seg}, X_{label}) \quad (5)$$

where  $\lambda_{X2Y}$ ,  $\lambda_{Cyc}$  and  $\lambda_{Seg}$  denote the weights of translation loss from the X domain to the Y domain, cycle-consistency loss and segmentation loss, respectively.

### 3.3. Semisupervised Teacher-Student Pipeline

In traditional teacher-student paradigms aimed at reducing model complexity and computational expenses [38,40], the teacher model is typically a large-parameter model valued for its superior feature extraction capabilities, while the student model is designed with fewer parameters to maintain comparable performance with reduced computational and memory demands. However, this conventional approach may falter when the labeled dataset is limited, such as when it comprises only a few hundred images. To overcome this challenge, we propose a novel teacher-student framework that leverages a low-capacity CNN-based teacher model paired with a high-capacity student model featuring a CNN-Transformer hybrid architecture.

Our design choices are driven by the unique strengths of CNNs and Transformers, tailored to the constraints of a small labeled dataset:

- Advantages of CNNs in the Teacher Model: CNNs require fewer annotations for training due to their relatively small parameter count compared to Transformers. This is particularly beneficial in medical image analysis, where annotating pathological cell nuclei is labor-intensive and prone to errors. Complex teacher models often struggle to produce stable, high-quality pseudolabels under such conditions. In contrast, our CNN-based teacher model, AttCycle, excels at generating reliable pseudolabeled data even with minimal annotated samples.
- Benefits of a High-Capacity Student Model: Large models generally outperform smaller ones when trained on abundant data. Thus, we employ a student model, TransCycle, which integrates CNN and Transformer architectures. This hybrid design harnesses CNNs' prowess in preserving fine-grained details and Transformers' strength in modeling global contextual information, aiming to elevate overall performance.

Our approach unfolds in a structured, two-stage process:

### 1. Training the Teacher Model (AttCycle):

- We begin by training AttCycle, a low-capacity CNN-based teacher model, on a small local dataset. AttCycle enhances its focus on cell nucleus regions by incorporating attention gates from Attention U-Net into a multitask network, improving both stain normalization and cell nucleus segmentation.
- Owing to its CNN architecture, AttCycle reliably generates high-quality pseudolabels with limited annotated data, outperforming Transformer-based alternatives in this constrained setting.
- To ensure pseudolabel quality, we apply a sigmoid probability method: only pseudolabels where each confidential pixel's sigmoid probability exceeds a predefined threshold, and the proportion of such pixels surpasses a preset confidence ratio, are retained.

### 2. Training the Student Model (TransCycle):

- The augmented dataset, enriched with high-quality pseudolabels from AttCycle, is then used to train TransCycle, our high-capacity student model. TransCycle employs a hybrid CNN-Transformer encoder within the dual-decoder generators of a multitask network.
- With a greater capacity than AttCycle, TransCycle excels at recovering localized spatial details and enhancing finer features when trained on large datasets, leveraging the complementary strengths of CNNs and Transformers.

Our innovative teacher-student paradigm synergizes the efficiency of a low-capacity CNN-based teacher model with the robustness of a high-capacity CNN-Transformer student model. By addressing the limitations of small labeled datasets, this approach significantly enhances the performance of medical image analysis, offering a practical solution for scenarios where annotated data is scarce.

## 3.4. Architecture and Design Details

The proposed teacher and student models are shown in Figure 1(b) and can be decomposed into their corresponding multitask generators and discriminators.

### 3.4.1. Teacher Model Generators

The teacher model, AttCycle utilizes generators  $G'_{X_2Y_t}$  and  $G'_{Y_2X_t}$  with attention gates (AGs) incorporated into a 3-level upsample-downsample architecture, featuring one encoder and two decoders. AGs are used to emphasize salient features that are passed through the skip connections between the encoder and decoders at the corresponding level. This enables the teacher model to generate high-quality staining normalization and segmentation results, which is crucial for the student model's training. Specifically,  $G'_{X_2Y_t}$  translates images from X domain to Y domain in the teacher model, while  $G'_{Y_2X_t}$  performs the inverse translation. The structure details of  $G'_{X_2Y_t}$  are shown in Figure 1(b) of the teacher model.

AGs select features by using contextual information extracted at coarser scales. Coarse feature maps capture contextual information and emphasize the class and location of target objects. The output of the AG is a multiplication of input feature maps  $x_{i,c}^l$  and attention coefficients  $\alpha_i^l$ , where  $x^l$  denotes higher dimensional image representations, and  $i$  and  $c$  denote spatial and channel dimensions, respectively:

$$\hat{x}_{i,c}^l = x_{i,c}^l \cdot \alpha_i^l \quad (6)$$

A gating vector  $g_i \in \mathbb{R}^{F_g}$  is used for each pixel  $i$  to determine the focus region, and  $F_g$  corresponds to the number of feature maps in layer  $g$ . The gating vector contains contextual information to prune lower-level feature responses. The gating coefficient is obtained by additive attention. Additive attention is formulated as follows:

$$q_{att}^l = \psi^T \left( \sigma_1 (W_x^T x_i^l + W_g^T g_i + b_g) \right) + b_\psi \quad (7)$$

$$\alpha_i^l = \sigma_2 \left( q_{\text{att}}^l(x_i^l, g_i; \theta_{\text{att}}) \right) \quad (8)$$

where  $\sigma_1$  represents the ReLU activation function and  $\sigma_2(x_{i,c}) = 1/1 + \exp(-x_{i,c})$  corresponds to the sigmoid activation function. AG is characterized by a set of parameters  $\theta_{\text{att}}$  containing linear transformations  $W_x \in \mathbb{R}^{F_1 \times F_{\text{int}}}$ ,  $W_g \in \mathbb{R}^{F_g \times F_{\text{int}}}$ ,  $\psi \in \mathbb{R}^{F_{\text{int}} \times 1}$  and bias terms  $b_\psi \in \mathbb{R}$ ,  $b_g \in \mathbb{R}^{F_{\text{int}}}$ . The linear transformations are computed using channelwise  $1 \times 1$  convolutions for the input tensors.

### 3.4.2. Student Model Generators

The student model, TransCycle, uses a CNN-Transformer hybrid architecture as the backbone of its generators' encoder in the proposed multitask structure. This hybrid architecture combines the advantages of detailed high-resolution spatial information from CNN features and the global context encoded by transformers. Additionally, it has more parameters, which may result in better performance when trained on more data [38]. Inspired by U-Net, the CNN-Transformer hybrid model is employed as an encoder where CNN first extracts and generates a feature map for the input. Transformers are then used to obtain self-attention features, which are upsampled to be combined with different high-resolution CNN features skipped from the encoder. Two cascaded decoders consisting of multiple upsampling steps, are used to decode the hidden feature for stain normalization and nuclei segmentation tasks. These cascaded decoders incorporate the hybrid encoder merge feature at different levels via skip connections.

Specifically, this paper implements the hybrid CNN-Transformer encoder by combining ResNet-50 with 12 Transformer layers, illustrated in Figure 1(b) student model. Each Transformer layer comprises multihead self-attention (MSA) and multilayer perceptron (MLP) blocks. Following patch embedding [28], ResNet-50 extracts feature map  $M$  with input  $X \in \mathbb{R}^{H \times W \times C}$ , where  $H \times W$  denotes the spatial resolution and  $C$  denotes the number of channels.  $M$  is tokenized into a sequence of flattened  $1 \times 1$  patches  $M_p \in \mathbb{R}^{N \times (1 \times 1 \times C)}$ , where  $N = H \times W / 1 \times 1$  is the number of patches. Then, we employ positional information embeddings  $E_{\text{pos}}$  to retain the positional information of patches in the previous feature map and map the vectorized patches  $M_p$  into a latent  $D$ -dimensional embedding space  $E_D$  via a trainable linear projection as follows:

$$Z_0 = M_p \cdot E_D + E_{\text{pos}} \quad (9)$$

where  $E_D \in \mathbb{R}^{(1 \times 1 \times C) \times D}$  denotes the linear projection and  $E_{\text{pos}} \in \mathbb{R}^{N \times D}$  denotes the position embedding. In the  $\ell$ th transformer layer, the output  $Z_\ell$  is formulated as follows:

$$Z'_\ell = \text{MSA}(\text{LN}(Z_{\ell-1})) + Z_{\ell-1} \quad (10)$$

$$Z_\ell = \text{MLP}(\text{LN}(Z'_\ell)) + Z'_\ell \quad (11)$$

where  $\text{LN}(\cdot)$  denotes the layer normalization operator. The structure of one transformer layer is shown in Figure 1(d).

### 3.4.3. Pseudolabeled Dataset Selection

In this paper, we propose a semisupervised teacher-student framework for improving nuclei segmentation accuracy in histopathology images. In this framework, the teacher model generates segmentation results for a limited number of annotated images, which are then used as the training set for the student model. However, to minimize noise interference in the teacher output, we introduce a selection method to screen the pseudolabeled data for credibility.

