

VIRTUAL IMMUNOHISTOCHEMISTRY FOR BREAST CANCER BIOMARKER PREDICTION FROM H&E-STAINED IMAGES USING GENERATIVE NETWORK

SHUYING WU✉ AND SHIWEI XU

School of Artificial Intelligence, Wenzhou Polytechnic, Wenzhou City, 325035, Zhejiang Province, China
e-mail: wushuying0905@sina.com

(Received April 10, 2025; revised July 10, 2025; accepted July 31, 2025)

ABSTRACT

Immunohistochemistry (IHC) is essential in diagnostic pathology but is often constrained by cost, time, and limited tissue availability. Virtual IHC staining, which predicts IHC stains from standard hematoxylin and eosin (H&E) images, presents a promising alternative. This study introduces a novel Conditional Generative Adversarial Network (cGAN) architecture based on a U-Net with depthwise separable convolutions to enhance the accuracy and efficiency of virtual IHC staining. This architectural refinement improves computational efficiency while preserving high image quality. We trained and evaluated our model using the BCI and MIST datasets and compared its performance against established image-to-image translation techniques, including Pix2Pix, CycleGAN, and a U-Net variant with standard convolutions. Performance was assessed using quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Fréchet Inception Distance (FID). The results showed that our model outperformed these benchmarks, achieving higher PSNR and SSIM scores, lower MAE and RMSE values, and a significantly reduced FID, indicating superior image quality and closer resemblance to ground-truth IHC images. Furthermore, the integration of depthwise separable convolutions led to a notable decrease in inference time and model size, improving its feasibility for clinical applications. These findings highlight the potential of our method as a significant advancement in virtual IHC staining, offering improved accuracy, efficiency, and suitability for broader clinical use.

Keywords: Generative adversarial network, H&E, IHC, virtual staining.

INTRODUCTION

Breast cancer continues to be one of the most formidable global health challenges of our time, with epidemiological data from the Global Cancer Observatory (GLOBOCAN) revealing approximately 2.3 million new cases diagnosed in 2020 alone, accounting for almost 25% of all cancer diagnoses among women worldwide (Sung *et al.*, 2021). The disease demonstrates remarkable molecular and clinical heterogeneity, encompassing multiple distinct subtypes that vary significantly in their biological behavior, treatment responsiveness, and long-term outcomes. This biological diversity has made the accurate assessment of predictive and prognostic biomarkers an indispensable component of modern breast cancer management. Current international guidelines, including those from the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP), mandate routine immunohistochemical (IHC) evaluation of three critical biomarkers: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status (Allison *et al.*, 2020; Wolff *et al.*, 2018). These molecular markers not

only define clinically relevant subtypes but also serve as crucial therapeutic targets, with ER/PR-positive tumors typically demonstrating responsiveness to endocrine therapies such as selective estrogen receptor modulators (e.g., tamoxifen) or aromatase inhibitors (e.g., letrozole), while HER2-positive cancers derive substantial benefit from targeted therapies like trastuzumab and pertuzumab (Group *et al.*, 2015; Slamon *et al.*, 2001).

The conventional IHC workflow represents a complex, multi-step process that begins with tissue fixation and extends through sectioning, antigen retrieval, primary antibody incubation, secondary antibody application, chromogenic development, and final interpretation by a qualified pathologist (Taylor and Levenson, 2006). Each of these steps requires specialized laboratory infrastructure, expensive reagents, and highly trained technical personnel, often resulting in diagnostic delays ranging from several days to weeks in routine clinical practice (Howat *et al.*, 2014). Moreover, the process is vulnerable to numerous pre-analytical variables including tissue fixation time (with both under-fixation and over-fixation posing problems), processing methods,

storage conditions, and antibody lot variability, all of which can significantly impact staining quality and subsequent interpretation (Goldstein *et al.*, 2003; Engel and Moore, 2011). These technical challenges are particularly acute in low- and middle-income countries where access to consistent, high-quality IHC testing remains constrained by infrastructure limitations, reagent costs, and workforce shortages (Orlandini *et al.*, 2021). The development of robust alternative methods capable of accurately predicting biomarker status while reducing reliance on traditional IHC could therefore have transformative clinical impact, potentially improving diagnostic turnaround times, reducing costs, and making precision oncology more accessible in resource-limited settings.

Hematoxylin and eosin (H&E) staining represents the most fundamental and universally available technique in diagnostic pathology, having served as the cornerstone of histopathological diagnosis for over a century since its introduction in the late 1800s (Fischer *et al.*, 2008). This remarkably durable staining method provides comprehensive morphological information through differential coloration of nuclear (hematoxylin) and cytoplasmic (eosin) components, enabling pathologists to evaluate tissue architecture, cellular morphology, and pathological changes with exceptional clarity (Bancroft and Gamble, 2008). The universal adoption of H&E staining across pathology laboratories worldwide, combined with its standardized protocols, relatively low cost, and routine availability in both prospective and archival specimens, makes it an ideal substrate for computational analysis (Gurcan *et al.*, 2009). Recent revolutionary advances in artificial intelligence, particularly in deep learning and computer vision, have demonstrated that convolutional neural networks can extract latent information from H&E images that extends far beyond human visual perception, enabling remarkably accurate prediction of molecular features, tumor grade, metastatic potential, and clinical outcomes. These developments have given rise to the rapidly evolving field of computational pathology, which seeks to augment and potentially transform traditional diagnostic paradigms through quantitative, data-driven image analysis.

The emergence of generative artificial intelligence, particularly generative adversarial networks (GANs) and more recently diffusion models, has opened unprecedented new possibilities in computational pathology (Goodfellow *et al.*, 2020). These sophisticated deep learning architectures can learn complex, nonlinear mappings between H&E morphological patterns and corresponding protein expression profiles traditionally detected by IHC

through their unique ability to model high-dimensional data distributions (Rivenson *et al.*, 2019). Several pioneering studies have convincingly demonstrated the feasibility of virtual IHC (vIHC), in which remarkably fidelity synthetic IHC images can be generated directly from H&E-stained tissue sections without the need for physical staining procedures (BenTaieb and Hamarneh, 2017). This innovative approach could potentially eliminate the requirement for additional tissue sections and physical IHC staining, simultaneously reducing costs, shortening diagnostic turnaround times, and enabling biomarker assessment in cases where tissue quantity is limited. Recent work has shown particular promise in predicting ER status from H&E images, with some advanced models achieving area under the curve (AUC) scores exceeding 0.90 in independent validation cohorts, approaching the performance of actual IHC testing in some scenarios (Wilm *et al.*, 2022).

