Semantic-driven synthesis of histological images with controllable cellular distributions

Computer Methods and Programs in Biomedicine 261 (2025) 108621

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Semantic-driven synthesis of histological images with controllable cellular distributions

Alen Shahini [a,*], Alessandro Gambella [b], Filippo Molinari [a], Massimo Salvi [a]

[a] Biolab, PoliTo[BIO]Med Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[b] Pathology Unit, Department of Surgical Sciences and Integrated Diagnostics (DISC), University of Genoa, Genoa, Italy

ARTICLE INFO

Keywords:
Digital pathology
Generative adversarial networks
Image simulation
Instance-aware models
Multi-scale models

ABSTRACT

Digital pathology relies heavily on large, well-annotated datasets for training computational methods, but generating such datasets remains challenging due to the expertise required and inter-operator variability. We present SENSE (SEmantic Nuclear Synthesis Emulator), a novel framework for synthesizing realistic histological images with precise control over cellular distributions. Our approach introduces three key innovations: (1) A statistical modeling system that captures class-specific nuclear characteristics from expert annotations, enabling generation of diverse yet biologically plausible semantic content; (2) A hybrid ViT-Pix2Pix GAN architecture that effectively translates semantic maps into high-fidelity histological images; and (3) A modular design allowing independent control of cellular properties including type, count, and spatial distribution. Evaluation on the MoNuSAC dataset demonstrates that SENSE generates images matching the quality of real samples (MANIQA: 0.52 ± 0.03 vs 0.52 ± 0.04) while maintaining expert-verified biological plausibility. In segmentation tasks, augmenting training data with SENSE-generated images improved overall performance (DSC from 79.71 to 84.86) and dramatically enhanced detection of rare cell types, with neutrophil segmentation accuracy increasing from 40.18 to 78.71 DSC. This framework enables targeted dataset enhancement for computational pathology applications while offering new possibilities for educational and training scenarios requiring controlled tissue presentations.

1. Introduction

Digital pathology, which makes it possible to collect, organize, and evaluate pathology data in a digital setting, has become a game-changing technology in the healthcare industry in recent years. This change has created new opportunities to increase the efficiency of pathology workflows, decrease turnaround times, and improve diagnostic accuracy [1]. The subjective nature of manual evaluations, which can result in variability between and within observers [2], as well as the increased workload on pathologists [3], are some of the major obstacles this sector must overcome.

To address these challenges, artificial intelligence (AI) and deep learning (DL) algorithms have shown great promise in automating various image analysis tasks in digital pathology [4,5]. These algorithms may reduce the subjectivity by learning complex patterns from raw image data and generating predictions based on learned features [6]. However, the success of AI in digital pathology heavily relies on the availability of large, well-annotated datasets. Obtaining such datasets represents a major challenge due to the time-consuming and expensive nature of manual annotation by domain experts.

1.1. Related works

Several challenges in digital pathology have driven increased interest in generative models, particularly Generative Adversarial Networks (GANs) [7] and more recently diffusion models. These challenges include data scarcity, the need to improve diagnostic accuracy, the reduction of intra- and inter-operator variability, and the simulation of rare conditions.

GANs have been applied in a variety of contexts, including stain normalization [8], stain and domain adaptation [9,10], segmentation using supervised models, synthesis enabling weakly supervised and unsupervised learning [11], data generation [12] and augmentation [13] to enhance classification tasks [14]. Xue et al. [15] modified DCGAN architecture to create cervical cancer samples based on their class to improve the accuracy of their classification models. However,

* Corresponding author.
E-mail address: alen.shahini@polito.it (A. Shahini).

https://doi.org/10.1016/j.cmpb.2025.108621
Received 19 November 2024; Received in revised form 18 December 2024; Accepted 25 January 2025
Available online 26 January 2025

they did not provide objective metrics assessing the quality of the generated data. Krause et al. [16] introduced an innovative DCGAN approach specifically designed to synthesize histopathological images of colorectal cancer. Karimi et al. [17] proposed an application of cGAN and DCGAN to generate prostate samples based on the Gleason score. Liu et al. [18] demonstrated that training a dedicated ProGAN model for each class of brain tumor histopathological images and incorporating synthetic data into the dataset resulted in a 5 % increase in accuracy. Cheng et al. [12] implemented Pix2Pix architecture for simulation and segmentation purposes. Semantic masks are generated from random points, which are then translated into synthetic tissue samples, leading to improvements in segmentation performance. Another approach, presented by Swiderska-Chadaj et al. [13], proposed the creation of histological image patches through cropping and combining different patches. This approach offers an alternative way to generate synthetic data in the context of histological images and has shown promising results regarding the quality of the generated patches. Li et al. [19] proposed a multi-scale conditional GAN for high-resolution, large-scale histopathology image generation and segmentation, with each level dedicated to generating and segmenting images at a distinct scale. Golfe et al. [20] employed a conditional GAN to synthesize prostate histopathological tissue patches by specifying the desired Gleason score in input. The reality assessment of synthetic samples was performed by external validation of a group of expert pathologists. Quiros et al. [21] achieved significant results in terms of the FID metric by synthesizing colorectal and breast cancer samples using a novel conditional architecture based on StyleGAN [22] and BigGAN [23]. Jehanzaib et al. [24] proposed a particular and novel pipeline for the segmentation and synthesis of histological images. While FastGAN generates semantic masks, PathopixGAN uses synthetic masks to generate realistic histological images. Synthetic data is used to solve the data imbalance problem commonly found in histopathological datasets, improving the segmentation performance of the PathoSeg network.

Modern transformer architectures have also been used for histological image synthesis and augmentation without convolutions in GANs [25]. In addition, diffusion models have shown remarkable performance in generating high-quality, realistic data and can address specific challenges in the pathology domain. Jeong et al. [26] developed a score-based diffusion model for stain normalization. Xu et al. [27] demonstrated the effectiveness, for high-quality histopathology image synthesis, of a hybrid model, ViT-DAE, combining transformers and diffusion models, on three public datasets. Linmans et al. [28] proposed the use of diffusion models in the field of unsupervised anomaly detection, particularly for out-of-distribution detection in digital pathology, but there is still a significant performance gap with fully supervised learning. Osorio et al. [29] demonstrated that incorporating image-derived features into the textual prompt, rather than relying solely on healthy and cancerous labels, results in a significant and quantitative improvement in the quality of images created with diffusion models.

1.2. Contribution of our work

While current approaches to histology image simulation have shown promise, they often lack fine-grained control over the characteristics of the generated images. Many existing methods either provide no control over the semantic content or rely on fixed, predefined semantic inputs that are often limited in complexity (e.g., single-class or few-class representations) [16,17]. This lack of control limits the utility of these simulations for targeted training of AI models or for generating diverse, clinically relevant scenarios. Moreover, current simulation approaches struggle to capture the complexity and variability of histological images, often failing to accurately represent the diverse cellular structures and tissue organizations critical for realistic digital pathology simulations [8].

To address these limitations, there is a need for more sophisticated methods of semantic content generation in histology image simulation. Ideally, such methods would allow for precise control over multiple aspects of the generated images, including cell types, object counts, and background characteristics. This level of control, combined with the ability to generate realistic and diverse semantic content, would significantly enhance the utility of simulated images for training robust AI models and for other applications in computational pathology [24].

