Synthesis of biomedical images based on generative intelligence tools*

Oleh Berezsky[1,*,†], Petro Liashchynskyi[1,†], Grygoriy Melnyk[1,†], Maksym Dombrovskyi[1,†] and Mykola Berezkyi[1,†]

[1] *West Ukrainian National University, 11 Lvivska st., Ternopil, 46001, Ukraine*

Abstract

The paper substantiates the use of generative intelligence tools to generate biomedical images. Analysis of the literature is conducted on methods and techniques for generating images using GAN and diffusion models. A new GAN architecture and algorithm have been developed for synthesizing cytological images based on a diffusion model. The analysis focuses on established datasets used for training deep neural networks. The widely recognized metrics for evaluating the quality of synthetic images are being analyzed: IS, FID. Computer experiments were conducted for synthesis of cytological images based on GAN and Stable Diffusion. The following results were obtained: diffusion model - FID – 0.63, IS – 3.99, GAN – FID – 3.39, IS – 3.95.

Keywords

cytological images, generative intelligence, image generation, generative adversarial networks, data sets, diffusion model, IS metric, FID metric

1. Introduction

Generative intelligence has now become the pinnacle of research in artificial intelligence. Generative intelligence systems allow you to generate texts, images, sounds, etc. Generative intelligence systems are based on deep neural network models that are trained on large samples of data.

Consequently, a variety of generative intelligence systems have emerged that transform text into image, image into image, image into text, sound into text, text into sound, sound into sound. Text-to-image transformation takes place on a fixed set of data. For this purpose, a transformer was used, which autoregressively simulates text and graphic tokens [1]. The Codex GPT language model, which is trained on GitHub, makes it possible to write code in Python. The paper [3] analyzes the opportunities and risks of fundamental models, such as language, vision, reasoning. In addition, the analysis of technical principles - the architecture of models, learning algorithms, data, is carried out. The impact of generative intelligence on society has also been studied.

Other papers [4] investigated a family of neurospeech models for LaMDA dialogue applications. The model generates responses based on learning from known sources. The authors investigated the LaMDA system in education.

Generative intelligence has also found applications in medicine. The paper investigates the use of generative intelligence in oncology, in particular for generating cytological images of breast cancer.

Breast cancer is one of the most common cancers among women worldwide. Early diagnosis and accurate determination of the stage of disease development are key factors for successful treatment and reduction of mortality. Cytological, histological and immunohistochemical images are used to

IDDM'24: 7th International Conference on Informatics & Data-Driven Medicine, November 14 - 16, 2024, Birmingham, UK

[*] Corresponding author.

[†] These authors contributed equally.

ob@wunu.edu.ua (O. Berezsky); p.liashchynskyi@gmail.com (P. Liashchynskyi); mgm@wunu.edu.ua (G. Melnyk); m20052002@gmail.com (Dombrovskyi); mykolaberezkyy@gmail.com (M. Berezkyi)

0000-0001-9931-4154 (O. Berezsky); 0000-0002-3920-6239 (P. Liashchynskyi); 0000-0003-0646-7448 (G. Melnyk); 0009-0008-2287-416X (M. Dombrovskyi); 0000-0001-6507-9117 (M. Berezkyi).

detect pathologies. These images are a class of biomedical images. Cytological analysis of images of cell preparations is one of the diagnostic methods, which allows the detection of pathological changes at the cellular level [5].

To train automatic systems for diagnosing breast cancer, large and high-quality datasets are needed that reflect the variety of possible pathological changes. Datasets of cytological images of breast cancer have the following features:

- Diversity of cell structures: normal cells, different types of atypical and malignant cells.
- Image variability: changes in color, lighting, focus, etc.
- Annotations & markups: availability of expert markup for supervised learning.

The available datasets of real images are limited and poorly annotated.

Therefore, an actual problem is the generation of biomedical images in oncology. This provides the necessary accuracy in the classification of biomedical images. To solve this problem, the paper uses the means of generative intelligence: GAN and diffusion models.

2. Literature review

Researchers in their works have developed a number of approaches to solving the problem of generating biomedical images. In particular, the article discusses the problems of creating medically significant fine-grained images of pulmonary adenocarcinomas using Stable Diffusion models [6]. The authors show how these models can be used to generate images with a limited number of samples, which is important for medical research where data can be scarce.

Other papers present the analysis of diffusion models in medical imaging [7]. The authors consider modern methods and approaches in the processing of medical images using deep learning, in particular diffusion models, which can significantly improve the quality of diagnostics.

The paper [8] presents a novel generative model that uses Langevin dynamics to generate samples by estimating gradients in data distribution with the addition of Gaussian noise. This avoids problems with low-dimensional manifolds and improves sample quality.

The paper explores how computer vision models trained on large sets of images from the Internet automatically learn human social biases, such as racism and sexism [9]. This question becomes important in the context of the ethical use of generative models.

The authors' article [10] describes the process of synthetic data generation in digital pathology using diffusion models. The authors present a comprehensive approach to assessing the quality of the generated images, which can be useful for educational purposes.

An article by A. Radford, J.W. Kim, C. Hallacy and other authors describes the CLIP model, which is trained on large datasets of images and texts to perform a variety of computer vision tasks without special training for each task [11]. The model demonstrates the ability to zero-learn on many datasets, which opens up new possibilities for application.

The authors R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer describe latent diffusion models for high-resolution image generation [12]. They use autoencoders to reduce the dimensionality of the data, which allows for a reduction in computational costs without losing image quality.

The article [13] presents the Imagen model, which is a text-to-image diffusion model with a high level of photorealism. The model uses large language models to encode the text, which greatly improves the quality of the samples.

A paper by other researchers describes the use of diffusion probabilistic models for the synthesis of histopathological images, which is important for pathology research [14].