Despite these remarkable technological advancements, significant challenges must be rigorously addressed before virtual IHC can be implemented in routine clinical practice. The generalizability of the model across different institutions with varying staining protocols, scanner systems, and tissue processing methods remains a critical concern, as demonstrated by studies showing performance degradation when models trained on the data of one institution are applied to that of another (Holzinger *et al.*, 2019). The interpretability of predictions and identification of specific morphological features driving biomarker classification are essential both for gaining pathologist acceptance and for meeting increasingly stringent regulatory requirements for explainable AI in medical applications. Furthermore, comprehensive clinical validation studies involving large, multi-institutional cohorts with diverse patient populations will be required to demonstrate sufficient robustness and reliability to meet regulatory standards for diagnostic use. Addressing these challenges systematically will be crucial for translating computational pathology advancements into tangible clinical benefits and ensuring equitable global access to these transformative technologies.

In this study, we present a comprehensive deep learning framework for predicting breast cancer biomarker status directly from routine H&E-stained whole slide images. Our approach leverages state-of-the-art conditional generative adversarial networks (cGANs) with depthwise separable convolutions to enhance to produce high-quality virtual IHC images while maintaining robust performance across diverse datasets. Specifically, we propose to predict virtual

IHC staining in breast cancer using two publicly available datasets that focusing on predicting ER, PR, HER2, and Ki67 from H&E-stained images. We performed extensive experiments

RELATED WORK

Various unsupervised techniques have been developed for generating high-resolution histopathology images. For instance, Hou et al. (Hou *et al.*, 2019) proposed an unsupervised segmentation approach for histopathology images, where they synthesized diverse training patches to represent different tissue types. A key aspect of this method is a re-weighting strategy applied to the training loss, which reduces the bias in generalization across the true data distribution. This innovation allowed for the use of a random polygon generator to create synthetic cellular structures, such as nuclear masks, in cases where real data was insufficient for specific tissue types, a scenario where GAN-based methods are often limited. Furthermore, the authors introduced a hybrid synthesis pipeline that merges textures from actual histopathology patches with those generated by GAN models, addressing the challenges of tissue texture variability. This strategy significantly improved generalization, particularly for cancer types with limited available training data. In the recent study, DoanNgan et al., (DoanNgan *et al.*, 2022) proposed a deep learning-based virtual HER2 IHC staining method utilizing a GAN to convert autofluorescence images of breast tissue into bright-field equivalent images, effectively replicating standard chemical staining. This approach eliminates the need for labor-intensive and costly histotechnological processing, significantly reducing analysis time. Validation by board-certified pathologists demonstrated that the virtual staining method achieves accuracy comparable to conventional immunohistochemical staining.

Peng et al. (Peng *et al.*, 2024) introduced a GAN-based virtual staining approach that integrates domain-specific knowledge of HER2 scoring, focusing on nuclei distribution and membrane staining intensity. Their method incorporates a nuclei density estimator to enhance cell alignment between real and generated images and a dedicated branch to improve membrane staining consistency. Using the RegH2I dataset, which includes 2,592 paired H&E-IHC images, they demonstrated the model's effectiveness through extensive experiments on internal and external datasets. This approach addresses limitations in prior virtual staining methods, improving HER2 scoring accuracy and facilitating downstream analysis. Qu et

al. (Qu *et al.*, 2024) proposed a deep learning-based approach for synthesizing IHC-HER2 slides from H&E-stained tissue sections, addressing challenges in multi-magnification pathology image processing. Their model integrates attention mechanisms and a multi-magnification processing strategy to extract and utilize critical information effectively, enhancing image translation quality. The attention module further prioritizes essential features while reducing irrelevant details, improving HER2 biomarker visualization. Rigorous evaluation on a publicly available breast cancer dataset demonstrated superior performance over existing methods, establishing this model as a state-of-the-art solution for virtual IHC staining. Liu et al. (Liu *et al.*, 2020) developed a deep convolutional network model to predict Ki-67 expression directly from H&E-stained slides, demonstrating that molecular-level differences are encoded in tissue morphology. Using a dataset of Ki-67 positive, negative, and background cell images extracted from H&E whole slide images (WSIs), the model was trained and evaluated on both classification and quantification performance. The model achieved a precision of 0.9371 in distinguishing Ki-67 positive and negative cells, with a correlation coefficient of 0.80 between predicted Ki-67 quantification and IHC-derived measurements. This study highlights the potential of deep learning to bridge morphological and molecular information in histopathology.

In (Štepec and Skočaj, 2020), the authors explored and evaluated advanced high-resolution generative models originally designed for face synthesis, demonstrating their effectiveness in the complex domain of digital pathology. Their findings revealed significant improvements in image synthesis, with enhanced quality and resolution of generated images, outpacing traditional approaches compared to supervised models. In a related study, Ma et al. (Ma *et al.*, 2022) proposed using high-resolution RGB images as a guide for super-resolution reconstruction of hyperspectral images. They developed a simple, yet efficient, unsupervised network that combines spatial information from high-resolution RGB images with spectral data from low-resolution hyperspectral images. This technique not only reduces the acquisition time and storage requirements for hyperspectral images but also addresses issues with low-quality spectral bands, enabling the use of hyperspectral imaging in WSI and automated histopathological cancer detection.

In the study presented by (Rizvi *et al.*, 2022), the Histopathology DatasetGAN (HDGAN) framework was introduced as an extension of the semi-supervised DatasetGAN approach, specifically designed for

image generation and segmentation of large-resolution histopathology images. Key modifications were made to the original framework, including enhancements to the generative backbone, selective extraction of latent features from the generator, and a shift to memory-mapped arrays. These changes resulted in significant reductions in memory usage, making the framework more efficient and suitable for medical imaging applications. HDGAN's performance was evaluated using a high-resolution thrombotic microangiopathy tile dataset, demonstrating its strong capabilities in generating image annotations. Similarly, in (Li *et al.*, 2022), Li *et al.* introduced a multi-scale GAN for generating and segmenting large-scale, high-resolution histopathology images. This model employs a pyramid of GAN architectures, each focused on generating and segmenting images at different scales. Using semantic masks, the generative model excelled in synthesizing visually realistic histopathological images. More recently, a coarse-to-fine sampling strategy was proposed in (Harb *et al.*, 2024) to address the challenge of generating high-resolution whole-slide images (WSIs). This method involves starting with a low-resolution image and progressively increasing its resolution, incorporating finer details at each step using a diffusion model to refine the image quality.

