Proceedings of 2025 IEEE
International Conference on Mechatronics and Automation
August 3 - 6, Beijing, China

MASAN: Multi-Attention Based Sliced Adversarial Network for Improved B-cell Acute Lymphoblastic Leukemia Classification

1st Hao Wang*
College of Health Science and Environmental Engineering
Shenzhen Technology University
Shenzhen, China
wanghao100@126.com

2nd Wei Chen
School of Mathematics
South China University of Technology
Guangzhou, China
maweichen@mail.scut.edu.cn

3rd Wei Zhou
School of Computer Science and Technology
Beijing Jiaotong University
Beijing, China
mforzhou@bjtu.edu.cn

4th Bingding Huang
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China
huangbingding@sztu.edu.cn

5th Jiawan Zhang
College of Health Science and Environmental Engineering
Shenzhen Technology University
Shenzhen, China
2310417003@stumail.sztu.edu.cn

6th Peixuan Li
College of Health Science and Environmental Engineering
Shenzhen Technology University
Shenzhen, China
15175161193@163.com

2025 IEEE International Conference on Mechatronics and Automation (ICMA) | 979-8-3315-1426-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICMA65362.2025.11120678

***Abstract*—B-cell Acute lymphoblastic leukemia (B-ALL) is a prevalent malignant tumor disease, often diagnosed using peripheral blood smears (PBS). Deep learning technologies have emerged as powerful tools to assist pathologists in analyzing white blood cell counts and types via microscopy for B-ALL subtype determination. However, training these pathologists requires a significant number of expert-annotated pathological slice images, which is time-consuming and costly. In addition, the limitation in the number of available images for training hinders the enhancement of the model's diagnostic precision. In this paper, we propose MASAN, a GAN-constructed model to generate PBS images specific to B-ALL subtypes. Built upon an efficient backbone, MASAN incorporates an efficient multi-scale attention(EMA) module, leading to high-quality synthetic image generation. Our experimental results demonstrate that MASAN can generate highly realistic B-ALL subtype PBS images, thus improving the performance on subsequent downstream classification tasks. The overall FID value of the generated images is 50.03. Moreover, the accuracy of two distinct classifiers improved by approximately 8% when using these generated images.**

***Index Terms*—Generative Adversarial Network, Data Augmentation, Conditional Image Synthesis, Acute Lymphoblastic Leukemia Image Classification**

I. INTRODUCTION

Leukemia is a widespread blood cancer, with more than 486,777 new cases diagnosed worldwide each year [1]. It can generate malignant white blood cells (WBCs) in the human body. These abnormal blood cells affect the blood and bone marrow, making the immune system vulnerable. In the United States alone, it is estimated that there will be 62,770 new cases and 23,670 deaths in 2024 [2]. Leukemia is generally classified into two types: acute and chronic, based on the cell line of origin and the progression of the disease. Acute lymphoblastic leukemia is caused by the excessive proliferation of immature white blood cells in the bone marrow, which eliminates healthy cells. The rapid progression of the disease requires a prompt response. The incidence of ALL has risen from 49,100 cases in 1990 to 64,200 cases in 2017, accounting for approximately 12% of all leukemic cases [3]. ALL affects lymphoid cells and has a high probability of occurring in children.

B-cell acute lymphoblastic leukemia (B-ALL) is a variant of acute lymphoblastic leukemia that develops from abnormal B-cell progenitors. It encompasses a multiplicity of subtypes categorized based on different genetic, molecular, and immunophenotypic characteristics. The subtypes of B-ALL can be further classified into precursor, mature, common, and pro-B cell ALL.

Several diagnostic instruments and techniques are needed for an accurate diagnosis of leukemia. Bone marrow punctures are used in many hospitals to provide a conclusive diagnosis; nevertheless, these invasive procedures are expensive and might result in consequences like discomfort and bruises. Consequently, blood microscopy remains the main method for ALL screening and diagnosis in several developing nations. It entails spreading blood across the plate and examining it under a microscope. However, manual inspection is very subjective and time-consuming because it relies on the knowledge and expertise of medical specialists with formal training [4].

