Development of Preprocessing Stage for Early Cervical Cancer Detection Using UNET Diffusion Model

Parimala Tamang[1], Annet Thatal[2], Mousumi Gupta1[1,*], and Snehashish Bhattacharjee[3]

[1]Department of Computer Applications, Sikkim Manipal Inst of Tech, Sikkim Manipal University, Sikkim, India
[2]Department of Obstetrics and Gynecology, Gautam Buddha Chikitsa Mahavidyalaya, Dehradun, India
[3]Department of Computer Science and Engineering, University of Engineering and Management Kolkata, Newtown, India
Email: parimala.t@smit.smu.edu.in (P.T.); annetth1@gmail.com (A.T.); mousmigt@gmail.com (M.G.); snehashishbhattacharjee@gmail.com (S.B.)
*Corresponding author

Abstract—**Colposcopy, a medical procedure for diagnosing and treating cervical cancer, generates medical images that often contain Specular Reflections (SRs). These bright areas, caused by moisture and device lighting, can affect further image analysis steps like feature extraction and classification. In this work, a two-stage U-shaped convolutional neural network architecture (UNET) diffusion model is employed for SR detection and inpainting in colposcopy images. The first stage creates SR region masks using local thresholding and top-hat morphological operations on augmented and pre-processed images. The second stage employs a UNET diffusion model trained with Dreambooth, and inpaints Specular Reflection (SR) regions from the processed images. The International Agency for Research on Cancer dataset utilized in the overall work comprises 913 colposcopy images to train and evaluate the developed model. The UNET diffusion model achieved reconstruction with a minimal loss of 0.169 and a stable Structural Similarity Index Measure (SSIM) of 0.85. State-of-the-art methods for specular reflection detection and inpainting in colposcopy images evaluated on limited datasets, often lacking diversity in imaging conditions. In contrast, the presented approach is developed to work on colposcopy images taken under varying conditions, including green filter, iodine staining, and acetic acid application, which is required for accurate diagnosis. This enhances the model's robustness, enabling it to perform effectively across a diverse range of colposcopy images.**

Keywords—colposcopy, specular reflection, Unet diffusion

I. INTRODUCTION

According to the report generated by the World Health Organization (WHO) in 2022, cervical cancer is among the fourth most common cancer in women worldwide with around 6,60,000 new cases and 350,000 deaths [1]. Several gold standards exist for cervical cancer screening such as the Papanicolaou (Pap) Test, the Human Papillomavirus (HPV) Test, and colposcopic evaluation. Low-income and middle-income countries often lack access to cervical cancer screening methods such as pap smear tests and HPV tests due to complex and poor healthcare infrastructure, leading to increased incidence and mortality rates [2].

Various modalities exist in the healthcare industry to assist patients with different medical conditions or injuries. One such modality is a colposcopy device. A colposcopy is a medical procedure that evaluates cervical neoplasia after abnormal Pap test results and is considered a dependable and accurate method for detecting and treating pre-malignant cervical lesions [3]. This medical procedure uses a device, a colposcope, to inspect cervix tissue and detect potential cervical neoplasia. The device's bright camera light can create areas of intense brightness known as Specular Reflection (SR) regions, which occur due to the presence of mucus on the cervix surface [4].

During a colposcopic examination, a 3% to 5% acetic acid solution is applied to the cervix, turning normal tissue white and creating acetowhite epithelium. Clinicians assess the colour density of acetowhite epithelium to evaluate the lesion severity. Specular reflections can interfere with diagnosis and reduce the effectiveness of further image-processing steps [5].

The recent advancements in the healthcare industry, driven by advanced technologies based on artificial intelligence, machine learning, and deep learning approaches, have transformed the healthcare industry, enhancing diagnosis, medical image analysis, and overall medical infrastructure.

These innovations have significantly improved medical image analysis with accurate results, quick diagnosis, and the development of new and advanced image analysis approaches [6].

Among these advancements, a diffusion model has achieved notable success in various generative tasks, including medical image segmentation [7–9]. Furthermore, the fusion of these two powerful techniques, diffusion

Manuscript received November 27, 2024; revised December 25, 2024; accepted March 25, 2025; published May 20, 2025.

model and UNET architecture, has proven to be highly adaptable with excellent performance in medical image analysis [10–11]. The integration of these methods shows a good potential for improving the accuracy and efficiency of medical image analysis [12–13].

Over the years, various approaches have been introduced that aim to detect the specular reflection regions and inpaint the regions in colposcopy images, including Adaptive patching [14], Exemplar [15], Arnold [16], Dichromatic Reflection Model [17], and Thresholding with morphological operations [18]. These methods were applied to colposcopy images taken during the colposcopy procedure. In contrast, our presented method performs the detection and inpainting of SR regions in colposcopy images taken under various conditions: after normal saline application, after acetic acid, after acetic acid with a green filter, and after Lugol's iodine application. Our approach outperforms existing methods as it effectively handles colposcopy images taken under different conditions, with high-quality results accepted by medical experts. Furthermore, unlike existing methods, our presented model works effectively in diverse colposcopy datasets. To provide comprehensive insights into its performance, we evaluated the SSIM and loss values, ensuring a thorough assessment of its performance.