Although several deep learning methods for generating IHC images have been proposed, their effectiveness is limited by complex tissue architecture, variability in cellular morphology, potential for nonspecific staining, and differences in antibody reactivity. To overcome these challenges, we propose the novel cGAN model, which integrates generative adversarial networks with depthwise convolution techniques to enhance the generation of high-quality IHC images.

PROPOSED METHOD

In this study, we employed a cGAN to predict IHC staining images from H&E stained tissue images. The architecture of the cGAN is composed of two main components: the generator network (G) and the discriminator network (D). The generator is designed with an encoder-decoder architecture, consisting of seven encoding layers and seven decoding layers, as illustrated in Fig. 1-up. A key feature of this design is the use of depthwise separable convolutions in the encoder components, which substantially improve the model's computational efficiency and effectiveness.

In the encoder network, each layer is constructed using depthwise separable convolutions, which are

followed by batch normalization and a leaky ReLU activation function with a slope of 0.2. Depthwise separable convolutions break down the traditional convolution operation into two distinct steps: the depthwise convolution and the pointwise convolution (1×1). In the depthwise convolution, a separate filter is applied to each input channel independently. This operation captures the spatial features of each channel individually, making it more computationally efficient. Afterward, a pointwise convolution (1×1 kernel) is applied across all the channels, combining the outputs from the depthwise convolution. This step enables the model to capture inter-channel relationships, learning how different feature maps interact.

The use of depthwise separable convolutions provides significant advantages, particularly in terms of computational efficiency and reduced model size. By decomposing the standard convolution into two simpler operations, the number of parameters and computations is drastically reduced compared to traditional convolutional layers, especially when working with a large number of input channels. This reduction leads to faster training and inference times, as well as lower memory requirements. These factors make the model more suitable for deployment on resource-constrained devices, such as edge devices or systems with limited processing power and memory.

In this architecture, the first and last encoding layers (En_1 and En_7) do not incorporate batch normalization. This design choice differentiates them from the other encoding layers, which helps prevent overfitting and allows the model to generalize better. The decoding layers mirror the structure of the encoder but employ transposed depthwise separable convolutions to upsample the feature maps generated during encoding. The decoding layers are followed by batch normalization, dropout (applied only in the first three decoding layers, Dn_1 , Dn_2 , and Dn_3), and ReLU activation functions to further refine the image generation process.

The final decoding layer (Dn_7) does not include a skip connection, which is commonly used in encoder-decoder architectures to maintain high-resolution features across layers. Instead, this design choice aims to focus on producing the final output of the network without overly relying on previous layer features, which could lead to unnecessary complexity. The depthwise separable convolution layers in the encoder reduce the activation map size by a factor of 2, while the transposed depthwise separable convolutions in the decoder increase the map size by a factor of 2, effectively preserving spatial dimensions during the encoding and decoding processes.

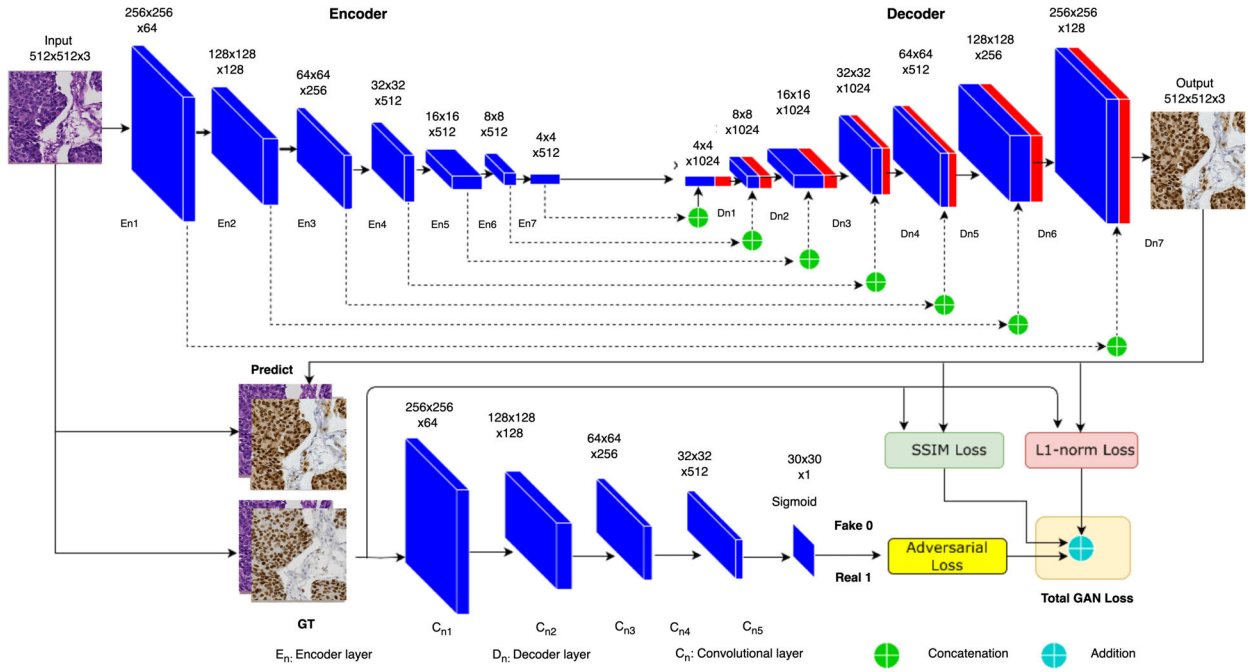


Fig. 1. Illustration of the proposed model.

