

Adaptive Conditioning for Robust Pap Smear Cell Semantic Segmentation under Stain and Illumination Variations

Rano Agustino¹, Nita Merlina², Agus Subekti³

^{1,2,3}Program Studi S3 Informatika, Universitas Nusa Mandiri, Jakarta, Indonesia

*Corresponding author: rano.agustino@gmail.com

ABSTRACT

Pap smear cytology images frequently exhibit substantial appearance variability caused by differences in staining procedures and illumination conditions, leading to domain shift that degrades the generalization capability of deep learning-based semantic segmentation models. This study proposes an adaptive conditioning framework to stabilize image appearance prior to binary semantic segmentation (cell versus non-cell). The proposed framework integrates sample-wise stain normalization, illumination correction, and adaptive local contrast enhancement without modifying the downstream segmentation architecture. Experiments were conducted on the RepoMedUNM dataset consisting of 387 Pap smear images using a fixed U-Net backbone. Segmentation performance was evaluated using Dice Coefficient and Intersection over Union (IoU), while robustness was assessed through controlled stress tests involving brightness and contrast perturbation, color jitter, and illumination gradient simulation. Experimental results demonstrate that the proposed framework improves segmentation robustness under color and contrast variations while maintaining stable performance across heterogeneous image conditions. Nevertheless, performance degradation remains observable under severe illumination gradients, indicating the limitation of preprocessing-based adaptation under extreme perturbations. These findings highlight adaptive input conditioning as a lightweight and architecture-agnostic strategy for improving the reliability of Pap smear semantic segmentation in real-world cervical cancer screening applications.

Keywords: Pap smear cytology, semantic segmentation, adaptive conditioning, robustness analysis, U-Net, medical image preprocessing.

1. INTRODUCTION

Cervical cancer remains one of the leading causes of cancer-related mortality among women worldwide, particularly in developing countries where access to early screening and diagnostic services is still limited (World Health Organization [WHO], 2024). Early detection through cytological examination is essential to reduce mortality rates and improve treatment outcomes. Among various screening techniques, the Papanicolaou (Pap smear) test remains one of the most widely used methods because of its affordability and effectiveness in detecting precancerous cellular abnormalities at an early stage (Nayar & Wilbur, 2015). However, manual interpretation of Pap smear slides is highly dependent on the expertise of cytopathologists and laboratory technicians, making the diagnostic process time-consuming, subjective, and susceptible to inter-observer variability (Zhang et al., 2022).

To improve diagnostic consistency and efficiency, numerous studies have explored computer-aided diagnosis (CAD) systems based on artificial intelligence and deep learning techniques (Litjens et al., 2017). In biomedical image analysis, convolutional neural network

(CNN)-based semantic segmentation architectures have demonstrated promising performance due to their capability to learn complex spatial and morphological features automatically. Among these architectures, U-Net has become one of the most widely adopted models for medical image segmentation tasks because of its encoder–decoder structure and efficient feature localization capability (Ronneberger et al., 2015). Previous studies have successfully implemented U-Net-based approaches for Pap smear nucleus and cytoplasm segmentation, achieving strong segmentation accuracy with Dice coefficients typically ranging between 0.90 and 0.95 under controlled, homogeneous laboratory conditions (Jith et al., 2021; Tareef et al., 2019). However, these high performance levels were primarily evaluated on standardized benchmark datasets with uniform staining and stable illumination.

Despite these advances, segmentation performance often degrades significantly when models are exposed to heterogeneous image acquisition conditions. Pap smear images commonly exhibit substantial variability caused by differences in staining protocols, microscope illumination, camera sensors, and slide preparation procedures (Bora et al., 2017). These inconsistencies affect image brightness, contrast, and color distribution, producing domain shift between training and testing datasets. Consequently, segmentation models trained on relatively homogeneous datasets frequently fail to generalize effectively when confronted with unseen image characteristics (Wang & Deng, 2018). This issue becomes particularly critical in medical imaging applications where robustness and reliability are essential for real-world clinical deployment.