In recent years, computer-aided diagnosis (CAD) technology has been increasingly utilized for the analysis, detection, and identification of leukemia. Using these methods, the speed and accuracy of subtype classification can be improved compared with the manual method. In addition, objective results are provided by computer-aided diagnosis systems.

However, to train a really robust model, these tasks require a large volume of tagged images. Large-scale medical picture collections have not been created with much effort for a number of reasons. For instance, highly skilled laboratory workers are needed to annotate medical data rather than just anybody. Furthermore, patient privacy concerns frequently restrict the data's accessibility. For all of these reasons, the annotation costs increase significantly.

To address data shortage, traditional classification tasks employ data augmentation techniques like rotation, scaling, flipping, etc. However, these techniques provide limited improvement due to the lack of diversity in the data. Baniukiewicz et al. [5] synthesized 2D/3D single-cell images using two different GAN designs. GANs were used by Bailo et al. [6] to synthesize microscope blood pictures and improve the performance of the detection and segmentation job. Xue et al. [7] proposed a new cGAN model for downstream classification application, which improves the classification performance of two histopathological datasets.

FID is a widely used metric to evaluate generative models, however in the context of medical image synthesis, particularly for high-complexity images like leukemic cell structures, FID scores tend to be higher than in natural image domains. Recent studies on similar tasks have reported FID scores ranging from 20 to 70, depending on the modality and difficulty of the dataset. For instance, Hazra et al. [8] reported an FID of 67.2 in bone marrow cell image synthesis, and Ji et al. [9] achieved scores around 24.09 generating cell images from cellProfiler representations.

In this paper, we propose a novel conditional generative adversarial network (GAN) called MASAN for synthesizing various B-ALL cytopathological images in this paper. To ensure stable and efficient training, the proposed MASAN is built upon CellGAN [10]. Additionally, MASAN leverages cross-layer feature information and benefits the generator architectures by incorporating an EMA module. Our experimental results demonstrate that the model can generate more realistic peripheral blood images, thereby enhancing the overall classification performance.

II. MATERIALS AND METHOD

The generation of medical images faces a contradiction between the need for large volumes of data and the quality of model training. To ensure high-quality image generation with a limited dataset, our proposed MASAN is based on CellGAN [10]. The objective is to efficiently train with smaller sample sizes and synthesize peripheral blood images of various B-ALL subtypes, which is crucial for subsequent classification tasks. The overall structure of MASAN is illustrated in Fig.3.

A. Generator

In the generator, the input is a noise image that passes through a backbone network. This process consists of seven upsampling modules that progressively transform the input noise into a fake image $G(z)$. These upsampling modules play a critical role by gradually scaling the feature map size from an initial 4×4 resolution to a final 256×256 resolution. Simultaneously, the number of feature channels decreases from 512 down to 3, allowing the network to efficiently compress and expand the representation. Specifically, each upsampling module doubles the spatial dimensions of the image while halving the number of feature channels, ensuring a balanced transformation of both the spatial and feature aspects of the image.

