Research Article Vol. 16, No. 11 / 1 Nov 2025 / Biomedical Optics Express 4243
Biomedical Optics EXPRESS

Deep-learning-assisted disposable bronchoscopy with enhanced image quality

ANG LI,[1] MING LEI,[1] ZACHARY BURNS,[1] GEORGE CHENG,[2] AND ZHAOWEI LIU[1,*]

[1]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA
[2]UC San Diego Health, San Diego, CA 92103, USA
[*]zhaowei@ucsd.edu

Abstract: The COVID-19 pandemic underscored the need for safer and more cost-effective bronchoscopy solutions, leading to increased interest in single-use bronchoscopes due to their elimination of cross-contamination risk. However, their widespread adoption has been limited by inferior image quality compared to the reusable counterparts. Addressing this need, we present in this study a deep-learning framework specifically designed to enhance the imaging performance of single-use bronchoscopes, improving the signal-to-noise ratio and image resolution while maintaining real-time processing capability. Our approach integrates a generative adversarial network architecture with attention mechanisms that focus on airway assessment. We also introduce a combined loss function to close the quality gap between reusable and disposable bronchoscopes. Additionally, we propose Fourier spectra comparison as a performance metric to evaluate the neural network performance on improving real bronchoscopy videos in the absence of ground truth images. The neural network model was initially trained and evaluated on a simulated dataset, achieving an 89% improvement in PSNR and a 16% increase in SSIM. When applied to real bronchoscopy videos, the trained model demonstrated a 1.8-fold increase in resolution based on the Fourier-spectra metric. The reconstruction speed was also computed for images of varying sizes, confirming the network's feasibility for real-time applications with large fields of view. This work not only presents a scalable AI solution that transforms disposable bronchoscopes into competitive alternatives for patients with respiratory disease but also establishes new methodological directions for evaluating and optimizing highly specialized medical imaging tasks in resource-constrained settings.

1. Introduction

Bronchoscopes are widely used medical devices that enable physicians to examine the trachea and bronchial subsegments, facilitating the diagnosis and management of various respiratory conditions including hemoptysis, airway obstructions and lung cancers [1,2]. The bronchoscopy procedure is particularly vital for critically ill patients, as the routine examinations are often an integrated part of their clinical care. One of the major challenges of traditional bronchoscopy is the risk of cross-infections, which imposes rigorous requirements on the reprocessing of used bronchoscopes [3]. A systematic review published in 2019 reported a 2.8% cross-contamination rate associated with reusable bronchoscopes in international bronchoscopy procedures [4]. In contrast, a recent study examining bronchoscopy procedures conducted shortly after the emergence of COVID-19 indicated an 8.69% cross-contamination rate based on publicly available data [5]. During the pandemic, the cross-contamination issue garnered increased attention and many pulmonary associations such as AABIP have recommended using single-use bronchoscopes for patients with suspected COVID conditions [6]. Despite their enhanced safety, portability and cost-effectiveness, the use of disposable bronchoscopes is still limited due to their inferior image quality compared to reusable bronchoscopes, specifically in terms of image resolution, noise level

#568633 https://doi.org/10.1364/BOE.568633
Journal © 2025 Received 29 May 2025; revised 28 Aug 2025; accepted 15 Sep 2025; published 2 Oct 2025

Biomedical Optics EXPRESS

and imbalanced color contrast [4,7,8]. The resolution of disposable bronchoscopes is constrained by the small number of camera pixels and the small lens size. These hardware limitations are for two reasons. First, the overall size of the bronchoscope probe must be compact enough to safely navigate human tubular organs and suction channels, minimizing the risk of potential injuries for patients. Second, the disposable bronchoscopes are designed to be cost-effective, maintaining their affordability and making them more competitive compared to the reusable models. To avoid increasing hardware dimensions or costs, we propose a deep-learning-based approach to enhance the image quality of disposable bronchoscopes to a level comparable to reusable bronchoscopes. While most deep-learning applications in endoscopy focus on tasks such as segmentation and polyp detection [9–13], some studies have demonstrated the efficacy of neural networks for resolution enhancement [14–17]. However, these techniques have not been applied to single-use bronchoscopes, as prior work primarily targets high-end endoscopes, which are typically diffraction-limited systems with relatively high signal-to-noise ratios (SNR). In contrast, single-use bronchoscopes have pixel-limited resolution, poor dynamic range, significant noise and imbalanced color contrast due to their small, inexpensive optics and detector. While these hardware constraints make the neural network reconstruction challenging, they also open vast opportunities for future improvement especially for medical images that require high accuracy.

Among various deep-learning models, the Generative Adversarial Network (GAN) has shown significant promise in super-resolution tasks due to its unique framework where two neural networks are trained to compete against each other to achieve final convergence. The first super-resolution GAN model was proposed by Ledig et al [18], along with the introduction of a perceptual loss function that has been implemented in numerous studies, especially in the field of medical images [19–24]. A widely adopted modification is the use of an encoder-decoder structure with an attention block as the generator [22,23]. This improved GAN architecture was later implemented by Almalioglu et al. for the image quality enhancement of capsule endoscopy images and proved to be superior compared to other convolutional networks [24]. Therefore, it is reasonable to assume that GANs can also assist with bronchoscopy, a type of endoscopy targeting the respiratory system.

Despite the promising results of GANs in other medical imaging contexts, we should still be very careful when generalizing the conclusions drawn from previous studies to the single-use bronchoscopes for several reasons. First, many studies have trained their models on open-source endoscopy datasets [14–17,23]. Endoscopes developed for diverse purposes, such as bronchoscopes and colonoscopes, can have significantly different optical systems characterized by distinct point spread functions (PSFs) and pixel sizes. Their different physical parameters imply that the blurring kernel learned by a neural network from one specific type of endoscope may not be translatable without introducing additional artifacts. Secondly, neural networks being trained on a specific endoscopy procedure can only learn the features significantly related to that procedure. For example, diseased tissues in GI endoscopy have very different types, shapes and sizes compared to bronchoscopy, and this mismatch may cause artifacts or amplification of noise rather than accurate features if we directly apply the models trained on endoscopy datasets to bronchoscopy images. Thirdly, many studies did not report the reconstruction speed of their trained networks. While their proposed neural networks showed impressive resolution and SNR improvements, complicated network architectures can lead to increased reconstruction time, making them impractical for real-life endoscopy procedures. Real-time speed is particularly important in all endoscopy procedures because patients undergo anesthesia, which can introduce additional medical risks with prolonged overall procedure times caused by delayed image processing.

