



EndoSRR: a comprehensive multi-stage approach for endoscopic specular reflection removal

Wei Li^{1,2} · Fucang Jia^{2,3} · Wenjian Liu¹

Received: 29 February 2024 / Accepted: 28 March 2024 / Published online: 20 April 2024
© CARS 2024

Abstract

Purpose Specular reflections in endoscopic images not only disturb visual perception but also hamper computer vision algorithm performance. However, the intricate nature and variability of these reflections, coupled with a lack of relevant datasets, pose ongoing challenges for removal.

Methods We present EndoSRR, a robust method for eliminating specular reflections in endoscopic images. EndoSRR comprises two stages: reflection detection and reflection region inpainting. In the reflection detection stage, we adapt and fine-tune the segment anything model (SAM) using a weakly labeled dataset, achieving an accurate reflection mask. For reflective region inpainting, we employ LaMa, a fast Fourier convolution-based model trained on a 4.5M-image dataset, enabling effective inpainting of arbitrarily shaped reflection regions. Lastly, we introduce an iterative optimization strategy for dual pre-trained models to refine the results of specular reflection removal, named DPMIO.

Results Utilizing the SCARED-2019 dataset, our approach surpasses state-of-the-art methods in both qualitative and quantitative evaluations. Qualitatively, our method excels in accurately detecting reflective regions, yielding more natural and realistic inpainting results. Quantitatively, our method demonstrates superior performance in both segmentation evaluation metrics (IoU, E-measure, etc.) and image inpainting evaluation metrics (PSNR, SSIM, etc.).

Conclusion The experimental results underscore the significance of proficient endoscopic specular reflection removal for enhancing visual perception and downstream tasks. The methodology and results presented in this study are poised to catalyze advancements in specular reflection removal, thereby augmenting the accuracy and safety of minimally invasive surgery.

Keywords Specular reflection · Reflection detection · Image inpainting · Endoscopic image · Transfer learning

Introduction

In minimally invasive surgery (MIS), endoscopy serves as a pivotal visual aid, facilitating precise lesion observation, diagnosis, and treatment. Nevertheless, the unique

imaging environment presents a challenge—harsh specular reflections inevitably occur during the procedure [1]. These reflections not only cause visual disturbances but also impede the performance of computer vision algorithms [2]. Therefore, effective removal of specular reflections from endoscopic images is essential and significant.

Specular reflection removal typically involves two stages: specular reflection detection and specular reflection region inpainting. In the specular reflection detection stage, traditional methods primarily depend on conventional image processing algorithms, falling into two categories—threshold-based methods and principal component analysis-based methods. Thresholding-based approaches often involve converting the image to HSV/YUV color space and subsequently employing double thresholding or adaptive thresholding to isolate the reflection region [3–6]. For instance, Arnold et al. [6] employed a detection method founded on a combination of nonlinear filtering and color thresholding. More recently,

✉ Fucang Jia
fc.jia@siat.ac.cn

✉ Wenjian Liu
andyliu@cityu.mo

Wei Li
w.li@siat.ac.cn

¹ Faculty of Data Science, City University of Macau, Macau, China

² Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

³ The Key Laboratory of Biomedical Imaging Science and System, Chinese Academy of Sciences, Shenzhen, China

Li et al. [7] introduced the concept of adaptive robust principal component analysis (AdaRPCA), and Pan et al. [1] proposed the accelerated adaptive non-convex robust principal component analysis (AANC-RPCA) for specular reflection detection. These principal component analysis-based methods typically execute sparse and low-rank decomposition of the endoscopic images to derive reflection masks. However, these conventional algorithms, reliant on fixed thresholds, often exhibit poor generalization ability, struggling to effectively identify reflection regions. The scarcity of specular reflection datasets constrains progress in deep learning methods. Monkam et al. [8] used a hybrid strategy, combining transfer learning and weakly supervised learning for training lightweight U-Net models with inaccurate labels, which, however, struggled with small reflective regions. Ali et al. [9] showed improved detection accuracy using RestNet50 with DeepLabv3+. As depicted in Fig. 1, the diverse forms, shapes, and sizes of specular reflections in endoscopic images pose unresolved challenges for current detection methods, leading to two predominant issues: over-segmentation (false positives) and under-segmentation (false negatives). These challenges are particularly pronounced when the representation of the reflective region closely resembles that of the organ surface tissue.

Furthermore, owing to the substantial variations in texture and color evident in endoscopic images, an unresolved issue arises during the reflective region inpainting stage: the inefficiency in accurately inpainting larger specular reflective regions using both global and local information from the image. Arnold et al. [6] proposed an inpainting technique involving an initial filling of the reflective region based on neighboring pixels and subsequent nonlinear decay along

the edges. Meanwhile, principal component analysis-based methods like AdaRPCA [7] and AANC-RPCA [1] utilize the low-rank image derived from matrix decomposition either directly or with outward attenuation as the inpainting result. Additionally, various traditional inpainting methods, including stochastic Bayesian estimation [10], specific Sobolev operators [4], and example-based methods [5], have been explored for handling reflective regions.