We utilised a U-shaped convolutional neural network architecture (UNET) diffusion model for inpainting Specular Reflection (SR) regions in colposcopy cervical cancer medical images, aiming to enhance the quality of these medical images, enabling more accurate and efficient post-processing.

In the following sections, we have discussed the existing techniques for detecting SR regions and present our proposed UNET diffusion model for inpainting SR regions in colposcopy medical images.

II. Related Works

Several studies exist that focus on the detection and inpainting of specular reflections in colposcopy images. Yue *et al.* [14] employed a method to detect and in-paint specular reflections in cervical images. The process starts by transforming the images to HSI colour space to identify reflections using low saturation and high intensity. Thresholding and gradient analysis are also employed to outline these reflections. For inpainting, patches from non-specular areas replace the reflections, with adaptive patch sizes and a similarity function ensuring the final image is clear and accurate. Gorantla *et al.* [15] employed Gaussian smoothing, intensity thresholding, and edge detection to identify bright regions as Specular Reflections (SRs). Morphological operations were then applied, followed by patch-based methods for image inpainting. Detection of specular reflections using pixel intensity, colour differences, and gradient edges is proposed by integration. To fill the detected specular areas, algorithms like Telea or Navier-Stokes are employed, which use texture information from the images. Wang *et al.* [16] detects and in-paint Specular Reflections (SR) from colposcopic images using a three-step approach: image smoothing via Arnold's method, SR detection through combined global and local thresholding, and exemplar-based inpainting for SR region reconstruction while preserving texture and structural details in colposcopic images for diagnosis. Oak *et al.* [17] compares three SR detection methods: the Dichromatic Reflection Model, which separates SR pixels but struggles with similar hues; RGB Plane Separation, which identifies SR pixels by separating colour planes but has difficulty with dimmer pixels; and the Morphological Top-Hat Transform, which uses a structuring element and Otsu's method but is sensitive to variations in the structuring element. Grajales *et al.* [18] employed a technique to detect SR regions by converting images to grayscale, applying Gaussian filtering, calculating gradients with adaptive thresholding, refining with morphological operations, and creating a binary mask, while inpainting uses patch-based or exemplar-based techniques to fill reflections with matching pixels or maintain texture and colour continuity. Author Susan *et al.* [19] employed a UNET architecture to generate a mask for the detection of SR regions and subsequently performing inpainting using neighbourhood pixel information. Martin *et al.* [20] proposed a neural network-based strategy to detect and inpaint specular reflections in colposcopic images. The method involves supervised training of the neural network to learn how to restore hidden regions under specular reflections, despite the absence of ground truth data for these regions.

State-of-the-art methods for Specular Reflection (SR) detection and inpainting in colposcopy images are primarily developed using limited datasets, often lacking variations in the dataset such as green filters, iodine staining, and acetic acid application. Most of these approaches depend on traditional image processing techniques, including thresholding, gradient analysis, and morphological operations, which may not provide effective results when applied to diverse colposcopy datasets, thus limiting overall performance. Furthermore, some studies have employed patch-based or exemplar-based inpainting techniques, often struggling to preserve fine texture and structural details, which are vital for accurate clinical diagnosis. Additionally, most studies focused on traditional feature extraction rather than utilizing artificial intelligence models designed to handle complex tissue patterns. Although Martin *et al.* [20] proposed a neural network-based approach, the absence of ground truth data for SR regions makes supervised training challenging, therefore limiting its effectiveness in real-world clinical applications.

Our study addresses these limitations by proposing a robust SR detection and inpainting method developed for colposcopy images captured under varying conditions rather than being limited to acetic acid images. These variations in colposcopy images are required for experts to analyse the images accurately for prognosis and diagnosis.

The presented UNET diffusion model is utilized to preserve the texture and structural details of the cervix, surpassing the preservation accuracy of traditional approaches. Additionally, our presented approach improves generalization offering a reliable solution for SR detection and inpainting in colposcopy images to support clinical decision-making. Furthermore, future work could focus on developing a segmentation model to segment regions of interest and non-regions of interest, along with lossless image compression to support remote clinical diagnosis. The details about the dataset set used and the types of approaches of the studies are given in Table I.

