




Endoscopic Artifact Inpainting for Improved Endoscopic Image Segmentation

Zhangyuan Yu ^{*1}, Chenlin Du ^{*2} , Hongrui Liang¹, Xiuqi Zheng¹, Zeyao Ma³, Mingjun Wu⁴, Mingwu Ao⁴, and Qicheng Lao¹  

¹ School of Artificial Intelligence, Beijing University of Posts and Telecommunications (BUPT), Beijing, China

qicheng.lao@bupt.edu.cn

² Department of Geriatric Dentistry, Peking University School and Hospital of Stomatology & National Center of Stomatology & National Clinical Research Center for Oral Diseases & Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health & NMPA Key Laboratory for Dental Materials, Beijing, China

³ Department of Orthodontics, School of Stomatology, Capital Medical University, Beijing, China

⁴ Ningbo Fregty Optoelectronics Technology Co., Ltd, Ningbo, China

Abstract. Endoscopic imaging plays a crucial role in modern diagnostics and minimally invasive procedures. However, artifacts caused by specular and diffuse reflections present significant challenges, particularly in tasks such as endoscopic image segmentation. Existing methods tackling endoscopic artifacts typically address only one type of reflection, failing to fully account for the non-Lambertian reflectance of endoscopic tissue structures. Therefore, inspired by the simplified Phong model for endoscopy, we propose a two-stage artifact inpainting framework. The first stage suppresses specular artifacts, while the second stage focuses on inpainting diffuse artifacts. Additionally, we introduce a weight map to control the handling of diffuse artifacts, ensuring a more precise enhancement. To evaluate its effectiveness, we focus on its impact on endoscopic image segmentation tasks. Extensive experiments on multiple colonoscopy and dental endoscopy datasets demonstrate that our framework can robustly improve the segmentation performance of endoscopic images, offering better enhancement than existing state-of-the-art methods. Particularly, for zero-shot SAM segmentation of teeth, a significant performance boost is observed after inpainting, with mDice and mIoU increasing from 51.5%/39.3% to 96.1%/93.0%. Code is available at [GitHub](#).

Keywords: Endoscopic Artifact · Specular Reflection · Diffuse Reflection · Image Inpainting · Endoscopic Image Segmentation.

1 Introduction

Endoscopy has become a cornerstone of modern medicine, offering clinicians a high-resolution, real-time view of the body’s intricate structures to detect, mon-

* Equal contribution.

itor, and treat various conditions while minimizing the risks of invasive procedures [1]. With the rapid development of deep learning, researchers are leveraging these endoscopic images for tasks like segmentation [2], diagnostic support [3], depth estimation [4], and 3D reconstruction [5]. The integration of AI with endoscopic imaging holds the potential to revolutionize medical procedures, leading to faster, more accurate, and less invasive interventions.

Despite these impressive strides, the unique anatomical characteristics of endoscopic tissue structures pose their own challenges. A significant number of artifacts arise during the imaging process, with reflection artifacts being one of the most common issues [6]. These artifacts not only impair immediate diagnostics but also compromise downstream quantitative analyses, such as video mosaicking and keyframe retrieval, leading to an increased need for repeat procedures and contributing to higher patient discomfort and healthcare costs [2]. Furthermore, small areas of sharp specular reflections can negatively impact image segmentation [7], while large overexposed areas caused by diffuse reflection may interfere with 3D reconstruction [8].

Some studies have attempted to address the issue of reflection artifacts in endoscopic imagery. However, most anatomical structures captured in such images exhibit non-Lambertian reflectance properties [8], meaning they simultaneously present both specular and diffuse reflections. Existing research addresses reflection artifacts in endoscopic imagery from three main approaches. One focuses on suppressing specular artifacts [7], but it overlooks diffuse artifacts. Another relies on iterative in-situ rendering [8], which is computationally expensive and unsuitable for real-time applications. A third approach utilizes video-based image processing techniques [9], but these methods often struggle with dynamic endoscopic environments. Moreover, the complex and varied anatomical structures in endoscopic images, such as the intricate tissue surfaces, make it challenging to acquire artifact-free images. This difficulty further complicates former effective methods like TSHRNet [10] and M2-Net [11]. Recently, larger-scale pretrained models such as LaMa [12] and StableDelight (SD) [13] offer potential for zero-shot inpainting of these artifacts, though their effective application to endoscopic imagery still requires further investigation.