Finally, the network employs a *TanH* activation function in the last layer to generate the predicted IHC images from the input H&E images. The *TanH* function is suitable for this task as it maps the output values to a continuous range between -1 and 1, which matches the expected range of pixel values in the generated IHC images. The overall design of this network allows for the efficient transformation of H&E images into high-quality virtual IHC images, offering a promising solution for automating the process of IHC staining while reducing the computational load and memory requirements compared to traditional methods.

The discriminator network (D), shown in Fig. 1-down, also benefits from the computational efficiency offered by depthwise separable convolutions, consisting of five such layers. These layers follow a structure similar to the encoding layers of the generator network. Specifically, each layer comprises a depthwise convolution followed by a pointwise convolution, batch normalization (applied after the second, third, and fourth convolutional layers, Cn_2 , Cn_3 , and Cn_4), and a leaky ReLU activation function with a slope of 0.2. However, the final layer in the discriminator is an exception, as it uses a sigmoid activation function instead of the leaky ReLU, transforming the output to a probability value between 0 and 1. The output of the discriminator is a 30×30 matrix, with each element representing the probability that a corresponding 70×70 patch from the input image is a genuine IHC image. This matrix indicates

whether a given patch is classified as real (from the actual IHC images) or fake (generated by the model). The decision made by the discriminator about the authenticity of each patch is a critical step in the adversarial training process, guiding the generator to improve its ability to produce realistic IHC images.

The generator loss function (ℓ_{Gen}) incorporates three key components: adversarial loss, L1 loss, and Structural Similarity Index Measure (SSIM) loss, each contributing to different aspects of image quality and realism in the generated outputs.

$$\begin{aligned} \ell_{Gen}(G, D) = & \mathbb{E}_{x,y,z} [-\log D(x, G(x, z))] \\ & + \lambda \mathbb{E}_{x,y,z} [\ell_{L1}(y, G(x, z))] \\ & + \alpha \mathbb{E}_{x,y,z} [\ell_{SSIM}(y, G(x, z))] \end{aligned} \quad (1)$$

Where x denote the input H&E image, y represent the corresponding ground truth IHC image, and z be the random noise input fed into the generator. The generator produces an IHC image $G(x, z)$ based on these inputs, while $D(x, G(x, z))$ indicates the discriminator's probability that the generated image is real. The loss function is weighted by λ and α , which control the relative importance of the L1 loss and the SSIM loss, respectively. The L1 loss term minimizes the pixel-wise differences between the generated and ground truth IHC images, promoting overall image similarity. The adversarial loss drives the generator to create images with realistic high-frequency details

that can deceive the discriminator. Meanwhile, SSIM loss focuses on enhancing the structural similarity and boundary sharpness of the generated images, capturing perceptual qualities that go beyond pixel-wise accuracy. In this study, the values of λ and α were set at 100 and 75, respectively, ensuring a balanced contribution of each loss component in optimizing the generator.

The discriminator loss function (ℓ_{Disc}) is defined as:

$$\ell_{Disc}(G, D) = \mathbb{E}_{x,y}[-\log D(x, y)] + \mathbb{E}_{x,y,z}[-\log(1 - D(x, G(x, z)))] \quad (2)$$

where $D(x, y)$ represents the discriminator's probability that the real IHC image is real. The loss function trains the discriminator to correctly tell apart real IHC images from generated ones. By minimizing $-\log(D(x, y))$, the discriminator learns to recognize real IHC images. At the same time, by minimizing $-\log(1 - D(x, G(x, z)))$, the discriminator learns to identify the generated images as fake. This back-and-forth training process helps the generator create more realistic IHC images over time.

EXPERIMENTAL RESULTS AND DISCUSSION

DATASET

In this study, we made use of two publicly available datasets to evaluate the proposed model. The first dataset, BCI (Liu *et al.*, 2022a) consists of 4,872 pairs of aligned H&E and IHC pathology image patches, specifically for the HER2 biomarker. The IHC images for 977 test samples are not available as they were not released by the challenge organizers. These image patches were sourced from the WSIs of more than 300 patients, as described in (Liu *et al.*, 2022a). To ensure a well-rounded evaluation, this dataset was randomly partitioned into three subsets: 3,396 pairs for training, 200 pairs for validation, and 300 pairs for testing.

This division allowed us to train and assess the model's performance on distinct sets of images. The second dataset used in this study is the MIST dataset (Li *et al.*, 2023), which includes image patches corresponding to four different breast cancer biomarkers: estrogen receptor (ER), progesterone receptor (PR), Ki67, and HER2. This dataset contains a total of 4,153, 4,139, 4,361, and 4,642 training patches for each respective biomarker, along with

1,000 testing patches derived from 64 WSIs. All the image patches in this dataset are non-overlapping and have a fixed size of 1024×1024 pixels. In the case of this dataset, the model was trained and evaluated separately for each biomarker, allowing for a focused analysis of its performance across different types of cancer biomarkers. Fig. 2 provides six example images from both datasets, showcasing H&E stained images alongside their corresponding IHC stained patches. This comparison helps visualize the differences between the two types of staining and provides a clear context for understanding the model's performance in generating IHC-like images from H&E stained slides.

IMPLEMENTATION DETAILS

In our proposed model, the input images were first resized to a standard size of 512×512 pixels to ensure consistency and facilitate processing. After resizing, the pixel values of the images, originally in the range of 0-255, were normalized to a scale of 0-1. This normalization process ensures that the model can learn the features of the images more effectively by working with values that are easier to handle and more consistent across different inputs. We chose the ADAM optimizer for training, starting with an initial learning rate of 0.0001. ADAM optimizer was selected due to its efficiency and ability to adapt the learning rate during training, which is particularly useful for optimizing complex models. The model was trained from scratch over the course of 500 epochs, with a batch size of 2 images per iteration. This choice of batch size allows for more frequent updates to the model's weights and helps with the convergence process, while also being computationally feasible given the available resources. The proposed model took seven hours to train and has nine million trainable parameters.