Specifically, we create a Sigmoid probability matrix for each segmentation result generated by the teacher model. To define the data used in our framework, we consider annotated images  $X = \{x_1, x_2, \dots, x_n\}$  and their corresponding labels  $X_{\text{label}} = \{x_{\text{label}_1}, x_{\text{label}_2}, \dots, x_{\text{label}_n}\}$ , teacher model segmentation results  $\tilde{X}_{\text{seg}} = \{x_{\text{seg}_1}, x_{\text{seg}_2}, \dots, x_{\text{seg}_n}\}$ , and unannotated images  $U = \{u_1, u_2, \dots, u_n\}$ . In the labeled dataset  $X$ , the images are categorized into two staining style domains, with an equal distribution of samples in each domain.

To determine the credibility of the segmentation results generated by the teacher model, we apply the Sigmoid activation function to obtain a classification probability map  $\alpha_i$  for each segmentation result  $p_i \in P$  generated for an unannotated image.  $n_i \in N$  is a staining normalization result generated by the teacher model with an unannotated image from  $U$ . We classify each pixel in  $\alpha_i$  as a confident pixel if the Sigmoid prediction value of that pixel is greater than the confidence threshold  $\theta = 0.7$ , and we increase the count of confident pixels by 1 if the confidence value of the pixel at position  $(i, j)$  is greater than  $\theta$ . Finally, we compute the ratio  $\mu_i$  of the number of confident pixels to the total number of pixels in the segmentation result for each unannotated image. If  $\mu_i$  is greater than the specified confidence pixel ratio threshold of  $\mu = 0.6$ , then the segmentation result of that image and its corresponding staining normalization result can be used as pseudolabeled dataset for the student model's training. The nucleus pseudolabels  $P(\mu_i > \mu)$  and pixel-wise nucleus label  $X_{\text{label}}$  are used for the training of student model's segmentation branch. And the pseudostaining normalization results  $N(\mu_i > \mu)$  and their corresponding unlabeled data  $U(\mu_i > \mu)$  work with labeled data  $X$  as the training dataset of the student's staining normalization branch.

Overall, our semi-supervised framework with the selection method for credible pseudolabeled data provide a promising approach for improving nuclei segmentation accuracy in histopathology images. The algorithm of pseudolabeled data selection is outlined as follows.

**Algorithm 1.** Teacher-student pseudolabel selection with confidence thresholding.

Algorithm 1 Pseudolabeled data selection

Input: Labeled data  $X = \{x_1, x_2, \dots, x_n\}$ , pixel-wise label  $X_{\text{label}} = \{x_{\text{label}_1}, x_{\text{label}_2}, \dots, x_{\text{label}_n}\}$ , teacher model segmentation output  $\tilde{X}_{\text{seg}} = \{x_{\text{seg}_1}, x_{\text{seg}_2}, \dots, x_{\text{seg}_n}\}$ , unlabeled data  $U = \{u_1, u_2, \dots, u_n\}$ , confidence threshold  $\theta$ , confidence pixel ratio  $\mu$ .

$$1: L_{\text{Seg}}(\tilde{X}_{\text{seg}}, X_{\text{label}}) = -E_{\tilde{x}_{\text{seg}} \sim \tilde{X}_{\text{seg}}, x_{\text{label}} \sim X_{\text{label}}} \left[ x_{\text{label}} \log(\tilde{x}_{\text{seg}}) + (1 - x_{\text{label}}) \log(1 - \tilde{x}_{\text{seg}}) \right]$$

2: For  $i = 1$  to  $n$  do

3:  $p_i, n_i = \text{teacher}(u_i)$

4:  $\alpha_i = \text{sigmoid}(p_i)$

5: If  $\alpha_{i_{h,w}} \geq \theta$  then

6:     number += 1

7: EndIf

8:  $\mu_i = \frac{\text{number}}{H \times W}$

9: EndFor

10: Training dataset of student model:  $S = X \cup U(\mu_i \geq \mu) \cup N(\mu_i \geq \mu)$ ,  $S_{\text{label}} = X_{\text{label}} \cup P(\mu_i \geq \mu)$

## 4. Results

### 4.1. Experimental Dataset Composition

Our dataset comprises 1,408 H&E-stained glioma WSIs from 1,408 distinct patients sourced from two medical centers: Huashan Hospital of Fudan University (HS; n=848, 60.2%) and the Chinese University of Hong Kong/Prince of Wales Hospital (CUHK; n=560, 39.8%). Glioma cases were retrospectively retrieved from the digital pathology archives of these institutions between January 2014 and December 2017. Cases were included if they had a confirmed glioma diagnosis according to WHO classification, were based on formalin-fixed paraffin-embedded (FFPE) tissue processed with standard H&E staining protocols, contained sufficient tumor content, and yielded technically adequate WSIs without major folds, tears, or severely out-of-focus regions. Slides with severe artifacts or uncertain diagnoses were excluded.

All included slides were FFPE sections captured at 40× magnification and digitized at 0.25 μm/pixel using institutional whole-slide scanners (Hamamatsu NanoZoomer) under default acquisition specifications. By WHO classification, the dataset included 360 grade II WSIs (25.6%), 312 grade III WSIs (22.2%), and 736 grade IV WSIs (52.3%), corresponding to 672 lower-grade (II–III) and 736 high-grade (IV) WSIs.

Tumor-rich regions of interest (ROIs) were selected by experienced neuropathologists. Regions dominated by normal parenchyma, severe tissue loss, pen marks, large hemorrhage, or scanning artifacts were excluded. Selected ROIs were tiled into a total of 52,680 non-overlapping 512×512-pixel patches at 40× magnification (128–768 patches per ROI). Three board-certified neuropathologists annotated glioma nuclei in 1,715 patches across varied staining styles using a multi-stage consensus protocol: the first nuclei performed initial annotations, the second reviewed them, and the third resolved any discrepancies. Structures were labeled as nucleus only if they presented a clear contour and sufficient chromatin contrast; vessels, necrosis, edema, hemorrhage, and other non-nuclear structures were excluded.

Training data labels were created exclusively for glioma tumor cell nuclei, which served as the only annotated structures used to supervise model training. Nuclei pixels were assigned a value of 0, with the background set to 1.

To quantify staining-style variability across the full cohort of 1,408 WSIs, we extracted stain matrices via sparse non-negative matrix factorization (SNMF) and computed optical density statistics for each WSI. A Gaussian mixture model (GMM) with the number of components selected by Bayesian Information Criterion (BIC) identified 101 stain-style clusters representing diverse chromatic profiles (Table 1).

**Table 1.** Summary of data-driven stain-style clusters.

Metric	Value
Total WSIs	1,408
Total stain-style clusters (GMM, BIC-selected)	101
Median WSIs per cluster (IQR)	11 (7–18)
Clusters present in both hospitals	84 (83.2%)
Clusters enriched in Hospital A only	9 (8.9%)
Clusters enriched in Hospital B only	8 (7.9%)
Mean within-cluster $\Delta E_{2000} \pm SD$	$3.4 \pm 0.9$
Mean between-cluster $\Delta E_{2000} \pm SD$	$8.7 \pm 2.1$

The full cohort of 1,408 WSIs was randomly split patient-wise in an approximate 3:1:1 ratio while maintaining balanced cross-domain directions ( $X \rightarrow Y$  and  $Y \rightarrow X$ ) for stain normalization and domain adaptation evaluation (Table 2). During model development, we utilized two stain-style domains: domain X contained patient-level staining variations, whereas domain Y represented a consistent reference style from a single patient. For the segmentation task, nuclei labels were paired with input

patches from domains X and Y. The teacher model was first trained on the limited manually annotated 1,715 patches. The teacher then automatically generated pseudolabels for numerous additional unlabeled patches (500×2, 2,000×2, and 5,000×2), which were used to train the student model in a semi-supervised teacher-student framework. Both models were validated using the same independent testing set.

**Table 2.** Dataset composition and data splits.

Medical Centers	Train (X→Y)	Train (Y→X)	Validate (X→Y)	Validate (Y→X)	Test (X→Y)	Test (Y→X)	Total
Huashan Hospital (HS)	312	320	104	112	104	96	848
CUHK	198	190	66	58	66	82	560
Total	510	510	170	170	170	178	1,408

#### 4.2. Evaluation Metrics and Comparison Methods

Our objective is to address staining style inconsistencies in pathological images by translating images from domain X to domain Y, thereby normalizing diverse staining appearances to a standardized style. To evaluate the quality of stain normalization, we compare the output images,  $\tilde{X}_{rs}$ , with reference images from the Y domain,  $X_{ref}$ , stained using the method proposed by Vahadane et al. [15]. We adopt three metrics for this assessment: Structural Similarity Index (SSIM), Root Mean Squared Error (RMSE), and Peak Signal-to-Noise Ratio (PSNR) [44]. For pixel-level evaluation of nuclei segmentation predictions,  $\tilde{X}_{seg}$ , we employ Mean Intersection over Union (mIOU), F1-score, and Aggregated Jaccard Index (AJI), as proposed by Kumar et al. [45].