TABLE I. A STATE OF ART FRAMEWORK OF EXISTING CONTRIBUTIONS WITH DATASET DETAILS

First author	Methodology	Approach			Sample size	Dataset source
		Supervised	Unsupervised	Digital/ traditional approach		
Yue	Threshold and improved exemplar-based image inpainting method	☒	☒	☑	3045	Hefei University of Technology
Gorantla	Flat field correction, Gaussian smoothing and local Laplacian filter	☒	☒	☑	1821	Kaggle
Wang	Exemplar-based method	☒	☒	☑	150	First Hospital of Quanzhou, Fujian Medical
Oak	Alsaleh and Kittler method	☒	☒	☑	100	National Cancer Institute (NCI)
Grajales	morphological operations	☒	☒	☑	12	Not mentioned
Martin	Supervised neural network	☑	☒	☒	22	Not mentioned

III. MATERIALS AND METHODS

We employed Stable Diffusion v1-5, a UNET diffusion model, and extended its training on our dataset of colposcopy images using Dreambooth for inpainting SR regions, detected using image processing techniques. Our method includes two stages: a) the creation of SR region masks from colposcopy images using local thresholding and white top-hat morphological operations and b) the existing UNET diffusion model additionally trained using Dreambooth [21] on our dataset for inpainting SR regions in colposcopy images.

The first stage includes pre-processing and augmentation of the sample size colposcopy images. This involves rescaling all colposcopy images to 512×512 pixels to match the UNET architecture requirements, then we apply vertical and horizontal flips to generate additional variations of each image. This augmentation helps to increase the dataset size and diversity. From the augmented dataset, masks were created of each image to locate the SR regions using finely tuned local thresholding and white top-hat morphological operations. The local thresholding is applied to the image to detect bright regions indicative of specular reflections.

The detection process involves setting a Threshold (T) for pixel Intensities (I), where pixels with I > T are identified as potential Specular Reflection (SR) regions. Mathematically, this can be expressed as:

$$\boldsymbol{SR_{mask}} = \{(\boldsymbol{x},\boldsymbol{y}) | \boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}) > \boldsymbol{T}\} \quad \text{where,}$$

- $\boldsymbol{SR_{mask}}$: represents the binary mask that identifies the detected Specular Reflection (SR) regions in the image. The mask assigns 1 (white) to pixels belonging to the SR region and 0 (black) to all other pixels.
- $(\boldsymbol{x},\boldsymbol{y})$: This denotes the coordinates of pixels in the image. Every pixel is defined by its position (x, y) in a 2D image space.
- $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y})$: This denotes the intensity value of a pixel at coordinate (x, y). In grayscale images, this intensity ranges from 0 (black) to 255 (white).
- $\boldsymbol{T}$: This is the threshold value set to differentiate SR pixels from non-SR pixels. Pixels with intensities higher than T are classified as part of the specular reflection region.
- $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}) > \boldsymbol{T}$: This condition checks whether a pixel's intensity is greater than the defined Threshold (T). If true, the pixel is considered part of the Specular Reflection (SR) region and included in SR_mask.

To further enhance and detect specular reflections, we apply the white top-hat morphological method, which highlights small bright spots in an image. This method works by subtracting the morphological opening of an image from the original image. Morphological opening is a process where the image undergoes erosion followed by dilation, effectively removing small bright structures. The white top-hat transform is mathematically represented as:

$$\boldsymbol{TopHat}(\boldsymbol{I}) = \{\boldsymbol{I} - \boldsymbol{Opening}(\boldsymbol{I})\} \text{ and } \boldsymbol{Opening}(\boldsymbol{I}) = \boldsymbol{Dilation}(\boldsymbol{Erosion}(\boldsymbol{I})) \text{ where,}$$

- TopHat(I): The top-hat transform highlights bright regions in an image that are smaller than the structuring element used in morphological operations.
- I: represents the original image where we want to detect bright spots (specular reflections).
- Opening(I): This is the morphological opening of the image, which removes small bright structures while preserving the overall background.
- I-Opening(I): Subtracting the opened image from the original image enhances bright, small regions, making specular reflections more prominent.

The second stage comprises training the UNET diffusion model on colposcopy images through Dreambooth with the keywords "Picture of SMIT colposcopy" to trigger the model so that its output would reflect the training data. To create the workflow for inpainting the detected SR regions we used the ComfyUI platform which provides a graphical user interface for creating and managing complex workflows using nodes and flowcharts. We used this platform to design and execute advanced diffusion pipelines. To further train the existing diffusion model, we used hyperparameters given in Table II.

Table II represents the hyperparameters employed in the UNET diffusion model. The model was trained on a dataset of 913 colposcopy images, which was expanded using image augmentation techniques from the original dataset of 1142 images. To avoid the risk of overfitting, a common concern with Dreambooth-based models, we have augmented the dataset as a pre-processing step while expanding the dataset. Furthermore, we limited the training to 800 optimization steps to ensure a balance between model performance and generalization. Additionally, we used a batch size of 1 to improve accuracy with no gradient accumulation. Based on the number of images (168) and optimization steps (800), an internal calculation yielded 5 epochs, this figure is less relevant in our hyperparameters as it is a result of the aforementioned values and can be omitted from the final analysis.