To address the aforementioned concerns, we adapted and modified the GAN architecture from previous work, implementing it with bronchoscopy-targeted loss function and trained on a large set of bronchoscopy video frames to extract and amplify clinically relevant features. The trained

Biomedical Optics EXPRESS

neural network was evaluated not only on the simulated training dataset, but also on the real images captured by the disposable bronchoscopes, ensuring that it can be effectively integrated into the post-processing workflow of single-use bronchoscopes.

2. Methods

2.1. Comparison between reusable and disposable bronchoscopes

Compared to other types of endoscopy, bronchoscopy presents a unique imaging challenge in that both high-contrast, large-scale features, such as the layered, branching airway structures, and low-contrast, fine features including blood vessels and subtle mucosal irregularities are equally critical for accurate navigation and diagnostic assessment. These requirements highlight the demand for high image quality, and thus underscore the limitations of disposable bronchoscopes compared to their reusable counterparts.

Figure 1(a) and (b) illustrates the differences for images captured by the single-use Olympus BCV1-02 H-SteriScope and the reusable Olympus BF-1TH190 video-bronchoscope, using an exemplary single frame from a similar location within the human bronchi. Clearly, the image from the reusable bronchoscope shows much better resolution, clarity, color balance, and contrast. The image from the disposable bronchoscope, shown in Fig. 1(a), displays a strong red hue. A comparison of the red channels for both images is presented in Fig. 1(c) and (d). Saturation artifacts in the red channel of the disposable bronchoscope image are a major limitation, leading to reduced color contrast and less distinction among red features, such as major blood vessels and inflammatory tissues. While the red channel was highlighted here because it most clearly demonstrates the loss of perceived contrast, all three color channels were used in the neural network training to ensure that no spectral information was overlooked. Figure 1(e) and 1(f) present the corresponding Fourier spectra for the grayscale version of each image. The Fourier spectrum comparison was proposed as an evaluation metric because the pixel sizes of the detectors in these bronchoscopes are not disclosed by the manufacturers and obtaining images from the two bronchoscopes at exactly the same location with identical angles, fields of view, and lens distortions is not feasible. These spectra are calculated using MATLAB's fast Fourier transform function and presented in log-scale. The radii of the areas containing frequencies with high intensities on the spectra are measured and compared. The cut-off frequency of the reusable bronchoscope is approximately 1.82 times greater than that of the disposable bronchoscope, providing a reasonable estimate of the resolution difference even though variations in contrast, perspective, and depth may slightly affect this estimation. In addition, images captured by the disposable bronchoscope suffer from low color contrast at over- and under-exposed regions, often accompanied with significant noise. Therefore, we have considered all these factors when generating the training dataset.

2.2. Formulation of super-resolution algorithms

Super-resolution techniques involve the reconstruction of high-resolution images I^{HR} from their degraded low-resolution counterparts I^{LR}. The degradation process is described by [25]:

$$I^{LR} = (I^{GT} \star k)_{\downarrow s} + n. \tag{1}$$

It includes the spatial convolution ($\star$) between the ground truth object I^{GT} and the point spread function k that characterizes the integral effect of the imaging system including diffraction, dispersion, distortion etc. The under-sampling process $\downarrow$s resulting from the detector's limited pixel size, number of pixels and additional noise n introduced by the recording system. For single-use bronchoscopes, the image quality is primarily affected by under-sampling, making it a pixel-size limited system. Additionally, the noise component also plays a significant role in the degradation process.

Biomedical Optics EXPRESS

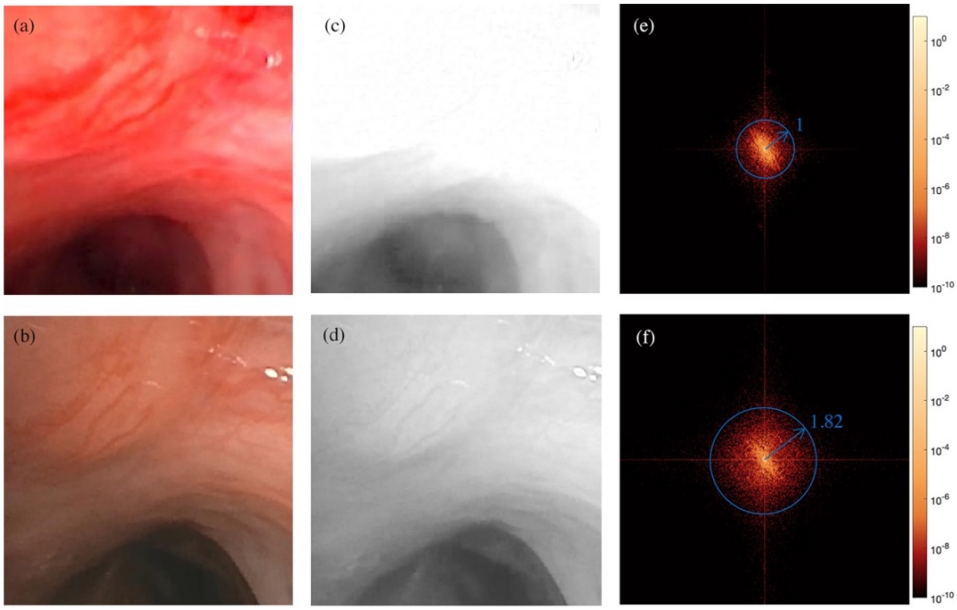

Fig. 1. Comparison between single-use and reusable bronchoscopes. (a), (b) shows the bronchoscopy image captured by a disposable bronchoscope and a reusable bronchoscope, respectively, at a similar location. (c) and (d) compares the red-channel image of (a) and (b), while (e) and (f) compares the Fourier spectra of (a) and (b), respectively. The Fourier spectra are presented in logscale for better contrast.