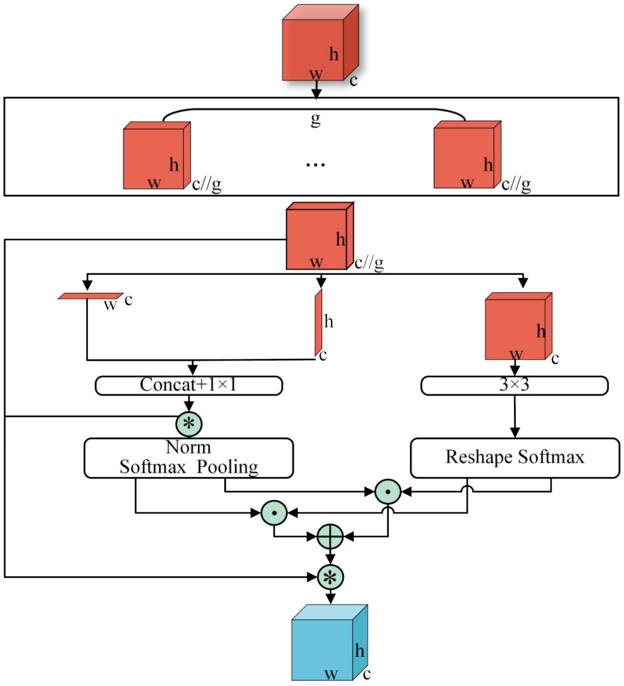

Fig. 1. The structure of the EMA module

To further refine the generated images and allow for greater diversity in the output, adaptive instance normalization (AdaIN) is applied at various intermediate layers of the network. AdaIN plays an important role by injecting class information, which guides the network to generate images from different categories based on the input conditions. This mechanism allows the generator to produce images across a wide range of classes with high fidelity.

In addition, previous research has shown the efficacy of incorporating visual attention mechanisms into deep learning models to enhance their performance. Attention mechanisms enable the model to focus on more informative regions in the input, improving both efficiency and accuracy. Therefore, we integrate the multi-scale attention (EMA) module [11], the structure of which is shown in Fig.1. The EMA module operates by attending to different spatial scales of the image, allowing the model to capture both fine-grained details and global features. The multi-scale attention mechanism can be described as follows:

$$A(x) = \sum_{i=1}^{N} \alpha_i \cdot f_i(x). \tag{1}$$

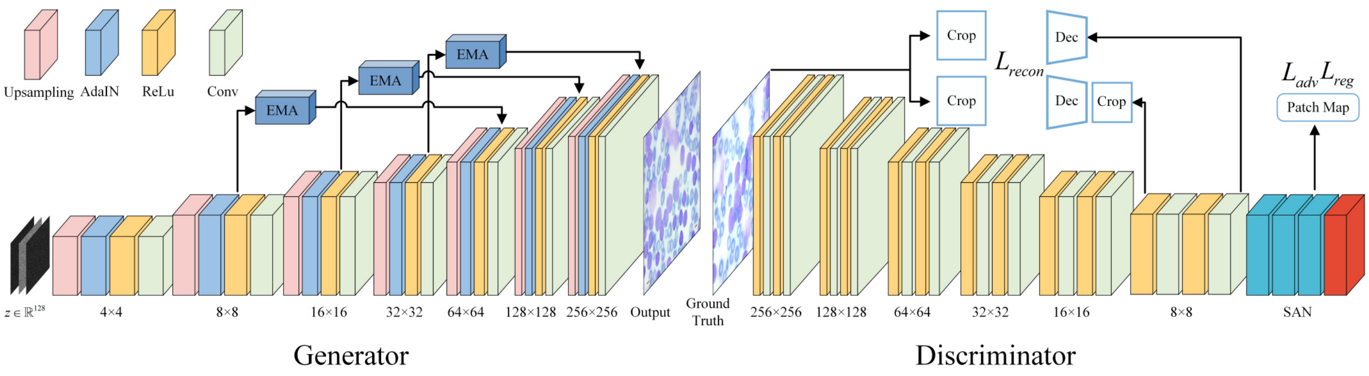

Fig. 2. Overall architecture of the proposed MASAN. The numbers indicate the feature map size

where, $A(x)$ represents the final attention map for the input x, $f_i(x)$ denotes the feature map at the i-th scale. α_i are the learnable attention weights for each scale, controlling the contribution of each scale, and N represents the total number of scales considered.

Each feature map $f_i(x)$ is generated by processing the input image through convolutional layers with different receptive field sizes, allowing the model to capture both local and global features. These multi-scale features are then weighted and aggregated to form a unified attention map that is used to modulate the input representation at each spatial location.