In our study, we benchmarked the performance of our proposed 3SGAN model against several established models in medical image analysis. These include traditional segmentation models such as U-Net [8], Attention U-Net [9], TransUNet [28], the Segment Anything Model (SAM) [31], and its medically optimized variant, MedSAM [32], as well as the image-to-image translation model CycleGAN [19]. U-Net, a foundational model for medical image segmentation, utilizes an encoder-decoder architecture with skip connections to enhance contextual feature learning. Attention-UNet extends U-Net by incorporating attention gates to prioritize salient features at coarser scales, improving target localization. TransUNet further augments U-Net with transformers to capture robust global contextual information. SAM, a versatile model pretrained on large-scale natural image datasets, excels in prompt-driven segmentation tasks but struggles with medical images due to domain differences. MedSAM [32], fine-tuned on a large-scale 2D medical image dataset, significantly improves segmentation accuracy for medical applications. CycleGAN, widely adopted for unpaired image-to-image translation, has been effectively applied to stain normalization tasks [46]. Our results demonstrate that 3SGAN outperforms these models in both stain normalization and segmentation tasks, achieving superior performance across the evaluated metrics.

#### 4.3. Implementation Details

The teacher model (AttCycle) and student model (TransCycle) were developed using the PyTorch 1.11.0 framework and trained for 200 epochs with a batch size of 16. Input patches and labels were sized at 256×256 pixels, and the patches were augmented through scaling, random cropping, and flipping. The initial learning rates for both the student and teacher models were 0.0002, which decayed linearly to 0 starting from the 100th epoch. The discriminators for both models had a learning rate of 0.0002. The weights for the translation loss from the X domain to the Y domain ( $\lambda_{X2Y}$ ), cycle-consistency loss ( $\lambda_{Cyc}$ ), and segmentation loss ( $\lambda_{Seg}$ ) were set to 2, 15, and 20, respectively. Adam optimizers were used with  $\beta_1=0.5$  and  $\beta_2=0.999$ . In the proposed multitask network, the output channel number of the stain normalization branch is set to 3, and the output channel number of nuclei segmentation is set to 1.

To facilitate faster convergence of the student model during training, the weights of its transformer layers were initialized with those of a transformer model that was pre-trained on ImageNet [47], which is a large-scale dataset of natural images. This initialization allows the student model to leverage the knowledge learned from the pre-trained model and initialize its parameters close to the optimal values, which can speed up the learning process and improve performance.

#### 4.4. Experimental Design and Results

In our work, we adopt a training scheme based on the teacher-student paradigm. We implement teacher and student models using the CycleGAN framework since the unavailability of pixel-level paired data. To demonstrate the effectiveness of our proposed approach in addressing the challenges of staining style differences and lack of pixelwise manual annotations, we conducted a series of experiments on the Joint Glioma dataset. Specifically, we compared traditional segmentation networks, normalization networks, and our proposed 3SGAN, as detailed in Table 3. Additionally, we studied the performance of multitask architecture and semisupervised teacher-student models on the Joint Glioma dataset within our proposed pipeline, revealing that our approach leads to further improvements in both tasks.

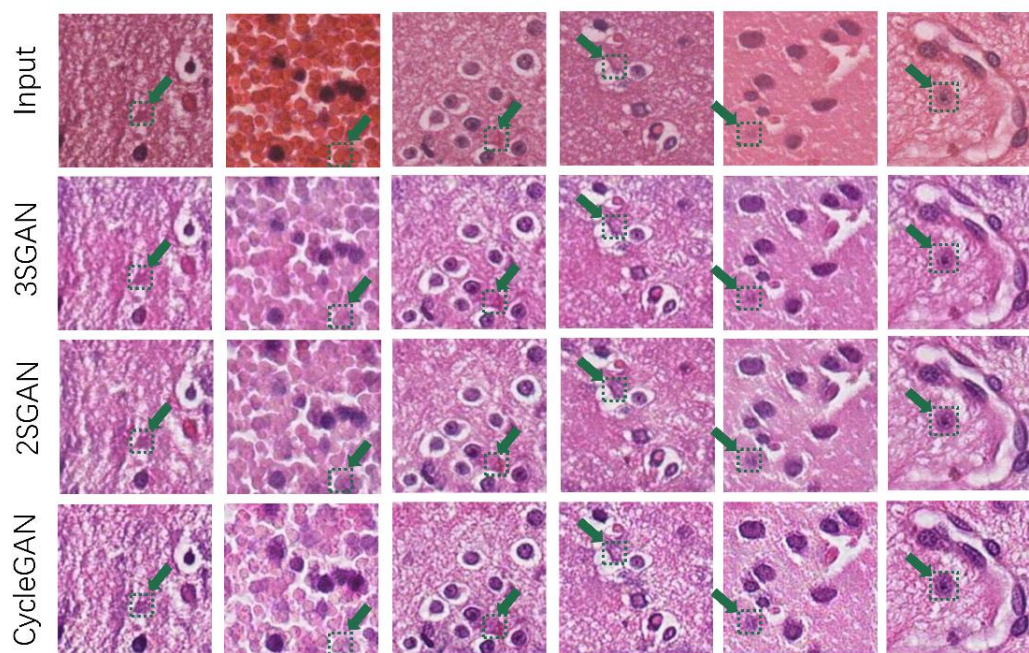
**Table 3.** Comparisons of 3SGAN with SOTA networks in nuclei segmentation and staining style normalization.

Tasks	Methods	Dataset		Metrics		
		Labeled	Pseudo-labeled	F1↑	mIOU↑	AJI↑
Nuclei segmentation	3SGAN	✓	✓(2000×2)	<b>0.8140±0.0042</b>	<b>0.8201±0.0040</b>	<b>0.6915±0.0045</b>
	SAM [31]			0.7325±0.0095	0.7518±0.0090	0.6011±0.0102
	MedSAM [32]			0.7781±0.0080	0.7724±0.0078	0.6325±0.0088
	TransUNet [28]	✓		0.7427±0.0088	0.7554±0.0085	0.5588±0.0105
	Attention-UNet [9]	✓		0.6602±0.0120	0.6947±0.0115	0.5971±0.0110
	U-Net [8]	✓		0.7115±0.0105	0.7358±0.0100	0.5728±0.0112
Stain normalization				<b>RMSE↓</b>	<b>PSNR↓</b>	<b>SSIM↑</b>
	3SGAN	✓	✓(2000×2)	<b>0.0908±0.0020</b>	<b>21.0615±0.1650</b>	<b>0.8556±0.0042</b>
	2SGAN	✓	✓(2000×2)	0.1008±0.0035	20.1358±0.2800	0.8474±0.0065
	CycleGAN [19]	✓		0.1018±0.0040	20.1150±0.3200	0.8140±0.0095

##### 4.4.1. Superiority of the Proposed Multitask Network

To demonstrate the effectiveness of our proposed multitask network in the 3SGAN, we conducted an ablation study by removing the nuclei segmentation branch, resulting in the 2SGAN, a semisupervised stain normalization network. The teacher and student models in the 2SGAN were trained identically to those in the 3SGAN, except for the absence of the segmentation branch.

The results, presented in Table 3, reveal that the 2SGAN's performance was inferior to that of the 3SGAN, highlighting the advantage of the multitask architecture in achieving superior stain normalization quality. Furthermore, visual comparisons of stain normalization outcomes between 3SGAN and 2SGAN are illustrated in Figure 3, where each column displays the results of applying different normalization techniques to the same input image. Green arrows indicate variations in specific regions across the normalized images. Our findings demonstrate that 3SGAN more effectively preserves the original stylistic characteristics of certain regions in the input image compared to 2SGAN and CycleGAN.



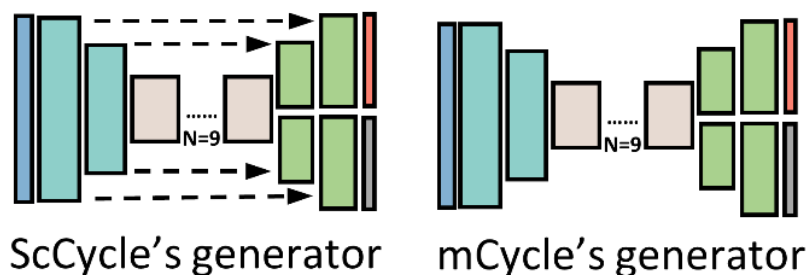
**Figure 3.** Results comparison between 3SGAN, 2SGAN and CycleGAN in stain normalization. Each column of images shows the results of applying different stain normalization methods to the same input image. The green arrows highlight the differences in the same area of the image after undergoing different stain normalization methods. It can be observed that the stain normalization results obtained by 3SGAN preserve the original style of some regions in the input image better than the other two methods. We can learn that with the assistance of a multitask architecture, 3SGAN can generate staining normalized histopathological images with more original information.