In this paper, we present a new framework for controllable semantic content generation in histology image simulation. The main contributions are listed as follows:

- We introduce SENSE (SEmantic Nuclear Synthesis Emulator), a novel method for simulating histological images that produces configurable semantic information. SENSE enables a high control over various image characteristics (cell type, numbers, etc.) while maintaining high levels of realism and diversity. This semantic input is then used by a generative model to create synthetic histological images.
- We develop and validate our method on a multi-class annotated dataset comprising diverse cell types with varying shapes and spatial densities. To recreate actual semantic content, SENSE learns the size, distribution, and properties of every semantic class in the dataset. This approach allows for the generation of complex, multi-cellular content that closely mimics real histological samples.
- We evaluate the performance of our approach using a three-fold validation process. First, we apply quantitative metrics to the synthethic images to measure their quality. Second, an expert pathologist conduct visual quality assessments of the generated images to ensure clinical relevance and realism. Finally, we assess the impact of these synthetic images during the training process of AI frameworks in computational pathology.

2. Materials and methods

Fig. 1 illustrates the overall framework of SENSE. The framework operates in two phases. During the training phase (Fig. 1a), the process follows a direct pipeline: starting from the original dataset with its semantic content, our method produces synthetic images using a GAN model. This phase enables the model to learn the transformation from semantic representations to realistic histological images. During the inference phase (Fig. 1b), new semantic content is first generated based on statistical features extracted from the dataset (such as number of objects per image and area distributions), which then serves as input for the trained generative model to produce new synthetic histological images.

2.1. Dataset

This study utilized retrospective data from the publicly available Multi-organ Nuclei Segmentation and Classification (MoNuSAC) challenge dataset [30]. The dataset consists of tiles of variable height and width extracted from H&E-stained whole slide images (WSIs) acquired at 40x magnification from four organs (breast, kidney, lung, and prostate), with expert annotations for four cell types: epithelial cells, lymphocytes, macrophages, and neutrophils. A total of 209 tiles from 46 patients were used for the Development Set (split into Training and Validation sets), while 101 tiles from 25 patients formed the Test Set.

Semantic ground truths are represented as 8-bit RGB images with distinct colors for each cell type: Epithelial [255, 0, 0], Lymphocyte [255, 255, 0], Macrophage [0, 255, 0], and Neutrophil [0, 0, 255]. Nuclear edges were included as an additional class to account for cellular overlap. White zones, corresponding to unstained tissue areas, are identified using LAB color space thresholding and added to provide additional semantic context.

All tiles and their corresponding semantic masks are cropped into 256 × 256 pixel patches with 50 % overlap. Table 1 summarizes the

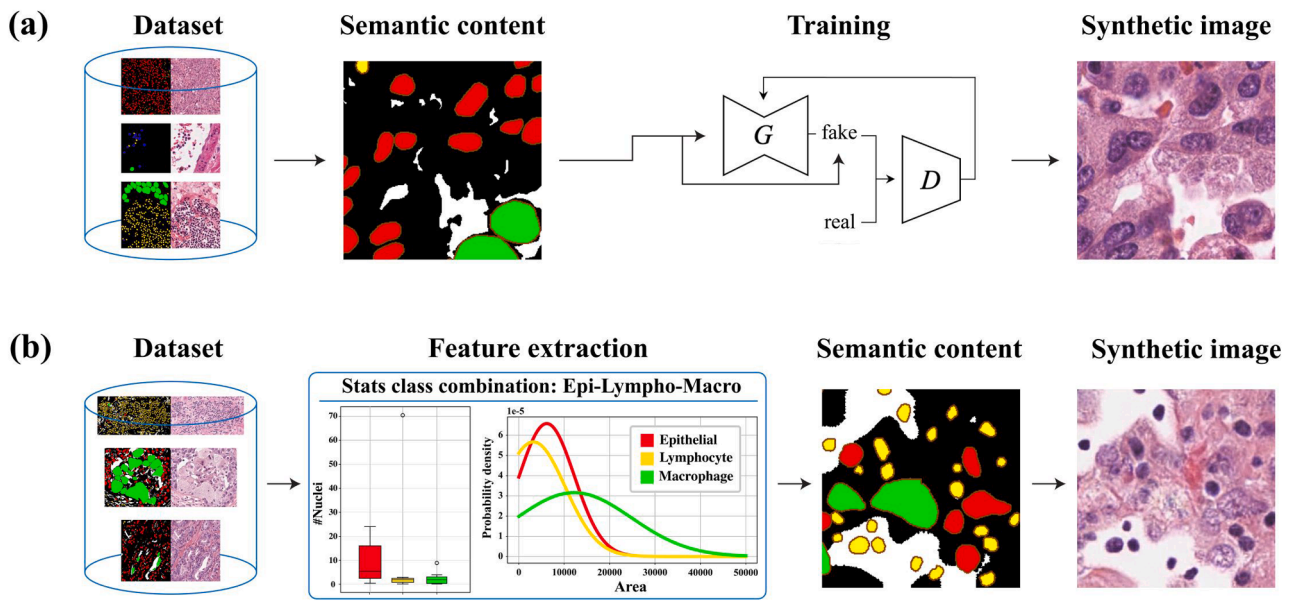

Fig. 1. Overview of the SENSE framework. (a) Training phase: simulation of synthetic images from real semantic content using a GAN model. (b) Inference: generation of new semantic content based on dataset features, followed by synthetic image creation using the trained GAN model.

Table 1
Dataset composition.

Subset	# Patches	Number of instances				
		Epithelial	Lymphocyte	Macrophage	Neutrophil	Total
Train	12,246	13,087	14,105	512	537	28,241
Val	1360	1452	1549	75	94	3170
Test	5601	7213	7806	307	172	15,498

distribution of patches and cell instances across training, validation, and test set.

2.2. Generative model

Given the paired nature of the dataset, we employ a hybrid ViT-Pix2Pix GAN as our generative model. Pix2Pix GAN [31] is designed for image-to-image translation tasks, consisting of a generator that converts semantic content into realistic histologic images and a discriminator that distinguishes between real and generated images. Through competitive training, where the generator tries to deceive the discriminator, the model learns to create high-quality image transformations.

Our custom architecture consists of a Pix2Pix architecture with a particular version of ViT as generator, called Restormer [32]. This network can generate high-resolution images by reducing the computational complexity of traditional ViT. The Restormer consists of four Transformer layers for both encoding and decoding, connected with skip connections. Each layer is composed by multiple blocks MDTA (Multi-Dconv Head Transposed Attention) and GDFN (Gated Dconv Feed-Forward Network) block. The discriminator implements a Patch Classifier, generating probability maps to assess the authenticity of the generated images.

2.3. Extraction of class-specific characteristics and semantic content generation

Our simulation algorithm generates synthetic semantic content containing multiple classes of cell nuclei with statistically controlled distributions of count and area. The process consists of three main stages: feature extraction, nuclei placement, and background generation.