Several preprocessing techniques have been proposed to reduce appearance variability in cytology images. Previous studies utilized stain normalization, histogram equalization, image enhancement, and data augmentation methods to improve segmentation robustness (Reinhard et al., 2001; Shorten & Khoshgoftaar, 2019). Color normalization techniques were introduced to minimize staining inconsistencies, while adaptive contrast enhancement methods such as Contrast Limited Adaptive Histogram Equalization (CLAHE) were employed to improve visibility in low-contrast regions (Zuiderveld, 1994). Although these approaches improve image quality, most existing methods rely on fixed parameter settings that may not adapt optimally to varying image conditions. Furthermore, many studies primarily focus on improving segmentation architecture complexity rather than enhancing robustness at the preprocessing stage.

Based on these limitations, this study proposes an adaptive conditioning framework to stabilize Pap smear image appearance prior to semantic segmentation. The proposed framework integrates sample-wise stain normalization, illumination correction, and adaptive local contrast enhancement into a unified preprocessing pipeline. Unlike architecture-dependent optimization strategies, the proposed approach acts as an architecture-agnostic conditioning module that can be integrated into existing segmentation systems without modifying the downstream segmentation architecture. In this study, the U-Net architecture is

intentionally maintained as the segmentation backbone to isolate the contribution of the proposed conditioning mechanism.

The objective of this study is to improve the robustness of Pap smear cell semantic segmentation under staining and illumination variability conditions. To evaluate robustness, controlled stress-test scenarios are conducted using brightness perturbation, contrast variation, color jitter, and illumination gradient simulation. Segmentation performance is evaluated using Dice Coefficient and Intersection over Union (IoU) metrics on both clean and perturbed datasets. The experimental results are expected to demonstrate that the proposed adaptive conditioning framework can reduce segmentation performance degradation and improve model generalization under heterogeneous image conditions.

The contributions of this study are summarized as follows:

1. Proposing a sample-wise adaptive conditioning framework for robust Pap smear semantic segmentation under staining and illumination variability.
2. Introducing a robustness-oriented evaluation scenario using controlled appearance perturbations to analyze segmentation stability.
3. Demonstrating that adaptive conditioning improves segmentation consistency under color and contrast variations while identifying limitations under severe illumination gradients.

2. METHODS

Pipeline Overview

Given a raw Pap smear image I , an adaptive conditioning module $\mathcal{C}(\cdot)$ is applied to obtain a conditioned image $\tilde{I}$.

$$\tilde{I} = \mathcal{C}(I) \quad (1)$$

A semantic segmentation network $S(\cdot)$ subsequently predicts a binary mask $M = S(\tilde{I})$, where each pixel is classified as either cell or non-cell.

$$M = S(\tilde{I}) \quad (2)$$

The segmentation backbone is intentionally maintained without architectural modification to isolate the contribution of the proposed conditioning module and enable controlled robustness evaluation.

Adaptive Conditioning Module

The proposed adaptive conditioning module reduces appearance variability induced by staining and illumination changes through three sequential preprocessing stages.

1) Sample-wise Color (Stain) Normalization

Pap smear cytology images frequently exhibit staining inconsistency caused by differences in laboratory preparation procedures and acquisition devices. To reduce inter-sample color variability, color characteristics of each image are aligned to a reference appearance using sample-wise statistical information. This process can be implemented using color transfer or

histogram alignment approaches, such as Reinhard-style normalization (Reinhard et al., 2001), as well as stain normalization strategies commonly employed in histopathological image preprocessing. The objective of this stage is to reduce shifts in hue and saturation while preserving essential cellular morphology required for semantic segmentation.