Therefore, in this paper, we propose a two-stage endoscopic artifact inpainting framework to address the challenges of non-Lambertian reflectance. The framework first suppresses high-intensity specular artifacts by locating and inpainting overexposed regions, restoring underlying tissue textures. It then addresses residual diffuse artifacts by adaptively blending intensity-guided refinements, harmonizing inconsistent illumination while preserving critical anatomical features. To evaluate the effectiveness of our approach, we perform extensive experiments across fully supervised learning and zero-shot adaptation on multiple endoscopic datasets. These real-world endoscopic scenarios demonstrate the robustness and generalizability of our method, showing significant improvements in segmentation accuracy, resilience to imaging artifacts, and adaptability to diverse anatomical structures.

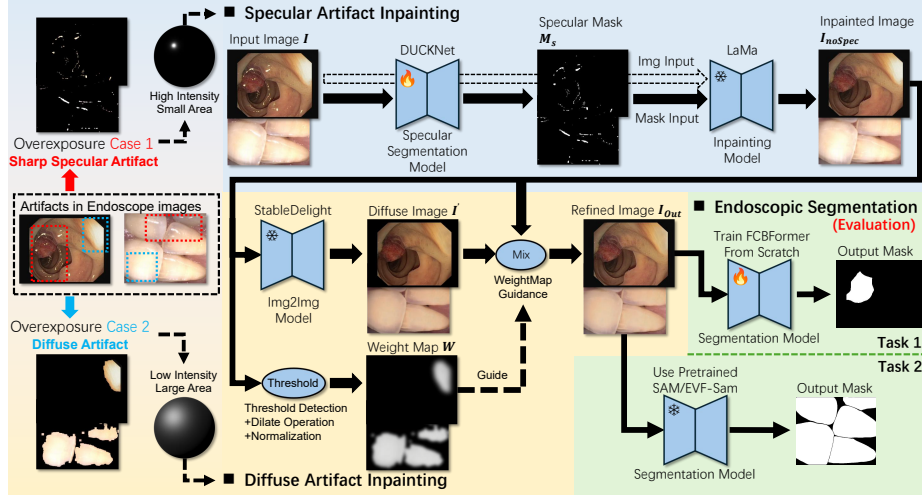


Fig. 1: Illustration of the proposed framework.

2 Methods

We propose a two-stage method to address endoscopic artifacts for improving endoscopic image segmentation. As illustrated in Fig. 1, our framework consists of two key components: **Specular Artifact Inpainting** (Sec. 2.2), which targets highly reflective specular regions, effectively mitigating intense glare and minimizing disruptive artifacts; and **Diffuse Artifact Inpainting** (Sec. 2.3), which refines areas with weaker reflections that exhibit Lambertian-like properties and appear closer to diffuse reflections, reducing residual artifacts and enhancing surface consistency. These enhancements make the processed endoscopic images more suitable for subsequent segmentation tasks.

2.1 Preliminary

To describe the reflection phenomenon in endoscopic images, we employ the Phong reflectance model [14], which is widely used to simulate the way light interacts with surfaces. This model is based on three components: ambient, specular, and diffuse reflection. Since endoscopy typically uses a single light source, and the light source and camera are positioned closely together [15], we assume the light \mathbf{L} is aligned with the viewing direction. Under these conditions, the equation for the observed endoscopic image I can be expanded as:

$$I = \sum k_a(x, y) |\mathbf{L}| + \underbrace{\sum k_s(x, y) |\mathbf{L}|}_{\text{specular artifact}} + \underbrace{\sum k_d(x, y) (\mathbf{N}(x, y) \cdot \mathbf{L})}_{\text{diffuse artifact}} \quad (1)$$

where $k_a(x, y)$ is the RGB value at pixel (x, y) , k_s is the specular reflection coefficient on smooth surface, e.g., tissue covered with mucus, k_d is the diffuse reflection coefficient on rougher surface, and $\mathbf{N}(x, y)$ represents surface normal.

