

STAIN IMPUTATION IN MULTIPLEX IMMUNOFLOUORESCENCE IMAGING (SIMIF) BASED ON RANDOM CHANNEL-WISE MASKING

Xingnan Li¹, Priyanka Rana¹, Tuba N Gide^{2,3,4}, Nurudeen A Adegoke^{2,3,4}, Yizhe Mao^{2,3,4},
Georgina Long^{2,3,4,5,6}, Richard A Scolyer^{2,7}, Shlomo Berkovsky¹, Enrico Coiera¹
James S Wilmott^{2,3,4}, Sidong Liu^{1,*}

¹ Centre for Health Informatics, Macquarie University, Sydney, NSW, Australia

² Melanoma Institute Australia, The University of Sydney, Sydney, NSW, Australia

³ Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

⁴ Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia

⁵ Royal North Shore Hospital, NSW, Australia

⁶ Mater Hospital, NSW, Australia

⁷ Department of Tissue Oncology and Diagnostic Pathology, Royal Prince Alfred Hospital,
and NSW Health Pathology, Sydney, NSW, Australia

ABSTRACT

Recent advances in digital imaging have fuelled interest in using multiplex immunofluorescence (mIF) images to study multiple biomarkers and their interactions within a single tissue of the tumour microenvironment. However, the mIF data remain less accessible due to the need for specialised equipment and costly reagents, which increases the technical complexity, expense and time required. Additionally, issues like misalignment and artefacts can result in unusable or incomplete data, highlighting the need for effective stain imputation methods. The current state-of-the-art (SOTA) stain imputation method relies on supervised generative deep-learning models with a fixed panel of biomarkers, which often fail when some of the required biomarkers are absent. To address this limitation, we propose a novel stain imputation method for mIF (SIMIF), which integrates the Wasserstein Generative Adversarial Network (GAN) with a random channel-wise masking (RCWM) training strategy. This method effectively enhances robustness in handling various input biomarkers, leading to a more stable optimisation process and improved synthetic image quality. Evaluated on a dataset of 33,265 image patches extracted from 36 mIF whole-slide images, SIMIF demonstrates clear superiority over the current SOTA, achieving marked improvements in imputation performance for both CD8 and PD-L1 biomarkers.

Index Terms—multiplex/multispectral immunofluorescence image, generative AI, stain imputation, WGAN

1. INTRODUCTION

Multiplex immunofluorescence (mIF) imaging has emerged as a powerful tool for visualising and analysing the interac-

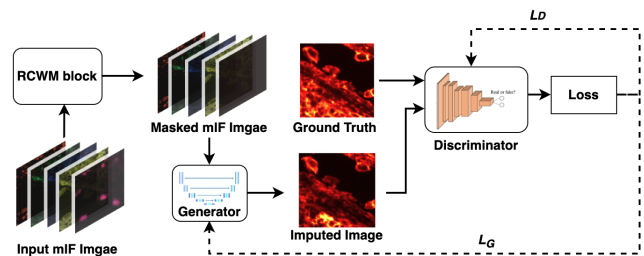


Fig. 1: The proposed SIMIF Method, integrating a generative adversarial network (GAN) model with a random channel-wise masking (RCWM) strategy for model training.

tions of various biomarkers within the tumour microenvironment at the cellular level. These images provide rich, multi-dimensional data that simultaneously capture the expression patterns of multiple biomarkers, offering critical insights into complex biological processes such as immune responses in cancer. However, despite their utility, mIF imaging is less accessible due to the need for specialised equipment and the extensive number of biomarkers required for comprehensive analysis. To address this limitation, stain imputation has become a crucial area of research, employing computational models to impute new or missing biomarkers from available biomarkers in mIF images. This approach maximises the utility of available tissue samples while minimising the need for repeated staining and the use of costly reagents.

Recent advances in medical image generation have primarily focused on radiology [1, 2], while studies in stain imputation in histopathology images, such as immunohistochemistry or mIF images, are limited. Recently, Shaban et

al. [3] introduced the MAXIM model specifically for mIF images, using a conditional Generative Adversarial Network (cGAN) [4] to generate the KI67 biomarker from other stains. While effective for stain imputation, MAXIM only supports a fixed panel of biomarkers, i.e., specific inputs for specific output. This approach sacrifices flexibility and is impractical, as it requires separate models for different combinations of input and output biomarkers.