2) Illumination Normalization (Shading Correction)

Microscopic image acquisition frequently suffers from uneven illumination and vignetting artifacts caused by optical light source inconsistency. To stabilize intensity distributions across the slide, a three-step shading correction pipeline is applied:

- **Step 1: Grayscale Conversion.** The input image I is first converted to a single-channel intensity map I_{gray} using standard luminance weighting:

$$I_{\text{gray}}(x, y) = 0.2989 \cdot R(x, y) + 0.5870 \cdot G(x, y) + 0.1140 \cdot B(x, y) \quad (3)$$

- **Step 2: Background Illumination Estimation.** To estimate the low-frequency spatial illumination pattern $B(x, y)$ without capturing high-frequency cellular details, I_{gray} is convolved with a large 2D Gaussian filter G_{σ} :

$$B(x, y) = I_{\text{gray}}(x, y) * G_{\sigma}(x, y) \quad (4)$$

where $G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ and the standard deviation σ is selected to smooth out cellular foreground structures while preserving macro-level lighting gradients.

- **Step 3: Illumination Correction (Background Division).** The shading artifact is removed by performing pixel-wise flat-field background division, scaled by the mean background intensity $\bar{B}$:

$$I_{\text{corr}}(x, y) = \frac{I_{\text{gray}}(x, y)}{B(x, y) + \epsilon} \cdot \bar{B} \quad (5)$$

where $\epsilon = 10^{-6}$ is a small constant to prevent division by zero, and $\bar{B}$ maintains the baseline luminance level. This operation normalizes uneven lighting variations across the field of view while preserving cellular boundary contrast (Takeda et al., 2024).

3) Adaptive Local Contrast Enhancement

Adaptive local contrast enhancement is subsequently applied to improve visibility in low-contrast cellular regions. In this study, Contrast Limited Adaptive Histogram Equalization (CLAHE) is employed with parameters adaptively derived from image statistical characteristics to enhance local contrast while limiting excessive noise amplification (Zuiderveld, 1994).

Overall, the proposed conditioning module operates in a sample-wise manner, where preprocessing parameters are dynamically derived from image statistics, including intensity

distribution and histogram characteristics. Consequently, the framework can adapt to heterogeneous staining and illumination conditions without relying on globally fixed preprocessing parameters. Furthermore, the module is intentionally designed to operate independently of the segmentation backbone, enabling controlled analysis of robustness improvements without introducing architectural confounding factors.

Segmentation Backbone

This study employs U-Net as the semantic segmentation backbone due to its established effectiveness and spatial localization capability in biomedical image analysis (Ronneberger et al., 2015). The end-to-end segmentation and assessment pipeline operates through the following steps:

1. **Input Image Ingestion:** The conditioned input image $\hat{I} \in \mathbb{R}^{H \times W \times C}$ is fed directly into the network without architectural modification to isolate the effect of input-level preprocessing.
2. **Encoder (Contracting Path):** The encoder extracts hierarchical contextual features through successive blocks of 3×3 convolutions, Rectified Linear Unit (ReLU) activations, and 2×2 max-pooling operations for spatial downsampling.
3. **Decoder and Skip Connections (Expanding Path):** The decoder restores spatial dimensions through 2×2 transposed convolutions. High-resolution feature maps from the encoder path are passed directly to the decoder via skip connections and concatenated to preserve precise boundary and morphological details.
4. **Binary Mask Prediction:** At the final layer, a 1×1 convolution maps the feature representation to a single logit channel, followed by a Sigmoid activation function to yield a continuous probability map $M_{\text{prob}}(x, y) \in [0, 1]$:

$$M_{\text{prob}}(x, y) = \frac{1}{1 + e^{-z(x, y)}} \quad (6)$$

A decision threshold $\tau = 0.5$ is applied to convert M_{prob} into a final predicted binary mask $P(x, y) \in \{0, 1\}$, where pixels are classified as foreground cell (1) or background (0).