To enhance the training process and prevent overfitting, data augmentation techniques were applied to artificially increase the size of the training dataset. These techniques included 90-degree rotations, horizontal flipping, and random scaling of the images with a probability of 0.5. By introducing such transformations, the model was exposed to a wider variety of image features, which improved its ability to generalize to unseen data and made it more robust in recognizing patterns across different image variations. For performance evaluation, we utilized three widely-recognized metrics in image generation tasks: Peak Signal-to-Noise Ratio (PSNR) in dB, Structural Similarity Index (SSIM), and Fréchet Inception Distance (FID). PSNR measures the quality of the generated images in comparison to the original,

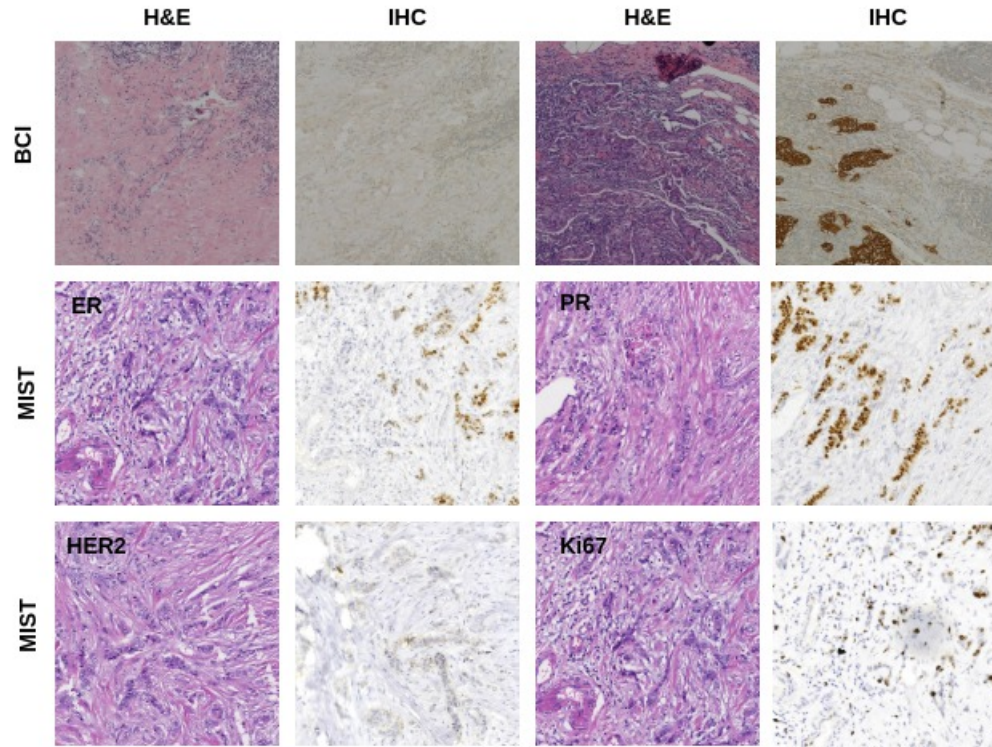


Fig. 2. Illustration of six example images from the both the datasets (BCI and MIST) showing H&E stained images alongside their corresponding patches stained with IHC.

SSIM evaluates the structural similarity between the images, and FID provides a measure of how similar the generated images are to real images in terms of feature distribution. These metrics are essential for quantitatively assessing the performance of image generation models. The model was implemented using the PyTorch framework, which provides flexibility and scalability for deep learning tasks. The training process was carried out on a system with 32GB of RAM and CUDA version 11.2 to leverage GPU acceleration. Specifically, we used an NVIDIA RTX2080Ti GPU with 11GB of video RAM to handle the computational demands of training and evaluation, allowing for efficient processing of the high-resolution images involved in this task.

RESULTS

The quantitative results of the virtual IHC staining models are presented in Table 1. We compared the performance of our proposed cGAN architecture with depthwise separable convolutions against several established image-to-image translation models, including Pix2Pix, Pyramid Pix2Pix, CycleGAN, and a U-Net variant employing standard convolutions. Performance was evaluated using a range of metrics, including PSNR, SSIM, MAE, RMSE, and FID.

As shown in Table 1, the proposed cGAN

model incorporating depthwise separable convolutions demonstrated higher performance across most metrics. To ensure the robustness of the proposed model, we applied five-fold cross-validation. The model achieved the highest PSNR (32.27 dB) and SSIM (0.90), indicating improved pixel-level accuracy and structural similarity to the ground truth IHC images. Furthermore, it shows the lowest MAE (0.04) and RMSE (0.08), confirming reduced pixel-wise differences and a better fit to the target distribution. The significantly lower FID score (31.8) for the proposed model with depthwise separable convolutions compared to the other models indicates a closer alignment of the generated IHC image feature distribution with that of the real IHC images, suggesting more realistic and visually plausible outputs.

The use of depthwise separable convolutions not only improved the quality of the generated images but also significantly enhanced the model's efficiency. The proposed model achieved the fastest inference time (0.15 s/image) among all tested models. This demonstrates the computational advantages of this approach, making it particularly suitable for real-time applications and deployment in resource-constrained environments. In contrast, the baseline models, Pix2Pix and CycleGAN, exhibited lower performance across all image quality metrics and required more

Table 1. Quantitative comparison of virtual IHC staining models on BCI dataset.

Model	PSNR (dB)	SSIM	MAE	RMSE	FID	Inference Time (s/image)
Pix2Pix (Baseline) (Henry <i>et al.</i> , 2021)	28.50	0.82	0.08	0.12	45.2	0.25
CycleGAN (Vasiljević <i>et al.</i> , 2022)	27.80	0.80	0.09	0.13	52.1	0.30
U-Net (with Std. Conv.) (Ronneberger <i>et al.</i> , 2015)	30.20	0.87	0.06	0.09	38.5	0.20
Pyramid Pix2Pix (Liu <i>et al.</i> , 2022b)	30.65	0.86	0.05	0.09	40.11	0.22
Proposed (with Depthwise Sep.)	32.27 $\pm$ 1.05	0.90 $\pm$ 1.94	0.04 $\pm$ 0.03	0.08 $\pm$ 0.06	31.8 $\pm$ 3.80	0.15 $\pm$ 0.08

computational resources, as evidenced by their higher inference times. The standard convolution-based U-Net, while performing better than the baseline models, was still outperformed by its depthwise separable counterpart in all aspects, highlighting the effectiveness of this specific architectural modification. We also evaluated the Pyramid Pix2Pix model, which showed lower performance 2% less in PSNR and 4% less in SSIM compared to the proposed model.