#### 4.4.2. Further Improvement Through the Teacher-Student Paradigm

To identify the most suitable teacher and student models for our proposed semisupervised framework, we conducted experiments evaluating various models within our designed multitask architecture. In this framework, the student model is expected to possess greater learning capacity than the teacher model when trained on large datasets. According to Cho and Hariharan [48], a higher number of model parameters correlates with enhanced learning capacity under conditions of abundant training data. To meet the requirements of our semisupervised paradigm, we compared the parameter counts of different models, all of which incorporate our multitask architecture and utilize the cycle-consistency loss from CycleGAN. These models—mCycle, ScCycle, and AttCycle—are distinguished by their unique dual-decoder generator designs.

The mCycle model enhances the single-encoder-single-decoder generator of CycleGAN by adopting a dual-decoder structure. Employing a hard parameter-sharing strategy in multitask learning, mCycle effectively leverages the complementary relationship between stain normalization and nuclei segmentation tasks. It is trained using a combination of supervised loss for nuclei segmentation and unsupervised loss for stain normalization, achieving outstanding performance even with limited annotated data for both tasks.

Building on mCycle, ScCycle also adopts a hard parameter-sharing multitask learning framework to capitalize on the synergy between stain normalization and nuclei segmentation. Additionally, ScCycle incorporates skip connections to better utilize multi-level feature information, thereby improving the quality of multitask outcomes. These technical advancements enable ScCycle to achieve superior accuracy in both stain normalization and nuclei segmentation, even with small-scale annotated datasets.



**Figure 4.** The structures of generators in ScCycle and mCycle. Either of them has a shared encoder and dual decoders. The generator of ScCycle integrates skip connections into mCycle' generator.

The selection of suitable teacher and student models is pivotal to the success of our semisupervised learning framework. In all designed multitask models, the discriminators share an identical structure. Consequently, our comparison of parameter counts focused solely on the generator parameters of each model. For the 3SGAN, we selected TransCycle as the student model due to its generator's high parameter count of approximately 116 million, as shown in Table 4. This substantial parameter capacity enables the student model to achieve optimal performance in stain normalization and nuclei segmentation tasks when provided with sufficient pseudolabeled data.

Based on the findings presented in Table 4, we selected AttCycle as the teacher model for 3SGAN, as it generates high-quality pseudostaining normalization results and nuclei segmentation pseudolabels, effectively supporting our semisupervised learning framework. To further enhance the framework's performance, we trained U-Net, attention-UNet, and TransUNet using the same nuclei-labeled data and trained CycleGAN with images from both source and target staining style domains.

Overall, the careful selection of teacher and student models in 3SGAN, combined with the generation of high-quality pseudolabeled data, is critical to the success of our proposed framework for medical image analysis. This approach has the potential to overcome the limitations of traditional supervised learning methods, enabling accurate analysis of medical images with limited labeled data.

**Table 4.** Generator parameter amounts of our designed multitask models. DD means using dual-decoder architecture, SC means using skip-connection, AT means using attention gates, TR means using Resnet50-transformer encoder architecture.

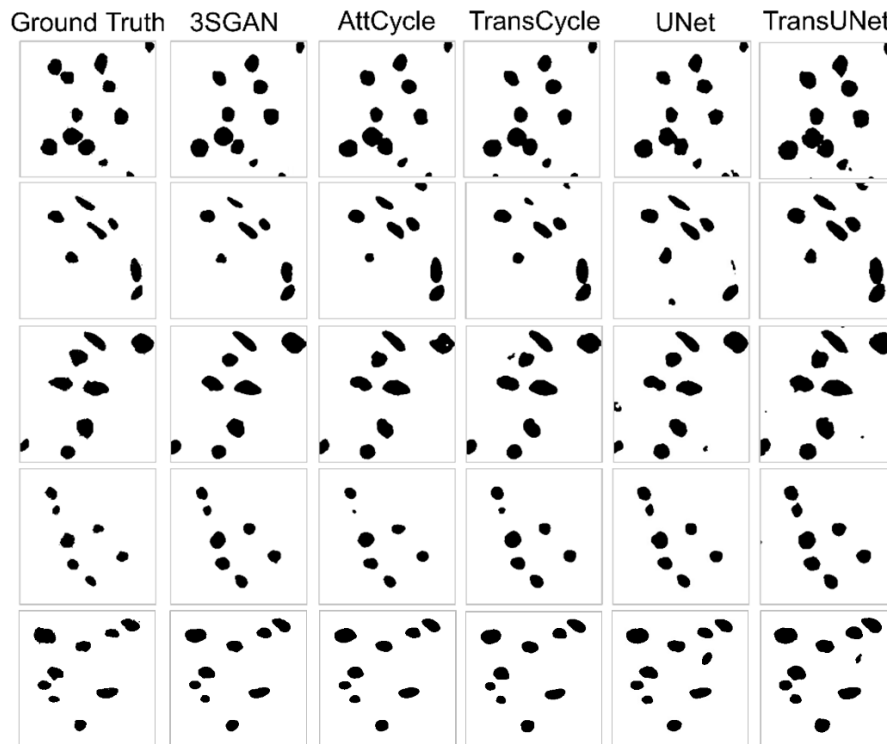
Models	Designs				Parameter Amounts
	DD	AT	SC	TR	
TransCycle	✓			✓	115,938,852
AttCycle	✓	✓			23,519,248
ScCycle	✓		✓		22,740,228
mCycle	✓				22,371,588

Table 5 demonstrates that our proposed semisupervised multitask network, 3SGAN, surpasses its baseline models, AttCycle and TransCycle, in both nuclei segmentation and stain normalization, owing to its effective teacher-student paradigm. This semisupervised framework significantly enhances performance in both tasks. All multitask networks we designed—AttCycle, TransCycle, ScCycle, and mCycle—outperformed traditional state-of-the-art (SOTA) networks in nuclei segmentation and stain normalization.

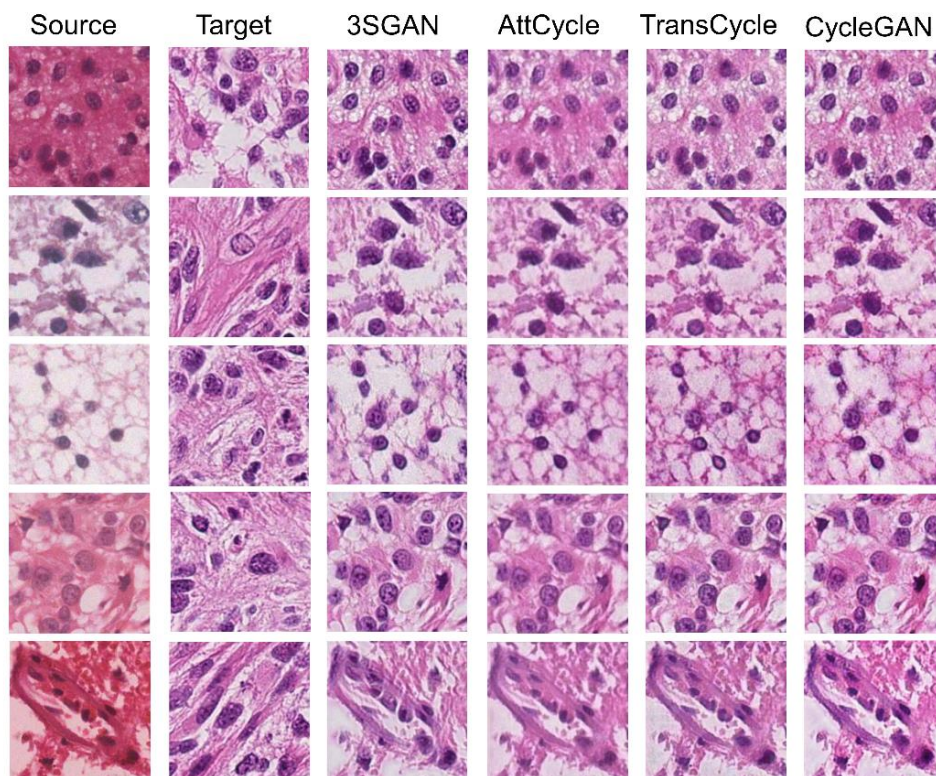
**Table 5.** Performances on nuclei segmentation and stain normalization of our designed multitask networks (AttCycle, TransCycle, ScCycle, mCycle), SOTA segmentation and style transfer networks. Based on their performance on the labeled dataset, we choose AttCycle and TransCycle to be teacher and student models in 3SGAN. Moreover, due to the semisupervised teacher-student pipeline, 3SGAN performs better than AttCycle and TransCycle.