TABLE II. UNET DIFFUSION MODEL HYPERPARAMETER

Hyper-parameter	Value
Num examples	168
Num batch per epoch	168
Num epochs	5
Total training batch size	1
Gradient accumulation steps	1
Total optimization steps	800

For the input to our diffusion pipeline, we used 913 pre-processed colposcopy images along with their masks and encoded them into the latent space using a variable auto-encoder so that the representation of the combination of the masked and pre-processed image can be used by our trained UNET model for inpainting. We also employed an IP adapter [22] to provide additional reference to our trained UNET model for better results.

The entire workflow is illustrated in the following Figs. 1 and 2.

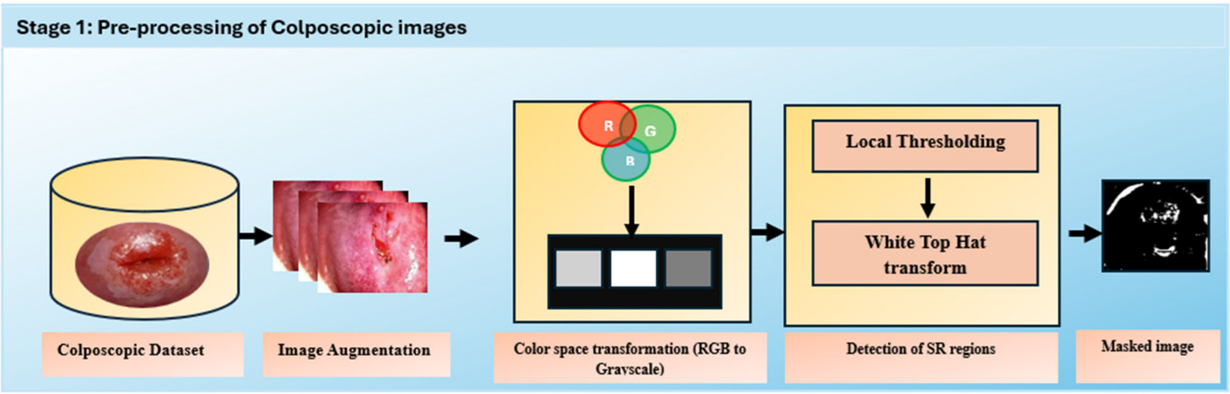


Fig. 1. Detection of SR regions and subsequent creation of a masked image.

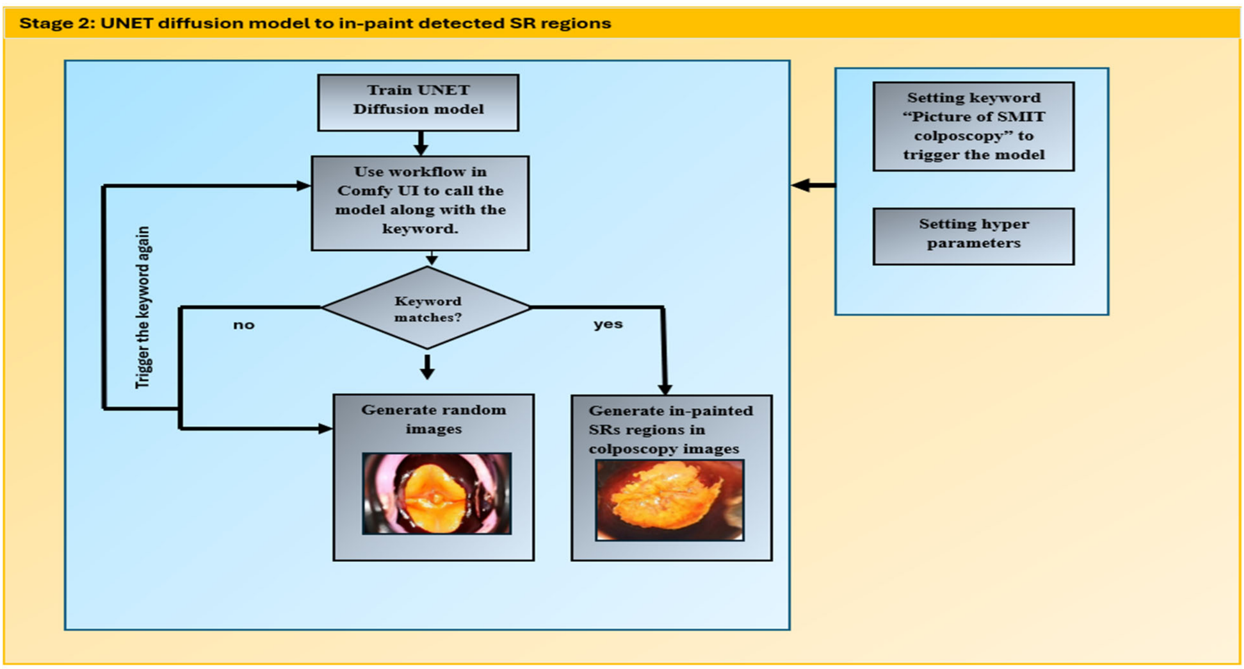


Fig. 2. Workflow of diffusion pipeline to perform inpainting on SR region.