The observed improvements with depthwise separable convolutions can be attributed to their efficient feature extraction capabilities. By decoupling spatial filtering and channel mixing, these convolutions effectively capture fine-grained textural details and complex morphological relationships within the histopathological images while significantly reducing computational complexity. This efficient feature extraction facilitates the accurate prediction of IHC stains from H&E images, leading to improved performance in terms of image quality and computational efficiency.

The results demonstrate that the proposed model offers a compelling approach for virtual IHC staining, achieving state-of-the-art performance in terms of both image quality and computational efficiency. This method has the potential to significantly improve the accessibility and efficiency of IHC analysis in clinical settings.

Table 2. Impact of input image size on virtual IHC staining performance using proposed model on BCI dataset.

Input Size	PSNR (dB)	SSIM	MAE	RMSE	FID
64 $\times$ 64	29.5	0.85	0.07	0.10	38.2
128 $\times$ 128	30.8	0.88	0.06	0.09	34.5
512 $\times$ 512	32.27	0.90	0.04	0.08	31.8
256 $\times$ 256	30.9	0.89	0.05	0.08	33.1

Table 2 presents the impact of input image size on the performance of our proposed model. We evaluated the model using input patch sizes of 64 $\times$ 64, 128 $\times$ 128, 256 $\times$ 256, and 512 $\times$ 512 pixels. As observed, increasing the input size generally led to improvements in image quality metrics. The input size 512 $\times$ 512 achieved the highest PSNR (32.27 dB), SSIM (0.90) and the lowest MAE (0.04), RMSE (0.08) and FID (31.8). This suggests that using larger

input patches allows the model to capture more contextual information, leading to more accurate and visually appealing IHC stain predictions. However, this improvement in quality comes at the cost of increased computational resources. For example, the 256 $\times$ 256 input size, while achieving comparable performance to the 512 $\times$ 512 size in terms of image quality. Therefore, a trade-off between image quality and computational resources must be considered when choosing the appropriate input size. In our experiments, we found that the input size 512 $\times$ 512 offers a good balance between performance and efficiency.

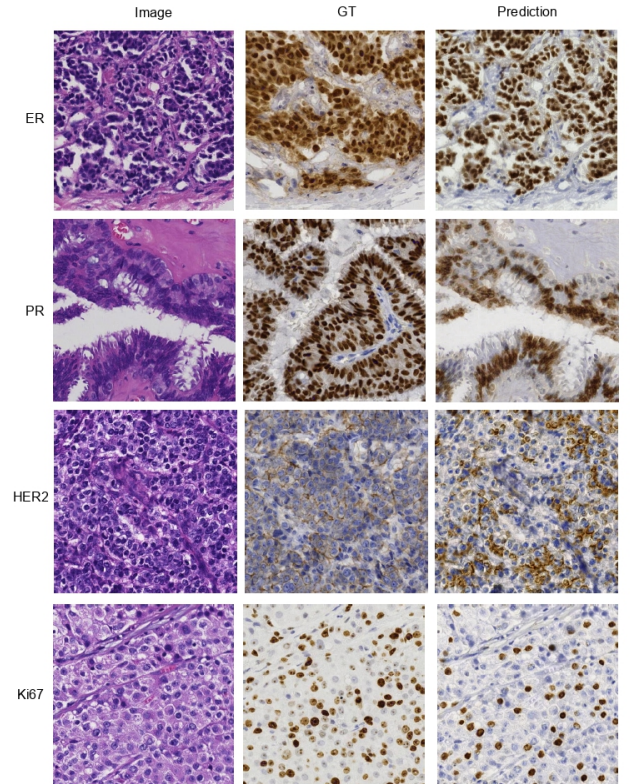


Fig. 3. Illustration of three high-quality virtually IHC-stained images generated by the proposed model on MIST dataset.

Table 3 evaluates the impact of different loss function combinations on the performance of our proposed model. We compared models trained with L1 loss alone, L1 combined with adversarial loss, L1

Table 3. *Impact of different loss functions on virtual IHC staining performance with proposed model using 512×512 input on BCI dataset.*

Loss Function	PSNR (dB)	SSIM	MAE	RMSE	FID	Inference Time (s/image)
L1 Only	29.8	0.86	0.065	0.095	40.1	0.15
L1 + Adversarial	30.5	0.875	0.055	0.085	35.8	0.15
L1 + SSIM	30.2	0.88	0.052	0.082	34.2	0.15
L1 + Adversarial + SSIM (Proposed)	32.27	0.90	0.04	0.08	31.8	0.15

combined with SSIM loss, and the full combination of L1, adversarial, and SSIM losses (our proposed approach). The combination of all three losses yielded the best results across all metrics. Using only L1 loss resulted in reasonable performance but exhibited slightly lower PSNR (29.8 dB), SSIM (0.86), and higher MAE (0.065), RMSE (0.095), and FID (40.1). Adding the adversarial loss improved the sharpness and realism of the generated images, leading to better scores (PSNR: 30.5 dB, SSIM: 0.875, FID: 35.8). Incorporating the SSIM loss further enhanced the structural consistency and reduced artifacts, resulting in further improvements (PSNR: 30.2 dB, SSIM: 0.88, FID: 34.2). The best performance was achieved when all three losses were combined (PSNR: 32.27 dB, SSIM: 0.90, MAE: 0.04, RMSE: 0.08, FID: 31.8), demonstrating the complementary effects of these loss terms in guiding the training process towards generating high-quality and structurally accurate IHC stains. The inference time remained consistent as the architecture did not change, only the loss function.

Table 4 presents the performance evaluation of the proposed model using the MIST dataset, comparing its results with the method by (Li *et al.*, 2023). The evaluation is conducted for four biomarkers: ER, PR, Ki67, and HER2. For each biomarker, three performance metrics are shown: PSNR, SSIM, and FID. Our proposed model achieved higher PSNR and SSIM values are better, as they indicate higher image quality and structural similarity, respectively. Lower FID values are desirable, as they suggest that the generated images are closer to real images. The results show that the proposed model performs well, with the highest PSNR and SSIM values observed for the Ki67 biomarker (PSNR: 25.21 dB, SSIM: 0.259) and the lowest FID value for HER2 (42.75). Compared to the method by (Li *et al.*, 2023), the proposed model achieves similar or better results in terms of PSNR and SSIM, with a lower FID in most cases. However, the HER2 biomarker in the proposed model shows slightly lower PSNR and SSIM values, which are highlighted in red. Overall, the proposed model demonstrates competitive performance in generating IHC images for various biomarkers.