The paper [15] presents diffusion probabilistic models used to generate high-quality images. These models demonstrate high-quality samples on various datasets, such as CIFAR10 and LSUN. Thus, the analysis of literature sources indicates significant progress in the development of image synthesis methods, in particular, through the use of diffusion models and GANs. This opens up new possibilities for improving the quality and diversity of synthesized images in medical imaging.

In the paper [16], researchers consider a deep learning approach using non-stationary thermodynamics. They represent diffusion probabilistic models that gradually break down the structure in the data through the diffusion process and then train the reverse process to reconstruct the structure, creating a flexible and computationally efficient generative model.

In the paper, the authors investigate diffusion models that are superior to generative adversarial networks (GANs) in image synthesis tasks [17]. They demonstrate that diffusion models can achieve high quality image samples, surpassing current generative models.

The paper [18] presents the use of cascading diffusion models to generate high-quality images. The cascade diffusion model consists of several stages, where each subsequent stage increases the resolution of the image.

The authors T. Karras, S. Laine, and T. Aila describe a new generator architecture for generative adversarial networks (GANs) that borrows ideas from stylistic transference [19]. This architecture allows for automatic and uncontrolled separation of high-level attributes from stochastic variations in generated images.

The paper describes a new approach to variational autoencoders (VAEs) for image generation [20]. The NVAE network uses deep-cut convolutions and batch normalization to improve the quality of generated images.

The paper [21] describes a novel approach to generative modeling that uses stochastic differential equations (SDE) to transform a data distribution to a simple noise distribution and vice versa. The model achieves high results in image generation and demonstrates the capabilities for solving inverse problems.

The authors of another paper [22] developed a method for filling images using diffusion probabilistic denoiseing models (DDPM). Based on this method, diverse and semantically meaningful images can be generated, surpassing current GAN-based methods

In [23], the authors describe improvements to diffusion probabilistic denoiseing models for image generation. They use accuracy and completeness metrics to compare images. Experiments have shown that diffusion models achieve higher completeness at similar values of the FID metric.

The authors of another publication [24] developed an algorithm for stochastic variational Bayesian inference. This approach allows you to train model parameters without using iterative inference schemes.

The authors of this publication have been analyzing biomedical images for over twenty years under the guidance of Professor Oleh Berezsky. A number of publications reflect methods, algorithms, and software tools for analyzing cytological, histological and immunohistochemical images [25-31]. This is the result of a creative collaboration of researchers from West Ukrainian National University and Ivan Horbachevsky Ternopil National Medical University.

3. Problem statement

Given: the set of real cytological images of I_C. Image synthesis will be carried out on the basis of GAN and networks that are built on *DMN* diffusion models. After generating by means of GAN, we get a set of I_{CG} images. Using *DMN*, we get a set of I_{CD} images. In addition, we are given two metrics: IS and FID.

It is necessary to find the M_{IS} and M_{FID} distances between the set of real I_C cytological images and the sets of I_{CG} and I_{CD} synthetic images using the IS and FID metrics, i.e.:

1. $M_{IS}(I_C, I_{CG})$ and $M_{FID}(I_C, I_{CG})$;
2. $M_{IS}(I_C, I_{CD})$ and $M_{FID}(I_C, I_{CD})$.

Compare:

3. $M_{FID}(I_C, I_{CD})$ and $M_{FID}(I_C, I_{CG})$;
4. $M_{IS}(I_C, I_{CD})$ and $M_{IS}(I_C, I_{CG})$.

4. Analysis of image datasets

When creating datasets of cytological images, it is important to standardize the annotation, as it ensures high quality, reliability and compatibility of data for their further use in machine learning and diagnostic processes. In addition, proper annotation increases the efficiency of training AI models, as well-defined labels reduce error rates in the learning process and help algorithms better recognize cell features and pathological changes. When segmenting and annotating objects on cytological images, it is important to adhere to the image annotation formats used in the PASCAL VOC [32] and COCO [33] datasets.

The APCData dataset [34] consists of cytological images of the cervix, developed in collaboration with the laboratory of anatomical pathology and cytology, located in Rivera, Uruguay. The set includes 425 images divided into 6 classes. The cells are labeled using bounding boxes and centers of the nuclei.

The dataset consists of 425 images of 2048 x 1532 pixels, corresponding to 73 diagnosed with Papanicolaou test. A total of 3619 cells were annotated. The images were taken using the Olympus CX40RF100 microscope and the Olympus LC30 Optical Microscope camera. Images are processed using Olympus L.Cmicro software. Bounding boxes were created for cells in an appropriate format for use with the YOLO convolutional neural network architecture.

The UFSC OCPap dataset [35] contains 9797 annotated images of 1200x1600 pixels in size, obtained from 5 slides with diagnosed oral tissue cancer and 3 healthy samples. The slides are provided by the Hospital Dental Center of the University Hospital of the Federal University of Santa Catarina. The dataset contains binary kernel masks and cell annotations in Json format. The images are divided into subsets of training, validation, and testing. The images were taken using an Axio Scan.Z1 microscope and a Hitachi HV-F202SCL camera. Dataset images are derived from virtual slides measuring 214,000 x 161,000 pixels (0.111 µm x 0.111 µm per pixel). For annotation, medical specialists used LabelMe and LabelBox tools.

The authors have developed a database of cytological images of breast cancer [36]. The image was obtained using a laboratory setup that includes a Delta Optical microscope, a Tucsen digital CMOS camera with a resolution of 8 megapixels. The sources of microscope slides and diagnostic information are provided by the Department of Pathological Anatomy with the Sectional Course of Forensic Medicine of the Ternopil National Medical University. The database consists of 14 related tables. The table of studies includes basic information about each study, its title, the object of the study, as well as references to the patient and doctor associated with this study.