Recent studies explored deep learning techniques for reflective region inpainting. Funke et al. [2], Ali et al. [9], and Daher et al. [11] employed a generative adversarial network-based approach. Monkam et al. [8] proposed the GatedResU-Net architecture, achieving more reasonable inpainting results. However, these approaches are marred by issues such as blurriness, a noticeable lack of texture, and an inability to seamlessly integrate with the surrounding texture. Such limitations compromise the meaningfulness of inpainting results for downstream computer vision applications. Consequently, developing a specular reflection removal system capable of accurately detecting reflective regions and realistically inpainting them concurrently poses a formidable challenge.

In this paper, we tackle three key challenges: (1) the scarcity of datasets; (2) the issues of over- and under-segmentation in specular reflection detection; and (3) the suboptimal inpainting results for specular reflection regions. While dataset labeling is a time-consuming and laborious task, we present a method for semi-automatic labeling of endoscopic specular reflection regions, resulting in a weakly labeled dataset. The proposed EndoSRR endoscope specular reflection removal framework, depicted in Fig. 2, utilizes this created dataset to fine-tune the adapted segment any-

Fig. 1 Different types of endoscopic specular reflections: **a** different reflection intensities; **b** different reflection halos; and **c** different reflection shapes and sizes

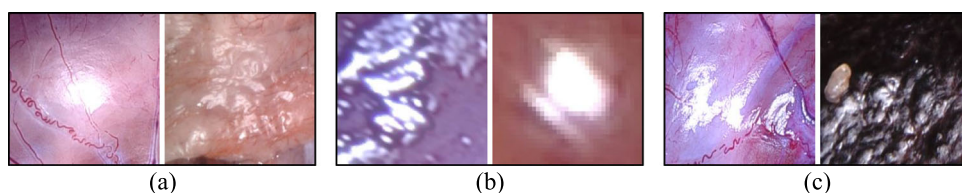
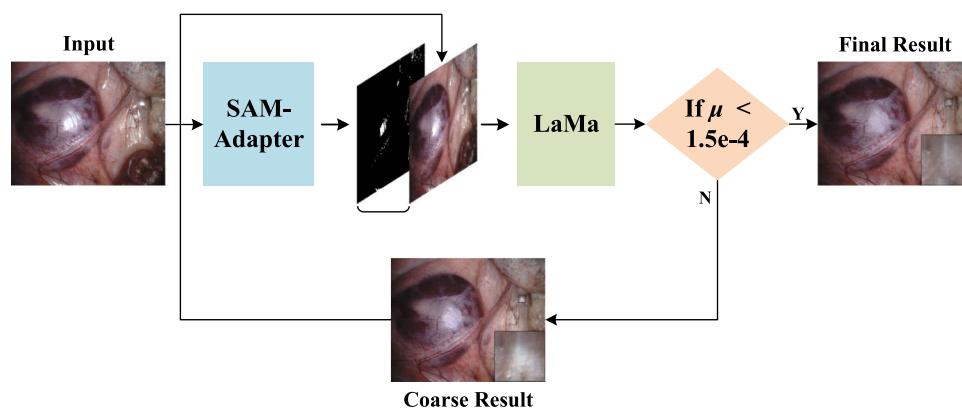


Fig. 2 EndoSRR: Overview of the proposed endoscopic specular reflection removal framework



thing model (SAM). The ensuing specular reflection masks serve as input for the resolution-robust large mask inpainting model (LaMa) to effectively inpaint specular reflection regions. Finally, we introduce a simple yet effective optimization strategy to further refine the specular reflection removal results. Through both qualitative and quantitative analysis and comparison, the outcomes of our proposed methods are optimal, excelling in both specular reflection detection and reflection region inpainting. Additionally, we directly apply the inpainting results to segment anything model, as well as visualize the inpainting results in 3D using the depth information provided by the SCARED-2019 dataset. The experimental results highlight that effective endoscopic specular reflection removal not only enhances downstream tasks but also alleviates the visual fatigue experienced by surgeons during prolonged surgical procedures.

The code is available at <https://github.com/Tobyzai/EndoSRR>.

Method

Creation of endoscopic specular reflection dataset

To fine-tune the SAM-adapter and acquire precise reflection masks, an endoscopic specular reflection weakly labeled dataset was meticulously crafted. Illustrated in Fig. 3, the process entails three main steps: (a) global k-means clustering of the image for initial coarse filtering of reflective regions, (b) local k-means clustering of the image to further refine reflective regions and encompass more regular halos via a dilation operation, and (c) manual outlining to meticulously refine irregular halos. The envisioned contribution of this dataset is to advance the field of reflection removal.

SAM-adapter for reflection detection

Capitalizing on SAM’s capabilities derived from massive corpora for the specific task of specular reflection detection prompts the question of how to effectively leverage them [12]. An efficient solution is to integrate Explicit Visual Prompts into the SAM model [13, 14]. In this study, we employ SAM-adapter to segment reflection regions based on a purpose-created, small dataset. Illustrated in Fig. 4, SAM-adapter comprises four modules.