The inverse problem of this formation is usually ill-posed, as there can be multiple possible solutions of I^{GT} because background noise is often not negligible. Deconvolution-based algorithms are commonly employed to tackle this complex, nonlinear problem. However, they tend to be computationally intensive and time-consuming, or fail to achieve the resolution required for medical applications. Therefore, deep-learning neural networks are preferred for the task. The reconstruction objective of the deep-learning-based approach is to restore the object information I^GT as accurately as possible, formulated as:

$$\theta = arg_{\theta} min(Loss(I^{Recon}, I^{GT}) + \lambda\Phi(\theta)). \tag{2}$$

where Loss (I^{Recon}, I^{GT}) denotes the loss function that guides the convergence of neural network training, Φ represents the regularization function, and λ is the tuning parameter to control the degree of regularity. Therefore, choosing an appropriate loss function is crucial for training a neural network [25,26].

2.3. Neural-network construction

The neural network architecture is adapted and modified from Almalioglu et al. [24], with fewer convolutional blocks in the middle of the U-net to make it more lightweight for both the training and the real-time application. Figure 2 illustrates the details of our GAN architecture, which includes a generator and a discriminator. The generator is responsible for producing high-resolution reconstructions utilizing a U-Net architecture. This architecture has proven effective across various super-resolution tasks in both diffraction-limited and pixel-limited imaging systems [27–29]. It integrates contextual information and retrieves precise details through an encoding-decoding process combined with layer concatenation. Two-layer convolutional blocks are a major component of this design, extracting image features at different levels. During down-sampling,

Biomedical Optics EXPRESS

the convolutional blocks increase the number of image channels, while the Max Pooling layer reduces the size of the feature map. After several rounds of down-sampling, the feature maps reach the bottleneck and are then up-sampled using bilinear interpolation. Each up-sampling step concatenates the output with the corresponding down-sampled block at the same level, thereby integrating both shallow and deep features to generate the final reconstruction. Additionally, after the first convolutional block, a spatial attention layer is inserted to prioritize the most significant features, as demonstrated in [23,24]. The discriminator, on the other hand, employs several convolutional blocks to extract deep image features and uses two fully connected layers to classify the images as either real or generated by the network. Its output is then incorporated into the loss function to further refine the generator.

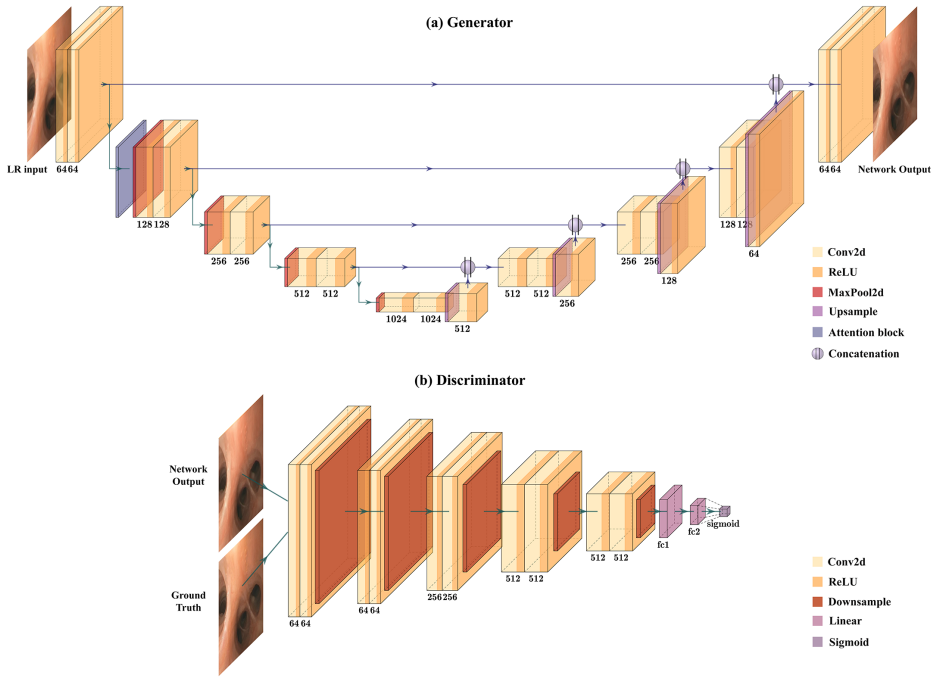

Fig. 2. GAN Architecture consists of (a) the U-net based generator with an attention block, which will pass its output to (b) the CNN-based discriminator for the computation of discriminator loss. All convolutional layers use a 3×3 kernel, with the desired output channels indicated below each layer.

We implemented our loss function with three components: pixel loss, perceptual loss and adversarial loss. The pixel loss, or the mean-squared-error (MSE), is crucial for guiding the neural network to generate images that are numerically close to the ground truth image. It is defined as the sum of pixel-wise differences between the ground-truth image (I^{GT}) and the neural network reconstructed image (I^{Recon}), formulated as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(I^{GT}(i) - I^{Recon}(i))^2. \tag{3}$$

The perceptual loss evaluates the similarity between the feature maps of the generated image and the ground truth image, ensuring the consistency in the high-level features of these two images. The feature extraction is generally achieved with the pre-trained Visual Geometry Group

Biomedical Optics EXPRESS

(VGG) network, formulated as:

$$Loss_{\frac{VGG}{i}j} = \frac{1}{{L_{i,j}}^2} \sum_{x=1}^{L_{i,j}} \sum_{y=1}^{L_{i,j}} (VGG_{i,j}(I^{HR})_{x,y} - VGG_{i,j}(I^{Recon})_{x,y})^2. \quad (4)$$

where L represents the dimension of the feature map and i, j specifies the number of convolutional layers being used from the pre-trained network architecture [30].

While the pre-trained VGG network has proven to be an excellent feature extraction tool for super-resolution tasks, it was originally trained on real-world images with compact details and high contrast. Given that bronchoscopy images often exhibit sparser features and lower contrast, we re-trained the VGG model on our bronchoscopy videos to address potential discrepancies between the training datasets and the specific characteristics of bronchoscopy images. The re-trained VGG model achieved a classification accuracy of over 99% in distinguishing between images captured by reusable and disposable bronchoscopes, demonstrating its proficiency in guiding the training of our GAN model as part of the loss function. We utilized all the layers before the fully connected layer, or the classifier, as the feature-extraction framework integrated into the perceptual loss function.