All images of cytological samples are divided into 4 classes. The database supports several user roles: physician, expert, administrator. The database contains information about the segmentation algorithm used. For each cell there are the following features: area, perimeter, contour height, contour width, contour circularity, center coordinates, main axis of inertia, minor axis of inertia, angle of inclination of the main axis, Feret diameter, coordinates of the bounding rectangle, roundness, compactness.

5. GAN-Based Artificial Image Synthesis

As you know, the architecture of modern GANs consists of a generator and a discriminator [37].

The generator and discriminator architectures are based on cells. A cell consists of nodes performing an append operation and operations between them.

The following operations are used in the generator cell: convolution by kernel 1×1, 3×3, 5×5; separable convolution by kernel 3×3; zero; skip connection. The cell architecture remains the same for the entire generator model.

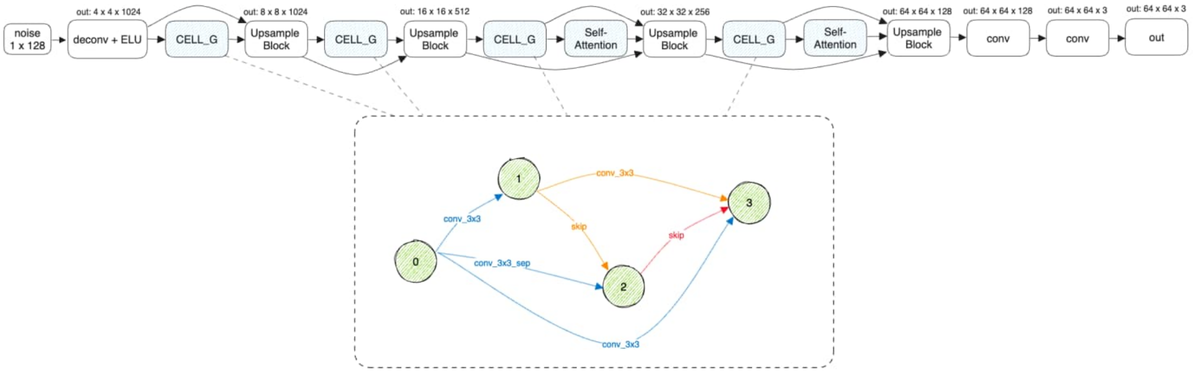

Figure 1: Generator Architecture

In contrast to the generator, the set of operations in the discriminator cell is extended by two operations: the maximum pooling by the kernel 3×3 and the average pooling by the 3×3 core. The architecture of the generator is shown in Figure 1 and described in Table 1 and Table 2.

Table 1
Generator Architecture

Layer	Options	Output= Form
L1: Input	Gaussian noise	1×128
L2: Transposed Conv + ELU activation	Kernel = 4, stride = 1, padding = 0	4×4×1024
L3: CELLG	Nodes = 4	4×4×1024
L4: L2 + L3		4×4×1024
L5: Upsample	Scale = 2	8×8×1024
L6: CELLG	Nodes = 4	8×8×1024
L7: L5 + L6		8×8×1024
L8: Upsample	Scale = 2	16×16×512
L9: CELLG	Nodes = 4	16×16×512
L10: Self Attention	Input channels = 512	16×16×512
L11: L8 + L10 + L9		16×16×512
L11: Upsample	Scale = 2	32×32×256
L12: CELLG	Nodes = 4	32×32×256
L13: Self Attention	Input channels = 256	32×32×256
L14: L11 + L13 + L12		32×32×256
L15: Upsample	Scale = 2	64×64×128
L16: Convolution	Kernel = 3, stride = 1, padding = 1	64×64×128
L17: Convolution	Kernel = 3, stride = 1, padding = 1	64×64×3
L18: Output		64×64×3

The discriminator architecture is shown in Figure 2 and described in Table 3 and Table 4.

The generator takes a noise vector from a Gaussian distribution of 1×128 as input, and outputs an image of 64×64×3.

The number of nodes in the generator and discriminator cells is 4 and 5 respectively. There are two *skip connection* operations in the generator cell, and 3 in the discriminator cell. There is also a *zero* operation in the discriminator cell, which is not present in the generator. The Self-Attention operation is applied 2 times in both the generator and the discriminator. However, in the generator, this operation is placed towards the end of the network, and in the discriminator, on the contrary, it is closer to the beginning.

Table 2
Generator *CELL_G* Cell and Upsample Block Structure

CELL$_G$ Cell Structure		
L0: Input		
L1: Conv → ELU → Batch Norm	Kernel = 3, stride = 1, padding = 1	
L2: L1 + Conv 3×3 → Conv 1×1 → ELU → Batch Norm	Conv 3x3 = (Kernel = 3, stride = 1, padding = 1), Conv 1x1 = (Kernel = 1, stride = 1, padding = 0)	
L3: L2 + Conv (L1) + Conv (L0)	Kernel = 3, stride = 1, padding = 1	
L0: Input		
L1: Conv → ELU → Batch Norm	Kernel = 3, stride = 1, padding = 1	
Upsample Block Structure		
L0: Input		H × W × C
L1: Upsample	Scale = 2, mode = nearest	(H × 2) × (W × 2) × C
L2: Convolution	Kernel = 3, stride = 1, padding = 1	(H × 2) × (W × 2) × C
L3: Conditional Batch Norm	Number of classes = 4	(H × 2) × (W × 2) × C
L4: Gated Linear Unit (GLU)	Dimension = 1	(H × 2) × (W × 2) × (C / 2)