Module-1: High-frequency components tune (First Column on the Top Left). Utilizing fast Fourier transform, the high-frequency component I_{hfc} is extracted from the image, resulting in a small patch I_{hfc}^P with the same format as SAM. To align with SAM’s dimensions, the patch undergoes projection to yield features F_{hfc} using a linear layer L_{hfc} . The primary objective of this module is to instill invariance in the pre-trained model to endoscopic image features through data

augmentation. The process is defined as follows:

$$F_{\text{hfc}} = L_{\text{hfc}}(I_{\text{hfc}}^P). \quad (1)$$

Module-2: Patch embedding tune (Second Column on the Top Left). This module is tailored to adjust the pre-trained patch embedding, aiming to shift its distribution from the pre-trained dataset to the endoscopic specular reflection dataset. I^P represents the frozen patch embedding output of SAM, traversing through a linear layer L_{pe} and projecting onto the features F_{pe} . The operational equation is defined as:

$$F_{\text{pe}} = L_{\text{pe}}(I^P). \quad (2)$$

Module-3: Adapter (Top Right). The purpose of each adapter is to dynamically integrate the features F_{hfc} and F_{pe} using their respective multilayer perceptron $\text{MLP}_{\text{tune}}^i$, activation functions GELU [15], and globally shared multilayer perceptron MLP_{up} , and attach the resulting output visual prompts P^i to their corresponding transformer layers. For the i -th adapter, the process is defined as follows:

$$P^i = \text{MLP}_{\text{up}}(\text{GELU}(\text{MLP}_{\text{tune}}^i(F_{\text{hfc}} + F_{\text{pe}}))). \quad (3)$$

Module-4: SAM (Bottom). SAM [13] serves as the backbone for the endoscopic specular reflection segmentation network, comprising an encoder and decoder. The encoder remains frozen, with each of its layers embedded with visual prompts P^i from adapters. Contrarily, the decoder does not intake any form of prompt information.

By complementing each other with the four modules of the SAM-adapter, task-specific knowledge is integrated with the general knowledge gained during the training process, thereby enhancing the utility of SAM for specular reflection detection task. The results of specular reflection segmentation are detailed in “Reflection detection” section.

LaMa for reflective region inpainting

After the detection of endoscopic specular reflections, the subsequent step involves inpainting the reflected region using LaMa, a state-of-the-art inpainting technique [16]. Illustrated in Fig. 5, the LaMa process comprises the following equations:

$$i' = \text{stack}(i, m), \quad (4)$$

$$\hat{i} = f_{\theta}(i'), \quad (5)$$

$$\mathcal{L}_{\text{final}} = \kappa \mathcal{L}_{\text{Adv}} + \alpha \mathcal{L}_{\text{HRFPL}} + \beta \mathcal{L}_{\text{DiscPL}} + \gamma R_1. \quad (6)$$

Initially, a 3-channel endoscopic image i and a 1-channel reflection mask m are stacked to form a 4-channel input i' . Subsequently, the feed-forward inpainting network $f_{\theta}(\cdot)$, which encompasses downscale, fast Fourier convolution

Fig. 3 Process for creating endoscopic specular reflection weakly labeled dataset

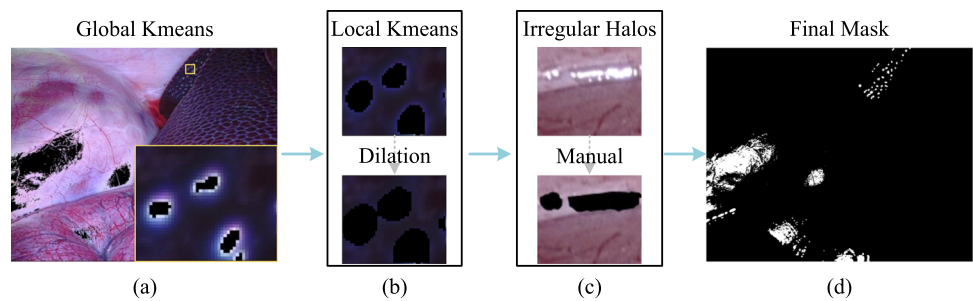


Fig. 4 SAM-adapter architecture consists of 4 modules: high-frequency components tune, patch embedding tune, adapter and SAM