This approach enables the model to focus on the most relevant regions of the image at varying scales, which is particularly beneficial in complex, multi-scale scenes. By leveraging the EMA module, our model achieves better generalization and robustness in handling objects of different sizes and spatial relationships.

B. Discriminator

After obtaining the generated image, the discriminator is employed to provide adversarial guidance to the generator, thereby directing the optimization in a reasonable direction for image generation. The discriminator network takes real and generated images as input and converts them into an $8 \times 8 \times 512$ feature map through a series of 6 down-sampling modules. Meanwhile, two decoders are used to reconstruct the 82 and 162 feature maps and compare them with the cropped original images to establish the loss. The following loss function is used:

$$L_{recon} = E_{f \sim Dis(x), x \sim I_{real}}[\|Dec(f) - \mathcal{T}(x)\|_{\ell 1}]. \tag{2}$$

where $\mathcal{T}$ denotes the image processing on real image I_{real} including downsampling and random cropping, and Dec represents the reconstruction decoder.

In the last layer before the output of the discriminator, the discriminator produces an 8×8 feature map. To improve the optimization process of the generator, inspired by the GAN sliced optimal transport method [12], we adopt sliced optimal transport for GAN optimization, aiming to reduce gradient discrepancies in the generator's optimization. Specifically, the discriminator $f : X \to \mathbb{R}$ is usually parameterized by a neural network, which combines linear operations and non-linear activation functions. We decompose the discriminator into a combination of a non-linear function $h(x)$ and the last linear layer $f(x) = w^T h(x)$, i.e., $f(x) = \langle w, h(x) \rangle$ [13], where w represents the direction vector. This decomposition aligns with the slicing mechanism described in [12], which applies the Radon transform to project high-dimensional features onto one-dimensional spaces for tractable computation. The sliced Wasserstein distance (SW) is then formulated as:

$$\mathrm{SW}_p(\mu, \nu) = \left(\int_{\omega \in \mathbb{S}^{d-1}} W_p^p(RI_\mu(\cdot, \omega), RI_\nu(\cdot, \omega)) d\omega \right)^{\frac{1}{p}}. \tag{3}$$

where $RI_\mu(\cdot, \omega)$ represents the Radon transform of the probability density I_μ and W_p is the Wasserstein-p distance in one dimension.

To address the challenges in gradient flow induced by non-optimal discriminators, they proposed conditions for metrizable discriminators, including directional optimality, separability, and injectivity. These conditions are necessary to ensure the discriminator accurately reflects a valid distance measure between distributions. For instance, the directional optimality ensures:

$$\omega^* = \arg \max_{\omega \in \mathbb{S}^{d-1}} d_{\langle \omega, h \rangle}(\mu_0, \mu_\theta). \tag{4}$$

where $d_{\langle \omega, h \rangle}$ represents the directional discrepancy measured in the feature space defined by $h(x)$.

The advantages of sliced Wasserstein and its augmentations are evident in reducing computational costs while maintaining robustness. The original and updated loss functions are as follows.

$$\max_f \mathcal{V}(f; \mu_\theta) \text{ and } \min_\theta \mathcal{J}(\theta; f). \tag{5}$$

$$\max_{\omega, h} \mathcal{L}^h(h; \omega, \mu_\theta) + \lambda \mathcal{L}^o(\omega; h, \mu_\theta) \text{ and } \min_\theta \mathcal{J}(\theta; \langle \omega, h \rangle). \tag{6}$$

where, μ_θ is the probability distribution of the generator, f is the discriminator, w is the direction vector, and h means the feature extraction function.