Adversarial loss is a critical component that makes GANs particularly effective in visual tasks. While the generator learns to produce images similar to the ground truth, the discriminator learns to differentiate between generated images and real images. The probability output by the discriminator—indicating whether an image is generated or real—is incorporated into this loss component to guide the generator in refining its images:

$$Loss_D = \sum_{i=1}^{N} -\log(D(I^{Recon})). \quad (5)$$

We also integrated the SSIM loss from the Pytorch Image Quality package to force the neural network to learn the overall structure and texture of our images. Eventually, the loss functions are combined with different weights assigned to each component, formulated as:

$$Loss_{total} = MSE + \alpha Loss_{VGG/i,j} + \beta Loss_D + \gamma Loss_{SSIM}. \quad (6)$$

The initial weights (α, β) are adopted from previous study [13] and then finely adjusted based on experimental results to optimize the Peak Signal-to-noise Ratio (PSNR) and the Structural Similarity Index (SSIM) until optimal performance is observed. The final best values selected based on our tests with different combinations of parameters are α=0.3, β=0.01, and γ=0.1.

2.4. Performance metrics

The performance of the GAN model will be evaluated and optimized on the training data based on several performance metrics including the quantitative and perceptual evaluation of the image quality improvement, and the image reconstruction speed. Commonly used quantitative metrics are PSNR and SSIM, which measures the structural and perceptual differences between the network outputs and the ground truths, defined by [31]:

$$PSNR = 10 \cdot log_{10}\left(\frac{max(I)^2}{MSE}\right). \quad (7)$$

$$SSIM(I^{Recon}, I^{HR}) = f(l(\cdot), c(\cdot), s(\cdot)). \quad (8)$$

Mathematically, the PSNR measures the maximum signal power relative to the noise level, which assumes the noise to be the absolute difference between the generated image and the ground truth. The SSIM evaluates the texture changes between the images with a function f(·) that takes into account multiple factors including luminance l(·), contrast c(·) and structure s(·), which are functions of I^{Recon} and I^{HR}.

Biomedical Optics EXPRESS

Beyond quantitative measurements, we also evaluate whether the trained model recovers depth-related information. Specifically, we use the MiDaS DPT-BEiT-Large model, a deep-learning-based reference-free depth estimator. Trained with mixed datasets on diverse natural images, the model can predict relative depth from a single image and has demonstrated strong cross-dataset generalization [32].

For the real bronchoscopy images that don't have corresponding ground truths, we propose two other metrics to estimate the resolution improvement of the neural network-reconstructed images by computing their Fourier spectra and comparing the information abundance, similar to the analysis shown in Fig. 1(e) and (f). This metric enables us to quantify the information abundance in the images and the additional high-frequency details revealed after enhancement, especially since the actual resolution of the bronchoscope is unknown without detailed knowledge of the hardware.

While the Fourier spectrum analysis targets resolution improvement, we also evaluate the improvement with non-reference methods: the Natural Image Quality Evaluator (NIQE) and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). These metrics are widely used for objective distortion assessment in clinical imaging and have supported contrast evaluation in X-ray studies [33]. NIQE estimates image quality from image statistics, so we fit the model to the reusable bronchoscopy images and then use it to evaluate the single-use bronchoscope images and their neural-network reconstructions. BRISQUE employs machine-learning models trained to capture a broad range of spatial distortions based on human-assigned scores. Together, NIQE and BRISQUE complement the Fourier analysis by reflecting perceptual improvements.

Lastly, we will report the reconstruction speed of our trained network while processing images of different sizes, which provides different fields of view for the physicians in clinical settings, to ensure the feasibility of real-time performance. For the single-use bronchoscope to serve as a viable substitute for reusable models, it must be capable of achieving real-time display.

2.5. Training dataset generation

In this study, the dataset used to train the neural network comprises three thousand data pairs, each consisting of a high-resolution bronchoscopy image captured with a high-end reusable bronchoscope as the ground truth, and a simulated low-resolution counterpart that mimics the image captured by a disposable bronchoscope in a low SNR setting.

Bronchoscopy image frames are carefully selected from six bronchoscopy videos, ensuring minimal motion blur and representation of key features. The images are preprocessed to smooth out the specular highlights, which are bright, sharp spots that can mislead the neural network during the training process, resulting in artifacts in the final reconstruction. These dots can be identified by checking for saturation across all color channels at each pixel. They are then smoothed using a blurring kernel, which averages the values of the surrounding pixels to reduce the intensity of these artifacts. The pre-processed images are used as the high-resolution ground truth. Their low-resolution counterparts are simulated by down-sampling followed by intensification of the red channel which produces color histograms similar to the images captured by disposable bronchoscopes. These low-resolution images are then scaled back to the same size as the high-resolution ones using bilinear interpolation, creating image pairs of identical dimensions that can be fed into the U-net architecture. We experimented with several down-sampling ratios and eventually selected a 6x down-sampling ratio as it most reasonably simulated images captured by disposable bronchoscopes. A smaller resolution difference between the image pairs would limit the neural network's ability to capture sufficient details, while a greater resolution difference could lead to overfitting and the amplification of image artifacts. Finally, Gaussian noise with varying signal-to-noise ratios is added to the low-resolution images to more accurately reflect the characteristics of disposable bronchoscopes.

Biomedical Optics EXPRESS

2.6. Neural network training setup

The SRGAN model has been initially trained on three thousand bronchoscopy frames, with a training-to-validation ratio of 8:2. For the initial training phase, we adopted network parameters from related work and our preliminary experiment, including an unbalanced learning rate of $1e^{-4}$ for the generator and $2e^{-5}$ for the discriminator, and a loss function introduced in the section 2.4. A batch size of 6 was chosen to balance the trade-off between image size and batch size within the constraints of limited GPU memory. Maintaining relatively large image size is important for preserving some macroscopic structural details of the bronchi, which are essential for detection of airway abnormalities. The training was performed on a computer with a NVIDIA RTX6000 Ada Generation graphics card.

3. Experiments and results

3.1. Performance evaluation on the simulation dataset

The GAN model was trained on the generated dataset for 50 epochs, while PSNR and SSIM converged to PSNR = 33.66 and SSIM = 0.97. Examples of the neural network prediction on the test subset of the training dataset is shown in Fig. 3, with the first column showing the low-resolution images, the second showing the GAN reconstruction, and the third showing the corresponding high-resolution ground truth images. The first example shows a PSNR increase from 15.68 to 28.83 and a SSIM increase from 0.70 to 0.96, while the second example demonstrates a PSNR improvement from 20.48 to 34.12 and a SSIM improvement from 0.93 to 0.99. Figure 3(d)–(f) shows the MiDaS depth maps corresponding to Fig. 3(a)–(c). The depth maps estimated from the reconstructions closely match those of the high-resolution images, whereas the low-resolution inputs exhibit clear deviations such as flattened structure and inconsistent depth gradients. These observations suggest that the model preserves and recovers relative depth information within the images.