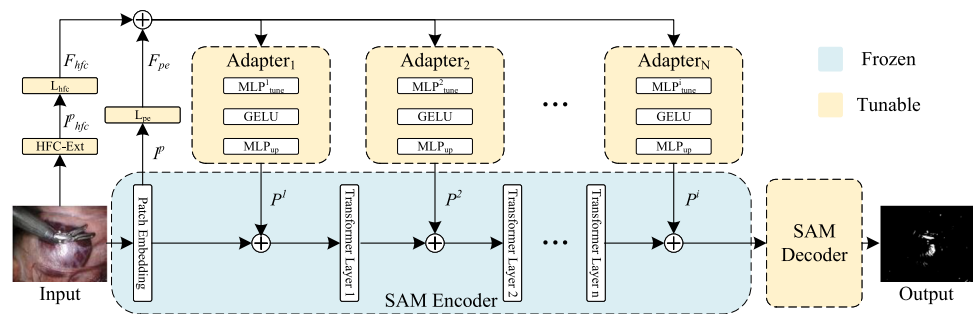
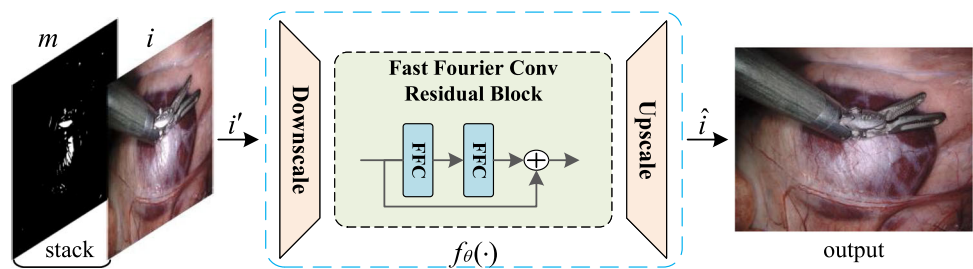


Fig. 5 LaMa architecture: mask and original image stacked as input to get reflection-free image



(FFC) [17], and upscale, processes the input i' in a fully convolutional manner, yielding the inpainted 3-channel image \hat{i} . Finally, the network parameters are inferred and optimized based on the $\mathcal{L}_{\text{final}}$ loss. Here, \mathcal{L}_{Adv} and $\mathcal{L}_{\text{DiscPL}}$ contribute to generating natural-looking local details, $\mathcal{L}_{\text{HRFPL}}$ oversees the global structure and supervised signal consistency, R_1 is the gradient penalty, and κ , α , β , γ are the weight values. Additional details about the LaMa inpainting model can be found in [16].

Given the absence of reflection-free endoscopic images, we address this challenge using a technique grounded in full transfer learning. Experimental findings reveal that the pre-trained model adeptly inpaints reflections in large regions effectively. The detailed inpainting results are presented in “Reflective region inpainting” section.

Optimization strategy

To address the limitations associated with weakly labeled datasets, we introduce an innovative dual pre-trained models iterative optimization strategy (DPMIO). The optimization process, detailed in Fig. 6’s pseudo-code, begins with the original image. SAM-adapter produces the reflection mask,

and, with the original, LaMa generates an inpainted image. This inpainted image is then iteratively fed back into SAM-adapter for mask refinement and LaMa for updated inpainting until specified criteria are met. The optimization strategy employs parameters μ ($1.5e-4$) and iter (5).

Despite its simplicity, this optimization strategy proves highly effective in refining inpainting results. Illustrated in Fig. 6, this strategy progressively enables the detection of challenging reflection regions, such as smaller or lighter reflections, leading to a more natural and plausible inpainting outcome. Ablation experiment is presented in “Reflective region inpainting” section.

Experiments

Implementation

The proposed method, EndoSRR, was implemented using PyTorch on an NVIDIA RTX 3090 GPU. The algorithm by Arnold et al. [6] was implemented in MATLAB R2021b on a system equipped with an AMD Ryzen 7 6800H 3.20 GHz processor, while the AdaRPCA [7] and AANC-RPCA

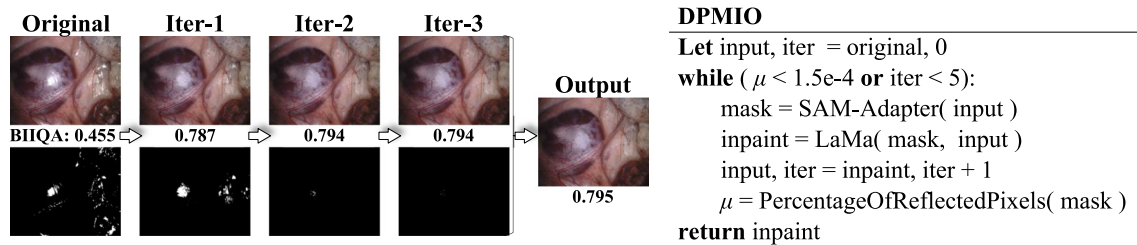


Fig. 6 Optimization process for reflection removal result

[1] algorithms were implemented using C++ on QT Creator 10.0.2. It is noteworthy that, apart from the aforementioned three endoscopic specular reflection removal algorithms, none of the other algorithms has released or shared their available implementations.

