Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 1990~2000
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp1990-2000

ClearNet: auto-encoder based denoising model for endoscopy images

Vikrant Shokeen[1], Sandeep Kumar[1], Vidhu Mathur[1], Amit Sharma[2], Indrajeet Gupta[3], Parita Jain[4]
[1]Department of Computer Science and Engineering, Maharaja Surajmal Institute of Technology, Delhi, India
[2]Department of Computer Science, IMS Engineering College, Ghaziabad, India
[3]Department School of Computer Science and AI, SR University, Warangal, India
[4]Department of Computer Science and Engineering, KIET Group of Institutions Ghaziabad, Delhi-NCR, India

Article Info

Article history:

Received Apr 5, 2024
Revised Dec 26, 2024
Accepted Feb 27, 2025

Keywords:

Disease classification
Encoder-decoder
Medical imaging
MRI
Transformers

ABSTRACT

Gastrointestinal (GI) endoscopy images play a crucial role in the detection and diagnosis of diseases within the digestive tract. However, the development of effective computer vision models for automated analysis and denoising of endoscopy images faces challenges arising from the diverse nature of abnormalities and the presence of image artefacts. In this work, the utilization of an encoder-decoder network for denoising GI endoscopy images using the HyperKvasir dataset has been analyzed. This approach involves training a custom encoder-decoder model on this extensive multi-class endoscopy image dataset and assessing its performance across 23 prevalent classes of digestive tract issues. Here experiments showcase the model's ability to learn robust visual representations from endoscopic data, enabling accurate disease prediction. The achieved promising results highlight the potential of encoder-decoder architectures as a foundational framework for computer-aided endoscopy analysis with a specific focus on denoising applications. Our model manages to increase the peak signal-to-noise ratio (PSNR) of original-noisy pair from 19.118954 to 69.892631 for original-reconstructed pair showcasing almost perfect reconstruction.

Corresponding Author:

Sandeep Kumar
Department of Computer and Engineering, Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi, India
Email: sandeep.jaglan@msit.in, san.jaglan@gmail.com

1. INTRODUCTION

Gastrointestinal (GI) endoscopy is a vital medical procedure allowing for the direct visualization of the digestive tract to screen and diagnose diseases. However, the quality of endoscopy images is often hampered by noise and artefacts, presenting a challenge for accurate diagnosis [1]. Unlike traditional classification models, this approach focuses on denoising endoscopy images obtained from the HyperKvasir dataset [2] which contains a diverse set of 10,662 images with 23 disease labels. GI diseases, including cancers, inflammation, and infections, impact millions globally each year.

The field of medical image analysis has seen rapid advancements with the advent of deep learning techniques, particularly convolutional neural networks (CNNs). These models have demonstrated remarkable success in various endoscopy image analysis tasks, from lesion detection to disease classification [3]-[5]. However, the majority of existing research has focused on classification and detection tasks, often overlooking the critical preprocessing step of image denoising.

Recent studies have begun to explore the potential of real-time artificial intelligence (AI) image processing systems for the diagnosis of upper GI malignancies Vania *et al.* [6]. These advancements contribute to the ongoing development of computer-aided diagnosis (CAD) systems for GI endoscopy, promising to enhance the accuracy and efficiency of disease detection and classification.

The primary contribution of this paper is the introduction of ClearNet, an innovative auto-encoder based denoising model specifically designed for endoscopy images. Our approach leverages the power of deep learning, particularly the encoder-decoder architecture inspired by recent advancements in vision transformer networks, to effectively remove noise and enhance image quality. By focusing on the crucial preprocessing step of denoising, ClearNet aims to improve the overall performance of subsequent analysis tasks, such as disease classification and lesion detection.

2. RELATED WORKS

Numerous studies have delved into leveraging deep learning for automated analysis in endoscopy images, initially concentrating on binary classification and detection tasks. A method proposed by [7] utilizes JDPCA for lesion detection in GI endoscopic images, achieving high accuracy rates for various conditions, surpassing traditional methods. He *et al.* [8] employs multiple pre-trained CNNs to improve classification of lesion types in colonoscopy images. For polyp detection, [9] indicates increased adenoma detection rates with AI computer-aided detection during colonoscopy but also highlights higher rates of unnecessary polyp removal. A thematic survey [10] on medical image segmentation using deep learning offers insights into supervised and weakly supervised learning methods. Tanwar *et al.* [11] introduce a deep learning approach for colorectal polyp detection and classification with 92% accuracy. Borgli *et al.* [2] hyperKvasir dataset facilitates multi-class classification, and [12] review imbalance problems in object detection, but studies using this dataset [13] typically focus on single-label classification.