Thus, the task of removing artifacts can be formulated as estimating the distribution of $\sum k_s(x, y)|\mathbf{L}|$ and $\sum k_d(x, y)(\mathbf{N}(x, y) \cdot \mathbf{L})$ across the image I , effectively separating the specular and diffuse artifact components.

2.2 Specular Artifact Inpainting

Since sharp specular artifacts primarily arise from strong directional reflections, they tend to exhibit high intensity, distinct color distributions, and strong spatial coherence, making them relatively easy to segment by appropriate masks.

As a result, we apply a pretrained mask-guided inpainting neural network LaMa [12], which leverages fast Fourier convolution blocks to generate missing structures within specified regions. Let $INP(\cdot, *)$ denote the inpainting model, we can obtain specular artifacts suppressed image I_{noSpec} with:

$$I_{noSpec} = INP(I_{masked}, M_s) \quad (2)$$

where $M_s \in \{0, 1\}^{H \times W}$ represents a pixel-wise binary mask indicting highly reflective regions across the input image I , and I_{masked} denotes the masked image generated by subtract operation, i.e., $I_{masked} = I - M_s \odot I$. To obtain M_s for the specular artifact inpainting, we train a segmentation network DUCKNet [16] from scratch, which is specially designed for endoscopic images.

Given that when specular reflections occur on sufficiently smooth tissue surfaces, the local intensity of reflected light can exceed the dynamic range of the imaging sensor, leading to overexposure and loss of structural details. Therefore, the distribution of $k_s(x, y)$ across the image I in Eq. (1) can be effectively estimated by using the binary specular mask M_s , i.e., $\sum k_s(x, y)|\mathbf{L}| = \sum M_s(x, y)|\mathbf{L}|$. As such, the specular mask-guided LaMa inpainting process can finally suppress the specular artifact component in Eq. (1).

2.3 Diffuse Artifact Inpainting

Note that even after specular artifacts have been suppressed, the resulting image I_{noSpec} still contains diffuse artifacts, which can be modeled as follows:

$$I_{noSpec} = \sum k_a(x, y)|\mathbf{L}| + \sum k_d(x, y)(\mathbf{N}(x, y) \cdot \mathbf{L}). \quad (3)$$

However, accurate pixel-wise estimation of $k_d(x, y)$ and $\mathbf{N}(x, y)$ across the specular artifact-free image I_{noSpec} is computationally inefficient. To overcome this limitation, a tailored stable-diffusion architecture for reflection removal, StableDelight [13], is utilized, which leverages the You-Only-Sample-Once (YOSO) method [17], enabling an effective image-to-image strategy that directly addresses the residual artifacts.

Therefore, given a well-pretrained VAE encoder $Enc(\cdot)$, control signal encoder f_ϕ , decoder blocks of U-Net $\mu_\theta(\cdot)$, and VAE decoder $Dec(\cdot)$ of StableDelight [13], this process could be described as:

$$I' = Dec(\mu_\theta(f_\phi(Enc(I_{noSpec})), t_+, \epsilon)) \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \quad (4)$$

where I' should be ideally equal to the artifact-free image. However, StableDelight was pretrained on multiple multi-light-source and non-endoscopic images including datasets used by TSHRNet [10]. As a result, pathological details in regions with fewer or no diffuse artifacts may be partially mistakenly removed.