The need for a robust and flexible method for stain imputation in mIF images remains unmet. To address this, our study introduces a framework, Stain Imputation in mIF images (SIMIF), that uses a Conditional Wasserstein GAN-gradient penalty (cWGAN-GP) [5] and incorporates a random channel-wise masking (RCWM) strategy for model training, as illustrated in Fig. 1. This allows the model to learn the joint distribution of various combinations of input biomarkers, facilitating the imputation of varying input biomarkers. Specifically, the two components in SIMIF address the restricted applicability of previous work [3] by enhancing stability in the training process, and flexibility during inference.

We have a dataset of mIF whole-slide images (WSIs) with 6 markers/channels and an autofluorescence (AF) channel. Given the clinical significance of CD8 and PD-L1 in predicting patients' response to immunotherapy, we show the effectiveness of SIMIF in imputing these biomarkers from combinations of other available biomarkers. The results show SIMIF offers greater flexibility in handling input biomarkers and superior image quality, compared to the baseline MAXIM, particularly when fewer biomarkers are available.

2. MATERIALS AND METHODS

2.1. Dataset

Our dataset consists of 36 mIF WSIs obtained from the Melanoma Institute Australia [6, 7], each acquired from a melanoma patient at an advanced stage. The images are presented with 7 channels corresponding to 6 biomarkers (CD68, SOX10, CD16, CD8, PD-L1, DAPI) and an additional autofluorescence (AF) channel, and have a resolution exceeding $20,000 \times 10,000$ pixels, scanned at $0.5 \mu\text{m}/\text{pixel}$.

Among these biomarkers, CD8 and PD-L1 are more directly linked to immune regulation and cancer therapy, making them critical for predicting a patient's immune response in the context of immunotherapy [8]. However, detecting CD8 and PD-L1 biomarkers poses technical challenges due to their uneven distribution and low expression levels, necessitating the use of high-resolution, advanced imaging systems for accurate detection and quantification [9]. By using other biomarkers to generate CD8 and PD-L1, the need for complex staining procedures may be alleviated. Accordingly, this study aims to apply SIMIF in imputing PD-L1 and CD8.

2.2. Pre-processing

The images are normalised to ensure consistent intensity levels. To eliminate large background areas, masks are generated using adaptive threshold with Otsu's method [10]. To address the challenge posed by low expression levels of PD-L1 and CD8 biomarkers and minimise empty patches, we combine the PD-L1 and CD8 masks for tissue segmentation.

The segmented tissue regions are divided into 224×224 pixel patches, ensuring a cleaner dataset for model training. To maintain consistency, we extracted 900 to 2,500 patches from each WSI, depending on the number produced by each. To improve training efficiency, we applied a 50% sampling rate, resulting in a final dataset of 33,265 patches.

2.3. Conditional WGAN-GP

A Generative Adversarial Network (GAN) learns the distribution of training data through the joint optimisation of two networks: the Generator, which aims to create realistic data, and the Discriminator, which attempts to distinguish between real and generated data. The two networks are trained alternately, focusing on maximising the probability of classifying real images as real and generated samples as synthetic/fake. However, training GAN can be challenging and is prone to several common failure modes, such as vanishing gradients, mode collapse, and unstable training [11].

To address these problems, Wasserstein GAN (WGAN) [12] was proposed which uses a new loss function based on Wasserstein distance instead of the Jensen-Shanon Divergence which is the standard objective function of GAN. This new loss function provides smoother and more informative gradients, addressing mode collapse and vanishing gradient problems. However, the computation of Wasserstein distance requires a k -Lipschitz constraint, which is a strict constraint. WGAN-GP [5], a variant of WGAN, employs gradient-penalty to enforce the Lipschitz constraint, resulting in more stable training and enhancing the quality of the generated synthetic images, producing more realistic outputs [13, 11]. The cWGAN-GP further integrates a conditioning mechanism that enables the model to generate data with greater control and precision, ensuring that the outputs correspond to specific categories or attributes [4].