This work addresses the task of denoising endoscopy images, considering various noise levels, presenting a sophisticated deep learning model tailored for this problem. While transformers like DETR have gained popularity for detection and classification (16), our study pioneers the use of an encoder-decoder transformer-based model for multi-label endoscopy image denoising, showcasing strong performance and generalizability. Image denoising remains a crucial challenge across medical imaging modalities, as highlighted in [14] survey on noise reduction techniques in lung cancer computed tomography (CT) scan images. Litjens *et al.* [15] provide a comprehensive review of deep learning applications in medical image analysis, while CNNs, such as the one utilized by [16], have shown promise for denoising medical images. The study [17] proposes the design of deep feed-forward denoising CNNs using residual learning and batch normalization to improve medical image denoising performance, particularly for small sample sizes.

Encoder-decoder models, exemplified by Ronneberger *et al.* U-net [18] and Tahmid *et al.* conditional adversarial training [19], are commonly employed for image restoration tasks. Lehtinen *et al.* [20] propose a novel approach to signal reconstruction using machine learning, demonstrating the ability to restore images from corrupted examples without clean data or explicit statistical models of corruption. Transformers, initially developed for natural language processing, have recently been applied to medical image analysis, demonstrating their capability to model long-range dependencies in image data. The research by He *et al.* [21] highlights the potential of transformers in improving various aspects of medical imaging, including segmentation, classification, and denoising tasks. These advancements could significantly enhance the performance of our auto-encoder based denoising model by leveraging the attention mechanisms inherent in transformer architectures.

Generative adversarial networks (GANs) have shown remarkable success in image denoising tasks due to their ability to learn and generate high-quality images. Alsaiari *et al.* [22] utilized a GAN-based approach for image denoising, achieving significant improvements in image clarity and noise reduction. Unlike their GAN-based method, our approach employs an autoencoder-style U-net architecture, which focuses on learning an efficient representation of the input data to achieve effective denoising.

3. METHOD

In this section the methodology of our paper. This section talks about the dataset prepration and model used in detail.

3.1. Dataset description

HyperKvasir contains 10662 GI tract images labelled with 23 common classes. Figure 1 shows sample images from the dataset demonstrating the diversity of visual patterns. With a primary emphasis on GI endoscopy, the dataset covers diverse segments of the digestive tract, such as the oesophagus, stomach, and colon. The distinguishing feature of the HyperKvasir dataset is its inclusiveness, encompassing not only typical endoscopy images but also a comprehensive array of images displaying a wide range of anomalies and GI disorders, Figure 2 shows the distribution of classes in the dataset.

Figure 1. Sample images from dataset

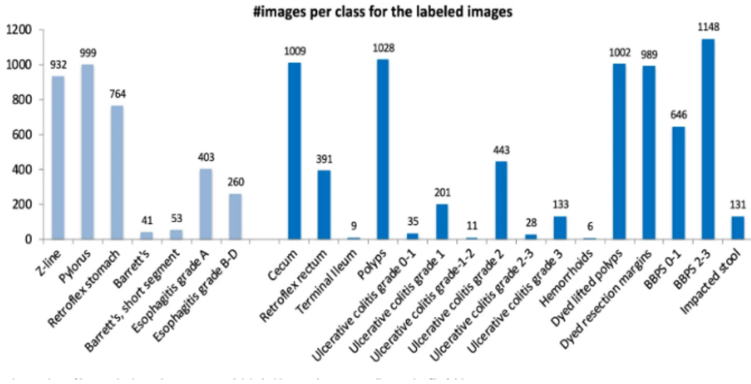

Figure 2. Dataset classes description

3.2. Model description

ClearNet is an encoder-decoder architecture inspired by recent vision transformer networks like DETR. The proposed model is tailored for denoising GI. As shown in Figure 3, the proposed model consists of an Inception-v3 encoder that extracts visual features, followed by a custom decoder for denoising. It has a U-Net-style architecture with some modifications. U-net is a CNN architecture that is commonly used for semantic segmentation, a task where the model must denoise the given image.