To address adjustments, we propose a WeightMap Guidance submodule, where the intermediate image I' from Eq. (4) is adaptively fused with the inpainted image I_{noSpec} in Eq. (2) to output the final refined image I_{out} :

$$I_{Out} = W \odot I' + (1 - W) \odot I_{noSpec} \quad (5)$$

where $W \in [0, 1]^{H \times W}$ denotes a newly calculated weight map, guiding the selective blending process. The weight map W plays a critical role in selectively refining regions affected by diffuse artifacts, particularly suppressing the excessively processed details by the StableDelight model.

To generate the critical weight map W , we leverage a threshold-based artifact detection method. First, we convert the specular artifact-suppressed image I_{noSpec} into a grayscale image G . An indicator function $IND(x \in G, \Theta)$ reserves only artifact regions with light intensity higher than threshold Θ , followed by dilation operation (kernel size = 15) $DILATE(\cdot)$ to refine the detection further. The resulting grayscale intensity map $DILATE(IND(G, \Theta)) \in [0, 255]^{H \times W}$ is then scaled to a continuous-valued weight map $W \in [0, 1]^{H \times W}$ using the following equation:

$$W = \frac{DILATE(IND(G, \Theta)) - \min(DILATE(IND(G, \Theta)))}{\max(DILATE(IND(G, \Theta))) - \min(DILATE(IND(G, \Theta)))} \quad (6)$$

where higher values of W indicate stronger diffuse artifact presence, ensuring that essential anatomical information is preserved while reducing artifacts.

3 Experiments

3.1 Experimental Setup

We evaluate our proposed method on endoscopic image segmentation tasks, including colonoscopy polyp segmentation and dental endoscopy tooth segmentation, where endoscopic artifacts negatively impact segmentation performance. We use both supervised and zero-shot settings to comprehensively evaluate the effectiveness of our approach. We report mDICE and mIOU for quantification.

Datasets 1) *Colonoscopy polyp segmentation* We use four polyp segmentation datasets including CVC-ClinicDB [18], Kvasir [19], CVC-ColonDB [20], and ETIS [21], which provide 612/1000/380/196 input-target pairs in total, respectively. We follow the data split (train:val:test = 8:1:1) and evaluation protocols in PraNet [22]. 2) *Dental endoscopy tooth segmentation* For the tooth segmentation task, we curate an in-house dataset of 195 dental endoscopic images captured using the RGB camera of an intraoral 3D scanner with human-annotated tooth masks.

Supervised and Zero-Shot Settings Two distinct learning paradigms were investigated: 1) *Supervised learning* for polyp segmentation, where FCBFormer [23], a model tailored for endoscopic applications, was trained end-to-end following [23] on artifact-free polyp images; 2) *Zero-shot adaptation* for tooth segmentation, leveraging foundation models without task-specific training. We employed SAM [24] with manual point prompts targeting diagnostically critical regions and EVF-SAM [25] using text prompts to infer processed artifact-free images.

Implementation Details We employ LaMa’s publicly-released pretrained weight **Big-LaMa** for specular artifact inpainting. To avoid the omission of artifacts, we set LaMa’s hyperparameter k_{size} to 15, expanding the coverage of specular mask M_s . We also leverage DUCKNet [16] trained on the CVC-ClinicSpecific dataset [26] for specular artifact detection and the acquisition of specular mask M_s . For all other hyperparameters related to LaMa and DUCKNet, we follow the official implementations for consistency and reproducibility. Lastly, threshold Θ is set to 150 in WeightMap guidance submodule.

Using an NVIDIA 4090 GPU and Pytorch 2.2.0 framework on python 3.10.16, we train DUCKNet with 16 filters for 100 epochs using a batch size of 8 and the RMSprop optimizer with an initial learning rate of $1e-4$. The sum of soft dice loss (smooth = 1) and binary cross entropy loss is used for the loss function.

For foundation models, we use SAM’s **SAM-ViT-H** and EVF-SAM’s **EVF-SAM-multimask** checkpoints. Note that EVF-SAM generates semantic-level masks only with a special token "[semantic]". Thus we utilize "[semantic] Tooth" for the text prompt.