Tasks	Methods	Dataset		Metrics		
		Label ed	Pseudo- labeled	F1↑	mIOU↑	AJI↑
Nuclei Segmenta- tion	3SGAN	✓	✓(2000×2)	<b>0.8140±0.0042</b>	<b>0.8201±0.0040</b>	<b>0.6915±0.0045</b>
	AttCycle	✓		0.8095±0.0065	0.8156±0.0062	0.6835±0.0070
	TransCycle	✓		0.7733±0.0088	0.7888±0.0085	0.6382±0.0092
	ScCycle	✓		0.8089±0.0068	0.8154±0.0065	0.6834±0.0072
	mCycle	✓		0.8008±0.0072	0.8081±0.0070	0.6770±0.0078
	SAM [31]	✓		0.7325±0.0095	0.7518±0.0090	0.6011±0.0102
	MedSAM [32]	✓		0.7781±0.0080	0.7724±0.0078	0.6325±0.0088
	TransUNet [28]	✓		0.7427±0.0088	0.7554±0.0085	0.5588±0.0105
	Attention U-Net [9]	✓		0.6602±0.0120	0.6947±0.0115	0.5971±0.0110
	U-Net [8]	✓		0.7115±0.0105	0.7358±0.0100	0.5728±0.0112
		✓		<b>RMSE↓</b>	<b>PSNR↑</b>	<b>SSIM↑</b>
Stain Normaliza- tion	3SGAN	✓	✓(2000×2)	<b>0.0908±0.0020</b>	<b>21.0615±0.1650</b>	<b>0.8556±0.0042</b>
	AttCycle	✓		0.1016±0.0038	20.0063±0.3100	0.8276±0.0085
	TransCycle	✓		0.1054±0.0042	19.7959±0.3500	0.8275±0.0090
	ScCycle	✓		0.0995±0.0035	20.2664±0.2900	0.8354±0.0080
	mCycle	✓		0.0984±0.0032	20.3820±0.2700	0.8197±0.0088
	CycleGAN [19]	✓		0.1018±0.0040	20.1150±0.3200	0.8140±0.0095

The results, illustrated in Figures 5 and 6, show that 3SGAN achieves the highest segmentation accuracy while producing stain normalization outputs that effectively preserve the original details of diagnostically critical regions, such as nuclei. Quantitatively, 3SGAN’s superior F1, mIoU, and AJI scores directly indicate minimal false-positive nuclei detection, as these metrics penalize over-segmentation. Our multi-task design inherently constrains the generator to maintain realistic nuclear structures, thereby greatly mitigating hallucinations compared to methods like standard CycleGAN.



**Figure 5.** Nucleus segmentation results of 3SGAN, AttCycle, TransCycle, U-Net, and TransUNet.



**Figure 6.** Staining style normalization results of 3SGAN, AttCycle, TransCycle, U-Net, and TransUNet. Different images from the source staining style domain are transferred to the target staining style domain. Target images are representative patches from the reference domain and are not paired with specific input patches. Among the methods, 3SGAN produces normalized outputs that most closely match the target staining style in terms of color tone and visual consistency.

#### 4.4.3. Effect of Different Quantities of Pseudolabeled Data

To evaluate the influence of pseudolabeled data quantity on the student model's performance within the proposed 3SGAN framework, we conducted experiments using varying amounts of unlabeled patches to generate pseudolabeled data, specifically 500×2, 2,000×2, and 5,000×2 patches. The results, shown in Table 6, indicate that increasing the volume of pseudolabeled data enhances the 3SGAN framework's performance in both stain normalization and nuclei segmentation tasks. However, a saturation point exists beyond which further increases in dataset size yield limited performance gains.

**Table 6.** Effect of Pseudolabeled Dataset on Student Model (TransCycle in 3SGAN).

Amount of Pseudo Labels	Metrics					
	Nuclei segmentation			Stain normalization		
	F1↑	mIOU↑	AJI↑	RMSE↓	PSNR↑	SSIM↑
0	0.7733	0.7888	0.6382	0.1054	19.7959	0.8275
	±0.0088	±0.0085	±0.0092	±0.0042	±0.3500	±0.0090
500×2	0.7923	0.8018	0.6722	0.1015	20.1354	0.8413
	±0.0065	±0.0062	±0.0070	±0.0035	±0.2800	±0.0075
2000×2	0.8140	0.8201	0.6915	0.0908	21.0615	0.8556
	±0.0042	±0.0040	±0.0045	±0.0020	±0.1650	±0.0042
5000×2	0.8153	0.8199	0.6928	0.1150	18.9607	0.8231
	±0.0045	±0.0043	±0.0048	±0.0055	±0.4200	±0.0105

#### 4.5. External Validation

To rigorously assess the generalization ability and robustness of 3SGAN beyond its training distribution, we conducted comprehensive external validation in two complementary settings: (i) cross-dataset evaluation on two established public multi-organ nuclei segmentation benchmarks, and (ii) testing on previously unseen, tumor-rich non-ROI regions extracted from our in-house WSIs. We jointly report nuclei segmentation performance (F1 score, mean IoU, and Aggregated Jaccard Index – AJI) and stain normalization quality (RMSE, PSNR, and SSIM), thereby providing a unified evaluation of structural fidelity and stain invariance.

##### 4.5.1. External Validation on MoNuSeg and PanNuke

We first evaluated 3SGAN on the MoNuSeg [49] and PanNuke [50] datasets, which are widely recognized for their substantial inter-organ heterogeneity, diverse nuclear morphologies, and pronounced staining variations. These characteristics make them stringent testbeds for evaluating cross-cohort transferability.

For nuclei segmentation, we compared 3SGAN against classical fully supervised architectures (U-Net, Attention-UNet, TransUNet) as well as recent large-scale foundation models (SAM and MedSAM). For stain normalization, comparisons were made with representative CycleGAN-based and multitask baselines (CycleGAN, AttCycle, TransCycle, ScCycle, and mCycle). Quantitative results are consolidated in Table 7. Across both datasets, 3SGAN consistently achieves the highest segmentation scores (F1, mIoU, AJI) while simultaneously yielding the best stain normalization performance (lowest RMSE, highest PSNR and SSIM), substantially outperforming all competing methods.

**Table 7.** External validation on the MoNuSeg and PanNuke datasets.

Tasks	Methods	MoNuSeg			PanNuke		
		F1↑	mIOU↑	AJI↑	F1↑	mIOU↑	AJI↑
Nuclei Segmentation	U-Net [8]	0.8000	0.6800	0.5900	0.7900	0.6600	0.5700
		±0.0080	±0.0085	±0.0092	±0.0085	±0.0090	±0.0098

	Attention U-Net [9]	0.8200 ±0.0072	0.7000 ±0.0078	0.6100 ±0.0085	0.8100 ±0.0078	0.6800 ±0.0083	0.5900 ±0.0090
	TransUNet [28]	0.8300 ±0.0065	0.7100 ±0.0070	0.6200 ±0.0076	0.8200 ±0.0070	0.6900 ±0.0075	0.6000 ±0.0082
	SAM [31]	0.7900 ±0.0092	0.6600 ±0.0098	0.5700 ±0.0105	0.7800 ±0.0095	0.6500 ±0.0102	0.5600 ±0.0110
	MedSAM [32]	0.8100 ±0.0078	0.6900 ±0.0082	0.6000 ±0.0090	0.8000 ±0.0080	0.6700 ±0.0087	0.5800 ±0.0095
	3SGAN	<b>0.8500</b> <b>±0.0038</b>	<b>0.7300</b> <b>±0.0040</b>	<b>0.6500</b> <b>±0.0043</b>	<b>0.8400</b> <b>±0.0040</b>	<b>0.7200</b> <b>±0.0042</b>	<b>0.6400</b> <b>±0.0045</b>
	<b>Methods</b>	<b>MoNuSeg</b>			<b>PanNuke</b>		
		<b>RMSE↓</b>	<b>PSNR↑</b>	<b>SSIM↑</b>	<b>RMSE↓</b>	<b>PSNR↑</b>	<b>SSIM↑</b>
<b>Stain Normalization</b>	CycleGAN [19]	0.1040 ±0.0048	19.6200±0. 3800	0.8100 ±0.0115	0.1060 ±0.0052	19.4200±0. 4100	0.8080 ±0.0120
	AttCycle	0.0990 ±0.0040	20.1200±0. 3200	0.8350 ±0.0098	0.1010 ±0.0045	19.9400±0. 3600	0.8300 ±0.0105
	TransCycle	0.1020 ±0.0042	19.8800±0. 3500	0.8280 ±0.0100	0.1040 ±0.0048	19.7100±0. 3700	0.8220 ±0.0108
	ScCycle	0.0980 ±0.0038	20.3400±0. 3000	0.8380 ±0.0095	0.1000 ±0.0042	20.0800±0. 3400	0.8340 ±0.0102
	mCycle	0.0970 ±0.0035	20.5100±0. 2800	0.8420 ±0.0090	0.0990 ±0.0040	20.2700±0. 3200	0.8400 ±0.0098
	3SGAN	<b>0.0920</b> <b>±0.0020</b>	<b>21.0300±0.</b> <b>1600</b>	<b>0.8580</b> <b>±0.0040</b>	<b>0.0930</b> <b>±0.0023</b>	<b>20.9200±0.</b> <b>1900</b>	<b>0.8550</b> <b>±0.0045</b>

#### 4.5.2. Generalization to Non-ROI Regions on In-House WSIs

To further examine real-world clinical robustness, we assessed 3SGAN on manually curated tumor-rich regions that were intentionally excluded from both the annotated training set and the pseudo-labeling pipeline (non-ROI regions). These areas typically exhibit complex tissue architecture and were never exposed to the model during training.