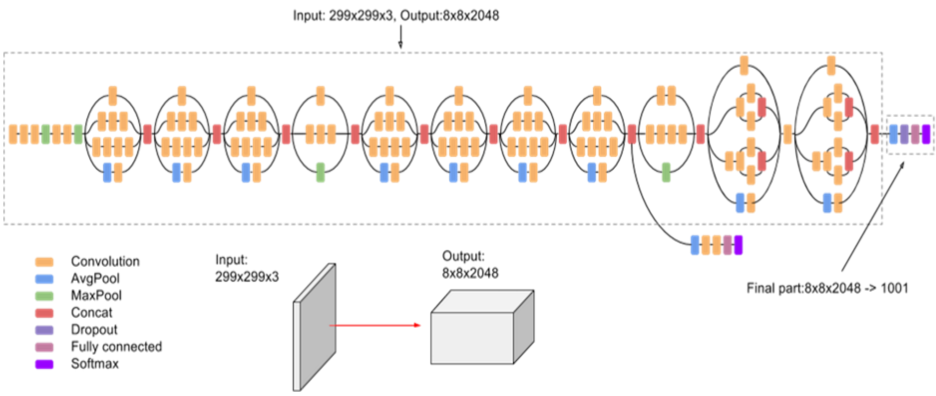

Figure 3. InceptionV3 architecture

DETR encoder-decoder: DETR (DEtection TRans- former) presented by [13] is a recent vision transformer model for object detection. It consists of a CNN encoder followed by a transformer decoder. The CNN encoder extracts feature representations from input images. The transformer decoder then performs denoising using these encoded features. A basic depiction of an autoencoder model is shown in Figure 4. The inspiration is drawn from DETR's overall encoder-decoder architecture for our endoscopy image denoising model. However, a custom decoder design tailored for the denoising of images is developed.

InceptionV3: we leverage a pre-trained Inception-v3 CNN as the encoder backbone, similar to DETR's use of a ResNet CNN backbone. Inception-v3 is an innovative CNN architecture proposed by [23]. The key feature of Inception-v3 is the Inception modules. These apply convolutional filters of different sizes in parallel and concatenate the outputs. Specifically, they use 1×1, 3×3, and 5×5 convolutions. This allows the network to capture multi-scale visual features from fine-grained patterns to abstract concepts.

The work hypothesizes that subtle texture changes in endoscopy images can indicate pathology, so this ability is significant. The addition of 1×1 convolutions also reduces dimensionality to improve computational efficiency compared to naive stacking of multi-size convolutions. Overall, the Inception-v3 architecture showed high performance on medical imaging tasks [24] while being optimized for efficiency.

Autoencoders (U-net): autoencoders represent a fundamental concept in deep learning, specifically designed for unsupervised learning tasks. The core idea behind autoencoders is to learn efficient representations of data, capturing essential features without explicit labels. Mathematically, the encoder operation $h=f_{decoder}(x)$ transforms the input x into a lower-dimensional representation, while the decoder operation $x\hat{}=f_{decoder}(h)$ reconstructs the input approximation x ̂.

$$h = \sigma(W\_encoder \cdot x + b\_encoder) \quad (1)$$

$$\hat{x} = \sigma(W\_decoder \cdot h + b\_decoder) \quad (2)$$

$$\mathcal{L}(x, \hat{x}) = MSE(x, \hat{x}) = 1/n \sum(x\_i - \hat{x}\_i)^2 \quad (3)$$

3.3. Architecture

A U-net style autoencoder network with a transformer-inspired structure with a CNN backbone for blind denoising endoscopy images is proposed. The CNN based autoencoder will work as depicted in Figure 5. The encoder leverages a pre-trained Inception-v3 model to extract hierarchical visual features from noisy inputs (24). Here's a representation of how a U-net works:

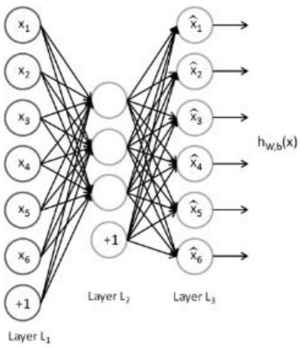

Figure 4. Basic autoencoder model

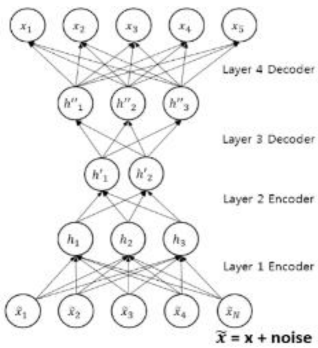

Figure 5. Convolutional autoencoder model

3.4. Algorithm

Here's a representation of how the U-net in our proposed architecture works:

1. Input layer:
   - X is the input image.
2. Contracting path:
   a. Apply convolutional layer with rectified linear unit (ReLU) activation:

$$Conv_i(X) = \sigma(Wi \cdot X + bi) \quad (4)$$

   b. Apply max-pooling layer:

$$Pool_i(X) = MaxPooling\big(Conv_i(X)\big) \tag{5}$$

c. Repeat the above steps to create a contracting path with multiple convolutional blocks and pooling layers. Save the intermediate feature maps:

$$Features_{contracting} = \{Conv_i, Pool_i\} \tag{6}$$

3. Bottleneck:
   a. Apply convolutional layer with ReLU activation:

$$Conv_{bottleneck}(X) = \sigma(Wb \cdot X + b_b) \tag{7}$$

   b. Perform up-sampling:

$$Upsample\big(Conv_{bottleneck}(X)\big) \tag{8}$$

4. Expansive path:
   a. Concatenate the feature maps from the contracting path:

$$Concat_{expansive} = Concat(Feat_{concat}, Upsample(Conv_{bottleneck}(X))) \tag{9}$$

   b. Apply convolutional layer with ReLU activation:

$$Conv_e\big(Concat_{expansive}\big) = \sigma\big(W_e * Concat_{expansive} + b_e\big) \tag{10}$$

   c. Apply up sampling:

$$Upsample\left(Conv_e\big(Concatenated_{expansive}\big)\right) \tag{11}$$

   d. Repeated the above steps to create an expansive path with multiple convolutional blocks and up sampling layers:

$$Features_{expansive} = \{Conv_e, Upsample_e\} \tag{12}$$

5. Output layer:
   - Apply convolutional layer with SoftMax activation to obtain the final de-noised image:

$$Y = SoftMax\big(W_{out} * Features_{expansive} + b_{out}\big) \tag{13}$$

Inception-v3, renowned for its multi-scale feature learning, employs parallel modules with convolutions of varying receptive field sizes, crucial for capturing both fine details and global context in medical imaging tasks. Its effectiveness on tasks with limited data makes it suitable for GI image denoising [25]. The decoder module utilizes transpose convolutions to upsample encoder embeddings for image prediction, aided by skip connections from encoder layers for local detail reconstruction. Our hybrid transformer-based U-net combines global reasoning with localization ability, benefiting denoising tasks. Optimization employs rms prop optimizer and mean squared error loss on paired real samples with simulated noise, enabling the model to handle diverse distortion types without explicit noise modeling. The flexible encoder-decoder scheme, coupled with localized skip connections, facilitates image restoration.

During inference, the trained network enhances noisy endoscopy images to minimize distortion, showcasing the synergy of CNN and transformer architectures in denoising. Thorough evaluation demonstrates significant improvement in tissue structure and lesion visualization amidst noise. The model (Figure 6) is trained end-to-end using MSE loss. Here is a breakdown of the architecture:

a. Input: the model takes a 256×256-pixel input image with 3 color channels red, green, and blue (RGB).
b. Inception V3: this model is pre-trained on the ImageNet dataset, which contains over 14 million images. The model has 10 layers and includes a variety of layers like convolutional layers, max-pooling layers, and batch normalization layers.
c. Convolutional and batch normalization layers: these layers are used for feature extraction and classification. After each convolutional layer, a batch normalization layer is applied to normalize the output.

$$Conv(X) = \sigma(W * X + b) \quad (14)$$

$$BatchNorm(X) = \frac{(X - \mu)}{\sqrt{(\sigma^2 + \varepsilon)}} \odot \gamma + \beta \quad (15)$$

d. Up-sampling layers: the up-sampling layers are used to upscale the output of the previous layer by a factor of 2. This process helps in obtaining the output of the same size as the input image.

$$UpSample(X) = Upscaling(X, factor = 2) \quad (16)$$

e. Convolutional layers: these layers are used for refining the output from the previous layer. They include 2D convolutional layers, each followed by a batch normalization layer.