3.2 Comparison with State-of-the-Art Methods

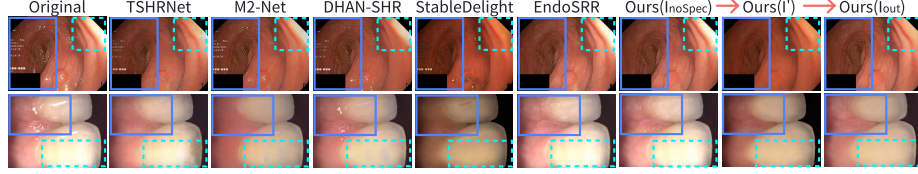
To thoroughly evaluate our proposed method, we compare the performance of our framework in enhancing the segmentation performance against SOTA models that can be adapted for endoscopic artifact inpainting. These include TSHR-Net [10], M2-Net [11], DHAN-SHR [27], EndoSRR [7] and StableDelight [13].

Table 1 shows primary assessment results on both colonoscopy polyp segmentation and dental endoscopy tooth segmentation tasks. The results of polyp segmentation are based on FCBFormer [23], whereas tooth segmentation results are obtained through prompting SAM/EVF-SAM [24,25]. As shown in the table, our proposed method outperforms all baseline methods across all metrics on the datasets, demonstrating the superior generalization and robustness of our approach compared to various established SOTA methods.

Colonoscopy polyp segmentation performance Although most artifact inpainting methods show improvements over the baseline with original images, these improvements are marginal and not guaranteed in all scenarios. In contrast, our proposed method consistently demonstrates improvements across all cases. For example, it achieves 81.0% mDice and 73.2% mIoU on the ETIS dataset, 1.5%/1.8% higher than the second-best performing EndoSRR ($p < 0.05$, Wilcoxon signed-rank test), indicating that these gains are statistically significant.

Table 1: Quantitative comparison of the proposed artifact inpainting method with state-of-the-art methods for segmentation performance enhancement (%)

| Methods | Polyp Datasets | | | | | | | | Tooth Dataset | | | |
|--------------------|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------|-------------|-------------|-------------|
| | CVC-ClinicDB | | | | Kavsir-SEG | | | | Teeth(Point) | | Teeth(Text) | |
| | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU |
| Original Image | 93.2 | 88.8 | 92.0 | 87.0 | 77.4 | 69.0 | 79.3 | 71.3 | 51.5 | 39.3 | 84.1 | 74.0 |
| TSHRNet [10] | 93.0 | 88.4 | 91.4 | 86.2 | 79.0 | 69.7 | 73.6 | 65.2 | 73.5 | 62.1 | 80.6 | 70.0 |
| M2-Net [11] | 93.6 | 89.4 | 92.5 | 87.4 | 80.9 | 73.0 | 78.2 | 70.9 | 75.7 | 65.6 | 84.6 | 74.9 |
| DHAN-SHR [27] | 93.6 | 88.5 | 92.1 | 87.0 | 79.4 | 70.2 | 77.7 | 68.6 | 65.1 | 52.9 | 85.1 | 75.4 |
| EndoSRR [7] | 94.1 | 89.2 | 92.7 | 87.6 | 79.6 | 70.0 | 79.5 | 71.4 | 95.8 | 92.5 | 86.3 | 77.3 |
| StableDelight [13] | 93.4 | 89.0 | 91.3 | 86.0 | 76.2 | 69.0 | 69.2 | 62.4 | 91.4 | 85.6 | 77.9 | 66.8 |
| Ours | 95.0 | 90.6 | 93.3 | 88.3 | 81.5 | 73.4 | 81.0 | 73.2 | 96.1 | 93.0 | 86.8 | 78.0 |

Fig. 2: Visual comparisons between original image, former methods, and our proposed method. *Blue* boxes: specular artifacts; *Green* boxes: diffuse artifacts.