IV. Results and Discussion

A. Experimental Setup

Our experiments were conducted on GPU: NVIDIA Tesla T4 with memory 16GB, CUDA Cores 2560, Tensor core 320 and system RAM 15GB.

We utilized ComfyUI to create diffusion pipelines and triggered the trained UNET diffusion model using the keyword "Picture of SMIT colposcopy," which inpainted the Specular Reflection (SR) regions and produced the output. ComfyUI is a powerful node-based interface for diffusion models that allows for creating complex workflows through a visual programming approach. It enables us to design, modify, and execute custom image processing pipelines by connecting various nodes representing different operations, models, and data transformations. This flexibility was crucial in implementing our two-stage approach for SR detection and inpainting in colposcopy images.

The ComfyUI workflow for our experiment consists of nodes for:

i. Loading and pre-processing the input colposcopy images
ii. Combining and converting the first-stage SR mask with its respective pre-processed image into the latentspace.
iii. Configuring and executing the trained UNET diffusion model for inpainting.
iv. Post-processing and saving the output image

This setup allowed us to iterate on our approach, fine-tune parameters efficiently, and process batches ofimages through our custom inpainting pipeline.

B. Dataset and Performance Metrics

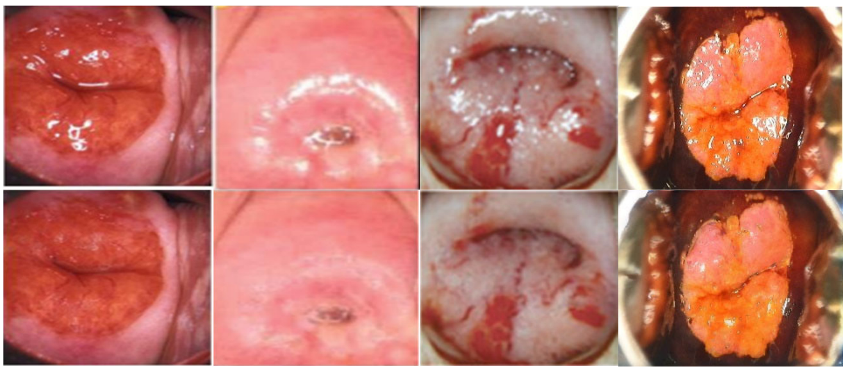
The dataset used in this study, obtained from The International Agency for Research on Cancer, consists of 913 colposcopy images taken under various conditions and image metadata. We augmented the dataset, where the UNET diffusion model performs Specular Reflection (SR) inpainting, achieving a minimal loss of 0.169 and maintaining a stable average Structural Similarity Index Measure (SSIM) of 0.85.

SSIM quantifies image similarity, mathematically expressed as: $SSIM(x,y)= [l(x,y)]\alpha\cdot[c(x,y)]\beta\cdot[s(x,y)]\gamma$ $SSIM(x,y)=[l(x,y)]\alpha\cdot[c(x,y)]\beta\cdot[s(x,y)]\gamma$ [23], Where: $l(x,y)$: Luminance, comparing brightness between images, $c(x,y)$: Contrast, measuring the range between brightest and darkest regions, $s(x,y)$: Structure, evaluating the similarity of local luminance patterns, α, β, γ: Positive constants weighing the relative importance of each component.

SSIM values range from 0 to 1, with 1 indicating perfect similarity. We calculated SSIM between the original and in-painted images. The achieved average SSIM of 0.85 suggests that the in-painted regions closely match the surrounding tissue in terms of brightness, contrast, and texture while leaving room for further improvement. Several methods have been proposed for Specular Reflection (SR) detection and inpainting in medical images, particularly in colposcopy. However, many of these approaches lack standard performance metrics such as the Structural Similarity Index (SSIM), making direct quantitative comparisons difficult. Instead, their effectiveness is evaluated based on qualitative results and dataset size, as reported in Fig. 3.

Fig. 3 presents the input images and the corresponding inpainted outputs obtained by various existing methods for specular reflection detection and inpainting. These include the DRM approach by Meslouhi *et al.* [5], which utilized a dataset of 286 colposcopy images, as well as the method proposed by Wang *et al.* [4], which employed local and global processing on 150 images. Furthermore, Wang *et al.* [16] applied the Exemplar-based method to a dataset of 150 images. In contrast, our presented model is trained and evaluated on a larger and more diverse colposcopy dataset comprising 913 colposcopy images taken under different conditions: after normal saline application, after acetic acid, after acetic acid with a green filter, and after Lugol's iodine application. Unlike other approaches reported in this study, we have evaluated our model's performance using both quantitative and qualitative evaluation. The performance metric such as SSIM (Structural Similarity Index Measure) is used to provide a more objective evaluation of inpainting quality while a subjective evaluation is conducted by the health care experts to visually assess the quality of the restored images.