Fig. 4 shows the visual comparison of real and proposed model generated IHC Staining. This

provides a qualitative assessment of our proposed model performance in generating virtual IHC images from H&E stained tissue. We also compared the proposed model with state-of-the-art methods, including Pix2Pix and Pyramid Pix2Pix. Fig. 4 presents a comparison between real IHC-stained samples (Ground Truth, GT) and the corresponding virtual IHC images generated by the proposed model (Prediction). Each row corresponds to a specific IHC marker—ER, PR, HER2, and Ki67 that highlighting the model's ability to reproduce distinct staining patterns. Visual inspection reveals strong alignment between the GT and generated images in terms of cellular morphology and staining distribution, indicating the model's potential to synthesize high-fidelity virtual IHC directly from H&E-stained images. This capability could significantly reduce the reliance on costly and time-intensive IHC procedures.

Fig. also showcases three representative regions of interest (labeled A through C) from the BCI dataset, each including the input H&E image, the corresponding real IHC image, and the predicted IHC image generated by the model. In examples (B) and (C), the spatial arrangement and staining intensity of nuclear features in the predicted images closely match those in the GT, suggesting that the model effectively learns the relationship between H&E morphology and IHC marker expression. However, subtle discrepancies remain—for instance, example (A) shows reduced staining intensity in the predicted image compared to the GT, indicating limitations in capturing the full range of expression levels.

These findings underscore the proposed model's promise in generating realistic virtual IHC images, while also pointing to areas where further refinement is needed to enhance predictive accuracy and robustness across diverse tissue samples. This digital staining approach holds considerable potential for applications where traditional IHC is impractical or unavailable. For comparison, the Pix2Pix model produced IHC images with evident blurriness and poorly defined boundaries, diminishing structural clarity. Although Pyramid Pix2Pix yielded somewhat better results, it still failed to preserve certain cellular structures relative to the proposed model. These comparisons

Table 4. Performance evaluation of the proposed model using the MIST dataset.

Biomarker	Proposed			(Li <i>et al.</i> , 2023)		
	PSNR (dB) $\uparrow$	SSIM $\uparrow$	FID $\downarrow$	PSNR (dB) $\uparrow$	SSIM $\uparrow$	FID $\downarrow$
ER	22.45	0.243	37.34	-	0.221	43.7
PR	22.97	0.254	41.25	-	0.240	44.8
Ki67	25.21	0.259	46.64	-	0.241	51.0
HER2	21.89	0.226	42.75	-	0.215	45.2

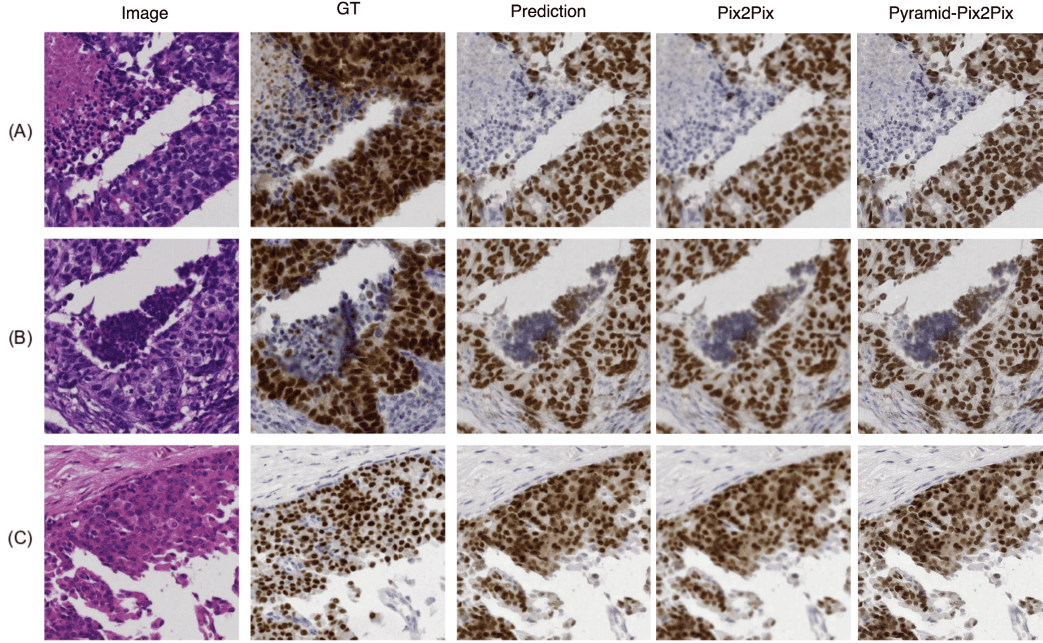


Fig. 4. Illustration of virtually IHC-stained images generated by the proposed model using the BCI dataset. Patches A, B, and C show three examples for qualitative comparison with state-of-the-art methods.

highlight the superior ability of our approach to retain fine morphological details in synthesized IHC images.

In conclusion, the use of depthwise separable convolutions is particularly well-suited for virtual staining tasks, given the inherent complexity of histopathological images. These images contain intricate textures and subtle variations in cellular and tissue architecture. By performing spatial filtering within individual channels, depthwise convolutions effectively capture fine-grained features, while pointwise convolutions integrate these channel-specific representations to model complex inter-channel relationships. This two-step approach enhances the network's ability to extract relevant morphological features—such as nuclear details, cytoplasmic patterns, and stromal structures—critical for accurate IHC prediction from H&E images.