Results are presented in Table 8. Even under this challenging out-of-distribution scenario, 3SGAN maintains superior segmentation accuracy and the strongest stain normalization quality, demonstrating that the proposed semi-supervised multitask framework effectively mitigates staining heterogeneity while preserving fine-grained nuclear structures.

**Table 8.** Nuclei segmentation and stain normalization performance on non-ROI regions from Nuclei segmentation and stain normalization performance on non-ROI regions from in-house WSIs.

<b>Tasks</b>	<b>Methods</b>	<b>Metrics</b>		
		<b>F1↑</b>	<b>mIOU↑</b>	<b>AJI↑</b>
<b>Nuclei Segmentation</b>	U-Net [8]	0.7900±0.0082	0.6700±0.0088	0.5800±0.0095
	Attention U-Net [9]	0.8100±0.0075	0.6900±0.0080	0.6000±0.0087
	TransUNet [28]	0.8200±0.0068	0.7000±0.0072	0.6100±0.0078
	SAM [31]	0.8000±0.0090	0.6800±0.0095	0.5900±0.0102
	MedSAM [32]	0.8200±0.0070	0.7000±0.0075	0.6100±0.0080
	3SGAN	<b>0.8700±0.0035</b>	<b>0.7400±0.0038</b>	<b>0.6600±0.0041</b>
<b>Stain Normalization</b>		<b>RMSE↓</b>	<b>PSNR↑</b>	<b>SSIM↑</b>
	CycleGAN [19]	0.1050±0.0050	19.8000±0.4200	0.8120±0.0130
	AttCycle	0.1010±0.0042	20.1800±0.3500	0.8320±0.0105
	TransCycle	0.1040±0.0045	19.9300±0.3800	0.8250±0.0112
	ScCycle	0.1000±0.0040	20.3000±0.3300	0.8370±0.0100
	mCycle	0.0990±0.0038	20.4500±0.3100	0.8230±0.0108

---

3SGAN	$0.0930 \pm 0.0025$	$20.9500 \pm 0.2000$	$0.8570 \pm 0.0045$
-------	---------------------	----------------------	---------------------

---

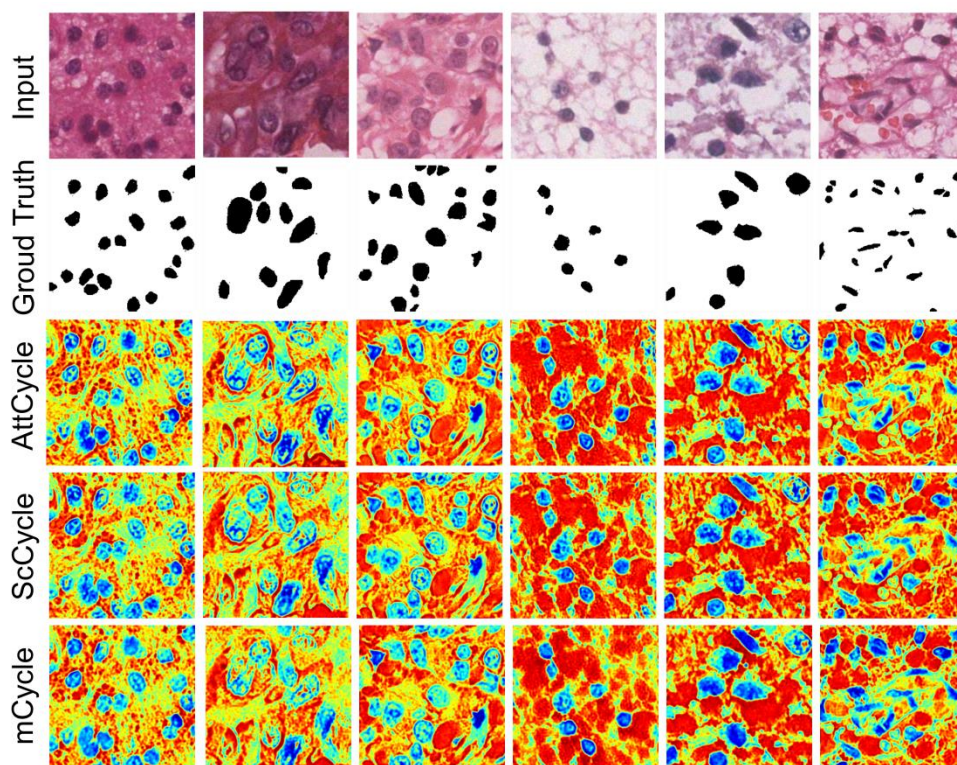
Collectively, the consistent superiority observed across public benchmarks and clinically realistic non-ROI regions underscores the efficacy of 3SGAN’s unified semi-supervised multitask design in delivering state-of-the-art performance with remarkable domain invariance and structural preservation.

## 5. Discussion

In this section, we examine four pivotal design aspects of the proposed 3SGAN framework: (1) the selection of teacher and student models, (2) the relationship between pseudo-labeled dataset size and performance saturation, (3) ablation studies comparing self-training baselines with the semi-supervised teacher-student paradigm, and (4) the role of explicit stain normalization in the era of large foundation models. We also discuss key limitations observed in our experiments and provide insights into potential directions for future work.

### 5.1. Teacher and Student Model Selection

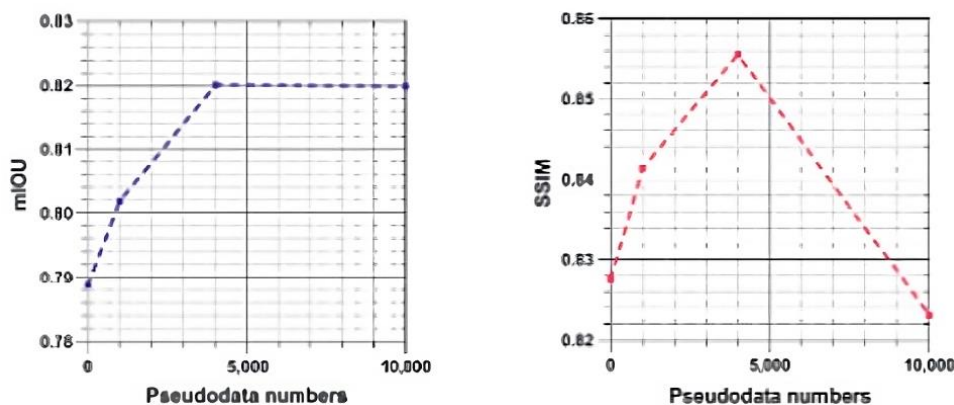
The choice of teacher and student models is critical to the success of our semisupervised framework. Beyond model capacity, preserving the structural integrity of target regions, such as nuclei, is a key consideration in model selection. For the stain normalization task, it is imperative that the model maintains nuclear features post-style translation. To evaluate this, we assessed three multitask models—AttCycle, ScCycle, and mCycle—by generating heatmaps of the output feature bypass to visualize their focus on target regions. As shown in Figure 7, AttCycle, equipped with attention gates, exhibited reduced focus on nuclear regions during stain normalization, thereby better preserving the original structure and color information. Additionally, AttCycle demonstrated superior performance in the nuclei segmentation task. Consequently, we selected AttCycle as the teacher model to generate high-quality pseudostaining normalization patches and corresponding nuclei pseudolabels. This strategic selection ensures that the student model benefits from high-quality pseudolabeled data, enhancing its performance in both nuclei segmentation and stain normalization.



**Figure 7.** Heatmap visualization of the stain normalization results. Red indicates a high attention score, and blue denotes a low score. Patches (first column) are put into AttCycle, ScCycle, mCycle, and the heatmap of the stain normalization result is shown in the third, fourth and fifth columns in turn. Corresponding annotations are illustrated in the second column.