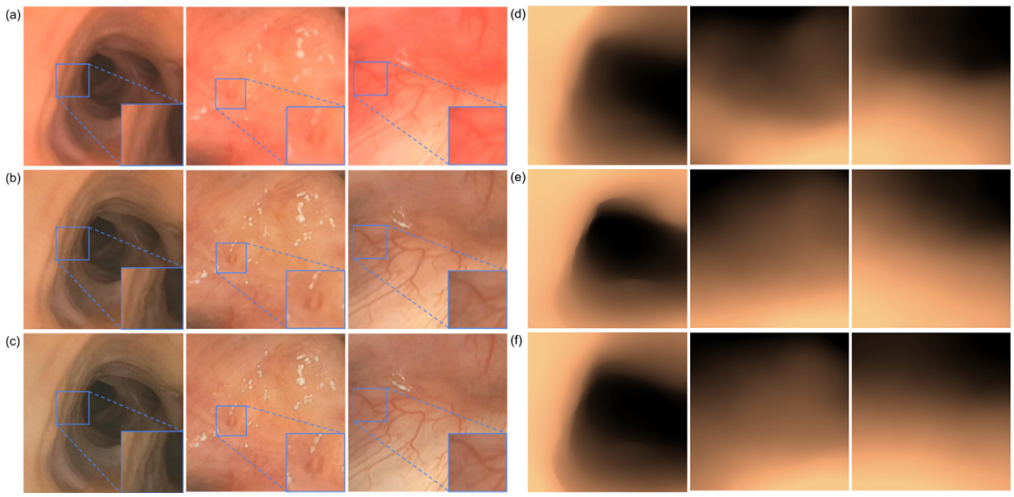

Fig. 3. Demonstration of neural network performance on the simulated dataset: (a) three examples low-resolution inputs; (b) the corresponding neural network reconstructions; (c) the ground-truth images. (d)-(f) MiDaS relative-depth maps corresponding to (a)-(c), respectively. Brighter colors indicate shallower depth.

Figure 4 presents the loss convergence of the training procedure, as well as the statistical distributions of PSNR and SSIM improvements on the test subset of the simulation dataset. The test subset contains 150 low-resolution images that involve features from high contrast tubular

Biomedical Optics EXPRESS

structures, and low contrast mucosal surfaces, with a wide variety of feature density and level of blurriness due to different focusing. The overall PSNR improvement is reported as 89% ± 19%, and the SSIM improvement is 16% ± 6%. These quantitative results indicate that deep-learning approaches hold significant promise for enhancing the image quality of bronchoscopy images. It can also be observed that the trained network is able to extract and amplify both macroscopic and microscopic features effectively.

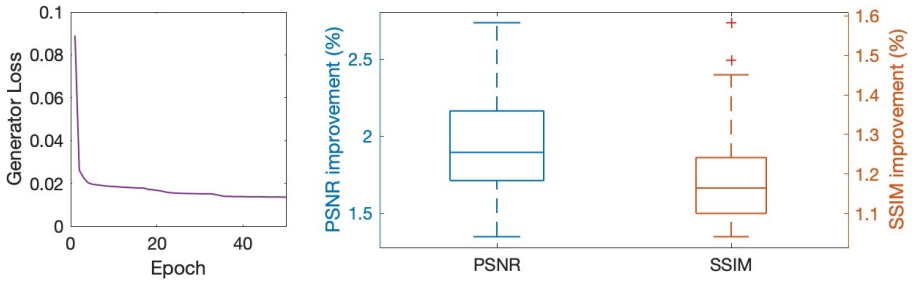

Fig. 4. Left: Generator loss convergence during the training process. Right: statistics of PSNR and SSIM improvement on the test subset of the simulation dataset.

3.2. Performance evaluation on real bronchoscopy images

The trained model is then applied to bronchoscopy images captured by the single-use bronchoscope. An example demonstrating its performance is shown in Fig. 5, with the first and third row displaying the images captured by a single-use bronchoscope and the second and forth row presenting the corresponding reconstructed images produced by the trained network. It can be observed from Fig. 5(b) and (c) that the enhanced images exhibits significantly improved color contrast, making detailed variations in the surrounding tissues more visible. Two pairs of zoomed-in comparison of the blue-marked areas are shown in Fig. 5(e)&(f) and (g)&(h), displaying the sharpened edges and overall morphology of the bronchi entry. Since no ground truths exist for these images, PSNR and SSIM cannot be directly computed for quantitative evaluation of the reconstruction. Therefore, we assessed image quality improvement with two other metrics: Fourier spectra analysis and non-reference image quality evaluation. The expansion of the high-frequency components in Fig. 5(j), compared to (i), as well as (l) compared to (k), both demonstrate approximately 1.8-fold enhancement in image quality, which is consistent with Fig. 1. This also indicates that the image quality of the deep-learning assisted disposable bronchoscope is comparable to that of a high end reusable bronchoscope.

The two no-reference image quality metrics also support the improvement, as shown in Fig. 6. The NIQE and BRISQUE scores are computed for 30 single-use bronchoscopy images and their corresponding reconstructions. The reconstructed images achieved lower scores than the single-use inputs on both metrics, indicating improved perceptual quality. Although contrast distortion in single-use bronchoscopy can reduce the magnitude of NIQE/BRISQUE enhancement [34], these metrics nevertheless indicate less image distortions in the reconstructed images, supporting our conclusion about the neural-network's capacity of improving image quality.