In the reflection detection stage, all modules of SAM-adapter are tunable, excluding the SAM encoder, which remains frozen, as illustrated in Fig. 4. ViT-B served as the pre-training parameter, AdamW as the optimizer, binary cross-entropy (BCE) and IoU loss as the loss functions, with a learning rate of $2e-4$, and the model underwent fine-tuning for 300 epochs. The Big-LaMa pre-trained model was employed for inpainting the reflection region, and κ , α , β , and γ were set to 10, 30, 100, and 0.001, respectively.

Datasets

Utilizing all the keyframes from the SCARED-2019 [18] dataset, we annotated the specular reflection dataset. Datasets 1 to 7 constitute the training set with 70 images, while datasets 8 and 9 form the testing set comprising 20 images. In the reflection detection stage, images are resized from 1280×1024 to 1024×1024 to match the SAM model, while in the inpainting stage, the original image size is retained.

Results and comparisons

Reflection detection

Quantitative comparison

Table 1 demonstrates that our proposed reflection detection method surpasses other methods in segmentation evaluation metrics, including accuracy, IoU, precision, and E-measure. The E-measure is defined as $E\text{-measure} = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h \phi_{FM}(x, y)$, where ϕ_{FM} is the enhanced alignment matrix, and h and w are the height and width of the map, respectively. Additional information on the E-measure can be found in [19]. In comparison to the state-of-the-art method AANC-RPCA [1], the proposed method achieves a higher IoU of 0.5888 compared to 0.5223. IoU's advantage lies in its ability to penalize both false negatives and false

positives, indicating that the proposed method effectively addresses the challenges of over- and under-segmentation, resulting in more accurate segmentation outcomes.

Qualitative comparison

As depicted in Fig. 7, the method proposed by Arnold et al. [6] excels in detecting smaller reflection regions but struggles with larger reflection regions, contributing to an under-segmentation problem. Conversely, AdaRPCA [7] and AANC-RPCA [1] tend to misclassify lighter-colored organ tissues as reflections, resulting in an over-segmentation problem. In contrast, the proposed method demonstrates a balanced approach, mitigating the challenges of both under- and over-segmentation.

Reflective region inpainting

Quantitative comparison

For a meaningful quantitative comparison, each method conducts inpainting on identical non-reflective regions, and the inpainting results are assessed using peak signal-to-noise ratio (PSNR), structured similarity indexing method (SSIM), and mean square error (MSE) metrics with the original image as a reference. Table 2 illustrates that the proposed method outperforms other methods significantly across all evaluation metrics. This superiority is attributed to the robust generalization ability of the model, enabling the combination of local and global information for optimal restoration of missing information in the image.

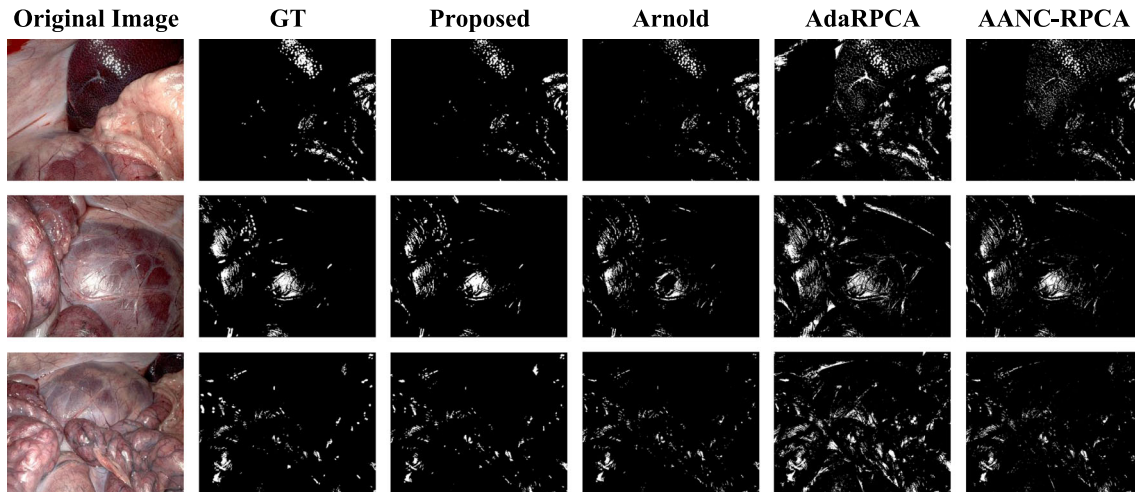
We further evaluated reflective region inpainting using the blind image inpainting quality assessment metric (BIIQA) [20], emphasizing local feature continuity. BIIQA is defined as $BIIQA = \alpha \bar{Q}_e + \beta \bar{Q}_t + \gamma \bar{Q}_s$, where α , β , and γ represent the percentage of edge patches, texture patches, and smooth patches, respectively. Additionally, \bar{Q}_e , \bar{Q}_t , and \bar{Q}_s denote the mean values of the edge, texture, and smooth scores, respectively. Table 3 confirms our method's superior performance. In the ablation experiments, our optimization strategy (DPMIO) improves inpainting (BIIQA: 0.686 to 0.693) but extends runtime (0.92s to 2.61s). While real-time capability is a limitation, our method remains the quickest.