The SIMIF model leverages cWGAN-GP and utilises the input biomarkers as conditional information to guide the generation of target biomarkers. The objective function (discriminator loss L_D) for our model is shown as follows:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_r(x)} [D(x|c)] - \mathbb{E}_{\hat{x} \sim P_g(\hat{x})} [D(\hat{x}|c)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x}|c)\|_2 - 1)^2] \quad (1)$$

where $x \sim P_r(x)$ represents real samples, while $\hat{x} \sim P_g(\hat{x})$ refers to generated samples. $D(x|c)$ and $D(\hat{x}|c)$ are the discriminator's scores for real and generated data conditioned on

c. The discriminator maximises the score difference between real and generated images, while the generator minimises it, making generated data indistinguishable by minimising the loss $L_G = -\mathbb{E}_{\hat{x} \sim P_g(\hat{x})}[D(\hat{x}|c)]$. The parameter λ regulates the gradient penalty for improved training stability.

We employed U-Net [14] as the generator, since its encoder-decoder architecture with skip connections captures multi-scale spatial features, preserving both high-level context and low-level details, critical for accurate stain imputation. The discriminator is a convolutional neural network, following the Pix2Pix design [15].

2.4. Random Channel-Wise Masking (RCWM)

To enable the model to impute target biomarker images with flexible input biomarkers, we propose an RCWM strategy. This approach allows the model to effectively learn the joint distribution of multiple input biomarkers and impute target stains, even when minimal biomarkers are available (e.g., DAPI and AF). Given the importance of CD8 and PD-L1, they are used as target biomarkers for imputation. DAPI and AF serve as basic input biomarkers, while we experiment with including other biomarkers in various input combinations.

The model is initially trained using all the available input biomarker images. This initial training ensures that the model can accurately impute the target biomarkers in the presence of all available input biomarkers, and learn a complete view of their relationships. Following that, random masking is applied to the input biomarkers based on an empirically predefined probability of masking one or more biomarkers out during training. Specifically, we initially set the probabilities for randomly masking 0, 1, 2, and 3 channels to 70%, 20%, 10%, and 0%, respectively. That means, for the CD68, SOX10 and CD16 biomarkers, at each epoch, the probability of no masking (p_0) is 70%, while the probability of masking one channel (p_1) is 10%, and so on. These probabilities are updated every 20 epochs during training to encourage the model to adapt to varying input biomarkers and reduce dependency on any specific biomarkers. The updates follow these steps: 1) p_0 is decreased by 10%, while p_1 and p_2 are each increased by 5% every 20 epochs until p_0 reaches its target; 2) p_1 is then reduced by 10%, while p_2 and p_3 are increased by 5% until p_1 reaches its target; and 3) p_2 and p_3 are finally adjusted in a similar manner. The final target probabilities in this study are set to 0%, 20%, 30%, and 50%.

Notably, the initial probability settings help the model progressively adapt to missing input biomarkers, mitigating early training challenges. In contrast, the target settings ensure the model comprehensively understands different biomarker combinations, preventing bias toward data with fewer input biomarkers. Additionally, these values may require further hyperparameter tuning to optimise performance.

2.5. Implementation

To prevent data leakage, we split the dataset at the patient level while ensuring that WSIs with varying numbers of patches are evenly distributed across the dataset. Specifically, we use stratified splitting to assign 20% of the data as a test set (7 patients), 20% as a validation set (8 patients), and 60% as a training set (21 patients). The input sets for both CD8 and PD-L1 are identical, containing 5 input markers (CD68, SOX10, CD16, DAPI, and AF).

All models are trained for 550 epochs with a learning rate of 1×10^{-5} for the discriminator and 1×10^{-4} for the generator, using the Adam optimiser and a batch size of 16 on Tesla V100 GPUs. The learning scheduler, optimiser, and initial hyperparameters are fixed across all experiments.

3. EXPERIMENTAL RESULTS

We conducted the following evaluations. Firstly, we used all available biomarkers to compare the proposed SIMIF with MAXIM, the current SOTA method for mIF stain imputation. Secondly, we compared the two models when they were trained on five biomarkers but used only DAPI and AF for imputation during inference. Thirdly, to evaluate the significance of RCWM, we compared SIMIF and MAXIM with and without RCWM in an ablation study. The imputed biomarker images are evaluated using the Pearson Correlation Coefficient (Corr), Structural Similarity Index Measure (SSIM), and Mean Absolute Error (MAE). For qualitative evaluation, we visualised the imputed CD8 and/or PD-L1 biomarker images derived from different methods in Fig. 2 and Fig. 3.