The original and inpainted outputs from these existing methods are compared to our proposed approach in Fig. 3.



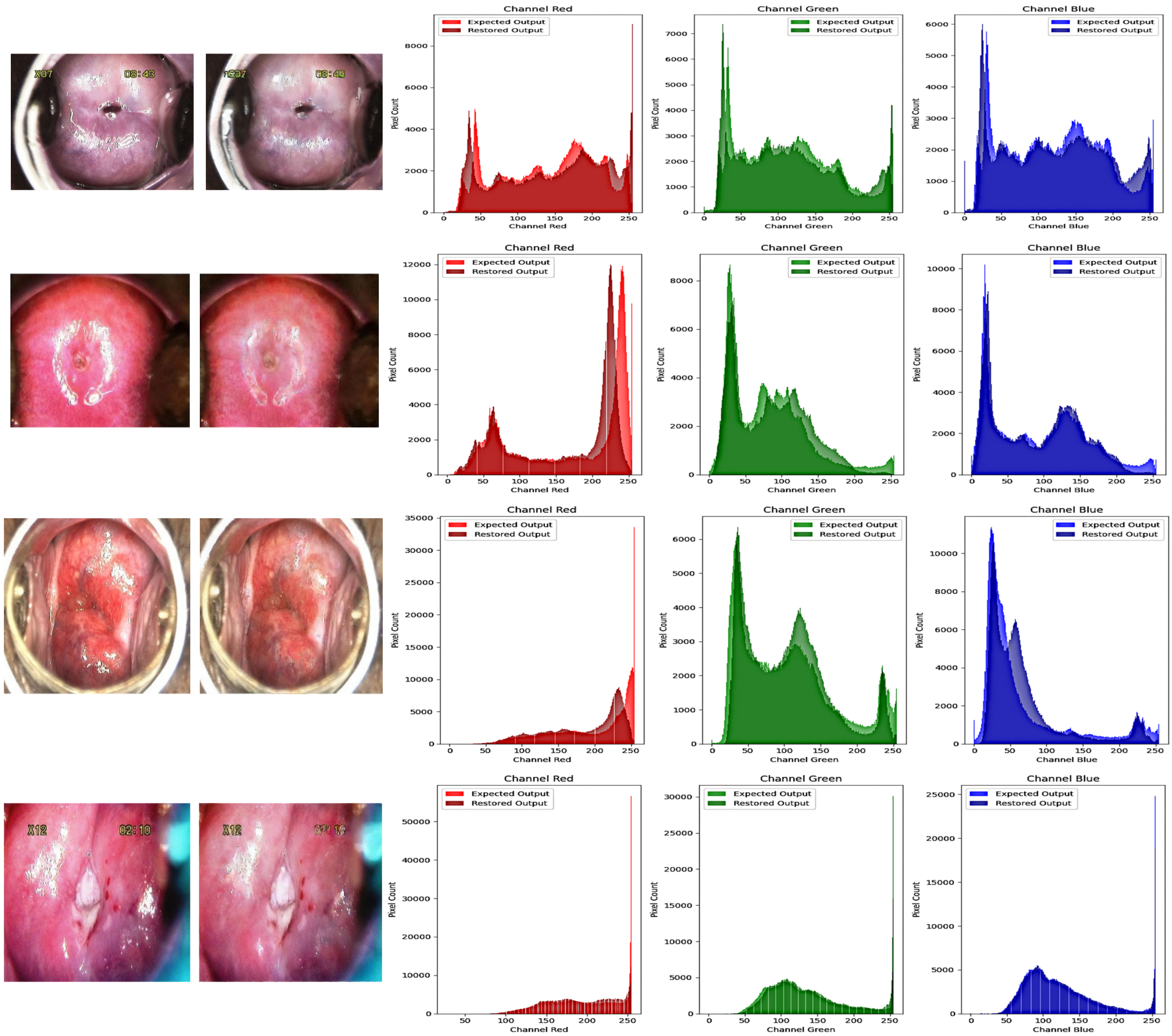
(a) Meslouhi [5] (b) Wang [4] (c) Wang [16] (d) Our presented approach