$$Conv2D(X) = \sigma(W * X + b) \quad (17)$$

$$BatchNorm(X) = ((X - \mu))/\sqrt{(\sigma^\wedge 2 + \varepsilon)} \odot \gamma + \beta \quad (18)$$

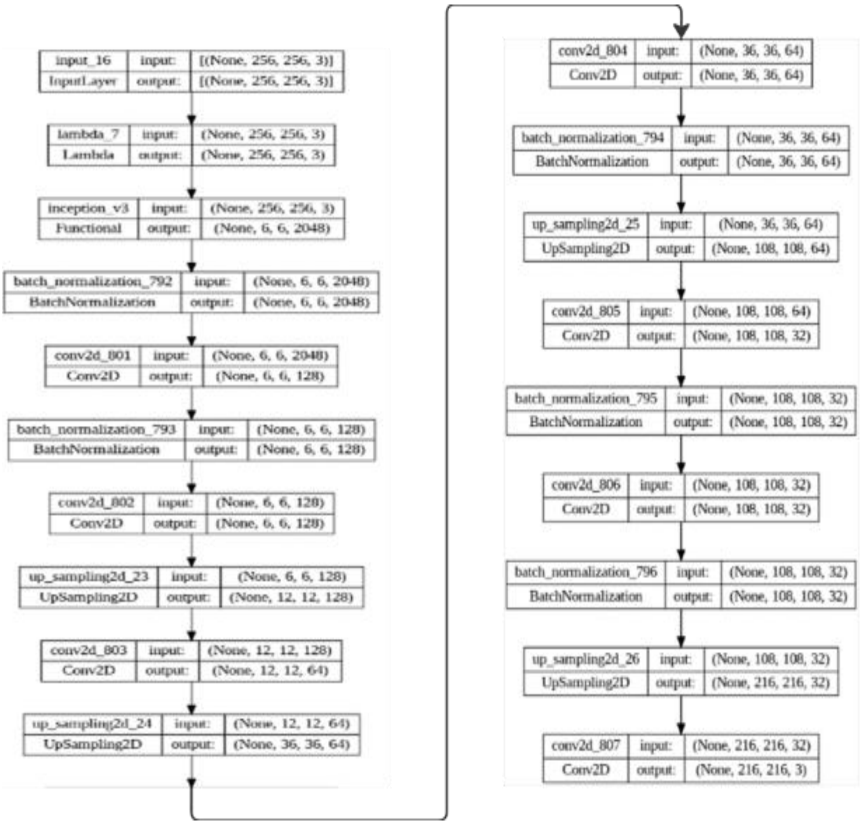

Figure 6. Proposed architecture (a-left to b-right)

3.5. Algorithmic expression for model

This encoder-decoder model can be explained in the following steps:
Overall model architecture: the overall model architecture can be represented as:

$$m = Encoder(x) + Decoder(z) \quad (19)$$

where x is the input image, z is the feature vector, and m is the output denoised image.

4. RESULTS AND DISCUSSION

The ClearNet is evaluated on the HyperKvasir dataset [2] with 10,662 images from 23 classes, split 80/20 into train and validation sets. Peak signal-to-noise ratio (PSNR), a metric quantifying reconstruction quality compared to the original, is used to assess denoising fidelity as shown in Figure 7. To rigorously evaluate noise robustness, specialized training and validation datasets are constructed.

The average PSNR of original and noisy images is 19.118954 and the PSNR of original and reconstructed image is 69.892631. The increase in PSNR from 19.12 dB to 69.89 dB shows that the noise reduction process was highly effective. The reconstructed image has far less noise and is much closer in quality to the original image. A PSNR above 50 dB generally indicates near-perfect reconstruction or an almost imperceptible difference between the original and the reconstructed image. The closer the PSNR is to 0 dB, the more dissimilar the noisy image is from the original, meaning more noise or distortion is present.