Dental endoscopy tooth segmentation performance Similarly, our proposed method also achieves remarkable improvements in enhancing zero-shot segmentation accuracy on the tooth dataset, reaching 96.1%/93.0% with SAM and 86.8%/78.0% with EVF-SAM, a dramatic improvement over the baseline by +44.6%/+53.7% and +2.7%/+4.0% in the mDice and mIoU.

Visualizations Fig. 2 demonstrates our framework can effectively inpaint both specular and diffuse artifacts, addressing challenges where existing models either omit specular artifacts or excessively handle diffuse artifacts. For example, M2-Net, DHAN-SHR and StableDelight fail to suppress specular artifacts (*blue boxes*), while TSHRNet and EndoSRR fail to inpaint diffuse artifacts (*green boxes*). However, our proposed method can accurately locate and inpaint both types of artifacts while preserving the original details with minimal distortion.

Fig. 3 illustrates how our proposed method overcomes the impact of reflection artifacts on segmentation results, providing more well-defined boundaries in the segmentation maps compared to other baseline methods.

3.3 Ablation Study

Key components for artifacts inpainting We conduct ablation studies to evaluate the effectiveness of key components in Table 2. In the specular artifact inpainting module, *Detect* refers to specular artifact detection using DUCKNet,

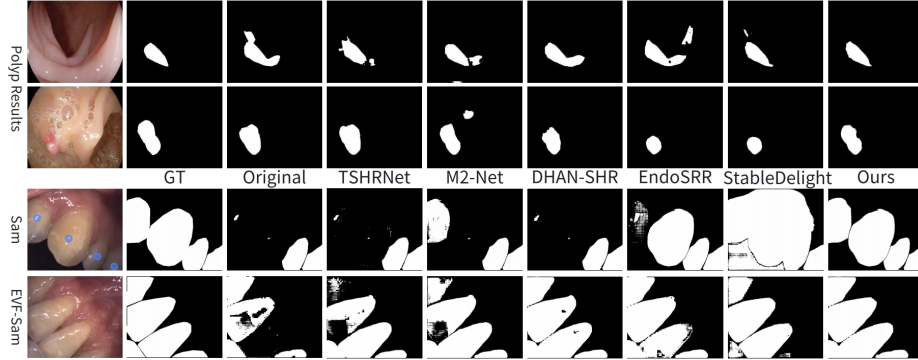


Fig. 3: Segmentation result comparisons between our method and baselines.

Table 2: Ablation on key components for artifact inpainting (%).

| Modules | | | | Datasets | | | | | | | | | | | |
|---------------|----------------|-----------|--------------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
| Specular | | Diffuse | | CVC-ClinicDB | | Kvasir-SEG | | CVC-ColonDB | | ETIS | | Teeth(Point) | | Teeth(Text) | |
| <i>Detect</i> | <i>Inpaint</i> | <i>SD</i> | <i>Guide</i> | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU |
| × | × | × | × | 93.2 | 88.8 | 92.0 | 87.0 | 77.4 | 69.0 | 79.3 | 71.3 | 51.5 | 39.3 | 84.1 | 74.0 |
| ✓ | × | × | × | 94.5 | 89.9 | 92.3 | 87.2 | 78.4 | 69.5 | 76.5 | 68.1 | 49.5 | 37.4 | 83.6 | 73.2 |
| ✓ | ✓ | × | × | 94.6 | 90.0 | 93.0 | 88.0 | 81.8 | 73.8 | 80.0 | 72.2 | 95.8 | 92.7 | 86.5 | 77.5 |
| ✓ | ✓ | ✓ | × | 94.4 | 89.6 | 91.1 | 86.4 | 76.8 | 69.7 | 77.7 | 70.5 | 94.5 | 90.5 | 83.4 | 73.6 |
| ✓ | ✓ | ✓ | ✓ | 95.0 | 90.6 | 93.3 | 88.3 | 81.5 | 73.4 | 81.0 | 73.2 | 96.1 | 93.0 | 86.8 | 78.0 |