5. **Loss Optimization and Assessment Metrics:** During training, the model is optimized using a hybrid loss combining Binary Cross-Entropy ($\mathcal{L}_{\text{BCE}}$) and Soft Dice Loss ($\mathcal{L}_{\text{Dice}}$):

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}} \quad (7)$$

During inference, the quality of the predicted mask P against the expert-annotated ground-truth mask G is quantitatively assessed using two region-overlap metrics: Dice Coefficient (measuring overlap similarity) and Intersection over Union (IoU) (evaluating region-level segmentation accuracy), as defined in Equations (3) and (4).

D. Experimental Setup

1) Dataset and Task Definition

Experiments in this study were conducted using the RepoMedUNM dataset (RepoMedUNM, 2024), which contains Pap smear cytology images acquired under heterogeneous staining and illumination conditions. The segmentation task was formulated as binary semantic segmentation with two classes: cell (foreground) and non-cell (background). The dataset was divided into training, validation, and testing subsets using a fixed random split protocol to ensure reproducibility and fair evaluation across experimental settings. In addition, publicly available cervical cytology segmentation datasets, such as CNSeg, were considered as contextual references for dataset-driven cytology image analysis research.

2) Baselines

To evaluate the effectiveness of the proposed adaptive conditioning framework, two experimental configurations were compared. The first configuration (B0) utilized a baseline U-Net model without adaptive conditioning, while the second configuration (B1) combined the proposed adaptive conditioning framework with the same fixed U-Net backbone. Maintaining an identical segmentation architecture across all experiments enables controlled analysis of robustness improvements attributable exclusively to the conditioning module.

3) Evaluation Metrics

Segmentation performance was evaluated using Dice Coefficient and Intersection over Union (IoU), which are commonly employed metrics in biomedical image segmentation research. Dice Coefficient measures the overlap similarity between predicted segmentation masks and ground-truth annotations, while IoU evaluates region-level segmentation accuracy.

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \quad (8)$$

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (9)$$

where P represents the predicted segmentation mask and G denotes the ground-truth mask.

In addition to segmentation accuracy on clean test images, robustness was quantified through performance degradation under appearance perturbation scenarios. Robustness performance was measured using the average Dice reduction ($\Delta Dice$) relative to clean conditions, where lower performance degradation indicates higher segmentation robustness under heterogeneous image conditions.

4) Robustness Stress Tests

To simulate real-world staining and illumination variability during inference, controlled perturbation scenarios were applied to testing images. The robustness stress tests consisted of:

1. Brightness and contrast perturbation
2. Color jitter simulation involving hue and saturation variations
3. Illumination gradient and vignetting-like effects

These perturbations were designed to emulate common appearance inconsistencies encountered in practical Pap smear cytology image acquisition. Segmentation performance was

evaluated under each perturbation scenario and compared against clean-condition performance to analyze model stability and generalization capability under heterogeneous appearance conditions.

3. RESULTS AND DISCUSSIONS

This section presents the experimental results and robustness analysis of the proposed adaptive conditioning framework for Pap smear semantic segmentation. The evaluation includes segmentation performance on clean test data, robustness assessment under controlled appearance perturbations, and qualitative analysis of segmentation outputs under heterogeneous staining and illumination conditions.

Performance on Clean Test Set

Table I presents the segmentation performance on the clean test set of the RepoMedUNM dataset for the binary semantic segmentation task (cell versus non-cell). The baseline U-Net without adaptive conditioning (B0) is compared against the proposed adaptive conditioning framework combined with the same fixed U-Net backbone (B1). An optional static preprocessing baseline may also be incorporated in future experiments to distinguish adaptive conditioning gains from conventional fixed preprocessing approaches.