TABLE I
COMPARISON OF CELLGAN AND MASAN UNDER DIFFERENT AUGMENTATION CONDITIONS

Model Setting		FID↓				
Aug	Model	Benign	Early	Pre	Pro	Mean
Non	CellGAN	364.79	398.24	329.54	392.78	371.34
	MASAN	311.71	407.98	356.91	397.48	368.52
Color	CellGAN	51.50	35.99	43.45	81.39	53.08
	MASAN	49.83	40.65	44.26	65.41	**50.03**
Color,Translation	CellGAN	73.22	72.22	83.45	105.18	83.52
	MASAN	61.88	47.05	65.98	99.25	68.54
Color,Cutout	CellGAN	114.86	100.93	112.64	132.43	115.21
	MASAN	62.71	42.86	51.35	94.51	62.86
Color,Cutout,Translation	CellGAN	73.57	56.81	56.59	115.68	75.66
	MASAN	64.76	54.08	58.12	110.31	71.82

C. Dataset

In this study, we used the open-source B-ALL dataset [14], created by the Bone Marrow Laboratory of Taleqani Hospital in Tehran. This dataset contains 3,256 peripheral blood smear (PBS) images from 89 patients suspected of having ALL, including 25 patients evaluated as benign and 64 diagnosed with one of the ALL subtypes. The dataset is categorized into four groups: ‘Benign’ class and three subtypes of B-Cell ALL, namely ‘Early pre-B’, ‘Pre-B’ and ‘Pro-B’, with 504, 985, 963, and 804 sample images, respectively. Blood samples were prepared and stained by laboratory personnel and accurately classified by experts. Each image is resized to 256×256 before being fed into the model.

Consistent hyperparameters were applied throughout the training process across experiments. We use the learning rate of 2.0×10^{-4}, batch size of 64, and SGD optimizer [15]. Each image is resized to 256×256 before being fed into the model and trained to 200K iterations per model.

III. RESULTS

A. Image Generation Comparative Analysis

We conducted comparative training experiments using the CellGAN and MASAN algorithms, applying various augmentation techniques during training. As shown in Table I, we compared models without any augmentation and then with the addition of augmentation techniques such as ‘Color’, ‘Translation’, and ‘Cutout’. The results show that MASAN consistently achieves lower FID scores compared to CellGAN across all augmentation methods. Notably, with the ‘Color’ augmentation, MASAN achieved an average FID score of 50.03.

In addition, we also performed a qualitative comparison using MASAN with CellGAN and Latent Diffusion Model (LDM). As shown in Fig.3, CellGAN tends to generate cells with blurred contours and darker colors, and the boundaries of cells are not clear; the color distribution in some areas appears unnatural.The images generated by LDM are clearer than those of CellGAN, with more distinct cell boundaries and less background blurring. for benign red blood cells also suffer from unclear contours, however, in some images, the color distribution of purple-stained areas appears unnatural, and the generated cell shapes are occasionally less diverse compared to real images. In contrast, MASAN demonstrates the ability to synthesize visually realistic cells for each category, producing sharper and more plausible images. The background processing and cell boundaries are closer to real images.

To verify the effect of components on MASAN, we performed ablation studies, as shown in Table II.

TABLE II
QUANTITATIVE ABLATION STUDY OF THE FOLLOWING KEY COMPONENTS: (A) COLOR AUGMENTATION, (B) EMA, (C) SLICED OPTIMAL TRANSPORT.

Model Setting			FID↓				
(a)	(b)	(c)	Benign	Early	Pre	Pro	Mean
-	-	-	364.79	398.24	329.54	392.78	371.34
✓	-	-	51.50	35.99	43.45	81.39	53.08
✓	✓	-	45.99	37.70	42.29	75.71	50.42
✓	✓	✓	49.83	40.65	44.26	65.41	**50.03**

It can be seen that using the Color augmentation method significantly enhances the model’s generation performance. Additionally, incorporating EMA could further decrease the model’s FID score. After adding sliced optimal transport, the average FID value also shows a slight decrease, indicating that the discriminator has a weak effect on improving the generator’s gradient optimization.

B. Evaluation of Augmentation Effectiveness

In order to verify the data augmentation capability of the proposed MASAN, we used three key evaluation parameters to assess the cell classification performance of two classifiers (ResNet34 [16], Resnext [17]]) conducting 5-fold cross-validations, and four distinct training data settings were designed for comprehensive comparison: (1) real data only (the Baseline); (2) baseline + CellGAN synthesized images; (3) Baseline + LDM synthesized images; (4) Baseline + MASAN synthesized images. Thus, this setup enabled a direct comparison of how synthetic images from different models impact the performance of the classifiers.