First, we extract class-specific features from the MoNuSAC, which contains 310 tiles from 71 patients. For each nucleus in each class, we record: patient identifier, tile information (identifier and size), nucleus coordinates within the tile, and pixel area occupied. Each nucleus receives a unique identification code for subsequent positioning. At the tile level, we calculate the total number of nuclei and area occupied by each cell class. To standardize these measurements across different tile sizes, we normalize all features to a 300 $\times$ 300 pixel patch. For tiles with similar cell combinations, we compute comprehensive statistics (mean, standard deviation, percentiles [5th, 25th, 50th, 75th, 95th], minimum, and maximum) of nuclei counts and areas.

For synthetic content generation, we iterate through all cell combinations present in the MoNuSAC dataset. For each class within a combination, the number of nuclei is randomly determined based on the extracted statistical distributions. Each nucleus is assigned an identification code and its original coordinates are selected from the source tile. These coordinates are resized to fit inside a patch of 300 by 300 pixels; any coordinates that are larger than the patch size are moved to the patch boundary. Realistic spatial connections are maintained while the nuclei are positioned at random over the patch. During placement, the algorithm checks for excessive overlap with previously placed nuclei. If a suitable position cannot be found within a defined time interval, a new nucleus is selected. The placement process continues until either all nuclei are successfully positioned or the total area of placed nuclei reaches the statistically derived maximum threshold. Finally, we generate background white areas by applying random noise to a black mask, followed by Gaussian blur and thresholding operations. These areas undergo morphological processing to achieve realistic shapes and distributions. The final semantic content is created by combining the placed nuclei and processed white areas, followed by central cropping to

256 × 256 pixels to ensure realistic edge characteristics. Fig. 2 shows the entire simulation process of a semantic content.

2.4. Evaluation metrics

Our evaluation strategy encompasses three key aspects to assess the quality and utility of our generation method. Quantitatively, we employ both reference and no-reference metrics. As a reference metric, we use FID (Fréchet Inception Distance) [33] to measure the similarity between real and generated image distributions. We computed three no-reference metrics: NIQE (Naturalness Image Quality Evaluator), MANIQA (Multi-dimension Attention Network for No-Reference) [34], and PAQ2PIQ (From Patches to Pictures) [35]. These metrics measure the similarity between real and generated image distributions and assess perceptual similarity based on learned features.

For clinical validation, an expert pathologist performs a visual assessment of image quality and realism using a two-score system on a scale of 1 to 3. Finally, to demonstrate practical utility in computational pathology, we evaluate our method's impact on nuclear segmentation performance with and without our augmented data.

3. Results

3.1. Image generation and quantitative metrics

We tracked the FID score during the training phase to identify the best model for image generation. As seen in Fig. 3, while both validation and test FID curves showed similar trends, there was a notable difference in their absolute values. This gap is relatively common in generative models and could potentially be reduced with larger training datasets. The model selected at epoch 42 showed the lowest FID value for the validation set. Despite this difference between validation and test FID scores, the overall convergence pattern indicates that our generative model maintains good generalization capabilities, which is further supported by the pathologist evaluation results.

The quality of the images produced by our framework is shown in Fig. 4. This figure demonstrates the ability of the first part of the SENSE framework to capture the diversity and heterogeneity of tissue characteristics of real images. The generation model is capable of recreating, by taking real input, synthetic images with the same semantic content as real images. Using class-specific annotations, the model effectively captures the realism and heterogeneity of nuclear textures. The synthesis of realistic nuclear morphologies is possible by the class-aware generation used by SENSE. Notably, in cases where certain semantic features are not annotated, our method maintains consistency by not generating these undefined structures.

The quantitative evaluation using no-reference metrics (Table 2) supports the visual quality of our generated images. Images generated using SENSE achieve comparable or better scores across all metrics compared to both real images and those generated from real semantic

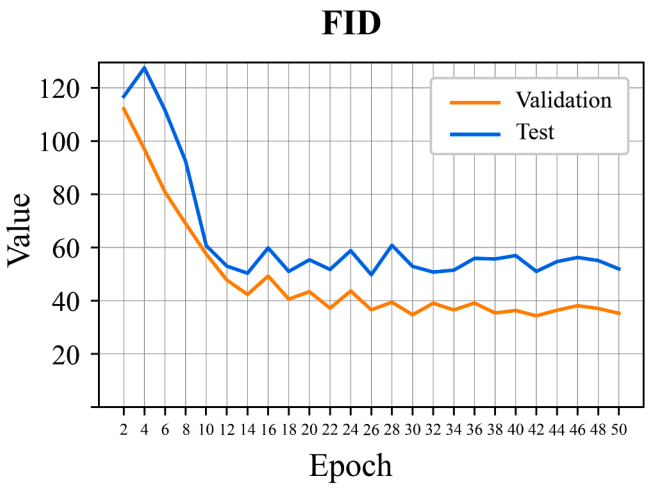

Fig. 3. FID scores for validation and test sets across training epochs.

content. Values are reported as mean ± standard deviation. Particularly noteworthy is the consistency in MANIQA scores (0.52 ± 0.03) matching real images (0.52 ± 0.04), and the improved PAQ2PIQ score (70.22 ± 1.76) compared to real images (69.39 ± 2.35). The lower standard deviations in our generated images suggest more consistent quality across the dataset.

A key feature of our simulation framework is its ability to generate images with precise control over semantic content while maintaining diversity. As demonstrated in Fig. 5a, even with identical statistical parameters, the framework generates distinct semantic contents, ensuring variety in the synthetic dataset. Fig. 5b illustrates the framework's capability to control specific features, such as the percentage and distribution of lipid or empty (white) areas, while maintaining the same cell population. This controllable generation process enables explicit manipulation of various histological structures, including cell nuclei, lipids, and white spaces, in terms of their number, size, shape, and spatial distribution.

3.2. Pathologist evaluation

For clinical validation, an expert pathologist conducts a visual assessment comparing three types of images (50 images of each type): real images from the test set, images generated from real semantic content, and images generated using our synthetic semantic content. Each image is assessed with two separate scores, each on a scale of 1 to 3. The first score evaluates the overall image realism, while the second assesses the consistency and realism of cellular structures' distribution. On this scale, scores of 1, 2 and 3 correspond to not very realistic, moderately realistic, and realistic image, respectively.

As shown in Fig. 6a, the realism assessment demonstrates that

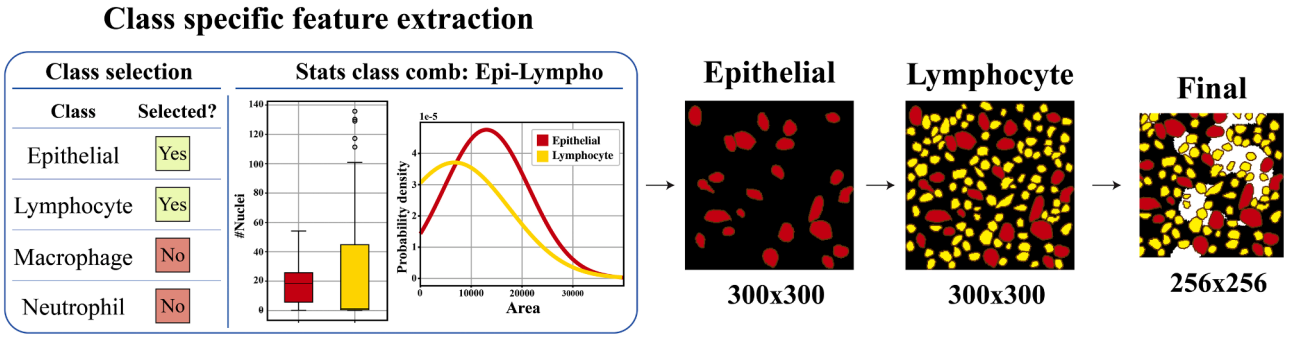

Fig. 2. Overview of the semantic content generation method. Once the class cell combination to be used has been determined, its extracted features are used to define the instances to be placed within the semantic content.