In addition, we compared the results of our reconstruction with other models trained on general endoscopy datasets to support our assumption that, despite some similarities in task, significant differences in optical systems and clinically relevant features make those models non-generalizable to the specific task of enhancing single-use bronchoscope images – which has low SNR, imbalanced color contrast, and pixel-limited detection system. Based on the reconstruction details shown in Fig. 7(b)-(c), it can be observed that models trained on high-resolution, low-noise endoscopy datasets often fail to resolve useful information from images

Biomedical Optics EXPRESS

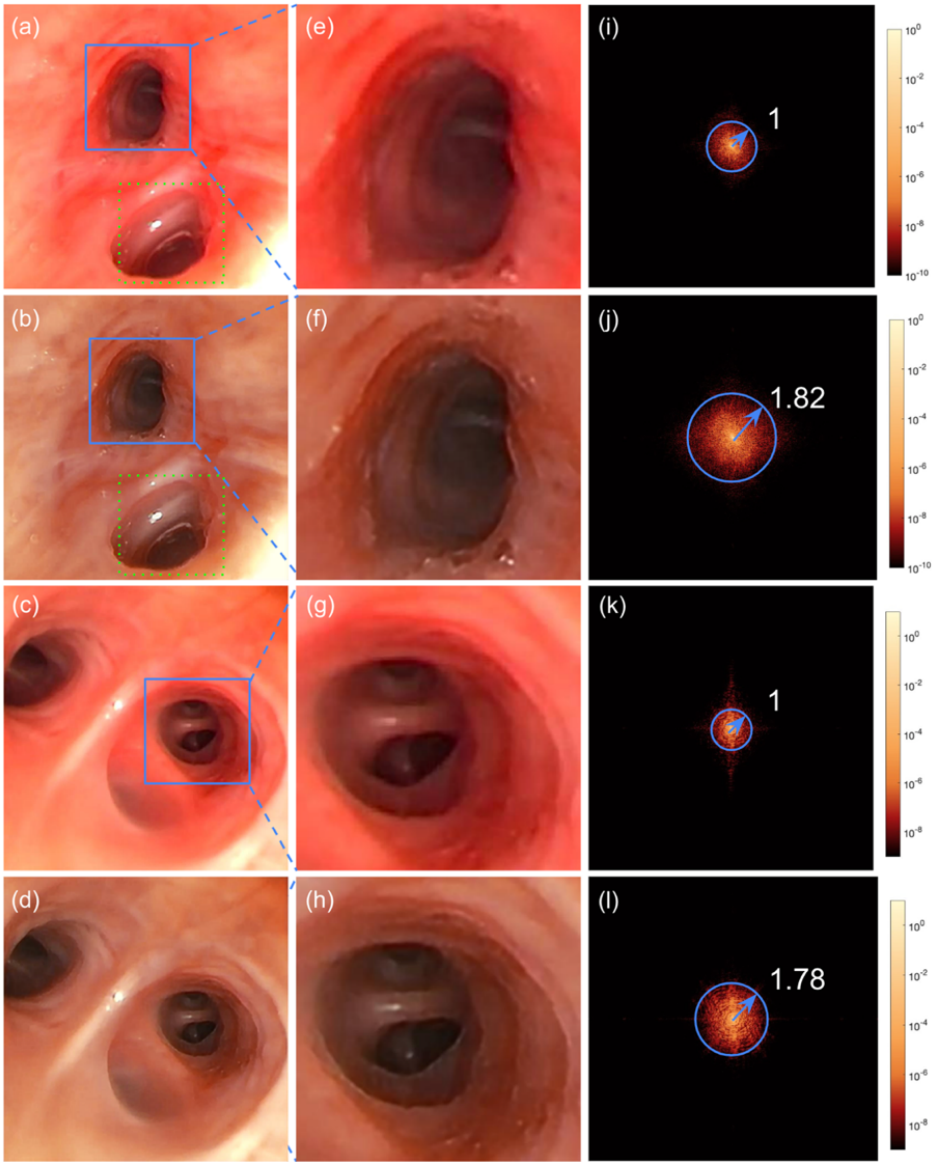

Fig. 5. Examples of improved image quality of single-use bronchoscope based on neural network reconstruction. (a), (c): The original single-use bronchoscopy images versus (b), (d): the neural network reconstructions with improved image quality. (c)-(h): Comparisons of the zoomed blue-box areas in (a)-(d). (i)-(l): The respective Fourier spectra of (a)-(d). The numbers are the ratio of radius increment after image quality improvement.

Biomedical Optics EXPRESS

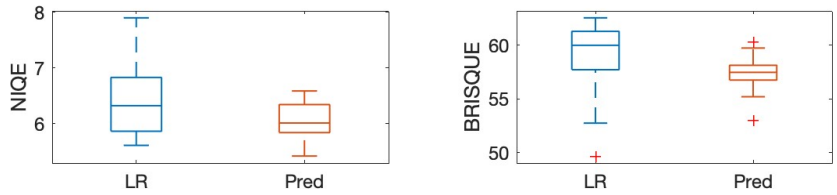

Fig. 6. Comparison of the NIQE and BRISQUE scores among the low-resolution disposable bronchoscope images (LR) and their corresponding predicted high-resolution images (Pred).

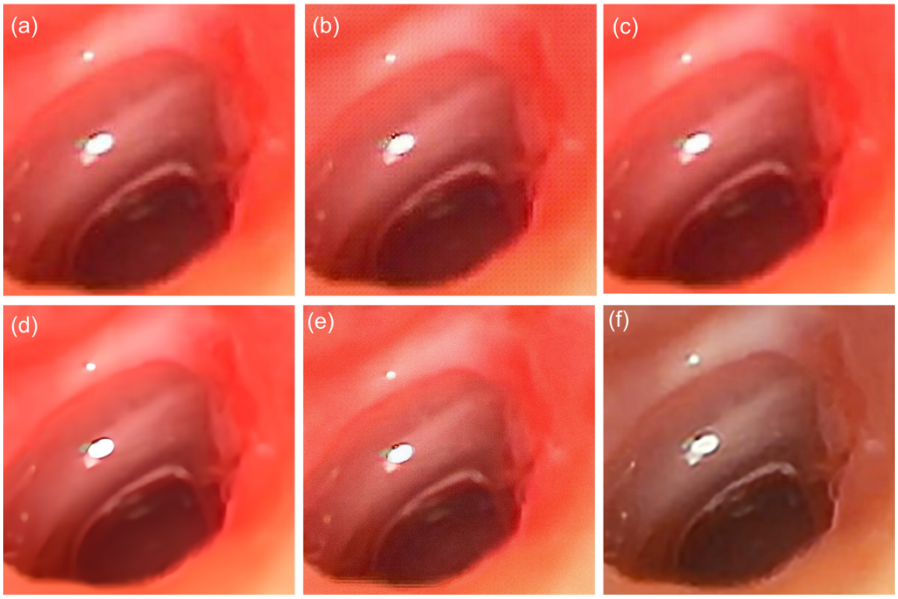

Fig. 7. Comparison of other neural networks with our specialized model for disposable bronchoscopy. (a) Region cropped from the green-dotted box in Fig. 5: low-resolution input from the single-use bronchoscope. Reconstructions by (b) EndoL2H [24], (c) DCSSR [15], (d) SwinIR [35], (e) LDM-4 [36], and (f) our model.