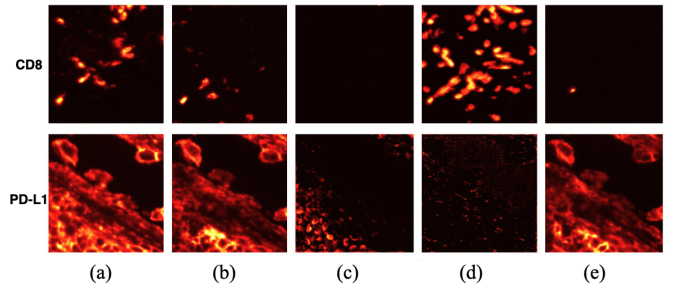


Fig. 2: Visualisation of real and imputed images for CD8 (top row) and PD-L1 (bottom row) by different methods: (a) ground truth; (b) SIMIF; (c) SIMIF (w/o RCWM); (d) MAXIM; and (e) MAXIM (w/ RCWM).

Our experimental results (Table 1) demonstrate that when generating target biomarker images using all five input biomarkers, SIMIF achieved comparable performance to the MAXIM method, with a SSIM score of 0.687 and a Corr value of 0.752 for CD8, and 0.613 and 0.754 for PD-L1. However, when using only DAPI and AF as input, SIMIF substantially outperformed MAXIM, with an increase of 0.282

Table 1: Performance comparison across different input biomarker combinations

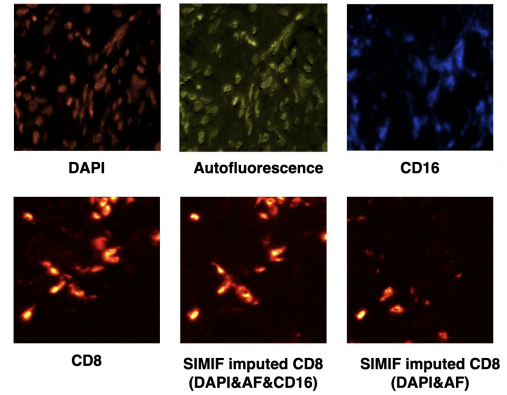
Study	Input Biomarkers					Imputed CD8			Imputed PD-L1		
	DAPI	AF	CD16	CD68	Sox10	SSIM \uparrow	Corr \uparrow	MAE \downarrow	SSIM \uparrow	Corr \uparrow	MAE \downarrow
MAXIM	✓	✓	✓	✓	✓	0.701 \pm 0.008	0.750 \pm 0.005	7.06	0.594 \pm 0.005	0.732 \pm 0.005	17.29
SIMIF (Ours)	✓	✓	✓	✓	✓	0.687 \pm 0.006	0.752 \pm 0.008	7.38	0.613 \pm 0.005	0.754 \pm 0.005	18.70
MAXIM	✓	✓	×	×	×	0.317 \pm 0.008	0.063 \pm 0.004	13.22	0.04 \pm 0.003	-0.03 \pm 0.003	38.59
SIMIF	✓	✓	×	×	×	0.599\pm0.007	0.160\pm0.007	10.73	0.522\pm0.005	0.563\pm0.007	18.64
MAXIM w/ RCWM	✓	✓	×	×	×	0.489 \pm 0.007	0.130 \pm 0.006	13.12	0.493 \pm 0.005	0.536 \pm 0.008	19.18
SIMIF w/o RCWM	✓	✓	×	×	×	0.323 \pm 0.009	0.029 \pm 0.002	12.22	0.057 \pm 0.003	0.072 \pm 0.006	37.93

Table 2: Performance comparison between MAXIM and SIMIF on stain imputation using three input biomarker images