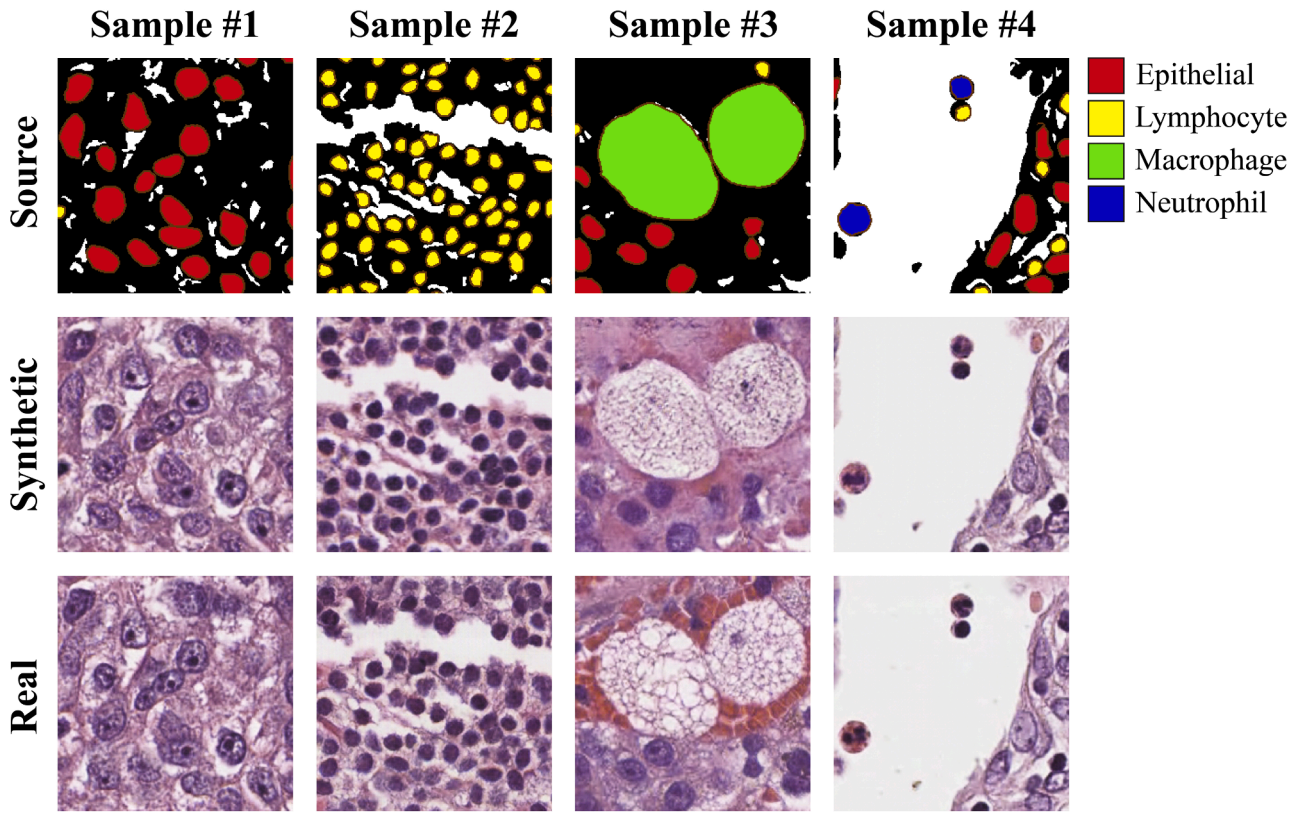

Fig. 4. Examples of synthetic images generated on the test set of MoNuSAC. Using the class labels of individual cells as input, the first part of the SENSE framework captures the diversity and heterogeneity of the nuclear textures.

Table 2
Quantitative evaluation using no-reference image quality metrics. Comparison between (i) real images from the MoNuSAC test set, (ii) images generated from real semantic content (Fake images), and (iii) images generated using our SENSE framework with synthetic semantic content (equal number of patches for all sets). All values are reported as mean ± standard deviation.

Subset	NIQE	MANIQA	PAQ2PIQ
Test Set MoNuSAC	8,39 ±	0,52 ±	69,39 ±
Real images ($n = 5601$)	2,51	0,04	2,35
Test set MoNuSAC	9,65 ±	0,50 ±	69,13 ±
Fake images ($n = 5601$)	2,52	0,04	2,16
Image generated using SENSE ($n = 5601$)	8,94 ± 1,62	0,52 ± 0,03	70,22 ± 1,76

images generated using real semantic content (Synthetic Test Images) achieved scores comparable to real images (Real Test Images), with average scores of 2.26 and 1.92 respectively. Images generated using our synthetic semantic content (SENSE-generated Images) also maintained convincing realism with a average score of 2.06. The evaluation of semantic content consistency (Fig. 6b) revealed interesting patterns in how our framework reproduces cellular distributions. Images produced using our SENSE framework retained good biological plausibility with a shift toward moderate and high scores (21 images scored as 2 and 24 as 3), whereas images generated with real semantic content exhibited a strong tendency toward the highest score (39 images scored as 3). Notably, only five images were rated as not realistic, suggesting that our framework reliably generates cellular configurations that are biologically plausible.

3.3. Impact in computational pathology

To evaluate the practical utility of our framework, we assessed its impact on nuclear segmentation performance. We conducted experiments comparing three training scenarios:

1. Baseline: using only real images from the original training set and validation set.
2. 50 % augmentation: original training set enhanced with synthetic images equal to 50 % of the original size
3. 100 % augmentation: original training set doubled in size with synthetic image.

For the synthetic data generation, we prioritized underrepresented classes (macrophages and neutrophils) to address dataset imbalance (Table 1). Performance was evaluated using both overall Dice Similarity Coefficient (DSC) and class-specific DSC metrics (Table 3). The results demonstrate that augmenting the training data with synthetic images improves overall segmentation performance. The 50 % augmentation scenario showed a notable improvement in overall DSC from 79.71 to 81.43 compared to the baseline. More significantly, class-specific metrics revealed substantial improvements for underrepresented classes, with macrophage DSC increasing from 81.63 to 84.43 and neutrophil DSC showing a dramatic improvement from 40.18 to 66.89. Further augmentation to 100 % yielded additional gains, raising the overall DSC to 84.86 and further improving performance on rare cell types, with neutrophil DSC reaching 78.71.