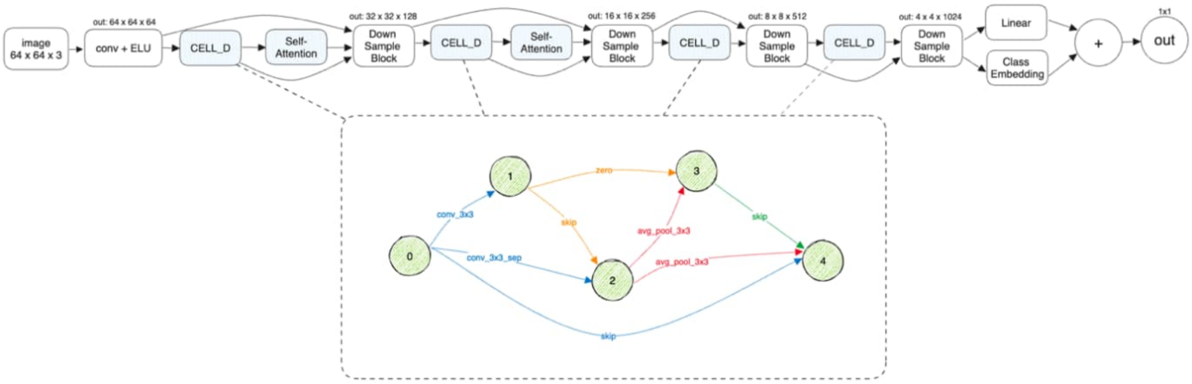

Figure 2: Discriminator architecture

6. Image Synthesis Based on Diffusion Model

Images based on the diffusion model are generated in the Stable Diffusion software environment. The basic Stable Diffusion model is trained on a large dataset of images. Training on the basis of its dataset takes place in the Hypernetwork neural network environment. This network adjusts the weights of the base model. The algorithm for generating images based on the diffusion model consists of the following steps:

1. training based on its dataset of images in the Hypernetwork environment;
2. the process of making noise of the initial I_C dataset;
3. noise reduction process.

Let's detail the steps. The initial dataset is transformed to a latency space: $I_C \to Z_{0C}$. Based on Z_{0C}, we calculate the noise value at each step t as follows:

$$Zt = \sqrt{\alpha_t} Z_{0C} + \sqrt{1-\alpha_t} \varepsilon_t,$$

where a_t is the coefficient that determines the noise rate at step *t*. The value of step *t* is selected from the range $t \in [0, T]$ where *T* is the number of steps; ε_t – is the value of random Gaussian noise at step *t*. Value ε_t calculated according to the expression:

$$\varepsilon_t : N(E, D),$$

where *N* is a normal distribution law with a expected value of *E* = 0 and a variance of D = 1.

Table 3
Discriminator architecture

Layer	Options	Output Form
L1: Input	Image	64×64×3
L2: Conv + ELU activation	Kernel = 3, stride = 1, padding = 1	64×64×64
L3: CELLD	Nodes = 5	64×64×64
L4: Self Attention	Input channels = 64	64×64×64
L5: L2 + L4 + L3		64×64×64
L6: Downsample	Scale = 2	32×32×128
L7: CELLD	Nodes = 5	32×32×128
L8: Self Attention	Input channels = 64	32×32×128
L9: L6 + L8 + L7		32×32×128
L10: Downsample	Scale = 2	16×16×256
L11: CELLD	Nodes = 5	16×16×256
L12: L10 + L11		16×16×256
L13: Downsample	Scale = 2	8×8×512
L14: CELLD	Nodes = 5	8×8×512
L15: L13 + L14		8×8×512
L16: Downsample	Scale = 2	4×4×1024
L17: Linear(Sum(L16))		1×1
L18: Sum(Multiply(Sum(L16), Embedding))	Number of classes = 4	1×1
L19: L17 + L18		1×1
L20: Output		1×1

The noise reduction value is calculated according to the expression:

$$Z_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(Z_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\widehat{\varepsilon_t}\right),$$

where $\widehat{\varepsilon_t}$ is the estimated noise value at step *t*; $\overline{\alpha_t}$ – the coefficient that determines the noise level in the previous step *t*; β_t is a coefficient that controls the level of noise reduction.

After performing the noise reduction process (after traversing *t*=*T* steps), a Z_{1C} vector is formed in the latency space. The encoder then transforms Z_{1C} into a set of I_{CD} images, with $I_{CD} \gg I_C$. The quality of the generated images is checked by IS and FID metrics.

7. Metrics for Synthesized Image Evaluation

Two main metrics are used to assess the quality of synthesized images: the IS metric and the FID *metric.*

The IS metric is based on the Google Inception V3 neural network model for color image classification. This metric was tested on the ImageNet dataset with a capacity of 1.2 million RGB images, which are divided into 1000 classes.

The analytic expression for the metric is as follows:

$$IS(G) \approx exp(E_{x\sim p_g}[D_{KL}(p(y|x) \,||\, p(y))]),$$

where E is the math expected value; $x\sim p_g$ shows what x an image synthesized from a distribution - $p_g(generator\ distribution)$; D_{KL} is the Kullback-Leibler distance between the conditional probability distribution and the marginal distribution $p(y)$ [38].

Table 4
Discriminator *CELL_D* Cell and Downsample block structure

CELL_D Cell Structure	
L0: Input	
L1: Conv → ELU → Batch Norm	Kernel = 3, stride = 1, padding = 1
L2: L1 + Conv 3×3 → Conv 1×1 → ELU → Batch Norm	(Kernel = 3, stride = 1, padding = 1), (Kernel = 1, stride = 1, padding = 0)
L3: AvgPool 3× 3 (L2)	Kernel = 3, stride = 1
L4: L0 + L3 + AvgPool 3× 3 (L2)	Kernel = 3, stride = 1

Downsample block structure		
L0: Input		H × W × C
L2: Convolution	Kernel = 3, stride = 1, padding = 1	H × W × (C × 2)
L3: Pixel Rearrange → Convolution	Kernel = 1, stride = 1, padding = 0	(H / 2) × (W / 2) × (C × 2)
L4: ELU		(H / 2) × (W / 2) × (C × 2)

The IS metric measures the average Kullback-Leibler distance between a conditional distribution $p(y|x)$ and a marginal class distribution $p(y)$. The minimum value of the metric is 1, and the maximum value is the number of classes.