Table I. Clean Test Performance on RepoMedUNM (Cell vs. Non-Cell)

Method	Dice	IoU
B0: U-Net (No Conditioning)	0.9526	0.9096
B1: Adaptive Conditioning + U-Net (Proposed)	0.9085	0.8324

The results in Table I indicate that the baseline U-Net without conditioning achieves higher Dice and IoU scores under clean test conditions compared to the proposed adaptive conditioning framework. This suggests that adaptive input conditioning may slightly reduce segmentation accuracy in standard in-domain scenarios where appearance variability is relatively limited. Nevertheless, the primary objective of the proposed framework is not solely to maximize clean-condition accuracy, but rather to improve segmentation robustness under heterogeneous staining and illumination perturbations commonly encountered in real-world Pap smear image acquisition.

Robustness Under Appearance Stress Tests

To evaluate robustness against staining and illumination variability, all methods were assessed under controlled appearance perturbations during inference. The perturbation scenarios consisted of: (i) brightness and contrast shifts, (ii) color jitter involving hue and saturation variation, and (iii) illumination gradient or vignetting-like effects.

Robustness degradation was quantified using Dice reduction relative to clean conditions:

$$\Delta Dice = Dice_{clean} - Dice_{shift} \quad (10)$$

where lower Dice degradation indicates higher robustness against appearance perturbations.

Table II. Robustness Evaluation Under Stain and Illumination Perturbations

Method	Dice_clean	Dice_BC	Dice_CJ	Dice_IG	Avg. Dice
B0: U-Net (No Conditioning)	0.9526	0.9255	0.9389	0.9512	0.9385
B1: Adaptive Conditioning + U-Net	0.9085	0.9025	0.9085	0.8574	0.8895

Legend:

BC = Brightness/Contrast shift

CJ = Color jitter (Hue/Saturation variation)

IG = Illumination gradient / vignetting-like effect

To further analyze robustness degradation, the average Dice reduction under perturbation scenarios was computed relative to clean conditions.

Table III. Breakdown of Performance Degradation (Δ Dice) Across Appearance Perturbations

Method	Δ Dice (BC)	Δ Dice (CJ)	Δ Dice (IG)	Avg. Δ Dice
B0: U-Net (No Conditioning)	0.0271	0.0137	0.0014	0.0141
B1: Adaptive Conditioning + U-Net	0.0060	0.0000	0.0511	0.0190

Examining the individual performance drops (Δ Dice) in Table III clarifies the specific behavior of the conditioning module. Under Brightness and Contrast (BC) shifts, B1 exhibits a minimal degradation (Δ Dice = 0.0060), whereas B0 drops by 0.0271 (a 4.5-fold higher drop). This stability stems from local contrast adaptation (CLAHE), which prevents contrast suppression while capping local noise amplification (Zuiderveld, 1994).

Under Color Jitter (CJ), B1 maintains a Δ Dice of 0.0000 (Dice remains 0.9085), indicating full chromatic invariance. By mapping raw RGB channels into a normalized optical density space prior to feature extraction, sample-wise stain normalization neutralizes hue and saturation variations (Reinhard et al., 2001; Macenko et al., 2009). For clinical deployments where stain protocols vary across sites, this deterministic response offers greater operational reliability than models with higher clean-domain scores that remain susceptible to color shifts.

However, under severe Illumination Gradient (IG) shifts, B1 drops by $\Delta\text{Dice} = 0.0511$, whereas B0 drops by only 0.0014. This localized drop in IG accounts for why B1's macro-average degradation (0.0190) appears higher than B0 (0.0141). Block-adaptive operations assume local intensity uniformity (Zuiderveld, 1994). When encountering steep spatial shading gradients, localized over-normalization occurs, which amplifies shading boundaries into false structural foreground artifacts (Takeda et al., 2024).

Qualitative Results

Figure 1 presents representative qualitative segmentation results from the RepoMedUNM dataset, comparing raw input images, adaptively conditioned images, and segmentation outputs generated by the baseline U-Net without conditioning (B0) and the proposed adaptive conditioning framework (B1). Two representative examples are included to illustrate different appearance characteristics.