4. Discussion

Generative modeling for image synthesis has become a crucial tool in digital pathology, supporting applications from data augmentation to clinical decision support [14,36,37]. However, existing approaches often lack fine-grained control over semantic content, limiting their utility in generating diverse and targeted histological images. Current models struggle to precisely dictate the number, shape, and

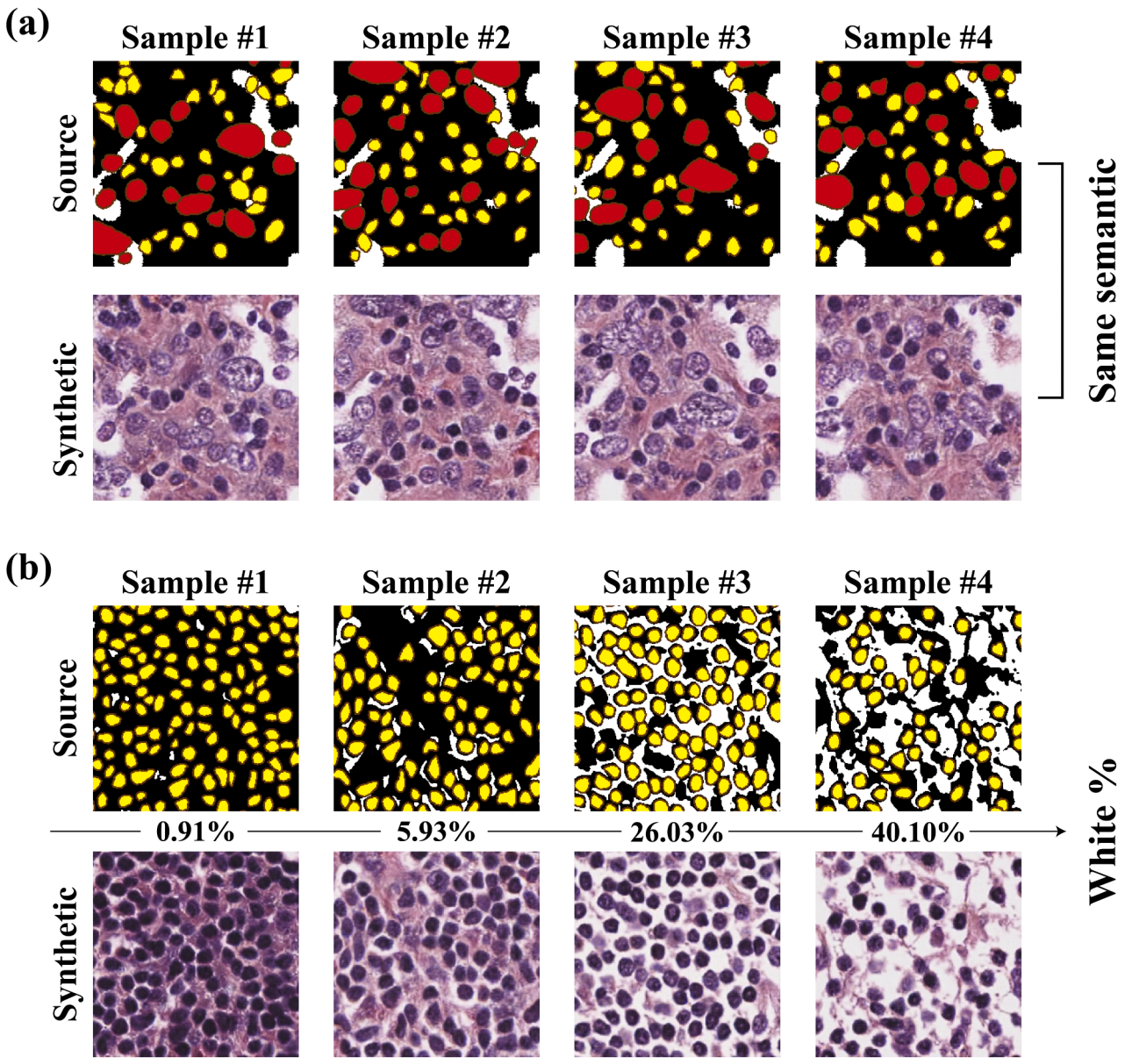

Fig. 5. Control over the semantic content of the SENSE simulation framework. (a) Different semantic contents generated with the same simulation parameters. (b) Generation of images with the same cell population but a greater percentage of white areas.

configuration of cellular structures, resulting in realistic-looking images but with uncontrolled tissue components [16]. While recent advances in generative techniques, such as diffusion models, have improved image quality and realism, the fundamental challenge of precise control over histological features and structures remains a key limitation [38].

Our work addresses these limitations through the SENSE framework, which enables explicit manipulation of histological structures and precise control over cellular distributions (Fig. 4). Our two-stage framework divides the generation process into two separate steps: first, statistical techniques are used to create semantic content, and then GANs are used to turn this content into realistic images. The GAN's work is greatly simplified by this division since it only needs to focus on translating the semantic layouts into realistic appearances rather than managing both structure and appearance at the same time. The quantitative evaluation demonstrates the effectiveness of our approach, with comparable MANIQA scores (0.52 ± 0.03) to real images and improved PAQ2PIQ scores (70.22 ± 1.76), indicating high perceptual quality. The visual assessment by expert pathologists further validates our method's capability to generate realistic images, with most synthetic samples receiving moderate to high scores for both overall realism and cellular distribution plausibility (Fig. 6). As shown in Fig. 6a, images from the real test set are more frequently classified as unrealistic than the generated synthetic images. This counterintuitive finding can be explained by several factors. Histological patches, derived from larger slides at a given magnification, often contain artifacts arising from sample preparation and digitalization protocols. Our simulation tool, which uses a network designed for image restoration, tends to minimize these artefacts while preserving the semantic content of the created image. Consequently, it is possible for an artefact to be present in a real image, but not in the corresponding simulated image. It also important to note that the three different sets (real, synthetic, and SENSE-generated) shown to the pathologist contained different samples.

Notably, our framework is able to address dataset imbalances in computational pathology tasks. According to the segmentation experiments, performance across cell types is significantly enhanced when our synthetic images are added to training data, with high gains for