The FID metric compares the distributions of original and synthetic data. Based on this metric, the distance between images is calculated as follows:

$$d^2((m_r C_r), (m_g C_g)) = \left\|m_r - m_g\right\|^2 + Tr(C_r + C_g - 2(C_r C_g)^{\frac{1}{2}}),$$

where $(m_r C_r)$ and $(m_g C_g)$ are the average and covariance of the real and synthesized data distributions respectively,

Tr — sum of the diagonal elements of the matrix.

Therefore, the smaller the value of the metric, the smaller the distance between the distributions, that is, the images are more similar to each other [39]. The FID metric is sensitive to distortion in images (shift, noise, etc.).

8. Computer experiments

Computer experiments on the synthesis of cytological images were carried out using GAN and Stable Diffusion.

To conduct computational experiments, a training set of cytological images was used, which was published on the Zenodo platform [40].

8.1. Computer experiments with GAN

Images from the training dataset have been transformed to a resolution of 64×64 pixels (the original resolution is 3264×2448). The initial number of images is around 100, which is not enough. Therefore, the dataset is expanded to 800 images by applying affine transformations. By applying this technique, the dataset was balanced – it contains the same number (200 images) for each class. To extend the training dataset, Rudi own library with default parameters [41] was used. Images are randomly rotated, flipped, scaled. All operations were applied with a probability of 50%.

Hardware. The Python programming language and the Pytorch framework were used to write the code. A virtual machine with the following configuration was used for experiments: 16 GB RAM, 10 vCPU x 2.2 GHz, Nvidia Tesla V100 GPU 16 GB (13.2 TFLOPS).

Training Options. In experiments, Hinge Loss was used as a loss function and Adam optimizer (betas = 0.5, 0.999). A technique called the Two Time-scale Update Rule is also used, which involves the use of different learning norms for the generator and the discriminator. Accordingly, the learning rate of the generator is 0.0001, and the discriminator is 0.0004. For all convolutional, deconvolutional, and linear layers, the spectral normalization technique was applied in both models, which allows to stabilize the learning process. Batch size – 128, number of iterations – 100,000. Training time ~13.6 GPU hours.

Experiment results. The FID metric value is 3.39 (Class 1 – 3.42, Class 2 – 3.42, Class 3 – 3.35, Class 4 – 3.37), and the IS metric value is 3.95

Examples of synthesized images are shown in Figure 3.

8.2. Computer experiments in Stable Diffusion environment

Stable Diffusion is a powerful AI model for generating images from text prompts that operates in a compressed latency space.

The main features of Stable Diffusion are as follows:

1. model Type: Latent Diffusion Text-to-Image Model;
2. training: Dataset "laion-aesthetics v2 5+";
3. architecture: Encoder, CLIP ViT-L/14 text encoder, UNet core model with cross-attention;
4. optimization: AdamW, 32 x 8 x A100 GPU.

Training Options. To train the model, the Linear loss function and the Adam optimizer were used. 768, 1024, 320, 640, 1280 layers with linear activation and initialization of Normal weights were chosen as the hypermodel structure. Batch size was set to 1 and Gradient Accumulation Steps to 1. Gradient Clipping with a value of 0.1 was used to stabilize learning. The training took place with a learning norm for the hypermodel of 0.00001. The total number of iterations was 20,000 steps, and the size of the images was fixed at 512x512 pixels. The training was carried out using text prompts based on a style_filewords.txt template. The intermediate results of the images were saved in the log directory every 100 steps.

Hardware. For the experiments, the infrastructure from Jarvis Labs was used, which has the following computing resources:

1. GPU: 1 x A6000 Ampere (CUDA 12.3);
2. Processors: 7 CPUs;
3. RAM: 32 GB RAM;
4. Video memory: 48 GB VRAM;
5. Linux system version: 22.04.

This configuration provides high performance for creating AI-generated images, allowing you to effectively use the capabilities of the Stable Diffusion model to generate high-quality results.

Experiment results. FID metric value – 0.63 (class 1 – 0.54, class 2 – 0.6, class 3 – 0.7, class 4 – 0.68). The value of the IS metric is 3.99.

An example of real images is shown in Figure 4. An example of synthetic images is shown in Figure 5.