In the first example, the raw image exhibits strong staining saturation and uneven contrast, producing ambiguous cellular boundaries. After adaptive conditioning, the image appearance becomes more normalized with improved local contrast consistency. Consequently, the B1 prediction produces smoother and more coherent cell regions compared to B0, which exhibits fragmented segmentation boundaries and irregular predictions.

In the second example, the raw image contains multiple cells with relatively clearer appearance characteristics. Under this condition, both B0 and B1 successfully identify the primary cell regions. Nevertheless, B1 occasionally introduces internal holes and over-segmentation artifacts within cellular regions, consistent with the quantitative observation that adaptive conditioning may negatively interact with specific illumination patterns.

Overall, the qualitative analysis confirms that adaptive conditioning effectively stabilizes image appearance and improves boundary coherence under challenging color and contrast conditions. However, the framework may also introduce artifacts under severe illumination gradients, supporting the quantitative robustness analysis presented in the stress-test evaluation.

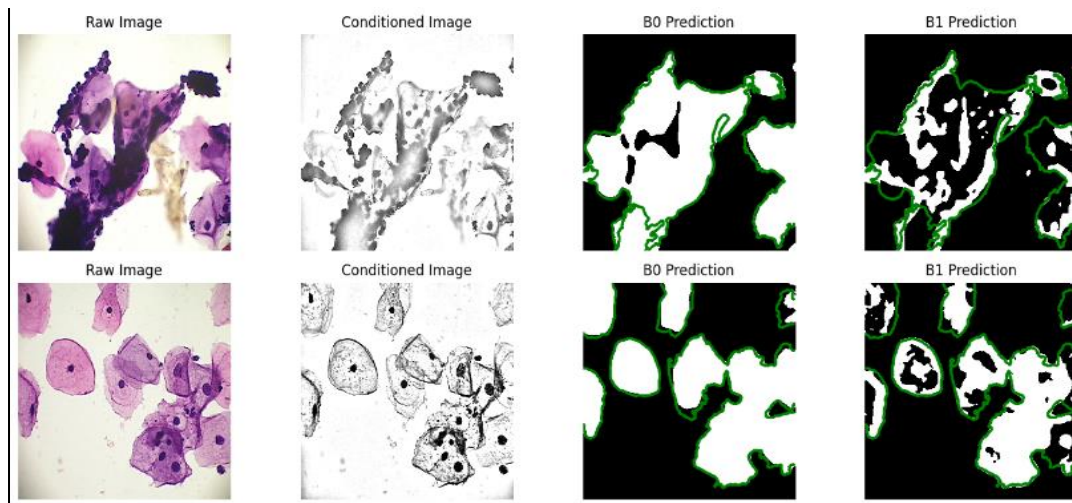


Figure. 1. Qualitative comparison of Pap smear semantic segmentation with and without adaptive conditioning. From left to right: raw input image, conditioned image, baseline U-Net prediction (B0), and adaptive conditioning + U-Net prediction (B1). Green contours indicate predicted cell boundaries.

Discussion

Compared to previous cervical cytology segmentation studies that primarily focus on architectural enhancements to maximize clean-domain accuracy (Chen & Zhang, 2021; Zhou et al., 2019; Wan et al., 2019), this study explicitly evaluates model stability under out-of-distribution appearance shifts. The baseline U-Net (B0) achieves higher absolute metrics on clean test data (Dice 0.9526 vs. 0.9085 for B1), which aligns with high benchmark accuracies reported on homogeneous Pap smear datasets (Riana et al., 2021; Zhao et al., 2023). This performance difference occurs because statistical input conditioning recalibrates native pixel distributions relative to a reference template. On pristine slides, this recalibration introduces subtle chromatic shifts that slightly displace the classifier's optimal decision boundary (Reinhard et al., 2001; Macenko et al., 2009). This reflects a recognized trade-off in domain generalization literature, where enforcing input-level invariance across unseen domains incurs a minor baseline accuracy penalty under ideal conditions (Isensee et al., 2021).