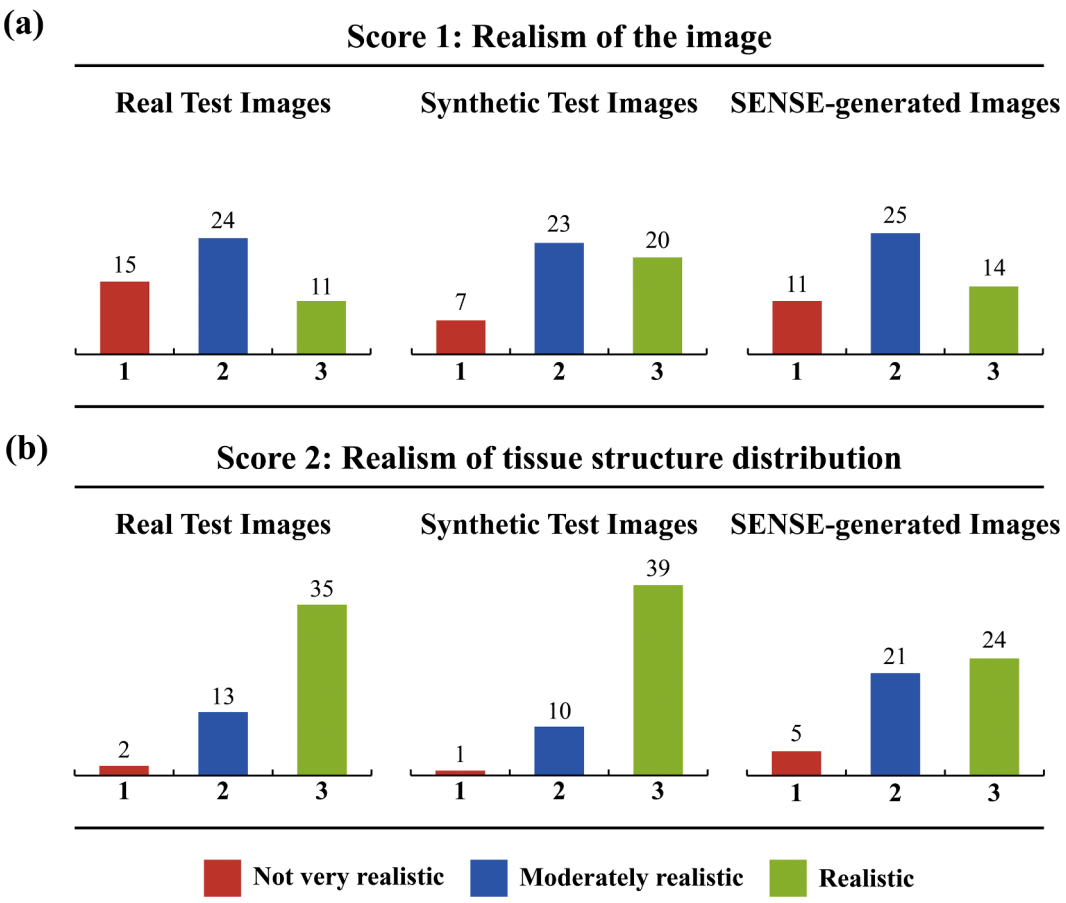

Fig. 6. Evaluation of the SENSE framework. (a) Distribution of image realism scores from expert pathologist evaluation across 50 images per category: real test set images, synthetic images from real semantic content, and synthetic images from SENSE-generated semantic content. (b) Distribution of semantic content scores evaluating the biological plausibility of cellular distributions and structural arrangements.

Table 3
Nuclear segmentation performance comparison across different training scenarios. Overall DSC measures general segmentation accuracy, while class-specific DSC values indicate performance for each cell type.

Subset	Overall segmentation	Class-specific metrics			
		Epithelial	Lymphocyte	Macrophage	Neutrophil
Original images	79.71	92.74	86.83	81.63	40.18
Original images + 50 % synthetic images	81.43	89.85	83.40	84.43	66.89
Original images + 100 % synthetic images	84.86	91.47	86.41	85.47	78.71

underrepresented classes. The most notable improvement was the increase in macrophage detection from 81.63 to 85.46 DSC and the improvement in neutrophil segmentation from 40.18 to 78.71 DSC with full augmentation.

Our framework's modular design provides a number of significant benefits. In the first place, it allows for fine-grained control over semantic content, allowing for highly accurate manipulation of the number, size, and location of cellular structures (Fig. 5). Importantly, as demonstrated in Fig. 4, SENSE exhibits versatility in handling different types of cellular annotations - from pure nuclear structures to complete cell boundaries including cytoplasm, as seen in the case of macrophages. This flexibility allows the framework to adapt to various cellular features of interest, as long as they are consistently annotated in the training data. Second, as demonstrated by the pathologist's evaluation scores, the statistical modeling of class-specific features guarantees that produced images retain biological plausibility while displaying natural variation. It is important to note that the quality of generated images is dependent on staining intensity variations, with our method performing optimally for standard H&E staining intensities typical of well-prepared slides. Extreme variations in staining in the training set, particularly very weak staining or oversaturated regions, can impact the quality of the synthesized images.

SENSE also exhibits promise in educational settings in addition to computational applications. Pathology training programs could be improved by the ability to create customized content with particular cellular distributions and arrangements, giving students access to a greater variety of cases than are usually offered in conventional educational settings.

However, several limitations should be acknowledged. Currently, our method focuses primarily on nuclear structures and operates at the patch level (256 × 256 pixels) rather than full-resolution microscopy tiles. This means that SENSE cannot yet model the full complexity of tissue architecture across broader spatial scales. Extending this approach to larger tiles (e.g., 3000 × 3000 pixels) and ultimately to WSIs would provide a more comprehensive view of tissue architecture and cellular interactions across different regions. Additionally, our current

method primarily controls object overlap, number, and shape, without explicitly modeling the spatial relationships between cells. The cellular microenvironment and spatial organization of different cell types are crucial aspects of histological evaluation that could provide additional biological insights. Further exploration is needed to incorporate more sophisticated spatial distribution characteristics of individual objects within the generated images. Furthermore, our current implementation is limited to the cell types present in the MoNuSAC dataset, and extending it to other tissue types would require additional expert annotations. These improvements would enhance the realism and diversity of the synthesized histological samples.

An important extension of our work would involve validating our approach on additional histopathology datasets. While we focused on the MoNuSAC dataset due to its comprehensive multi-class cell annotations, future studies should explore the application of SENSE to other datasets such as CryoNuSeg [39] and CoNiC [40]. This would require collaboration with expert pathologists to enrich existing binary annotations (nuclei vs. background) with detailed cell type classifications. Such extension would not only validate the generalizability of our framework but also contribute to the creation of new annotated datasets for the computational pathology community.

Different future directions can be explored for our work. In the future, our goal is to expand our methodology to more intricate structures like glands and tubules, as well as to tissue-level classification tasks (e.g., grading tumoral versus healthy tissue). Beyond pathology, our semantic-driven generation concept shows potential for broader medical imaging applications. Recent research has shown that immunohistochemistry and multiplexed imaging methods can improve ground truth annotations in histological images [41]. Such methods could be combined with our framework in a synergistic way, for example, by training more accurate semantic controls for cell type-specific features using annotations derived from IHC. Similar to how MAPS [41] and Nimbus [42] allow automated classification of cellular expression in multiplexed imaging data, our approach could be modified for fluorescence microscopy applications to simulate multicolor immunofluorescence images with controlled nuclear arrangements and marker expression patterns. Examples include simulating time-lapse fluorescence microscopy sequences with specific cell behaviors or creating synthetic training data for FISH image analysis, where exact control over nuclear organization and signal patterns is essential. Additionally, our approach may help close the gap between various imaging modalities by allowing the creation of corresponding views of the same tissue in H&E, IHC, and IF, as shown by Zurek et al. [43]. This would provide useful multi-modal training data for computational pathology. These extensions would further demonstrate the versatility of our semantic-driven approach in addressing various healthcare imaging challenges while contributing to the growing ecosystem of tools for automated histological analysis.