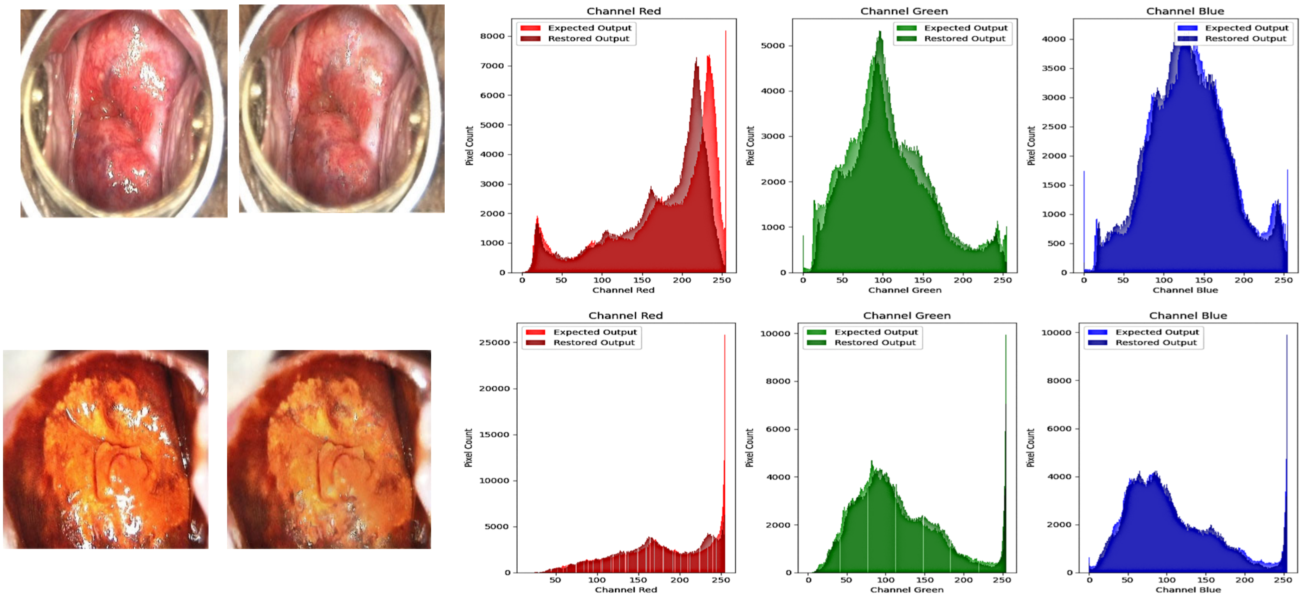
Fig. 3. Colposcopy input images and the corresponding inpainted outputs obtained by various methods.

Furthermore, to evaluate the results of our proposed method, we have compared the RGB values of the original image with those of the inpainted output image, as illustrated in Fig. 4. Fig. 4, section (c) illustrates a histogram comparison of the RGB channels between the original colposcopy image and the inpainted output image. The histograms demonstrate that the inpainted image closely preserves the colour distribution of the original image across all three channels. Each graph represents the Red, Green, and Blue (RGB) channels separately, displaying two distributions: Original Image and Restored Output (Inpainted Image).

In the Red Channel, pixel intensity values range from 0 to 255, with the Original Image represented in light red and the Restored Output (Inpainted Image) in dark red. A high peak in the distribution suggests a strong red intensity across both images. For the Green Channel, the Original Image is shown in light green, while the Restored Output (Inpainted Image) is in dark green. The distribution indicates good preservation of green intensity values, with only minor variations observed. Similarly, in the Blue Channel, the Original Image is displayed in light blue, and the Restored Output (Inpainted Image) is in dark blue. The histograms for both images remain closely aligned, reflecting minimal deviation in pixel intensities.

The slight variations in some intensity regions highlight the model's ability to effectively reconstruct missing areas while maintaining the overall colour integrity. The strong peak structures across all three channels demonstrate the model's effectiveness in preserving the original colposcopy image's colour characteristics, ensuring a seamless inpainting result.





(a) (b) (c)

Fig. 4. Result obtained from the model. (a): Original image; (b) inpainted image (output); (c) the histograms of RGB colour intensities per channel corresponding to the input and expected output image (in red, green, and blue).

V. Conclusion

In this study, the two-stage UNET diffusion model used a customized diffusion pipeline implemented model for the detection and inpainting of Specular Reflections (SR) in colposcopy images. The development is done through ComfyUI, which yields good results in addressing the challenging task of SR removal in medical imaging. The proposed model, trained on a dataset of 168 colposcopy images with user-defined hyperparameters to prevent overfitting, achieved a minimal loss of 0.169 and a stable Structural Similarity Index Measure (SSIM) of 0.85. The results indicate that the proposed method effectively removes specular reflections and preserves the diagnostic features of the surrounding tissue in colposcopy images. Future work will focus on deploying the model as a practical tool for automated colposcopy image enhancement, ensuring its usability in real-world medical settings. Furthermore, the SSIM value can be improved by fine-tuning the model, training it on a diverse and large colposcopy dataset, and capturing various features from colposcopy images before detecting SR regions. The inpainting process can be enhanced by filling the detected regions with previously captured features in the SR-affected regions, ensuring better structural preservation and image quality. This paper contributes to the ongoing research to detect and in-paint SR regions in colposcopy images to enhance the colposcopy medical image quality and reliability of cervical cancer screening and diagnosis.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Author PT conducted the research and model development; Author AT provided clinical expertise and analyzed the model outputs; Authors PT, MG, and SB contributed to writing and revising the manuscript. All authors reviewed and approved the final version of the paper.