Study	Input Biomarkers					Imputed CD8			Imputed PD-L1		
	DAPI	AF	CD16	CD68	Sox10	SSIM \uparrow	Corr \uparrow	MAE \downarrow	SSIM \uparrow	Corr \uparrow	MAE \downarrow
MAXIM	✓	✓	✓	×	×	0.473 \pm 0.008	0.441 \pm 0.010	10.31	0.324 \pm 0.007	0.479 \pm 0.008	26.45
SIMIF	✓	✓	✓	×	×	0.644\pm0.010	0.485\pm0.007	9.36	0.540\pm0.005	0.590\pm0.007	18.48
MAXIM	✓	✓	×	✓	×	0.553 \pm 0.008	0.125 \pm 0.008	11.45	0.080 \pm 0.004	0.047 \pm 0.005	37.16
SIMIF	✓	✓	×	✓	×	0.580\pm0.007	0.152\pm0.007	10.72	0.532\pm0.005	0.582\pm0.007	18.14
MAXIM	✓	✓	×	×	✓	0.395 \pm 0.009	0.047 \pm 0.005	11.87	0.486 \pm 0.006	0.527 \pm 0.008	21.81
SIMIF	✓	✓	×	×	✓	0.591\pm0.008	0.236\pm0.009	10.51	0.526\pm0.005	0.541\pm0.008	19.21

in SSIM and 0.097 in Corr for CD8, and 0.518 in SSIM and 0.569 in Corr for PD-L1. SIMIF remains robust with DAPI and AF, whereas MAXIM struggled with the same inputs, resulting in much lower SSIM scores (0.317 for CD8 and 0.040 for PD-L1) and Corr values (below 0.1). The imputed biomarker images, as shown in Fig. 2(d), are predominantly noisy or empty. Removing RCWM from SIMIF led to a marked decline in the quality of imputed images. In contrast, incorporating RCWM into MAXIM substantially improved its performance, with SSIM scores increasing by 0.172 and 0.453 for CD8 and PD-L1, respectively. However, MAXIM with RCWM still remained inferior to SIMIF. These findings demonstrate that the RCWM strategy effectively enhances the model’s performance and robustness when only a limited number of biomarkers are available. SIMIF also relaxes the constraints of fixed input and output biomarkers without requiring training separate models, addressing a key limitation of current SOTA methods.

When comparing the imputation performance of PD-L1 and CD8, CD8 generally achieves a lower Corr value (below 0.2). All models struggled to generate accurate CD8 biomarker images by using only DAPI and AF (Fig. 2 top row). This difficulty arises as CD8 is located on the cell membrane [16], while DAPI stains the nucleus [17], with no direct relationship between them. Additionally, AF often provides weak and noisy information, further complicating the task (Fig. 3 top row). To address this issue, we introduce an additional biomarker image to help the model impute CD8. We found that adding CD16 significantly improves our model’s performance, resulting in a 0.325 increase in Corr, as shown in Table 2 and Fig. 3 bottom row. However, adding an additional biomarker only results in marginal improvement for PD-L1, as DAPI and AF likely capture most of the PD-L1 regions.

**Fig. 3:** Visualisation of DAPI, AF and CD16 (top row) and comparison between real and imputed CD8 with and without the additional CD16 biomarker image (bottom row).

4. CONCLUSION

This work introduces SIMIF, a method for stain imputation in multiplexed immunofluorescence images. Compared to the current state-of-the-art, SIMIF effectively handles varying input biomarkers while maintaining high imputation performance. Tested on a dataset of 33,265 mIF image patches, SIMIF demonstrates robust performance in producing high-quality PD-L1 images using only DAPI and autofluorescence images as input. We also underscore the importance of including CD16 for accurate CD8 imputation. Future work will focus on extending the model to impute additional biomarkers and further enhance the quality of imputed images.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study involves human participants and was approved by the Sydney Local Health District Human Research Ethics Committee (Protocol No. X20-0086 and 2020/ETH00426), with informed consent obtained from all participants.

6. CONFLICTS OF INTEREST

G.V.L. is consultant advisor for Aduro Biotech Inc, Amgen Inc, Array Biopharma inc, Boehringer Ingelheim International GmbH, Bristol-Myers Squibb, Hexel AG, Highlight Therapeutics S.L., Merck Sharpe & Dohme, Novartis Pharma AG, OncoSec, Pierre Fabre, QBiotics Group Limited, Regeneron Pharmaceuticals Inc, SkylineDX B.V., and Specialised Therapeutics Australia Pty Ltd.