Table 1 Quantitative comparison of reflection detection

Method	Accuracy	IoU	Precision	E-measure
Arnold et al. [6]	0.974 ± 0.019	0.401 ± 0.148***	0.836 ± 0.147	0.788 ± 0.147***
AdaRPCA [7]	0.958 ± 0.021***	0.403 ± 0.098***	0.497 ± 0.151***	0.857 ± 0.122**
AANC-RPCA [1]	0.976 ± 0.010	0.522 ± 0.069*	0.699 ± 0.111***	0.945 ± 0.044
Proposed	0.981 ± 0.012	0.589 ± 0.085	0.836 ± 0.061	0.947 ± 0.051

Bold values indicates the best performance among different metric

*, **, and *** indicate p value less than 0.05, 0.01, and 0.001, respectively

**Fig. 7** Qualitative comparison of reflection detection between the proposed method with Arnold et al [6], AdaRPCA [7] and AANC-RPCA [1]**Table 2** Quantitative comparison of inpainting results with reference assessment metrics

Method	PSNR	SSIM	MSE
Arnold et al. [6]	33.18 ± 2.64***	0.965 ± 0.023***	39.02 ± 31.95***
AdaRPCA [7]	26.33 ± 2.91***	0.944 ± 0.038***	200.11 ± 186.93***
AANC-RPCA [1]	31.79 ± 3.23***	0.951 ± 0.036***	61.88 ± 70.35**
Proposed	42.47 ± 4.00	0.990 ± 0.008	6.06 ± 7.22

Bold values indicates the best performance among different metric

** and *** indicate p value less than 0.01 and 0.001, respectively

Qualitative comparison

For qualitative comparison, all methods perform inpainting on the specified reflection regions. As shown in Fig. 8, the interpolation-based method [6] and the low-rank decomposition method [1, 7] exhibit ineffective inpainting of the reflection region. The inpainting traces of these methods are

conspicuous, and the results appear blurred. In contrast, the proposed method yields results that seamlessly blend with the background texture, presenting a more natural appearance while effectively minimizing the loss of organ texture information. As depicted in the final row of Fig. 8, our method exhibits a limitation in effectively addressing subtle

Table 3 Quantitative comparison of inpainting results with non-reference assessment metric

Method	Arnold et al. [6]	AdaRPCA [7]	AANC-RPCA [1]	wo/DPMIO	w/DPMIO
BIIQA	0.664 ± 0.081	0.60 ± 0.153*	0.630 ± 0.134	0.686 ± 0.067	0.693 ± 0.065
Time(s)	25.84 ± 26.09***	9.48 ± 9.10**	5.99 ± 0.58***	0.92 ± 1.23***	2.61 ± 1.34

Bold values indicates the best performance among different metric

*, **, and *** indicate p value less than 0.05, 0.01, and 0.001, respectively

Fig. 8 Qualitative comparison of reflective region inpainting between the proposed method with Arnold et al [6], AdaRPCA [7] and AANC-RPCA [1]

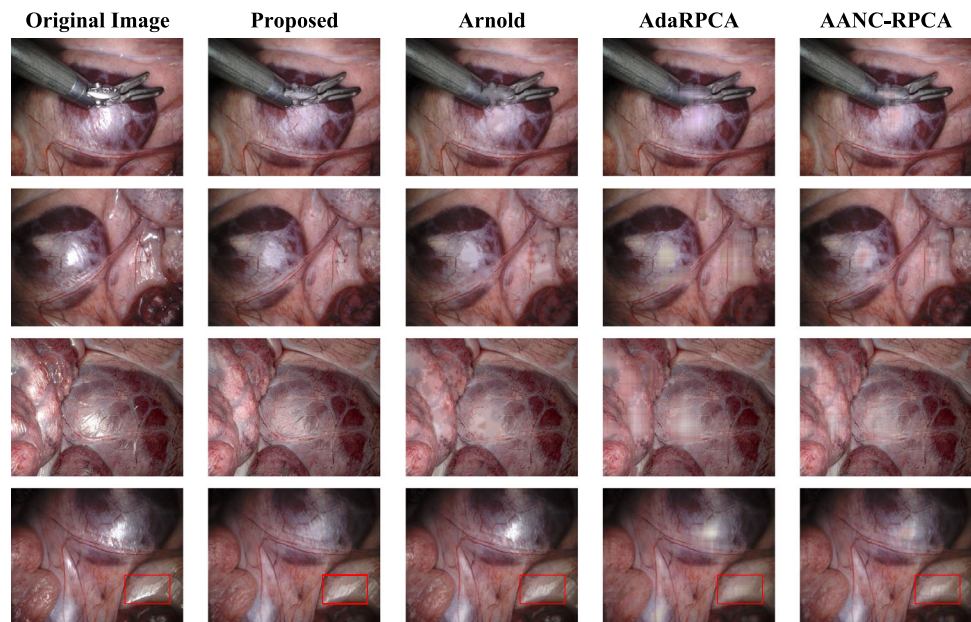
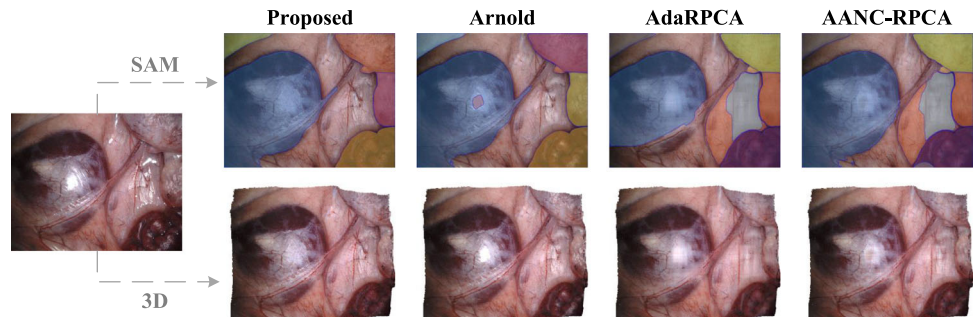