References

[1] World Health Organization, *Global Strategy to Accelerate the Elimination of Cervical Cancer as a Public Health Problem*, Google Books, World Health Organization, 2020.

[2] N. Ebrahimi *et al.*, "Human papillomavirus vaccination in low-and middle-income countries: Progression, barriers, and future prospective," *Front Immunol.*, vol. 14, 2023. doi: 10.3389/FIMMU.2023.1150238

[3] K. Nam, "Colposcopy at a turning point," *Obstet. Gynecol. Sci.*, vol. 61, no. 1, 2018. doi: 10.5468/OGS.2018.61.1.1

[4] X. Wang *et al.*, "Integration of global and local features for specular reflection inpainting in colposcopic images," *Journal of Healthcare Engineering*, 5401308, 2021. doi: 10.1155/2021/5401308

[5] O. E. Meslouhi, M. Kardouchi, H. Allali, T. Gadi, and Y. A. Benkaddour, "Automatic detection and inpainting of specular reflections for colposcopic images," *Central European Journal of Computer Science*, vol. 1, no. 3, pp. 341–354, 2011. doi: 10.2478/s13537-011-0020-2

[6] N. R. Shetty, N. H. Prasad, and H. C. Nagaraj, "Advances in communication and applications," *Proceedings of ERCICA*, vol. 2, 2023.

[7] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[8] J. Wu *et al.*, "Medsegdiff: Medical image segmentation with diffusion probabilistic model," *Medical Imaging with Deep Learning*, pp. 1623–1639, 2024.

[9] Z. Xing, L. Wan, H. Fu, G. Yang, and L. Zhu, "Diff-UNet: A diffusion embedded network for volumetric segmentation," arXiv preprint, arXiv:2303.10326, 2023.

[10] M. J. Susan and P. Subashini, "Detection of specular reflection on the smart colposcopy images using fine-tuned U-Net model," *Engineering, Science, and Sustainability*, pp. 16–21, 2023.

[11] J. Susan and P. Subashini, "Deep learning inpainting model on digital and medical images—A review," *Int. Arab J. Inf. Technol.*, vol. 20, no. 6, pp. 919–936, 2023.

[12] G. J. Chowdary and Z. Yin, "Diffusion transformer u-net for medical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer Nature Switzerland, pp. 622–631, 2023.

[13] Y. Yang, H. Fu, A. I. Aviles-Rivero, C. B. Schönlieb, and L. Zhu, "Diffmic: Dual-guidance diffusion network for medical image classification," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer Nature Switzerland, pp. 95–105, 2023.

[14] Z. Yue, S. Ding, X. Li, and S. Yang, "Automatic acetowhite lesion segmentation via specular reflection removal and deep attention network," *IEEE Journal of Biomedical and Health Informatics*, 2021. https://ieeexplore.ieee.org/abstract/document/9372756/

[15] R. Gorantla, R. Singh, R. Pandey, and M. Jain, "Cervical cancer diagnosis using cervixnet—A deep learning approach," in *Proc. 2019 IEEE 19th International Conference on Bioinformatics and Biomedicine*, 2019. doi: 10.1109/BIBM.2019.00000

[16] X. Wang, P. Li, Y. Lv, H. Xue, T. Xu, Y. Du, and P. Liu, "Integration of global and local features for specular reflection inpainting in colposcopic images," *Journal of Healthcare Engineering*, vol. 1, p. 5401308, 2021.

[17] P. Oak and B. Iyer, "Specular reflection detection for early prediction of cervix cancer," *Emerging Trends in Electrical, Communications, and Information Technologies*, vol. 569, pp. 683–691, 2020. doi: 10.1007/978-981-13-8942-9_58

[18] A. Marquez-Grajales, H. G. Acosta-Mesa, E. Mezura-Montes, and R. Hernández-Jiménez, "Cervical image segmentation using active contours and evolutionary programming over temporary acetowhite patterns," *IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 3863–3870.

[19] M. B. Susan and P. Subashini, "Improvising grading of cervical cancer using quality assessment method in smart colposcopy images," *Measurement: Sensors*, 2023. https://www.sciencedirect.com/science/article/pii/S2665917423001241

[20] L. Jimenez-Martin, D. A. V. Pérez, A. M. S. Asteasuainzarra, L. Leonard, and M. L. B. Díaz-Romañach, "Specular reflections removal in colposcopic images based on neural networks: Supervised training with no ground truth previous knowledge," arXiv preprint, arXiv:2106.02221, 2021.

[21] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2023, pp. 22500–22510.

[22] H. Ye, J. Zhang, S. Liu, X. Han, and W. Yang, "Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models," arXiv preprint, arXiv:2308.06721, 2023.

[23] Z. Wang and H. R. Sheikh, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004. doi: 10.1109/TIP.2003.819861