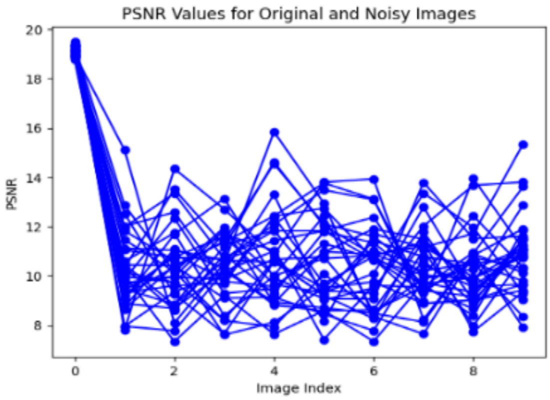

Figure 7. PSNR values

4.1. Generation of noisy images

To enable the training and evaluation of image-denoising models, custom datasets is generated by artificially corrupting images from the HyperKvasir GI endoscopy dataset. Specifically, random gaussian noise is added to the original images to simulate noisy acquisition conditions. Gaussian noise is statistically generated from a normal distribution with mean μ and variance σ^2. It is commonly used to model noise from natural sources like sensor noise. The formula for a gaussian noise distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{2(x-\mu)^2}{2\sigma^2}} \quad (20)$$

To rigorously evaluate noise robustness, specialized training and validation datasets are constructed. For the training set, gaussian noise with zero mean and 20% noise level is added to each image. The validation set uses a higher 30% noise level, 10% higher than training, to evaluate model generalization. This simulated noise generation provides a controllable way to create paired noisy and clean images for training denoising models. Exposing models to varying noise levels during training helps learn robust representations transferable to handling real-world noise and distortions, enabling rigorous evaluation and advancement of endoscopy image denoising techniques.

Using a noisier validation set rigorously tests the model’s generalization and robustness capabilities. If the model learns noise-invariant representations during training, it should handle higher noise levels at inference, preventing overfitting to the specific training noise level. Figures 8 and 9 show examples from the noisy training (20% noise) and validation (30% noise) sets, respectively, with noise manifesting as grainy speckles obscuring details and edges, presenting a major challenge for image processing algorithms. Figure 10 displays ClearNet’s denoised reconstructions for the noisy validation inputs from Figure 9, effectively filtering out noise and recovering cleaner images with restored key features and structures while suppressing spurious noise. Qualitatively, the model produces natural-looking reconstructions close to the original ground truth images, highlighting its denoising capabilities and ability to learn robust representations capturing real-world image structures, distinguishing signal from arbitrarily corrupted noise during testing. Quantitative analysis uses PSNR, a metric quantifying the quality of a reconstructed signal compared to its original version as shown in Figure 7, to assess the fidelity of denoised images. PSNR and perceptual similarity to ground truth validate the qualitative observations.

4.2. Generation of noisy images

For the training set, gaussian noise with zero mean and 20% noise level is added to each image. The validation set uses a higher 30% noise level to evaluate model generalization, 10% higher than training, enabling rigorous testing of robustness and preventing overfitting to the training noise level. Figures 8 and 9 show examples from the noisy training (20% noise) and validation (30% noise) sets, respectively, with the noise manifesting as grainy speckles obscuring details and edges, presenting a major challenge. Figure 10 displays the denoised reconstructions by ClearNet for the noisy validation inputs from Figure 9. ClearNet effectively filters out the noise, recovering cleaner images with restored key features and structures while suppressing spurious noise. Qualitatively, the model produces natural-looking reconstructions close to the original ground truth images, highlighting its denoising capabilities and ability to learn robust representations that capture real-world image structures, distinguishing signal from arbitrarily corrupted noise during testing. Quantitative analysis of reconstruction error and perceptual similarity to the ground truth validates these qualitative observations.

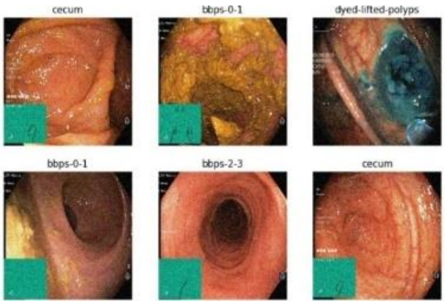

Figure 8. 20% noise images

Figure 9. 30% noise images

Figure 10. Reconstructed images

4.3. Discussion and limitations of research

Our model successfully increases the average PSNR value of the sample images from 19.12 dB to 69.89 dB, which showcases its effectiveness in denoising the image. The PSNR value of 69.89 dB shows that the reconstructed images are very similar to the original images. This denoising has potential to become a very important part of medical diagnosis of GI diseases. This research also paves way for automated systems which can be created to help doctors in classifying diseased better and faster.