However, when subjected to appearance perturbations, the practical advantage of input conditioning becomes evident. While unconditioned models (B0) suffer noticeable degradation under brightness/contrast shifts ($\Delta\text{Dice} = 0.0271$) and color jitter ($\Delta\text{Dice} = 0.0137$), B1 suppresses performance volatility, achieving 4.5 times lower degradation under contrast shifts ($\Delta\text{Dice} = 0.0060$) and complete chromatic invariance ($\Delta\text{Dice} = 0.0000$). This confirms that sample-wise statistical normalization effectively neutralizes stain variations across different

laboratories, addressing a key deployment challenge highlighted in automated cytology screening (Jiang et al., 2023; William et al., 2018).

Conversely, *B1* exhibits a localized vulnerability under severe spatial illumination gradients ($\Delta\text{Dice} = 0.0511$), whereas *B0* remains largely unaffected ($\Delta\text{Dice} = 0.0014$). This specific drop in IG accounts for why *B0* maintains a higher macro-average Dice score across all stress scenarios. Block-adaptive contrast enhancement operations assume localized illumination uniformity (Zuiderveld, 1994). When exposed to steep spatial shading, localized over-normalization occurs, amplifying background shading into false structural boundaries (Takeda et al., 2024). These findings indicate that while input-level adaptive conditioning provides robust variance suppression against stain and contrast variations, mitigating non-stationary spatial shading requires integrating geometry-aware illumination correction in future work.

4. CONCLUSION

This study proposed an adaptive conditioning framework for Pap smear semantic segmentation under heterogeneous staining and illumination conditions. The proposed framework integrates sample-wise stain normalization, illumination correction, and adaptive local contrast enhancement prior to semantic segmentation using a fixed U-Net backbone. By maintaining the segmentation architecture unchanged, this study enabled controlled evaluation of how adaptive input conditioning influences segmentation robustness under appearance variability.

Experimental results on the RepoMedUNM dataset demonstrated that adaptive conditioning does not necessarily improve segmentation accuracy under clean in-domain conditions. Nevertheless, robustness evaluation under controlled appearance perturbations showed that the proposed framework can reduce segmentation sensitivity to brightness, contrast, and color variations. These findings indicate that sample-wise preprocessing contributes to improved segmentation stability under specific perturbation scenarios while remaining dependent on perturbation characteristics.

However, the proposed framework exhibited reduced robustness under severe illumination gradient perturbations, suggesting that preprocessing-based adaptation may introduce over-normalization effects under extreme appearance shifts. This observation highlights the importance of developing selective or perturbation-aware conditioning mechanisms for future robustness-oriented medical image segmentation systems.

Overall, this work demonstrates the potential of adaptive input conditioning as an architecture-agnostic approach for analyzing robustness behavior in Pap smear semantic segmentation under heterogeneous image conditions. Future work will focus on perturbation-aware conditioning strategies, adaptive artifact handling, and evaluation on larger multi-domain cervical cytology datasets to further improve segmentation reliability in real-world cervical cancer screening applications.