Table 3: Ablation on loss functions(%)

| Dice | BCE | CVC-ClinicSpecific | |
|------|-----|--------------------|--------------|
| | | mDice | mIoU |
| ✓ | × | 79.08 | 67.03 |
| × | ✓ | 82.78 | 71.62 |
| ✓ | ✓ | 83.44 | 72.61 |

Table 4: Ablation on dilate kernel size(%)

| k_{size} | CVC-ClinicDB | |
|------------|--------------|--------------|
| | mDice | mIoU |
| 12 | 94.66 | 90.12 |
| 15 | 95.02 | 90.63 |
| 18 | 93.75 | 89.66 |

Table 5: Ablation on threshold selection(%)

| θ | CVC-ClinicDB | |
|----------|--------------|--------------|
| | mDice | mIoU |
| 140 | 94.10 | 89.37 |
| 150 | 95.02 | 90.63 |
| 160 | 94.06 | 89.50 |

and *Inpaint* stands for the inpainting process. In the diffuse artifact inpainting module, *SD* stands for the inpainting with StableDelight, and *Guide* is the WeightMap guidance. The results demonstrate the effectiveness of each component in our proposed method across most datasets. However, it should be noted that on the CVC-ColonDB dataset, which contains blurry images with significant dispersion, the contribution of the diffuse artifact inpainting is less pronounced compared to other scenarios.

Loss function for specular artifact detection Two loss functions are evaluated for specular artifact detection including soft dice loss (Dice) and binary

cross entropy loss (BCE). Table 3 shows that the best performance of DUCKNet is achieved by combining both loss functions in the training process.

Dilate kernel size for specular artifact inpainting Table 4 examines the impact of LaMa’s configurable hyperparameter k_{size} on specular artifact inpainting performance in the CVC-ClinicDB dataset. Experimental findings indicate that setting this parameter to 15 yields optimal inpainting results.

Threshold selection for WeightMap Guidance Table 5 demonstrates the effect of Θ on segmentation enhancement using the CVC-ClinicDB dataset. The results suggest that 150 is more suitable for guiding the mixture of I' and I_{noSpec} .

4 Conclusion

We propose a novel two-stage framework for endoscopic artifact inpainting, addressing both specular and diffuse artifacts in non-Lambertian anatomical structures. Extensive experiments on polyp and dental datasets demonstrate significant improvements of the proposed method in downstream endoscopic tasks, achieving state-of-the-art performance in supervised segmentation and robust zero-shot segmentation for endoscopic images. Future work will focus on evaluating and adapting the proposed method for more downstream tasks, as well as integrating real-time capabilities to facilitate clinical deployment.

Disclosure of Interests. No conflicts of interests to be declared.

References

1. Li, H., Hou, X., Lin, R., Fan, M., Pang, S., Jiang, L., Liu, Q., Fu, L.: Advanced endoscopic methods in gastrointestinal diseases: a systematic review. *Quantitative Imaging in Medicine and Surgery* **9**(5), 905 (2019)
2. Ali, S., Dmitrieva, M., Ghatwary, N., Bano, S., Polat, G., Temizel, A., Krenzer, A., Hekalo, A., Guo, Y.B., Matuszewski, B., et al.: Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy. *Medical image analysis* **70**, 102002 (2021)
3. Klang, E., Barash, Y., Margalit, R.Y., Soffer, S., Shimon, O., Albshesh, A., Ben-Horin, S., Amitai, M.M., Eliakim, R., Kopylov, U.: Deep learning algorithms for automated detection of crohn’s disease ulcers by video capsule endoscopy. *Gastrointestinal endoscopy* **91**(3), 606–613 (2020)
4. Han, J.J., Acar, A., Henry, C., Wu, J.Y.: Depth anything in medical images: A comparative study. *arXiv preprint arXiv:2401.16600* (2024)
5. Ozyoruk, K.B., Gokceler, G.I., Bobrow, T.L., Coskun, G., Incetan, K., Almalioglu, Y., Mahmood, F., Curto, E., Perdigoto, L., Oliveira, M., et al.: Endoslam dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos. *Medical image analysis* **71**, 102058 (2021)
6. Ali, S., Zhou, F., Daul, C., Braden, B., Bailey, A., Realdon, S., East, J., Wagnieres, G., Loschenov, V., Grisan, E., et al.: Endoscopy artifact detection (ead 2019) challenge dataset. *arXiv preprint arXiv:1905.03209* (2019)