5. Conclusion

This work introduces SENSE, a novel framework that enables controlled synthesis of histological images through precise semantic content generation. Our method offers fine-grained control over cellular distributions, thereby addressing important limitations in existing histology image synthesis. While statistical modeling guarantees that the generated images maintain realistic tissue organization patterns, the modular design permits flexible manipulation of nuclear characteristics and spatial arrangements. While current implementation focuses on nuclear-level features, the SENSE framework provides a foundation for extension to more complex tissue structures and broader medical imaging applications.

Data availability

The data used in this study are available from the corresponding author upon reasonable request.

Ethics statement

Ethical approval was not required for this study as it utilized only publicly available data from published literature and open source datasets. No human subjects, animals, or clinical trials were involved in this research.

CRediT authorship contribution statement

Alen Shahini: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Alessandro Gambella:** Writing – review & editing, Formal analysis, Data curation. **Filippo Molinari:** Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Massimo Salvi:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2025.108621.

References

[1] M.G. Hanna, V.E. Reuter, J. Samboy, C. England, L. Corsale, S.W. Fine, N. P. Agaram, E. Stamelos, Y. Yagi, M. Hameed, D.S. Klimstra, S.J. Sirintrapun, Implementation of digital pathology offers clinical and operational increase in efficiency and cost savings, Arch. Pathol. Lab. Med. 143 (2019) 1545–1555, https://doi.org/10.5858/arpa.2018-0514-OA.

[2] J. Persson, U. Wilderäng, T. Jiborn, P.N. Wiklund, J.-E. Damber, J. Hugosson, G. Steineck, E. Haglind, A. Bjartell, Interobserver variability in the pathological assessment of radical prostatectomy specimens: findings of the Laparoscopic Prostatectomy Robot Open (LAPPRO) study, Scand. J. Urol. 48 (2014) 160–167, https://doi.org/10.3109/21681805.2013.820788.

[3] D.M. Metter, T.J. Colgan, S.T. Leung, C.F. Timmons, J.Y. Park, Trends in the US and Canadian pathologist workforces From 2007 to 2017, JAMA Netw. Open. 2 (2019) e194337, https://doi.org/10.1001/jamanetworkopen.2019.4337.

[4] M. Salvi, F. Molinari, U.R. Acharya, L. Molinaro, K.M. Meiburger, Impact of stain normalization and patch selection on the performance of convolutional neural networks in histological breast and prostate cancer classification, Comput. Methods Programs Biomed. Update 1 (2021) 100004, https://doi.org/10.1016/j.cmpbup.2021.100004.

[5] E.A. Rakha, M. Toss, S. Shiino, P. Gamble, R. Jaroensri, C.H. Mermel, P.-H.C. Chen, Current and future applications of artificial intelligence in pathology: a clinical perspective, J. Clin. Pathol. 74 (2021) 409–414, https://doi.org/10.1136/jclinpath-2020-206908.

[6] M. Salvi, F. Molinari, S. Iussich, L.V. Muscatello, L. Pazzini, S. Benali, B. Banco, F. Abramo, R. De Maria, L. Aresu, Histopathological classification of canine cutaneous round cell tumors using deep learning: a multi-center study, Front. Vet. Sci. 8 (2021), https://doi.org/10.3389/fvets.2021.640944.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144, https://doi.org/10.1145/3422622.

[8] F.G. Zanjani, S. Zinger, B.E. Bejnordi, J.A.W.M. van der Laak, P.H.N. de With, Stain normalization of histopathology images using generative adversarial networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 573–577, https://doi.org/10.1109/ISBI.2018.8363641.

[9] N. Zhou, D. Cai, X. Han, J. Yao, Enhanced cycle-consistent generative adversarial network for color normalization of H&E stained images, in: 2019: pp. 694–702. https://doi.org/10.1007/978-3-030-32239-7_77.

[10] Z. Xu, X. Huang, C.F. Moro, B. Bozóky, Q. Zhang, GAN-based virtual re-staining: a promising solution for whole slide image analysis, 2019, https://doi.org/10.48550/arXiv.1901.04059.

[11] A. Rana, G. Yauney, A. Lowe, P. Shah, Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 828–834, https://doi.org/10.1109/ICMLA.2018.00133.

[12] J. Cheng, Z. Wang, Z. Liu, Z. Feng, H. Wang, X. Pan, Deep adversarial image synthesis for nuclei segmentation of histopathology image2021 2nd Asia Conference on Computers and Communications (ACCC), IEEE (2021) 63–68, https://doi.org/10.1109/ACCC54619.2021.00017.

[13] Z. Swiderska-Chadaj, E. Stoelinga, A. Gertych, F. Ciompi, Multi-Patch Blending improves lung cancer growth pattern segmentation in whole-slide images, in: 2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE), IEEE, 2020, pp. 1–4, https://doi.org/10.1109/CPEE50798.2020.9238710.

[14] M.E. Tschuchnig, G.J. Oostingh, M. Gadermayr, Generative adversarial networks in digital pathology: a survey on trends and future potential, Patterns 1 (2020) 100089, https://doi.org/10.1016/j.patter.2020.100089.

[15] Y. Xue, Q. Zhou, J. Ye, L.R. Long, S. Antani, C. Cornwell, Z. Xue, X. Huang, Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification, 2019, https://doi.org/10.48550/arXiv.1907.10655.

[16] J. Krause, H.I. Grabsch, M. Kloor, M. Jendrusch, A. Echle, R.D. Buelow, P. Boor, T. Luedde, T.J. Brinker, C. Trautwein, A.T. Pearson, P. Quirke, J. Jenniskens, K. Offermans, P.A. Brandt, J.N. Kather, Deep learning detects genetic alterations in cancer histology generated by adversarial networks, J. Pathol. (2021), https://doi.org/10.1002/path.5638 path.5638.

[17] D. Karimi, G. Nir, L. Fazli, P.C. Black, L. Goldenberg, S.E. Salcudean, Deep learning-based gleason grading of prostate cancer from histopathology images—role of multiscale decision aggregation and data augmentation, IEEE J. Biomed. Health Inform. 24 (2020) 1413–1426, https://doi.org/10.1109/JBHI.2019.2944643.

[18] S. Liu, Z. Shah, A. Sav, C. Russo, S. Berkovsky, Y. Qian, E. Coiera, A. Di Ieva, Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning, Sci. Rep. 10 (2020) 7733, https://doi.org/10.1038/s41598-020-64588-y.

[19] W. Li, J. Li, J. Polson, Z. Wang, W. Speier, C. Arnold, High resolution histopathology image generation and segmentation through adversarial training, Med. Image Anal. 75 (2022) 102251, https://doi.org/10.1016/j.media.2021.102251.

[20] A. Golfe, R. del Amor, A. Colomer, M.A. Sales, L. Terradez, V. Naranjo, ProGleason-GAN: conditional progressive growing GAN for prostatic cancer Gleason grade patch synthesis, Comput. Methods Programs Biomed. 240 (2023) 107695, https://doi.org/10.1016/j.cmpb.2023.107695.

[21] A.C. Quiros, R. Murray-Smith, K. Yuan, PathologyGAN: learning deep representations of cancer tissue. https://doi.org/10.48550/arXiv.1907.02644.