7. ACKNOWLEDGMENTS

This work was supported by the NHMRC, the Centre for Research Excellence in Digital Health, and the NHMRC Investigator Grant. Data generation was funded by the NHMRC, the Melanoma Research Alliance, Cancer Council NSW, and the CINSW Transnational Program Grant. T.N.G. is supported by a CINSW Early Career Fellowship (2020/ECF1244). G.V.L. and R.A.S. are supported by NHMRC Investigator Grants (GNT2007839, GNT2018514), and J.S.W. by an NHMRC Fellowship (APP1174325), a Cancer Council NSW project grant (RG19-15), and a CINSW Translational Program Grant (TPG 2021/TPG2114). G.V.L. is also supported by the Melanoma Foundation of the University of Sydney.

8. REFERENCES

- [1] Cao B., Bi Z.W., Hu Q.H., Zhang H., Wang N.N., Gao X.B., and Shen D.G., "Autoencoder-driven multimodal collaborative learning for medical image synthesis," *Int. J. Comput. Vision*, vol. 131, no. 8, pp. 1995–2014, 2023.
- [2] Dalmaz O., Yurt M., and Cukur T., "Resvit: Residual vision transformers for multimodal medical image synthesis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, Oct. 2022.
- [3] Shaban M., Lassoued W., Canubas K., Bailey S., Liu Y., Allen C., Strauss J., Gulley J. L., Jiang S., Mahmood F., Zaki G., and Sater H. A., "Deep learning model imputes missing stains in multiplex images," *bioRxiv*, 2023.
- [4] Mirza M. and Osindero S., "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.
- [5] Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., and Courville A. C., "Improved training of wasserstein gans," *Neural Information Processing Systems*, 2017.
- [6] Yaseen Z., Gide T.N., Conway J.W., Potter A.J., Quek C., Hong A.M., Long G.V., Scolyer R.A., and Wilmott J.S., "Validation of an accurate automated multiplex immunofluorescence method for immuno-profiling melanoma," *Frontiers in Molecular Bioscience*, vol. 9, no. e810858, 2022.
- [7] Gide T.N., Quek C., Menzies A.M., Tasker A.T., Shang P., Holst J., Madore J., Lim S.Y., Velickovic R., Wongchenko M., Yan Y., Lo S., Carlino M.S., Guminski A., Saw R.P.M., Pang A., McGuire H.M., Palendria U., Thompson J.F., Rizzos H., Da Silva I.P., Batte M., Scolyer R.A., Long G.V., and Wilmott J.S., "Distinct immune cell populations define response to anti-pd-1 monotherapy and anti-pd-1/anti-ctla-4 combined therapy," *Cancer Cell*, vol. 35, no. 2, pp. 238–255, 2019.
- [8] Deng M., Li S. H., Fu X., Yan X. P., Chen J., Qiu Y. D., and Guo R. P., "Relationship between pd-1 expression, cd8+ t-cell infiltration and prognosis in intrahepatic cholangiocarcinoma patients," *Cancer cell international*, vol. 21, no. 1, pp. 371, 2021.
- [9] Nimmagadda S., "Quantifying pd-1 expression to monitor immune checkpoint therapy: Opportunities and challenges," *Cancers*, vol. 12, no. 11, pp. 3173, 2020.
- [10] Otsu N., "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [11] Rana P., Sowmya A., Meijering E., and Song Y., "Data augmentation with improved regularisation and sampling for imbalanced blood cell image classification," *Scientific Reports*, vol. 12, no. 1, pp. 18101, 2022.
- [12] Arjovsky M., Chintala S., and Bottou L., "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, Doina Precup and Yee Whye Teh, Eds. 06–11 Aug 2017, vol. 70 of *Proceedings of Machine Learning Research*, pp. 214–223, PMLR.
- [13] Rana P., Sowmya A., Meijering E., and Song Y., "Imbalanced cell-cycle classification using wgan-div and mixup," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–4.
- [14] Ronneberger O., Fischer P., and Brox T., "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.
- [15] Isola P., Zhu J. Y., Zhou T., and Efros A. A., "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016.
- [16] Raskov H., Orhan A., Christensen J. P., and Gögenur I., "Cytotoxic cd8+ t cells in cancer and cancer immunotherapy," *British Journal of Cancer*, vol. 124, no. 2, pp. 359–367, 2021.
- [17] Tarnowski B. I., Spinale F. G., and Nicholson J. H., "Dapi as a useful stain for nuclear quantitation," *Biotechnic & Histochemistry*, vol. 66, no. 6, pp. 296–302, 1991.