7. Li, W., Jia, F., Liu, W.: Endosrr: a comprehensive multi-stage approach for endoscopic specular reflection removal. *International Journal of Computer Assisted Radiology and Surgery* **19**(6), 1203–1211 (2024)
8. Zhu, B., Yang, Y., Wang, X., Zheng, Y., Guibas, L.: Vdn-nerf: Resolving shape-radiance ambiguity via view-dependence normalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 35–45 (2023)
9. Zhang, F.X., Chen, S., Xie, X., Shum, H.P.: Depth-aware endoscopic video inpainting. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 143–153. Springer (2024)
10. Fu, G., Zhang, Q., Zhu, L., Xiao, C., Li, P.: Towards high-quality specular highlight removal by leveraging large-scale synthetic data. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 12857–12865 (2023)
11. Huang, Z., Hu, K., Wang, X.: M2-net: multi-stages specular highlight detection and removal in multi-scenes. *arXiv preprint arXiv:2207.09965* (2022)
12. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 2149–2159 (2022)
13. Stable-X: Stabledelight: Revealing hidden textures by removing specular reflections. <https://github.com/Stable-X/StableDelight> (2025), accessed: 2025-02-10
14. Tan, P.: Phong reflectance model. *Computer Vision: A Reference Guide* pp. 1–3 (2020)
15. Okatani, T., Deguchi, K.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer vision and image understanding* **66**(2), 119–131 (1997)
16. Dumitru, R.G., Peteleaza, D., Craciun, C.: Using duck-net for polyp image segmentation. *Scientific reports* **13**(1), 9803 (2023)
17. Ye, C., Qiu, L., Gu, X., Zuo, Q., Wu, Y., Dong, Z., Bo, L., Xiu, Y., Han, X.: Stablenormal: Reducing diffusion variance for stable and sharp normal. *ACM Transactions on Graphics (TOG)* (2024)
18. Bernal, Jorge, Vilarino, Fernando, Fernandez-Esparrach, Gloria, Gil, Debora, Javier, Sanchez: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society* **43**, 99–111 (2015)
19. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: *MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II* 26. pp. 451–462. Springer (2020)
20. David, V., Jorge, B., Javier, S.F., Gloria, F.E., M., L.A., Adriana, R., Michal, D., Aaron, C.: A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017, (2017-7-26) **2017**, 1–9 (2017)
21. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery* **9**(2), 283–293 (2013)
22. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranel: Parallel reverse attention network for polyp segmentation. In: *International conference on medical image computing and computer-assisted intervention*. pp. 263–273. Springer (2020)

23. Sanderson, E., Matuszewski, B.J.: Fcn-transformer feature fusion for polyp segmentation. In: Annual conference on medical image understanding and analysis. pp. 892–907. Springer (2022)
24. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv:2304.02643 (2023)
25. Zhang, Y., Cheng, T., Hu, R., Liu, L., Liu, H., Ran, L., Chen, X., Liu, W., Wang, X.: Evf-sam: Early vision-language fusion for text-prompted segment anything model. arXiv preprint arXiv:2406.20076 (2024)
26. Sánchez, F.J., Bernal, J., Sánchez-Montes, C., de Miguel, C.R., Fernández-Esparrach, G.: Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. *Machine Vision and Applications* **28**(8), 917–936 (2017)
27. Guo, X., Chen, X., Luo, S., Wang, S., Pun, C.M.: Dual-hybrid attention network for specular highlight removal. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 10173–10181 (2024)