[22] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 4396–4405, https://doi.org/10.1109/CVPR.2019.00453.

[23] A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis. https://doi.org/10.48550/arXiv.1809.11096.

[24] M. Jehanzaib, Y. Almalioglu, K.B. Ozyoruk, D.F.K. Williamson, T. Abdullah, K. Basak, D. Demir, G.E. Keles, K. Zafar, M. Turan, A robust image segmentation and synthesis pipeline for histopathology, Med. Image Anal. 99 (2025) 103344, https://doi.org/10.1016/j.media.2024.103344.

[25] M. Li, C. Li, P. Hobson, T. Jennings, B.C. Lovell, MedViTGAN: end-to-end conditional GAN for histopathology image augmentation with vision transformers, in: 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 4406–4413, https://doi.org/10.1109/ICPR56361.2022.9956431.

[26] J. Jeong, K.D. Kim, Y. Nam, C.E. Cho, H. Go, N. Kim, Stain normalization using score-based diffusion model through stain separation and overlapped moving window patch strategies, Comput. Biol. Med. 152 (2023) 106335, https://doi.org/10.1016/j.compbiomed.2022.106335.

[27] X. Xu, S. Kapse, R. Gupta, P. Prasanna, in: A. Mukhopadhyay, I. Oksuz, S. Engelhardt, D. Zhu, Y. Yuan (Eds.), Deep Generative Models. MICCAI 2023. Lecture Notes in Computer Science, 14533, Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-53767-7_7.

[28] J. Linmans, G. Raya, J. van der Laak, G. Litjens, Diffusion models for out-of-distribution detection in digital pathology, Med. Image Anal. 93 (2024) 103088, https://doi.org/10.1016/j.media.2024.103088.

[29] P. Osorio, G. Jimenez-Perez, J. Montalt-Tordera, J. Hooge, G. Duran-Ballester, S. Singh, M. Radbruch, U. Bach, S. Schroeder, K. Siudak, J. Vienenkoetter, B. Lawrenz, S. Mohammadi, Latent diffusion models with image-derived annotations for enhanced AI-assisted cancer diagnosis in histopathology, Diagnostics 14 (2024) 1442, https://doi.org/10.3390/diagnostics14131442.

[30] R. Verma, N. Kumar, A. Patil, N.C. Kurian, S. Rane, S. Graham, Q.D. Vu, M. Zwager, S.E.A. Raza, N. Rajpoot, X. Wu, H. Chen, Y. Huang, L. Wang, H. Jung, G.T. Brown, Y. Liu, S. Liu, S.A.F. Jahromi, A.A. Khani, E. Montahaei, M.S. Baghshah, H. Behroozi, P. Semkin, A. Rassadin, P. Dutande, R. Lodaya, U. Baid, B. Baheti, S. Talbar, A. Mahbod, R. Ecker, I. Ellinger, Z. Luo, B. Dong, Z. Xu, Y. Yao, S. Lv, M. Feng, K. Xu, H. Zunair, A. Ben Hamza, S. Smiley, T.-K. Yin, Q.-R. Fang, S. Srivastava, D. Mahapatra, L. Trnavska, H. Zhang, P.L. Narayanan, J. Law, Y. Yuan, A. Tejomay, A. Mitkari, D. Koka, V. Ramachandra, L. Kini, A. Sethi, MoNuSAC2020: a multi-organ nuclei segmentation and classification challenge, IEEE Trans. Med. ImAging 40 (2021) 3413–3423, https://doi.org/10.1109/TMI.2021.3085712.

[31] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967–5976, https://doi.org/10.1109/CVPR.2017.632. Honolulu, HI, USA.

[32] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M. Yang. Restormer: Efficient Transformer for High-Resolution Image Restoration, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5718–5729, https://doi.org/10.1109/CVPR52688.2022.00564. New Orleans, LA, USA.

[33] G. Parmar, R. Zhang, J.-Y. Zhu, On aliased resizing and surprising subtleties in GAN evaluation. https://doi.org/10.48550/arXiv.2104.11222.

[34] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, Y. Yang, MANIQA: multi-dimension attention network for no-reference image quality assessment. https://doi.org/10.48550/arXiv.2204.08958.

[35] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, From patches to pictures (PaQ-2-PiQ): mapping the perceptual space of picture quality. https://doi.org/10.48550/arXiv.1912.10088.

[36] L. Jose, S. Liu, C. Russo, A. Nadort, A. Di Ieva, Generative adversarial networks in digital pathology and histopathological image processing: a review, J. Pathol. Inform. 12 (2021) 43, https://doi.org/10.4103/jpi.jpi_103_20.

[37] M. Salvi, F. Branciforti, F. Molinari, K.M. Meiburger, Generative models for color normalization in digital pathology and dermatology: advancing the learning paradigm, Expert. Syst. Appl. 245 (2024) 123105, https://doi.org/10.1016/j.eswa.2023.123105.

[38] M. Pozzi, S. Noei, E. Robbi, L. Cima, M. Moroni, E. Munari, E. Torresani, G. Jurman, Generating and evaluating synthetic data in digital pathology through diffusion models, Sci. Rep. 14 (2024) 28435, https://doi.org/10.1038/s41598-024-79602-w.

[39] A. Mahbod, G. Schaefer, B. Bancher, C. Löw, G. Dorffner, R. Ecker, I. Ellinger, CryoNuSeg: a dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images, Comput. Biol. Med. 132 (2021) 104349, https://doi.org/10.1016/j.compbiomed.2021.104349.

[40] S. Graham, M. Jahanifar, Q.D. Vu, G. Hadjigeorghiou, T. Leech, D. Snead, S.E. A. Raza, F. Minhas, N. Rajpoot, CoNIC: colon nuclei identification and counting challenge 2022. https://doi.org/10.48550/arXiv.2111.14485.

[41] M. Shaban, Y. Bai, H. Qiu, S. Mao, J. Yeung, Y.Y. Yeo, V. Shanmugam, H. Chen, B. Zhu, J.L. Weirather, G.P. Nolan, M.A. Shipp, S.J. Rodig, S. Jiang, F. Mahmood, MAPS: pathologist-level cell type annotation from tissue images through machine learning, Nat. Commun. 15 (2024) 28, https://doi.org/10.1038/s41467-023-44188-w.

[42] J.L Rumberger, N.F. Greenwald, J.S Ranek, P. Boonrat, C. Walker, J Franzen, S. R. Varra, A Kong, C. Sowers, C.C. Liu, I Averbukh, H Piyadasa, R. Vanguri, I Nederlof, X.J Wang, D. Van Valen, M Kok, T.J Hollmann, D. Kainmueller, M Angelo, Automated classification of cellular expression in multiplexed imaging data with Nimbus, bioRxiv [Preprint] 3 (2024), https://doi.org/10.1101/2024.06.02.597062.

[43] N. Zurek, Y. Zhang, D.P.B. McGovern, A.E. Walts, A. Gertych, Immunohistochemistry annotations enhance AI identification of lymphocytes and neutrophils in digitized H&E slides from inflammatory bowel disease, Comput. Methods Programs Biomed. 257 (2024) 108423, https://doi.org/10.1016/j.cmpb.2024.108423.