captured by single-use bronchoscopes. This may be due to the distraction caused by a saturated red channel or the presence of features with very low contrast. These models tend to extract and intensify high-frequency components, which can highlight the edge of polyps or diseases in other types of endoscopies or biopsy procedures, but can also lead to exaggerated noise in low-quality input images. Additionally, we extended the comparison to state-of-the-art super-resolution approaches with different architectures. Specifically, we evaluated a Transformer-based model (SwinIR) [35] and a diffusion-based model (LDM-4) [36]. As shown in Fig. 7(d–e), both models demonstrated apparent resolution improvement. However, SwinIR tends to over-smooth tissue texture, and LDM-4 introduces high-frequency noise and other artifacts. They also failed to provide sufficient contrast for the layered structure on the right side, which can be clearly presented in our reconstruction due to its explicit handling single-use bronchoscopes' distorted contrast. Therefore, to make disposable bronchoscopes a safe and cost-effective option, we need to integrate them with specifically designed and trained neural network model into the post-processing workflow.

Finally, we compared the reconstruction speed across models. For 1024×1024 images, the average reconstruction times were approximately 0.25 s for SwinIR and 8.5 s for LDM-4 on an

Biomedical Optics EXPRESS

NVIDIA RTX6000 GPU, both insufficient to support real-time bronchoscopy. In contrast, our model required only ~60 ms per frame on the same GPU, corresponding to a video frame rate of ~16 fps. Splitting the computational workload across two GPUs in parallel can increase the reconstruction speed to about 30fps and achieve real-time display. By cropping the image to 512×512, similar to the frame size of the disposable bronchoscope, the reconstruction speed increases to 67 fps, though this comes with a reduced field of view. With better GPU capabilities, such as increased GPU memory or multiple GPUs running in parallel, real-time image quality improvement are achievable while maintaining a large field of view. Despite the trade-off, the reported reconstruction speeds have demonstrated that our neural network is capable of supporting real-time applications.

4. Conclusion and future work

In this study, a deep-learning-based super-resolution approach was developed specifically for enhancing the image quality of single-use bronchoscopes. We employed a U-net-based GAN architecture to address the challenges posed by the low resolution and low signal-to-noise ratio of these bronchoscopes. To guide the network effectively, a hybrid loss function with a bronchoscopy-targeted feature-extraction framework was implemented. It combines pixel-wise, perceptual, and adversarial loss components to ensure accurate reconstruction by enhancing high-frequency features while maintaining natural image details.

The effectiveness of our model was demonstrated through both qualitative and quantitative analyses. Image quality improvements are perceptually evident in terms of clarity and contrast, with enhanced visibility of detailed features in bronchoscopy images. Quantitatively, our model achieved approximately 89% improvement in PSNR and 16% improvement in SSIM on the simulation dataset, demonstrating its ability to significantly enhance image quality. When applied to real images from a single-use bronchoscope, the neural network demonstrated approximately a 1.8-fold increase of image resolution based on the Fourier spectra analysis indicating significant improvement in reveling image details. Non-reference metrics also suggested less image distortions in the reconstructed images.

Although this study is limited by not directly assessing the impact of the enhanced images on clinical diagnosis, evidence from other medical domains strongly suggests that improved resolution supports better disease detection. In ultrasound, deep-learning based super-resolution technique has enabled recovery of vascular profile for lesion differentiation and characterization [37]; in MRI, better image quality helped improve the segmentation accuracy of highly heterogeneous brain tumors [38]. Building on this precedent, we plan to address the limitation of our study in future work through clinical validation studies focused on disease detection using enhanced bronchoscopy images. We will also explore the model's potential in AI-enabled automatic disease identification, differentiation, and segmentation within bronchoscopy procedures. On the technical side, efforts will include optimizing the simulation algorithm for generating training datasets, including a more precise simulation of image characteristics by studying the hardware of both disposable and reusable bronchoscopes to better model their difference in pixel sizes and point spread functions. Additional efforts will be directed towards refining the model architecture and improving reconstruction speed.

Funding. Institute of Engineering in Medicine, UCSD.

Acknowledgment. The authors would like to thank the Galvanizing Engineering in Medicine (GEM) Program from the Institute of Engineering in Medicine (IEM) at UCSD for their grant support.

Disclosures. The authors declare no conflicts of interest.

Data availability. The code and example images used to test the neural network are available via GitHub Repository upon email request. Data used for neural network training process presented in this study were collected by Dr. George Cheng at UCSD Health. Due to the absence of explicit patient consent for public data sharing, the data are not publicly available at this time but may be obtained from the corresponding author upon reasonable request.