Fig. 9 Endoscopic specular reflection removal for Segmentation and 3D-Visualization



and faintly reflective regions, presenting a notable challenge for future improvements.

Application of specular reflection removal

Reasonable and natural specular reflection removal results are more helpful for downstream tasks. As shown in Fig. 9, it helps to enhance the segmentation accuracy of SAM across diverse tissue regions, as well as to improve the 3D visualization effect for better application in VR or AR based surgical navigation systems.

Discussion and conclusion

In this paper, we introduce EndoSRR, a novel algorithm for endoscopic specular reflection removal empowered by a large-scale model. Our approach begins with the creation of a weakly labeled dataset using a semi-automatic labeling tool. Subsequently, fine-tuning of the SAM-adapter accurately detects reflective regions. The reflective areas are then inpainted and optimized through a combination of the

state-of-the-art inpainting technique LaMa and a proposed optimization strategy. We validate the significant benefits of effective reflection removal for advancing downstream tasks and mitigating intraoperative visual fatigue in segmentation applications and 3D visualization applications. Our contributions include:

- **Creation of weakly labeled dataset:** We introduce a weakly labeled dataset, addressing the scarcity of endoscopic specular reflection datasets. This contribution is poised to advance the deep learning domain in the specular reflection removal task.
- **Optimization strategy:** We propose a simple yet effective optimization strategy, enhancing the naturalness and texture realism of the reflection removal results. This strategy, with its potential for application to similar tasks, serves as a valuable contribution.
- **Big model and transfer learning:** Our work pioneers specular reflection removal on a small-scale dataset by leveraging big models and transfer learning, resulting in significantly improved removal results. This approach

is particularly informative for the data-starved medical field.

Despite achieving superior results in both reflection detection and reflection region inpainting compared to state-of-the-art methods, our proposed EndoSRR method has certain limitations:

- **Color and texture restoration:** The algorithm struggles to fully restore the real color and texture information of the image, a challenge shared by existing methods in the field.
- **Limited weakly labeled datasets:** Due to the complexity and distribution scattering of endoscopic specular reflections, our weakly labeled datasets are limited in number. Further improvements in segmentation results and rigorous quantitative evaluation are necessary.
- **Real-time performance:** The algorithm does not achieve real-time performance, necessitating optimization and enhancements for practical use.

In conclusion, endoscopic specular reflection removal remains a formidable challenge. This work aims to contribute to the ongoing development of this field, ultimately enhancing the performance of computer vision downstream tasks and advancing the safety of surgical procedures. In forthcoming research endeavors, our aim is to expand the dataset of reflection masks, enabling precise localization of reflection regions in every frame of the video. This augmentation will facilitate the utilization of temporal information across various frames within the video sequence, enhancing the ability to restore the authentic texture and color details of the reflection regions.

Acknowledgements The present study was supported in part by the National Natural Science Foundation of China (No.62172401), the Guangdong Natural Science Foundation (Nos. 2022A1515010439 and 2022A0505020019), the Shenzhen Key Basic Research Grant (No. JCYJ20220818101802005), the Zhuhai Science and Technology Program (No. ZH22017002210017PWC), the Shenzhen Key Laboratory Program (No.ZDSYS201707271637577), and the Education Science Planning Program of Guangdong Department of Education (Higher Education Special Project) (No. 2022GXJK325).

Declarations

Conflict of interest The authors declare that there are no conflicts of interest with regard to this study.

Ethical approval The data used in this paper are a public dataset.

Informed consent The data used in this paper are a public dataset.