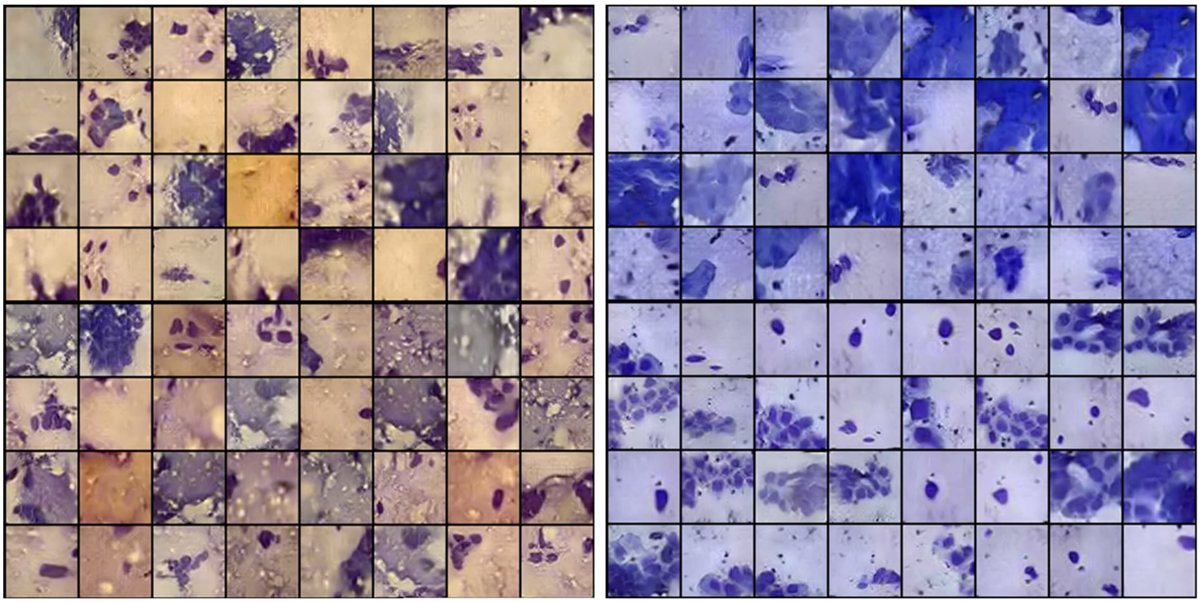

Figure 3: Examples of synthesized images

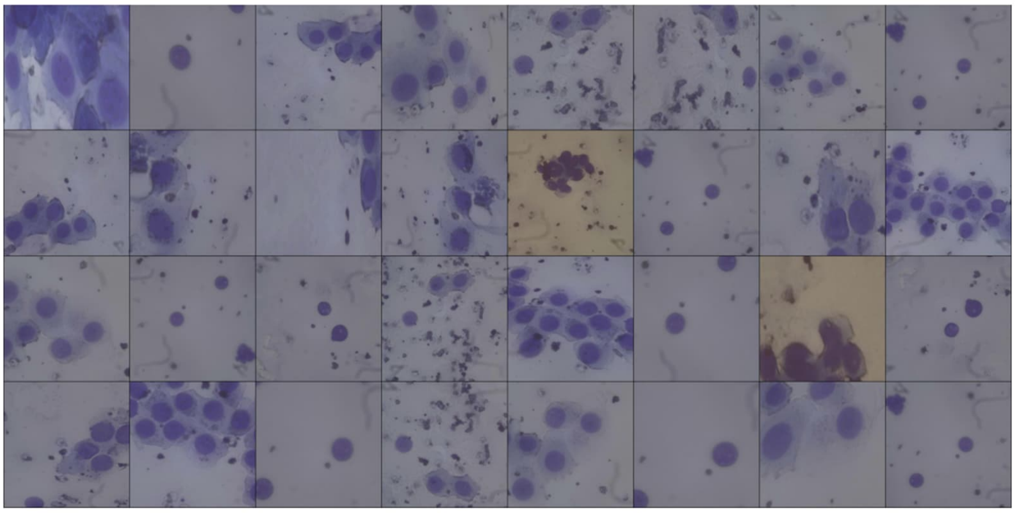

Figure 4: Example of Real Images

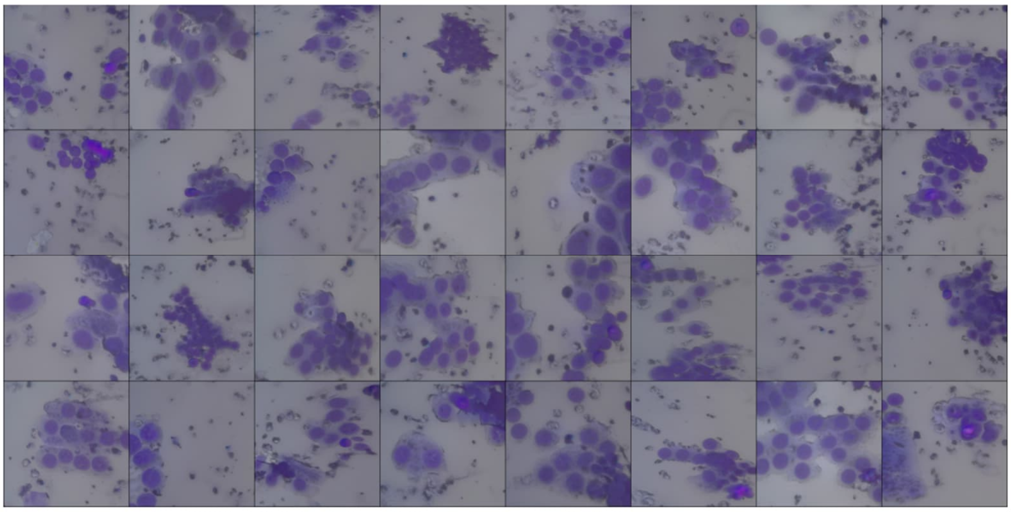

Figure 5: Example of synthetic images

9. Discussions

Let's analyze the conducted computer experiments using GAN and Stable Diffusion. The results of comparison of synthesized cytological images quality using the developed GAN architecture and other known architectures are given in Table 5.

Table 5
Results of comparison with other GAN architectures

Method	FID
DCGAN	12,67
WGAN	12,72
WGAN-GP	19,09
BGAN	10,03
BEGAN	15,32
Developed architecture	3,39

Consequently, the developed GAN architecture provided better results in terms of FID metrics than other well-known architectures.

Let's analyze the advantages and disadvantages of generating images based on GAN and based on diffusion models.