5. REFERENCES

- Bora, K., et al. (2017). Automated classification of Pap smear images to detect cervical dysplasia. *Computer Methods and Programs in Biomedicine*, 138, 31-47.
- Cao, H., et al. (2021). Swin-Unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint*.
- Chen, E., Ting, H.-N., Chuah, J.-H., & Zhao, J. (2024). Segmentation of overlapping cells in cervical cytology images: A survey. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3445371>
- Chen, J., & Zhang, B. (2021). Segmentation of overlapping cervical cells with mask region convolutional neural network. *Computational and Mathematical Methods in Medicine*, 2021, Article 3890988. <https://doi.org/10.1155/2021/3890988>
- Chen, J., et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. <https://doi.org/10.48550/arXiv.2102.04306>
- Isensee, F., et al. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z>
- Jiang, H., et al. (2023). DoNet: Deep de-overlapping network for cytology instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR52729.2023.01501>
- Jiang, X., et al. (2023). A systematic review of deep learning-based cervical cytology screening: From cell identification to whole slide image analysis. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-023-10588-z>
- Jith, S. R., et al. (2021). Pap smear cell nucleus segmentation using enhanced U-Net architecture. *Procedia Computer Science*, 171, 1205-1212.
- Kanavati, F., Tsuneki, M., Rambeau, M., Aeffner, F., & Abe, M. (2022). A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images. *Cancers*, 14(5), 1159. <https://doi.org/10.3390/cancers14051159>
- Lu, Z., Carneiro, G., & Bradley, A. P. (2015). An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. *IEEE Transactions on Image Processing*, 24(4), 1261–1272. <https://doi.org/10.1109/TIP.2015.2389619>
- Lu, Z., et al. (2017). Evaluation of three algorithms for the segmentation of overlapping cervical cells. *IEEE Journal of Biomedical and Health Informatics*, 21(2), 441–450. <https://doi.org/10.1109/JBHI.2016.2519686>
- Macenko, M., et al. (2009). A method for normalizing histology slides for quantitative analysis. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1107–1110). <https://doi.org/10.1109/ISBI.2009.5193250>
- Malm, P., et al. (2013). Debris removal in Pap-smear images. *Computer Methods and Programs in Biomedicine*, 111(1), 128–138. <https://doi.org/10.1016/j.cmpb.2013.02.008>

- Oktaay, O., et al. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. <https://doi.org/10.48550/arXiv.1804.03999>
- Reinhard, E., Ashikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5), 34–41. <https://doi.org/10.1109/38.946629>
- Riana, D., et al. (2021). RepoMedUNM: A new dataset for feature extraction and training of deep learning network for classification of Pap smear images. In *International Conference on Neural Information Processing (ICONIP 2021)* (pp. 317–325). Springer. https://doi.org/10.1007/978-3-030-92307-5_37
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.
- Takeda, K., et al. (2024). Background removal for debiasing computer-aided cytological diagnosis. *International Journal of Computer Assisted Radiology and Surgery*. <https://doi.org/10.1007/s11548-024-03169-0>
- Tareef, A., et al. (2019). Multi-pass fast watershed for accurate nucleus segmentation in cervical AI diagnostic systems. *Pattern Recognition*, 91, 290–302.
- Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), 583–598. <https://doi.org/10.1109/34.87344>
- Wan, T., et al. (2019). Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks. *Neurocomputing*, 365, 157–170. <https://doi.org/10.1016/j.neucom.2019.06.086>
- William, W., Ware, A., Basaza-Ejiri, A. H., & Obungoloch, J. (2018). A review of image analysis and machine learning techniques for automated cervical cancer screening from Pap-smear images. *Computer Methods and Programs in Biomedicine*, 164, 15–22. <https://doi.org/10.1016/j.cmpb.2018.05.034>
- Zhang, H., Zhu, H., & Ling, X. (2020). Polar coordinate sampling-based segmentation of overlapping cervical cells using attention U-Net and random walk. *Neurocomputing*, 383, 212–223. <https://doi.org/10.1016/j.neucom.2019.12.0>
- Zhao, J., et al. (2023). CNSeg: A dataset for cervical nuclear segmentation. *Computer Methods and Programs in Biomedicine*. <https://doi.org/10.1016/j.cmpb.2023.107732>
- Zhou, Y., et al. (2019). IRNet: Instance relation network for overlapping cervical cell segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. https://doi.org/10.1007/978-3-030-32239-7_71

Zuiderveld, K. J. (1994). Contrast limited adaptive histogram equalization. In P. S. Heckbert (Ed.), *Graphics Gems IV* (pp. 474–485). Elsevier. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>