For each cell category in the dataset, we randomly selected 500 images and randomly assigned 400 images as the training set and 100 images as a test set. For the generated group, we generated 2000 images for each cell category and added them to each fold’s training data. We trained all the classifiers for

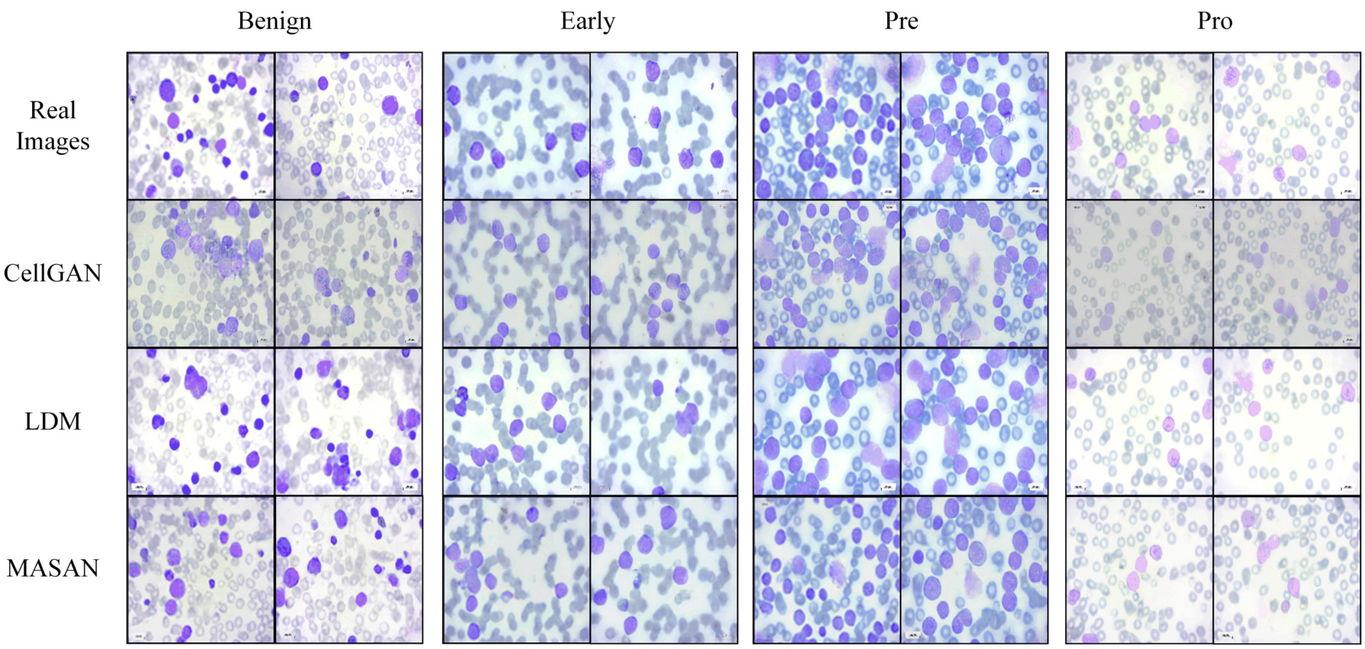

Fig. 3. Comparison Between Generated Images and Images Drawn from Dataset

30 epochs. At the same time, random flipping was used as a data augmentation method.