The advantages of GAN are as follows:

1. The ability to generate high-quality, realistic images, video, and audio.
2. The ability to control the synthesis process (from the smallest details to common features in the image).
3. Relatively high speed of image synthesis, which is synthesized in one pass (forward pass) of the neural network.

The disadvantages of GAN are as follows:

1. Significant computing resources and the need for expertise to learn effectively, making them less accessible.
2. Collapse mode, where the generator begins to produce a limited number of images, which reduces the variety of synthetic images.
3. The learning process is complex and long because GAN consists of two neural networks competing with each other.

The advantages of diffusion models are as follows:

1. The ability to produce high-quality images that often surpass GAN in terms of realism and variety.
2. The ability to work with complex data distributions, which makes diffusion models universal for different areas.
3. A simpler learning process compared to GAN, which avoids the problem of collapse.

The disadvantages of diffusion models are as follows:

1. Significant computing resources for training and generation, which may limit the availability of use.
2. Data generation using an iterative process is quite resource-intensive compared to the forward pass method used by GAN.

Diffusion models transform noise distribution into data distribution through a diffusion process, gradually improving the generated image. This process provides a high degree of control over the generation process, as the model can be stopped at any point to obtain different levels of detail.

However, GANs generate data in a single step, where the generator creates the image and the discriminator evaluates it. Although this process is faster, it can lead to collapse mode, where the generator produces a limited number of images.

Consequently, GAN is built using the concept of competition between a generator and a discriminator to create realistic images, while diffusion models transform noise into images through an iterative process of diffusion (noise reduce). Diffusion models involve careful tuning of hyperparameters and longer training times. In addition, both approaches require a large amount of training data to perform optimally.

10. Conclusions

As a result, the tools for synthesizing cytological images have been developed and their comparison has been conducted in the work.

At the same time, the following results were obtained:

1. A new GAN architecture has been developed, which, unlike existing architectures, uses the Self-Attention mechanism in the generator and discriminator, which made it possible to improve the quality of synthesized images. The developed architecture for image synthesis supports the mechanism of image synthesis by labels (conditional generation), which is not relevant for the above architectures and approaches.
2. A new algorithm for the synthesis of cytological images based on diffusion models has been developed. In the Stable Diffusion environment, an algorithm for synthesizing cytological images was implemented, which made it possible to synthesize a sufficient sample of images for CNN training.
3. Computer experiments based on the diffusion model in the Stable Diffusion environment were carried out, and the following results were obtained: the value of the FID metric is 0.63 (class 1 – 0.54, class 2 – 0.6, class 3 – 0.7, class 4 – 0.68), and the value of the IS metric is 3.99. Generating based on GAN provided the following results: FID – 3.39 (class 1 – 3.42, class 2 – 3.42, class 3 – 3.35, class 4 – 3.37), IS – 3.95.

Consequently, generation based on the diffusion model in the Stable Diffusion environment showed better results compared to generation based on GAN.

Therefore, further research will be the development of new diffusion models for generating histological and immunohistochemical images.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] A. Ramesh, M. Pavlov, G. Goh, Zero-Shot Text-to-Image Generation, arXiv preprint (2021). doi:10.48550/arXiv.2102.12092.

[2] M. Chen, J. Tworek, H. Jun, Evaluating Large Language Models Trained on Code, arXiv preprint, (2021). doi:10.48550/arXiv.2107.03374.

[3] R. Bommasani, D.A. Hudson, E. Adeli, On the Opportunities and Risks of Foundation Models, arXiv preprint (2021). doi:10.48550/arXiv.2108.07258.

[4] R. Thoppilan, D. De Freitas, J. Hall, LaMDA: Language Models for Dialog Applications, arXiv preprint (2022). doi:10.48550/arXiv.2201.08239.

[5] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram, A. Jemal, Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin (2024). doi:10.3322/caac.21834.

[6] Y. Xu, J. Liang, Y. Zhuo, L. Liu, Y. Xiao, L. Zhou, TDASD: Generating Medically Significant Fine-Grained Lung Adenocarcinoma Nodule CT Images Based on Stable Diffusion Models with Limited Sample Size, Computer Methods and Programs in Biomedicine 248 (2024) 108103. doi:10.1016/j.cmpb.2024.108103.

[7] A. Kazerouni, E. Khodapanah Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, D. Merhof, Diffusion Models in Medical Imaging: A Comprehensive Survey, Medical Image Analysis 88 (2023) 102846. doi:10.1016/j.media.2023.102846.

[8] Y. Song, S. Ermon, Generative Modeling by Estimating Gradients of the Data Distribution, NeurIPS 2019 (Oral) 2019. doi:10.48550/arXiv.1907.05600.

[9] R. Steed, A. Caliskan, Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), ACM, 2021 pp. 701–713. doi:10.1145/3442188.3445932.

[10] M. Pozzi, S. Noei, E. Robbi, L. Cima, M. Moroni, E. Munari, E. Torresani, G. Jurman, Generating Synthetic Data in Digital Pathology Through Diffusion Models: A Multifaceted Approach to Evaluation, bioRxiv preprint (2023). doi:10.1101/2023.11.21.23298808.

[11] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever, Learning Transferable Visual Models From Natural Language Supervision, arXiv preprint (2021). doi:10.48550/arXiv.2103.00020.

[12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, arXiv preprint (2021). doi:10.48550/arXiv.2112.10752.

[13] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S.K. Seyed Ghasemipour, B. Karagol Ayan, S.S. Mahdavi, R. Gontijo Lopes, T. Salimans, J. Ho, D.J. Fleet, M. Norouzi, Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, arXiv preprint (2022). doi:10.48550/arXiv.2205.11487.

[14] P. Azadi Moghadam, S. Van Dalen, K.C. Martin, J. Lennerz, S. Yip, H. Farahani, A. Bashashati, A Morphology Focused Diffusion Probabilistic Model for Synthesis of Histopathology Images, arXiv preprint, (2022). doi:10.48550/arXiv.2209.13167.