CONCLUSIONS

This study highlights the promising potential of utilizing a cGAN combined with a U-Net architecture and depthwise separable convolutions for virtual immunohistochemical (IHC) staining. The proposed model demonstrates exceptional performance, not only in producing high-quality images but also in optimizing computational efficiency. By leveraging these advanced deep learning techniques, our model provides a powerful tool that can enhance the accessibility and speed of IHC analysis, reducing the time, cost, and complexity associated with traditional IHC staining methods. The ability to generate high-fidelity virtual IHC images from H&E stained tissue samples could be transformative, particularly in settings with limited access to specialized equipment or resources. Our model's state-of-the-art performance, as demonstrated in the experimental results, suggests that it holds substantial promise for advancing the field of

digital pathology and improving diagnostic workflows. However, while these results are promising, further research is necessary to fine-tune the model and ensure its robustness across a wider variety of tissue types, staining protocols, and clinical conditions. Additionally, clinical validation through collaboration with medical professionals is essential to assess the model's reliability and its potential impact in real-world diagnostic settings. In the future, we anticipate that with the right clinical integration and validation, this technology could become a valuable asset in clinical pathology, aiding pathologists in more efficient and accurate diagnoses. It also opens the door to further innovations, such as automating the IHC staining process, enhancing biomarker discovery, and potentially providing tools for personalized medicine.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, Hayes DF, Lakhani SR, Chavez-MacGregor M, Perlmutter J, *et al.* (2020). Estrogen and progesterone receptor testing in breast cancer: Asco/cap guideline update. *J Clin Oncol* 38:1346–66.
- Bancroft JD, Gamble M (2008). *Theory and practice of histological techniques*. Elsevier health sciences.
- BenTaieb A, Hamarneh G (2017). Adversarial stain transfer for histopathology image analysis. *IEEE Trans Med Imaging* 37:792–802.
- DoanNgan B, Angus D, Sung L, *et al.* (2022). Label-free virtual her2 immunohistochemical staining of breast tissue using deep learning. *BME Front*.
- Engel KB, Moore HM (2011). Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med* 135:537–43.
- Fischer AH, Jacobson KA, Rose J, Zeller R (2008). Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harb Protoc* 2008:pdb-prot4986.
- Goldstein NS, Ferkowicz M, Odish E, Mani A, Hastah F (2003). Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *Am J Clin Pathol* 120:86–92.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020). Generative adversarial networks. *Commun ACM* 63:139–44.
- Group EBCTC, *et al.* (2015). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* 386:1341–52.
- Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B (2009). Histopathological image analysis: A review. *IEEE Rev Biomed Eng* 2:147–71.
- Harb R, Pock T, Müller H (2024). Diffusion-based generation of histopathological whole slide images at a gigapixel scale. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*.
- Henry J, Natalie T, Madsen D (2021). Pix2pix gan for image-to-image translation. *Research Gate Publication* :1–5.
- Holzinger A, Langs G, Denk H, Zatloukal K, Müller H (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 9:e1312.
- Hou L, Agarwal A, Samaras D, Kurc TM, Gupta RR, Saltz JH (2019). Robust histopathology image analysis: To label or to synthesize? In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Howat WJ, Lewis A, Jones P, Kampf C, Pontén F, van der Loos CM, Gray N, Womack C, Warford A (2014). Antibody validation of immunohistochemistry for biomarker discovery: recommendations of a consortium of academic and pharmaceutical based histopathology researchers. *Methods* 70:34–8.
- Li F, Hu Z, Chen W, Kak A (2023). Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer.
- Li W, Li J, Polson J, Wang Z, Speier W, Arnold C (2022). High resolution histopathology image generation and segmentation through adversarial training. *Med Image Anal* 75:102251.
- Liu S, Zhu C, Xu F, Jia X, Shi Z, Jin M (2022a). Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

- Liu S, Zhu C, Xu F, Jia X, Shi Z, Jin M (2022b). Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Liu Y, Li X, Zheng A, Zhu X, Liu S, Hu M, Luo Q, Liao H, Liu M, He Y, *et al.* (2020). Predict ki-67 positive cells in h&e-stained images using deep learning independently from ihc-stained images. *Front Mol Biosci* 7:183.
- Ma L, Rathgeb A, Mubarak H, Tran M, Fei B (2022). Unsupervised super-resolution reconstruction of hyperspectral histology images for whole-slide imaging. *J Biomed Opt* 27:056502.
- Orlandini LF, Antonio MVdN, Espreafico Jr CR, Bosquesi PL, Poli-Neto OB, de Andrade JM, Dos Reis FJC, Tiezzi DG (2021). Epidemiological analyses reveal a high incidence of breast cancer in young women in brazil. *JCO Glob Oncol* 7:81–8.
- Peng Q, Lin W, Hu Y, Bao A, Lian C, Wei W, Yue M, Liu J, Yu L, Wang L (2024). Advancing h&e-to-ihc virtual staining with task-specific domain knowledge for her2 scoring. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer.
- Qu L, Zhang C, Li G, Zheng H, Peng C, He W (2024). Advancing h&e-to-ihc stain translation in breast cancer: A multi-magnification and attention-based approach. In: 2024 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE International Conference on Robotics, Automation and Mechatronics (RAM). IEEE.
- Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y, Günaydın H, Zuckerman JE, Chong T, Sisk AE, *et al.* (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng* 3:466–77.
- Rizvi S, Cicalese P, Seshan S, Sciascia S, Becke J, Nguyen H (2022). Histopathology datasetgan: Synthesizing large-resolution histopathology datasets. In: 2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE.
- Ronneberger O, Fischer P, Brox T (2015). U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. Springer.
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, *et al.* (2001). Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses her2. *N Engl J Med* 344:783–92.
- Štepec D, Skočaj D (2020). Image synthesis as a pretext for unsupervised histopathological diagnosis. In: Simulation and Synthesis in Medical Imaging: 5th International Workshop, SASHIMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 5. Springer.
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71:209–49.
- Taylor C, Levenson RM (2006). Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment ii. *Histopathology* 49:411–24.
- Vasiljević J, Nisar Z, Feuerhake F, Wemmert C, Lampert T (2022). CycleGAN for virtual stain transfer: Is seeing really believing? *Artif Intell Med* 133:102420.
- Wilm F, Benz M, Bruns V, Baghdadian S, Dextl J, Hartmann D, Kuritsyn P, Weidenfeller M, Wittenberg T, Merkel S, *et al.* (2022). Fast whole-slide cartography in colon cancer histology using superpixels and cnn classification. *J Med Imaging* 9:027501.
- Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JM, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, *et al.* (2018). Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline focused update. *Arch Pathol Lab Med* 142:1364–82.