TABLE III
DATA AUGMENTATION COMPARISON BETWEEN THE PROPOSED MASAN AND OTHER SYNTHESIS-BASED METHODS USING RESNET AND RESNEXT CLASSIFICATION MODEL

Classifier	Method	Accuracy_top-1↑	Precision↑	F1_score↑
ResNet	Baseline	77.35±3.50	76.98±3.78	76.74±3.69
	+CellGAN	86.35±1.65	86.88±1.70	86.28±2.47
	+LDM	86.90±1.21	87.24±1.24	86.87±1.12
	+MASAN	**90.25±1.50**	**90.53±1.20**	**90.15±1.43**
ResNext	Baseline	86.40±1.40	86.39±1.24	86.22±1.22
	+CellGAN	90.20±1.50	90.29±1.42	89.96±1.18
	+LDM	90.45±0.54	90.51±0.53	90.28±0.63
	+MASAN	**94.00±2.25**	**94.14±2.11**	**93.92±1.90**

The accuracy, precision, and F1_score values of the experimental data are shown in Table III. It is shown that both classifiers achieve the best scores in all metrics using the additional synthesized data from MASAN. Three indicators of the classification task using images generated by MASAN have been improved to varying degrees compared with the baseline, indicating that our synthesized data can significantly enhance the overall classification performance.

IV. DISCUSSION

In this paper, we propose a novel GAN architecture designed to efficiently generate high-quality B-ALL images, The proposed model is capable of generating cellular tissue maps that are virtually indistinguishable to the naked eye, and it offers improved accuracy for subsequent classification tasks. Our experiments show that the performance of generating B-ALL subtype images is significantly enhanced by our model.

Compared with the existing literature, a large amount of work has focused on image generation for CT, MRI, and other imaging modalities, but the generation of pathology images has not been sufficiently studied. Similar to the works of W-DRAG [18], Fetal-brainawarenet [19], etc., our model follows an hourglass structure, wherein the generator gradually synthesizes the latent vector z from a 128-size vector to the full resolution of the image. To better capture the diversity in cellular spatial relationships, we incorporate the EMA between different layers to capture cross-layer feature information. Additionally, we integrate SAN at the final stage of the discriminator to optimize the gradient flow. We also noticed that each model in W-DRAG includes one generator and seven discriminators, which is a similar approch to Projected GANs [20]. The use of multiscale discriminators enhance the utilization of deeper features in the network, improving overall model performance. We plan to explore incorporating this concept into our subsequent work.

Despite the improvements demonstrated by MASAN, the model does have some limitations. First, the generated images tend to appear darker compared to the original images, likely due to the stained cells. The model seems to focus more on the purple-stained regions, leading to darker outputs. Additionally, the original dataset includes a small amount of labeled information in the bottom right corner of each image, which are used to note parameters during manual annotation. The model has learned this feature and includes it in the generated images, despite it being irrelevant for our purposes. Addressing this issue could be a focus of future research to prevent the model from generating these artifacts. Furthermore, since bone marrow aspiration remains the gold standard for diagnosing B-ALL, we consider using bone marrow aspiration data for pre-training the model or applying cross-validation for classification tasks in future studies. Lastly, while MASAN shows significant improvements in FID and classification accuracy, there is still room to improve the FID score of the generated images. More research is needed to further enhance the effectiveness of the model.

Meanwhile, we noticed that Jayasumana.et.al. [21] point out the drawbacks of FID score, and propose a new metric called CMMD, which provides a more robust and reliable assessment of image quality. Future work could involve comparative validation using CMMD to assess the quality of generated images more effectively.

V. CONCLUSION

In this paper, we have developed a method MASAN to synthesize photorealistic B-ALL subtype cell by utilizing conditional generative adversarial networks. These synthetic images are used alongside real data to meaningfully increase small datasets.

In summary, MASAN is built upon CellGAN for training efficiently, incorporating the EMA model and using it to transmit features across different layers. We also add SAN based on sliced optimal transport theory in order to guide the generator's optimization direction. Qualitative and quantitative validate the semantic realism as well as the data augmentation effectiveness of the synthesized images from MASAN.

REFERENCES

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, 2024.

[2] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024.," *CA: a cancer journal for clinicians*, vol. 74, no. 1, 2024.