[15] J. Ho, A. Jain, P. Abbeel, Denoising Diffusion Probabilistic Models, arXiv preprint (2020). doi:10.48550/arXiv.2006.11239.

[16] J. Sohl-Dickstein, E.A. Weiss, N. Maheswaranathan, S. Ganguli, Deep Unsupervised Learning Using Nonequilibrium Thermodynamics, arXiv preprint (2015). doi:10.48550/arXiv.1503.03585.

[17] P. Dhariwal, A. Nichol, Diffusion Models Beat GANs on Image Synthesis, arXiv preprint (2021). doi:10.48550/arXiv.2105.05233.

[18] J. Ho, C. Saharia, W. Chan, D.J. Fleet, M. Norouzi, T. Salimans, Cascaded Diffusion Models for High Fidelity Image Generation, arXiv preprint (2021). doi:10.48550/arXiv.2106.15282.

[19] T. Karras, S. Laine, T. Aila, A Style-Based Generator Architecture for Generative Adversarial Networks, arXiv preprint (2019). doi:10.48550/arXiv.1812.04948.

[20] A. Vahdat, J. Kautz, NVAE: A Deep Hierarchical Variational Autoencoder, arXiv preprint (2020). doi:10.48550/arXiv.2007.03898.

[21] Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-Based Generative Modeling through Stochastic Differential Equations, arXiv preprint (2021). doi:10.48550/arXiv.2011.13456.

[22] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, L. Van Gool, RePaint: Inpainting using Denoising Diffusion Probabilistic Models, arXiv preprint (2022). doi:10.48550/arXiv.2201.09865.

[23] A. Nichol, P. Dhariwal, Improved Denoising Diffusion Probabilistic Models, arXiv preprint (2021). doi:10.48550/arXiv.2102.09672.

[24] D.P. Kingma, M. Welling, Auto-Encoding Variational Bayes, arXiv preprint (2013). doi:10.48550/arXiv.1312.6114.

[25] O. Berezsky, P. Liashchynskyi, O. Pitsun, I. Izonin, Synthesis of Convolutional Neural Network architectures for biomedical image classification, Biomedical Signal Processing and Control 95 (2024) 106325. doi:10.1016/j.bspc.2024.106325.

[26] O. Berezsky, P. Liashchynskyi, O. Pitsun, G. Melnyk, Method and Software Tool for Generating Artificial Databases of Biomedical Images Based on Deep Neural Networks, CEUR Workshop Proceedings, 2023 pp. 15–26.

[27] O. Berezsky, O. Pitsun, G. Melnyk, T. Datsko, I. Izonin, B. Derysh, An Approach toward Automatic Specifics Diagnosis of Breast Cancer Based on an Immunohistochemical Image, Journal of Imaging 9.1 (2023) 12. doi:10.3390/jimaging9010012.

[28] O. Berezsky, O. Pitsun, P. Liashchynskyi, B. Derysh, N. Batryn, Computational Intelligence in Medicine, In: S. Babichev, V. Lytvynenko (eds.), Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022, volume 149 of Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham, 2023. doi:10.1007/978-3-031-16203-9_28.

[29] O. Berezsky, P. Liashchynskyi, O. Pitsun, M. Berezkyy, Comparison of Deep Neural Network Learning Algorithms for Biomedical Image Processing, IDDM-2022, CEUR Workshop Proceedings, 2022, pp. 135–145.

[30] O. Berezsky, Y. Batko, G. Melnyk, S. Verbovyy, L. Haida, Segmentation of cytological and histological images of breast cancer cells, 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2015, pp. 287–292. doi:10.1109/IDAACS.2015.7340745.

[31] O. Berezsky, S. Verbovyy, T. Datsko, The intelligent system for diagnosing breast cancers based on image analysis, 2015 Information Technologies in Innovation Business Conference (ITIB), Kharkiv, Ukraine, 2015 pp. 27–30. doi:10.1109/ITIB.2015.7355067.

[32] The PASCAL Visual Object Classes Homepage, 2014. URL: http://host.robots.ox.ac.uk/pascal/VOC/

[33] Common Objects in Context dataset, 2021. URL: https://cocodataset.org

[34] P. Cuña Cabrera, APCData cervical cytology cells, 2024. URL: https://data.mendeley.com/datasets/ytd568rh3p/1. doi:10.17632/YTD568RH3P.1.

[35] André Victória Matias, UFSC OCPap: Papanicolaou Stained Oral Cytology Dataset (v4), 2022. URL: https://data.mendeley.com/datasets/dr7ydy9xbk/1. doi:10.17632/DR7YDY9XBK.1

[36] O. Berezsky, T. Datsko, G. Melnyk, V. Nykoliuk, O. Pitsun, S. Verbovyy. Database of Digital Histological and Cytological Images "BPCI2100". Database. Copyright registration certificate number 75359, December 14, 2017, bulletin No. 47 from January 26, 2018.

[37] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial networks, Communications of the ACM 63 (2014) 139–144.

[38] S. Barratt, R. Sharma, A Note on the Inception Score (2018). doi:10.48550/ARXIV.1801.01973.

[39] A. Borji, Pros and cons of GAN evaluation measures, Comput. Vis. Image Underst. 179 (2019) 41–65. doi:10.1016/j.cviu.2018.10.009.

[40] O. Berezsky, T. Datsko, G. Melnyk, Cytological and histological images of breast cancer, 2023. URL: https://doi.org/10.5281/zenodo.7890874. doi:10.5281/zenodo.7890874.

[41] Rudi, 2024. URL: https://github.com/liashchynskyi/rudi.