References

1. Pan J, Li R, Liu H, Hu Y, Zheng W, Yan B, Yang Y, Xiao Y (2023) Highlight removal for endoscopic images based on accelerated adaptive non-convex RPCA decomposition. *Comput Methods Programs Biomed* 228:107240. <https://doi.org/10.1016/j.cmpb.2022.107240>
2. Funke I, Bodenstedt S, Riediger C, Weitz J, Speidel S (2018) Generative adversarial networks for specular highlight removal in endoscopic images. In: *Medical imaging 2018: image-guided procedures, robotic interventions, and modeling*, vol 10576. SPIE, Houston, pp 8–16. <https://doi.org/10.1117/12.2293755>
3. Oh J, Hwang S, Lee J, Tavanapong W, Wong J, de Groen PC (2007) Informative frame classification for endoscopy video. *Med Image Anal* 11(2):110–127. <https://doi.org/10.1016/j.media.2006.10.003>
4. Alsaleh SM, Aviles AI, Sobrevilla P, Casals A, Hahn JK (2016) Adaptive segmentation and mask-specific sobolev inpainting of specular highlights for endoscopic images. In: *EMBC. IEEE, Lake Buena Vista*, pp 1196–1199. <https://doi.org/10.1109/EMBC.2016.7590919>
5. Wang X, Li P, Yongzhao D, Lv Y, Chen Y (2019) Detection and inpainting of specular reflection in colposcopy images with exemplar-based method. In: *ASID. IEEE, Xiamen*, pp 90–94. <https://doi.org/10.1109/ICASID.2019.8925202>
6. Arnold M, Ghosh A, Ameling S, Lacey G (2010) Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *EURASIP J Image Video Process* 2010:1–12. <https://doi.org/10.1155/2010/814319>
7. Li R, Pan J, Si Y, Yan B, Hu Y, Qin H (2019) Specular reflections removal for endoscopic image sequences with Adaptive-RPCA decomposition. *IEEE Trans Med Imaging* 39(2):328–340. <https://doi.org/10.1109/TMI.2019.2926501>
8. Monkam P, Wu J, Lu W, Shan W, Chen H, Zhai Y (2021) Easyspec: automatic specular reflection detection and suppression from endoscopic images. *IEEE Trans Comput Imaging* 7:1031–1043. <https://doi.org/10.1109/TCI.2021.3112117>
9. Ali S, Zhou F, Bailey A, Braden B, East JE, Lu X, Rittscher J (2021) A deep learning framework for quality assessment and restoration in video endoscopy. *Med Image Anal* 68:101900. <https://doi.org/10.1016/j.media.2020.101900>
10. Chwyl B, Chung AG, Wong A, Clausi DA (2015) Specular reflectance suppression in endoscopic imagery via stochastic bayesian estimation. In: *ICIAR. Springer, Canada*, pp 385–393. https://doi.org/10.1007/978-3-319-20801-5_42
11. Daher R, Vasconcelos F, Stoyanov D (2023) A temporal learning approach to inpainting endoscopic specularities and its effect on image correspondence. *Med Image Anal* 90:102994. <https://doi.org/10.1016/j.media.2023.102994>
12. Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y (2023) Segment anything model for medical image analysis: an experimental study. *Med Image Anal* 89:102918. <https://doi.org/10.1016/j.media.2023.102918>
13. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollar P, Girshick R (2023) Segment anything. *arXiv preprint arXiv:2304.02643*. <https://doi.org/10.48550/arXiv.2304.02643>
14. Liu W, Shen X, Pun CM, Cun X (2023) Explicit visual prompting for low-level structure segmentations. In: *CVPR*, pp 19434–19445
15. Hendrycks D, Gimpel K (2016) Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. <https://doi.org/10.48550/arXiv.1606.08415>
16. Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, Silvestrov A, Kong N, Goka H, Park K, Lempitsky V (2022) Resolution-robust large mask inpainting with Fourier convolutions. In: *WACV*, pp 2149–2159

17. Chi L, Jiang B, Mu Y (2020) Fast Fourier convolution. *Adv Neural Inf Process Syst* 33:4479–4488
18. Allan M, Mcleod J, Wang C, Rosenthal JC, Hu Z, Gard N, Eisert P, Fu KX, Zeffiro T, Xia W, Zhu Z, Luo H, Jia F, Zhang X, Li X, Sharan L, Kurmann T, Schmid S, Sznitman R, Psychogyios D, Azizian M, Stoyanov D, Maier-Hein L, Speidel S (2021) Stereo correspondence and reconstruction of endoscopic data challenge. arXiv preprint [arXiv:2101.01133](https://arxiv.org/abs/2101.01133), <https://doi.org/10.48550/arXiv.2101.01133>
19. Fan DP, Gong C, Cao Y, Ren B, Cheng MM, Borji A (2018) Enhanced-alignment measure for binary foreground map evaluation. In: *IJCAI*, pp 698–704, <https://doi.org/10.24963/ijcai.2018/97>
20. Rezki AM, Serir A, Beghdadi A (2022) Blind image inpainting quality assessment using local features continuity. *Multimed Tools Appl* 81(7):9225–9244. <https://doi.org/10.1007/s11042-021-11872-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.