[3] K. Dese, H. Raj, G. Ayana, T. Yemane, W. Adissu, J. Krishnamoorthy, and T. Kwa, "Accurate machine-learning-based classification of leukemia from blood smear images," *Clinical Lymphoma Myeloma and Leukemia*, vol. 21, no. 11, pp. e903–e914, 2021.

[4] R. B. Hegde, K. Prasad, H. Hebbar, B. M. K. Singh, and I. Sandhya, "Automated decision support system for detection of leukemia from peripheral blood smear images," *Journal of digital imaging*, vol. 33, pp. 361–374, 2020.

[5] P. Baniukiewicz, E. J. Lutton, S. Collier, and T. Bretschneider, "Generative adversarial networks for augmenting training data of microscopic cell images," *Frontiers in Computer Science*, vol. 1, p. 10, 2019.

[6] O. Bailo, D. Ham, and Y. Min Shin, "Red blood cell image generation for data augmentation using conditional generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 0–0, 2019.

[7] Y. Xue, J. Ye, Q. Zhou, L. R. Long, S. Antani, Z. Xue, C. Cornwell, R. Zaino, K. C. Cheng, and X. Huang, "Selective synthetic augmentation with histogan for improved histopathology image classification," *Medical image analysis*, vol. 67, p. 101816, 2021.

[8] D. Hazra, Y.-C. Byun, W. J. Kim, and C.-U. Kang, "Synthesis of microscopic cell images obtained from bone marrow aspirate smears through generative adversarial networks," *Biology*, vol. 11, no. 2, p. 276, 2022.

[9] Y. Ji, M. F. Cutiongco, B. S. Jensen, and K. Yuan, "Generating realistic single-cell images from cellprofiler representations," *Medical Image Analysis*, p. 103574, 2025.

[10] Z. Shen, M. Cao, S. Wang, L. Zhang, and Q. Wang, "Cellgan: Conditional cervical cell synthesis for augmenting cytopathological image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 487–496, Springer, 2023.

[11] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023.

[12] Y. Takida, M. Imaizumi, T. Shibuya, C.-H. Lai, T. Uesaka, N. Murata, and Y. Mitsufuji, "San: Inducing metrizability of gan with discriminative normalized linear layer," *arXiv preprint arXiv:2301.12811*, 2023.

[13] T. Shibuya, Y. Takida, and Y. Mitsufuji, "Bigvsan: Enhancing gan-based neural vocoders with slicing adversarial network," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 10121–10125, IEEE, 2024.

[14] M. Ghaderzadeh, M. Aria, A. Hosseini, F. Asadi, D. Bashash, and H. Abolghasemi, "A fast and efficient cnn model for b-all diagnosis and its subtypes classification using peripheral blood smear images," *International Journal of Intelligent Systems*, vol. 37, no. 8, pp. 5113–5133, 2022.

[15] H. Robbins and S. Monro, "A stochastic approximation method," *The annals of mathematical statistics*, pp. 400–407, 1951.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[17] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[18] C. Park, J.-W. Kang, D.-E. Lee, W. Son, S.-M. Lee, C. Park, and M. Kim, "W-drag: A joint framework of wgan with data random augmentation optimized for generative networks for bone marrow edema detection in dual energy ct," *Computerized Medical Imaging and Graphics*, vol. 115, p. 102387, 2024.

[19] A. Lasala, M. C. Fiorentino, A. Bandini, and S. Moccia, "Fetalbrainawarenet: Bridging gans with anatomical insight for fetal ultrasound brain plane synthesis," *Computerized Medical Imaging and Graphics*, p. 102405, 2024.

[20] A. Sauer, K. Chitta, J. Müller, and A. Geiger, "Projected gans converge faster," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17480–17492, 2021.

[21] S. Jayasumana, S. Ramalingam, A. Veit, D. Glasner, A. Chakrabarti, and S. Kumar, "Rethinking fid: Towards a better evaluation metric for image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9307–9315, 2024.