Biomedical Optics EXPRESS

References

1. J. H. Ahn, "An update on the role of bronchoscopy in the diagnosis of pulmonary disease," Yeungnam University Journal of Medicine **37**(4), 253–261 (2020).
2. T. S. Panchabhai and A. C. Mehta, "Historical perspectives of bronchoscopy: Connecting the dots," Annals of the American Thoracic Society **12**(5), 631–641 (2015).
3. D. Biondini, A. Pasini, A. Guarise, *et al.*, "The role of bronchoscopy in the diagnosis and management of patients with SARS-CoV-2 infection," Diagnostics **11**(10), 1938 (2021).
4. J. M. Mouritsen, L. Ehlers, J. Kovaleva, *et al.*, "A systematic review and cost-effectiveness analysis of reusable vs. single-use flexible bronchoscopes," Anaesthesia **75**(4), 529–540 (2020).
5. H. S. Travis, R. V. Russell, and J. Kovaleva, "Cross-contamination rate of reusable flexible bronchoscopes: A systematic literature review and meta-analysis," Journal of Infection Prevention **24**(3), 95–102 (2023).
6. E. Ho, A. Wagh, K. Hogarth, *et al.*, "Single-use and reusable flexible bronchoscopes in pulmonary and critical care medicine," Diagnostics **12**(1), 174 (2022).
7. N. Darrell, S. Grant, H. Abdurrahman, *et al.*, "Operator perception of the performance of multiple single-use bronchoscopes compared to standard re-usable bronchoscope," Am. J. Biomed. Sci. Res. **17**(2), 77–83 (2022).
8. A. M. Sweeney, G. Kavanagh, K. F. Deasy, *et al.*, "Single-use or disposable flexible bronchoscopy in advanced bronchoscopy procedures: Experience in a quaternary referral centre," Respiration **101**(10), 931–938 (2022).
9. Z. Ding, H. Shi, H. Zhang, *et al.*, "Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model," Gastroenterology **157**(4), 1044–1054.e5 (2019).
10. K. Pogorelov, A. Petlund, P. Halvorsen, *et al.*, "Deep learning and handcrafted feature-based approaches for automatic detection of angiectasia," *Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, USA, pp. 365–368, 2018.
11. D. Jha, S. Ali, N. K. Tomar, *et al.*, "Real-time polyp detection, localization, and segmentation in colonoscopy using deep learning," IEEE Access **9**, 40496–40510 (2021).
12. S.-L. Cai, W. Wang, H. Chen, *et al.*, "Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma," Gastrointestinal Endoscopy **90**(5), 745 (2019).
13. N. Rochmawati, C. Fatichah, B. Amaliah, *et al.*, "Deep learning-based lesion detection in endoscopy: A systematic literature review," IEEE Access **13**, 43532–43556 (2025).
14. X. Yang, Y. Chen, R. Tao, *et al.*, "Endoscopic image deblurring and super-resolution reconstruction based on deep learning," in *Proc. Int. Conf. Artif. Intell. Comput. Eng.* (ICAICE), 2020, pp. 168–172.
15. M. Hayat and S. Aramvith, "Saliency-aware deep learning approach for enhanced endoscopic image super-resolution," IEEE Access **12**, 83452–83465 (2024).
16. C. Zhou, K. Qiu, C. Chen, *et al.*, "Video super-resolution for wireless capsule endoscopy imaging sensor," IEEE Sens. J. **22**(17), 17283–17290 (2022).
17. J. Lin, G. Huang, J. Huang, *et al.*, "Quaternion attention multi-scale widening network for endoscopy image super-resolution," Phys. Med. Biol. **68**(7), 075012 (2023).
18. C. Ledig, L. Theis, F. Huszár, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), (2017), pp. 105–114.
19. M. Zhao, Y. Wei, and K. K. L. Wong, "A generative adversarial network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images," Magn. Reson. Imaging **85**, 153–160 (2022).
20. M. D. Robinson, S. J. Chiu, C. A. Toth, *et al.*, "New applications of super-resolution in medical imaging," in *Super-Resolution Imaging*, CRC Press, (2017), pp. 401–430.
21. W. Ahmad, H. Ali, Z. Shah, *et al.*, "A new generative adversarial network for medical images super resolution," Sci. Rep. **12**(1), 9533 (2022).
22. G. Gaál, B. Maga, and A. Lukács, "Attention U-Net based adversarial architectures for chest X-ray lung segmentation," arXiv (2020).
23. A. Sarvaiya, H. Vaghela, K. Upla, *et al.*, "Channel attention network for wireless capsule endoscopy image super-resolution," in H. Kaur, V. Jakhetiya, P. Goyal, P. Khanna, B. Raman, and S. Kumar, eds., *Computer Vision and Image Processing. CVIP 2023*, vol. 2010, Springer, 2024, pp. 581–594.
24. Y. Almalioglu, K. B. Ozyoruk, A. Gokce, *et al.*, "EndoL2H: Deep super-resolution for capsule endoscopy," arXiv (2020).
25. Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," IEEE Trans. Pattern Anal. Mach. Intell. **43**(10), 3365–3387 (2021).
26. M. K. Ng and N. K. Bose, "Mathematical analysis of super-resolution methodology," IEEE Signal Process. Mag. **20**(3), 62–74 (2003).
27. M. Lei, J. Zhao, J. Zhou, *et al.*, "Super resolution label-free dark-field microscopy by deep learning," Nanoscale **16**(9), 4703–4709 (2024).
28. T. Liu, K. de Haan, Y. Rivenson, *et al.*, "Deep learning-based super-resolution in coherent imaging systems," Sci. Rep. **9**(1), 3926 (2019).
29. O. Gulenko, H. Yang, K. Kim, *et al.*, "Deep-learning-based algorithm for the removal of electromagnetic interference noise in photoacoustic endoscopic image processing," Sensors **22**(10), 3961 (2022).
30. R. Sood, B. Topiwala, K. Choutagunta, *et al.*, "An application of generative adversarial networks for super resolution medical imaging," in *Proc. IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, 2018, pp. 326–331.

Biomedical Optics EXPRESS

31. Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process. **13**(4), 600–612 (2004).
32. R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 12179–12188 (2021).
33. H. M. S. S. Herath, H. M. K. K. M. B. Herath, N. Madusanka, *et al.*, "A systematic review of medical image quality assessment," J. Imaging **11**(4), 100 (2025).
34. E. Eldarova, V. Starovoitov, and K. Iskakov, "Comparative analysis of universal methods for no-reference quality assessment of digital images," J. Theor. Appl. Inf. Technol. **99**(9), 1977–1987 (2021).
35. J. Liang, J. Cao, G. Sun, *et al.*, "SwinIR: Image restoration using Swin Transformer," *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 1833–1840 (2021).
36. R. Rombach, A. Blattmann, D. Lorenz, *et al.*, "High-resolution image synthesis with latent diffusion models," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 10684–10695 (2022).
37. O. Bar-Shira, Y. Gotlib, R. Malka, *et al.*, "Learned super resolution ultrasound for improved breast lesion characterization," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021*, M. de Bruijne, eds., Lect. Notes Comput. Sci. 12907, 110–120 (Springer, Cham, 2021).
38. Z. Zhou, A. C. Bovik, and A. Pandey, "Super-resolution of brain tumor MRI images based on deep learning: Improved segmentation accuracy," Med. Phys. **23**(11), e13758 (2022).