### 5.2. Pseudolabeled Dataset Size Saturation Point.

The capacity of the implemented models influences their saturation points for pseudolabeled dataset size. As illustrated in Figure 8, the performance of 3SGAN improves with increasing pseudolabeled dataset size up to 2,000×2 patches. Beyond this threshold, further increasing the dataset to 5,000×2 patches yields no additional performance gains, indicating that 3SGAN reaches its saturation point at 2,000×2 patches. This suggests that the model's learning capacity is fully utilized at this dataset size, and additional pseudodata provides negligible benefits.



**Figure 8.** Effect of Pseudolabeled Dataset Size on Stain Normalization and Nuclei Segmentation Performance.

### 5.3. Self-Training Ablation Experiments on Teacher and Student Models

When the teacher and student models share the same architecture and parameters (both based on AttCycle), the semisupervised strategy reduces to self-training. In this scenario, the teacher model generates pseudolabeled data and subsequently trains itself with this data. We conducted a statistical analysis to compare the saturation point performance gains of the semisupervised teacher-student paradigm and the self-training approach. The results, presented in Table 9, show that self-training requires 5,000×2 pseudolabeled patches to reach its saturation point, compared to 2,000×2 for the teacher-student paradigm. Despite utilizing more pseudodata, self-training yields only marginal performance improvements compared to the teacher-student paradigm. This is likely due to the amplification of errors and noise in self-training, as the same model iteratively refines its own predictions. In contrast, the teacher-student paradigm leverages distinct models, mitigating error propagation and achieving superior performance.

**Table 9.** Saturation Point Performance Gains of Self-training and Teacher-student Paradigm. The 3SGAN is trained in the proposed teacher-student pipeline; self-training means that the teacher and student models are the same, and both are based on AttCycle.

Different training strategies	Metrics					
	Nuclei segmentation			Stain normalization		
	F1↑	mIOU↑	AJI↑	RMSE↓	PSNR↑	SSIM↑
Semisupervised	0.0407	0.0313	0.0546	-0.0150	1.2656	0.0281
	±0.0031	±0.0028	±0.0040	±0.0018	±0.1120	±0.0035
Self-training	0.0158	0.0140	0.0129	-0.0127	1.2403	0.0471
	±0.0042	±0.0035	±0.0048	±0.0022	±0.1350	±0.0052

#### 5.4. Relevance of Stain Normalization in the Era of Foundation Models

Although foundation models such as SAM and MedSAM are pretrained on vast and heterogeneous corpora, they remain vulnerable to performance degradation when exposed to rare staining protocols, atypical pre-analytic conditions, or site-specific chromatic variations that are underrepresented in their training data. As demonstrated in Table 5, 3SGAN consistently surpasses both SAM and MedSAM across F1, mIoU, and AJI on our nuclei segmentation benchmark, highlighting the continued efficacy of an explicit stain-normalization-plus-segmentation paradigm in domains characterized by pronounced stain variability.

From a methodological perspective, stain normalization retains considerable value when integrated with foundation models. By standardizing input color distributions, it effectively mitigates local domain shifts, reduces reliance on extensive fine-tuning in label-constrained scenarios, and enables the deployment of lightweight, task-specific pipelines (e.g., 3SGAN) that are readily adaptable to resource-limited or regulatory-constrained clinical environments. Moreover, it functions as a transparent and interpretable preprocessing module that harmonizes inputs prior to foundation-model inference, thereby enhancing generalization without costly retraining while preserving mechanistic insight into stain-induced effects on downstream predictions.

Thus, far from being supplanted, stain-normalization-based frameworks exemplified by 3SGAN remain highly relevant and complementary within the contemporary landscape dominated by foundation models.

## 6. Conclusions

In this study, we introduced 3SGAN, a novel semi-supervised multitask framework designed to simultaneously address two major challenges in digital pathology: stain style normalization and nuclei segmentation under severely limited pixel-level annotations. Trained and rigorously evaluated on a large-scale glioma H&E dataset from two institutions—characterized by extensive staining variability and minimal nucleus-level labels—3SGAN effectively alleviates staining inconsistencies while substantially improving segmentation performance and preserving diagnostically critical structural details.

By leveraging a semi-supervised teacher-student paradigm, our approach overcomes the constraints imposed by scarce annotations. Comprehensive ablation studies and comparisons with state-of-the-art methods confirm the superiority of 3SGAN over both traditional supervised segmentation models and contemporary foundation models. External validation on the MoNuSeg and PanNuke datasets further demonstrates excellent generalization beyond glioma histopathology, underscoring the framework's ability to robustly handle staining heterogeneity and complex nuclear morphologies with limited annotated data.

Despite these promising results, rigorous benchmarking on additional challenging histologies (e.g., lymphomas and epithelial malignancies exhibiting extremely dense, small, or highly pleomorphic nuclei) remains an important direction for future work. We also plan to quantitatively assess the preservation of diagnostically relevant structural features during stain normalization and to conduct expert pathologist evaluations of normalization outcomes for clinical acceptability. Finally, refining the pseudo-labeling process to better suppress noise in teacher-generated annotations will be critical for further enhancing the robustness and reliability of the student model.

Overall, 3SGAN represents a significant step toward practical, annotation-efficient computational pathology pipelines capable of delivering consistent, high-quality nuclei segmentation and stain-invariant analysis across diverse clinical settings.

**Author Contributions:** Conceptualization, J.Y.; methodology, Y.C. and Z.Y.; software, Y.C. and Z.Y.; validation, Y.C. and Z.Y.; data curation, Q.T., K.K.-W.L., H.-K.N., Z.S. and G.Z.; writing—original draft preparation, Y.C. and Z.Y.; writing—review and editing, Y.C., G.W., Q.T., K.K.-W.L., H.-K.N., Z.S., G.Z. and J.Y.; visualization, Y.C. and Z.Y.; supervision, Z.S. and G.Z.; project administration, J.Y.; funding acquisition, Z.S., G.Z. and J.Y.;

clinical guidance and data annotation, Q.T., K.K.-W.L., H.-K.N. and Z.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (91959127 and 82072020), and the Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01) and ZJLab.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Fudan University and Huashan Hospital of Fudan University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The MoNuSeg dataset used for external validation is publicly available at <https://monuseg.grand-challenge.org>. The PanNuke dataset used for external validation is publicly available at [https://warwick.ac.uk/fac/cross\\_fac/tia/data/pannuke](https://warwick.ac.uk/fac/cross_fac/tia/data/pannuke). The glioma Whole-Slide Images used in this research from Huashan Hospital of Fudan University and the Chinese University of Hong Kong/Prince of Wales Hospital are available from the corresponding author upon reasonable request.

**Acknowledgments:** During the preparation of this manuscript, ChatGPT-5 was used to assist with English language polishing. The authors have carefully reviewed and edited the output and take ultimate responsibility for the content of the publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* 2016, 33, 170–175. <https://doi.org/10.1016/j.media.2016.06.037>
2. Mahapatra, D.; Bozorgtabar, B.; Thiran, J.P.; Shao, L. Structure preserving stain normalization of histopathology images using self supervised semantic guidance. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*; Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L., Eds.; Springer: Cham, Switzerland, 2020; pp. 309–319. [https://doi.org/10.1007/978-3-030-59722-1\\_30](https://doi.org/10.1007/978-3-030-59722-1_30)
3. Srinidhi, C.L.; Ciga, O.; Martel, A.L. Deep neural network models for computational histopathology: A survey. *Med. Image Anal.* 2021, 67, 101813. <https://doi.org/10.1016/j.media.2020.101813>
4. Lyon, H.O.; De Leenheer, A.; Horobin, R.; Lambert, W.; Schulte, E.; Van Liedekerke, B.; Wittekind, D. Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents. *Histochem. J.* 1994, 26, 533–544. <https://doi.org/10.1007/BF00158587>
5. Ahmad, I.; Xia, Y.; Cui, H.; Islam, Z.U. Dan-nucnet: A dual attention based framework for nuclei segmentation in cancer histology images under wild clinical conditions. *Expert Syst. Appl.* 2023, 213, 118945. <https://doi.org/10.1016/j.eswa.2022.118945>
6. Majumdar, S.; Pramanik, P.; Sarkar, R. Gamma function based ensemble of cnn models for breast cancer detection in histopathology images. *Expert Syst. Appl.* 2023, 213, 119022. <https://doi.org/10.1016/j.eswa.2022.119022>
7. Sethanan, K.; Pitakaso, R.; Srichok, T.; Khonjun, S.; Thannipat, P.; Wanram, S.; Boonmee, C.; Gonwirat, S.; Enkvetchakul, P.; Kaewta, C.; et al. Double amis-ensemble deep learning for skin cancer classification. *Expert Syst. Appl.* 2023, 234, 121047. <https://doi.org/10.1016/j.eswa.2023.121047>
8. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
9. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. arXiv 2018, arXiv:1804.03999.

10. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv 2020, arXiv:2010.11929.
11. Yi, X.; Walia, E.; Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* 2019, 58, 101552. <https://doi.org/10.1016/j.media.2019.101552>
12. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph. Appl.* 2001, 21, 34–41. <https://doi.org/10.1109/38.946629>
13. Onder, D.; Zengin, S.; Sarioglu, S. A review on color normalization and color deconvolution methods in histopathology. *Appl. Immunohistochem. Mol. Morphol.* 2014, 22, 713–719. <https://doi.org/10.1097/PAI.0000000000000003>
14. Khan, A.M.; Rajpoot, N.; Treanor, D.; Magee, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans. Biomed. Eng.* 2014, 61, 1729–1738. <https://doi.org/10.1109/TBME.2014.2303294>
15. Vahadane, A.; Peng, T.; Sethi, A.; Albarqouni, S.; Wang, L.; Baust, M.; Steiger, K.; Schlitter, A.M.; Esposito, I.; Navab, N.; et al. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans. Med. Imaging* 2016, 35, 1962–1971. <https://doi.org/10.1109/TMI.2016.2529665>
16. Xu, J.; Xiang, L.; Liu, Q.; Gilmore, H.; Wu, J.; Tang, J.; Madabhushi, A. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imaging* 2015, 35, 119–130. <https://doi.org/10.1109/TMI.2015.2458702>
17. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q., Eds.; MIT Press: Cambridge, MA, USA, 2014.
18. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. <https://doi.org/10.1109/CVPR.2017.632>
19. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 2223–2232. <https://doi.org/10.1109/ICCV.2017.244>
20. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797. <https://doi.org/10.1109/CVPR.2018.00916>
21. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119. <https://doi.org/10.1109/CVPR42600.2020.00813>
22. BenTaieb, A.; Hamarneh, G. Adversarial stain transfer for histopathology image analysis. *IEEE Trans. Med. Imaging* 2017, 37, 792–802. <https://doi.org/10.1109/TMI.2017.2781228>
23. Shaban, M.T.; Baur, C.; Navab, N.; Albarqouni, S. Staingan: Stain style transfer for digital histological images. In *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 8–11 April 2019; pp. 953–956. <https://doi.org/10.1109/ISBI.2019.8759573>
24. Cong, C.; Liu, S.; Di Ieva, A.; Pagnucco, M.; Berkovsky, S.; Song, Y. Colour adaptive generative networks for stain normalisation of histopathology images. *Med. Image Anal.* 2022, 82, 102580. <https://doi.org/10.1016/j.media.2022.102580>
25. Jeong, J.; Kim, K.D.; Nam, Y.; Cho, C.E.; Go, H.; Kim, N. Stain normalization using score-based diffusion model through stain separation and overlapped moving window patch strategies. *Comput. Biol. Med.* 2023, 152, 106335. <https://doi.org/10.1016/j.compbiomed.2022.106335>
26. Yang, H.; Lyu, M.; Yan, S.; Zhong, T.; Li, J.; Xu, T.; Xie, H.; Liu, S. Sastaindiff: Self-supervised stain normalization by stain augmentation using denoising diffusion probabilistic models. *Biomed. Signal Process. Control* 2025, 107, 107861. <https://doi.org/10.1016/j.bspc.2025.107861>

27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
28. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. arXiv 2021, arXiv:2102.04306.
29. Zhang, T.; Wei, Q.; Li, Z.; Meng, W.; Zhang, M.; Zhang, Z. Segmentation of paracentral acute middle maculopathy lesions in spectral-domain optical coherence tomography images through weakly supervised deep convolutional networks. *Comput. Methods Programs Biomed.* 2023, 240, 107632. <https://doi.org/10.1016/j.cmpb.2023.107632>
30. Han, C.; Yao, H.; Zhao, B.; Li, Z.; Shi, Z.; Wu, L.; Chen, X.; Qu, J.; Zhao, K.; Lan, R.; et al. Meta multi-task nuclei segmentation with fewer training samples. *Med. Image Anal.* 2022, 80, 102481. <https://doi.org/10.1016/j.media.2022.102481>
31. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026. <https://doi.org/10.1109/ICCV51070.2023.00371>
32. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* 2024, 15, 654. <https://doi.org/10.1038/s41467-024-44824-z>
33. Griebel, T.; Archit, A.; Pape, C. Segment anything for histopathology. arXiv 2025, arXiv:2502.00408.
34. Zhang, Z.; Yang, L.; Zheng, Y. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9242–9251. <https://doi.org/10.1109/CVPR.2018.00963>
35. Ren, M.; Dey, N.; Fishbaugh, J.; Gerig, G. Segmentation-renormalized deep feature modulation for unpaired image harmonization. *IEEE Trans. Med. Imaging* 2021, 40, 1519–1530. <https://doi.org/10.1109/TMI.2021.3059726>
36. Tomczak, A.; Ilic, S.; Marquardt, G.; Engel, T.; Forster, F.; Navab, N.; Albarqouni, S. Multi-task multi-domain learning for digital staining and classification of leukocytes. *IEEE Trans. Med. Imaging* 2020, 40, 2897–2910. <https://doi.org/10.1109/TMI.2020.3046334>
37. Jin, S.; Yu, S.; Peng, J.; Wang, H.; Zhao, Y. A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning. *Sci. Rep.* 2023, 13, 6762. <https://doi.org/10.1038/s41598-023-33357-y>
38. Yalniz, I.Z.; Jégou, H.; Chen, K.; Paluri, M.; Mahajan, D. Billion-scale semi-supervised learning for image classification. arXiv 2019, arXiv:1905.00546.
39. Shaw, S.; Pajak, M.; Lisowska, A.; Tsaftaris, S.A.; O'Neil, A.Q. Teacher-student chain for efficient semi-supervised histology image classification. arXiv 2020, arXiv:2003.08797.
40. Lee, E.J.E. Corrective feedback preferences and learner repair among advanced ESL students. *System* 2013, 41, 217–230. <https://doi.org/10.1016/j.system.2013.01.022>
41. Zhang, D.; Chen, B.; Chong, J.; Li, S. Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images. *Med. Image Anal.* 2021, 70, 102005. <https://doi.org/10.1016/j.media.2021.102005>
42. Xiao, Z.; Su, Y.; Deng, Z.; Zhang, W. Efficient combination of cnn and transformer for dual-teacher uncertainty-guided semi-supervised medical image segmentation. *Comput. Methods Programs Biomed.* 2022, 226, 107099. <https://doi.org/10.1016/j.cmpb.2022.107099>
43. Lu, S.; Zhang, Z.; Yan, Z.; Wang, Y.; Cheng, T.; Zhou, R.; Yang, G. Mutually aided uncertainty incorporated dual consistency regularization with pseudo label for semi-supervised medical image segmentation. *Neurocomputing* 2023, 548, 126411. <https://doi.org/10.1016/j.neucom.2023.126411>
44. Sara, U.; Akter, M.; Uddin, M.S. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *J. Comput. Commun.* 2019, 7, 8–44. <https://doi.org/10.4236/jcc.2019.73002>
45. Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A.; Sethi, A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* 2017, 36, 1550–1560. <https://doi.org/10.1109/TMI.2017.2677499>

46. Zhao, B.; Han, C.; Pan, X.; Lin, J.; Yi, Z.; Liang, C.; Chen, X.; Li, B.; Qiu, W.; Li, D.; et al. Restainnet: a self-supervised digital re-stainer for stain normalization. *Comput. Electr. Eng.* 2022, 103, 108304. <https://doi.org/10.1016/j.compeleceng.2022.108304>
47. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, F.F. Imagenet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, Miami, FL, USA, 20–25 June 2009; pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
48. Cho, J.H.; Hariharan, B. On the efficacy of knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4794–4802. <https://doi.org/10.1109/ICCV.2019.00489>
49. Kumar, N.; Verma, R.; Anand, D.; Zhou, Y.; Onder, O.F.; Tsougenis, E.; Chen, H.; Heng, P.A.; Li, J.; Hu, Z.; et al. A multi-organ nucleus segmentation challenge. *IEEE Trans. Med. Imaging* 2019, 39, 1380–1391. <https://doi.org/10.1109/TMI.2019.2947628>
50. Gamper, J.; Alemi Koohbanani, N.; Benet, K.; Khuram, A.; Rajpoot, N. Pannuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*; Reyes-Aldasoro, C.C., Janowczyk, A., Veta, M., Bankhead, P., Sirinukunwattana, K., Eds.; Springer: Cham, Switzerland, 2019; pp. 11–19. [https://doi.org/10.1007/978-3-030-23937-4\\_2](https://doi.org/10.1007/978-3-030-23937-4_2)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.