In comparison to existing methods, our research introduces a custom encoder-decoder model for denoising GI endoscopy images, which significantly outperforms other approaches in this domain. Jifara *et al.* [17] utilized a CNN with residual learning for medical image denoising, achieving a PSNR of 41.6843, while Chen *et al.* [16] applied a CNN to low-dose CT scans, reporting a PSNR of 42.1514. These studies primarily focused on different imaging modalities like CT, with notable success in improving image quality but not specific to endoscopy. Our model, tailored for the HyperKvasir dataset, enhances the PSNR from 19.118954 to 69.892631, showcasing nearly perfect reconstruction and demonstrating superior denoising performance in the context of endoscopic imaging. This highlights the robustness and applicability of our approach in scenarios where precise image reconstruction is critical, thus offering a substantial improvement over traditional CNN-based model in both the endoscopic domain and general medical image denoising.

This study has several limitations that should be acknowledged. First, due to computational and resource constraints, the model could not be tested in real-world clinical environments, which may affect the generalizability of the results. All evaluations were conducted using the same dataset, which may not fully represent the variability encountered in practical medical imaging scenarios. Additionally, the field of AI-based denoising for medical image enhancement is relatively new, resulting in a limited amount of data and established benchmarks for comparison. Consequently, while our approach shows promise, further validation with diverse datasets and real-world applications is necessary to fully assess its efficacy.

5. CONCLUSION

ClearNet, a sophisticated denoising model with an InceptionV3 encoder-decoder architecture, demonstrated significant noise reduction and improved tissue visualization on the diverse HyperKvasir endoscopy dataset. Trained on images with 20% added Gaussian noise, it successfully removed even 30% noise during testing, highlighting its robustness to handle noise levels beyond training. Quantitative and qualitative results confirmed enhanced image quality and restored anatomical details obscured by noise. This pre-processing step aids gastroenterologists in disease detection by improving tissue visualization. Future work includes expanding evaluation across noise types and endoscopy modalities, clinical validation, and extending the model for joint denoising and anomaly segmentation. The study demonstrates the promise of transformer-based encoder-decoder networks for blind denoising of endoscopic data, potentially enhancing workflows and outcomes for GI disease screening.

ACKNOWLEDGMENTS

Authors are grateful to their respective institutions for all forms of support.

FUNDING INFORMATION

This research received no specific grant from any funding agency.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Vikrant Shokeen	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Sandeep Kumar	✓	✓			✓	✓		✓	✓	✓	✓	✓		
Vidhu Mathur	✓	✓	✓	✓	✓	✓			✓		✓		✓	
Amit Sharma	✓	✓		✓	✓		✓			✓				
Indrajeet Singh	✓			✓	✓		✓			✓		✓		
Parita Jain	✓			✓	✓	✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

INFORMED CONSENT

The HyperKvasir dataset used here, is the largest publicly released GI tract image dataset. It was collected with strict adherence to ethical guidelines, including obtaining informed consent from patients.

ETHICAL APPROVAL

The data collection in the dataset mentioned was conducted following ethical guidelines and was approved by the relevant institutional review board or ethics committee. This ensures that the research adhered to national regulations and institutional policies.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [Sandeep Kumar], upon reasonable request.

REFERENCES

[1] P. F. Niederer, J. Haefliger, P. Blessing, Y. Lehareinger, D. Doswald, and N. Felber, “Image quality of endoscopes,” in *Biomonitoring and Endoscopy Technologies*, Jan. 2001, vol. 4158, pp. 1–10, doi: 10.1117/12.413778.

[2] H. Borgli *et al.*, “HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy,” *Scientific Data*, vol. 7, no. 1, p. 283, Aug. 2020, doi: 10.1038/s41597-020-00622-y.

[3] M. F. Byrne *et al.*, “Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model,” *Gut*, vol. 68, no. 1, pp. 94–100, Jan. 2019, doi: 10.1136/gutjnl-2017-314547.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 2, pp. 1097–1105, 2012.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[6] M. Vania, B. A. Tama, H. Maulahela, and S. Lim, “Recent advances in applying machine learning and deep learning to detect upper gastrointestinal tract lesions,” *IEEE Access*, vol. 11, pp. 66544–66567, 2023, doi: 10.1109/ACCESS.2023.3290997.

[7] D.-Y. Liu *et al.*, “Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process,” *Medical Image Analysis*, vol. 32, pp. 281–294, Aug. 2016, doi: 10.1016/j.media.2016.04.007.

[8] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, “On optimization methods for deep learning,” *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 265–272, 2011.

[9] C. Hassan *et al.*, “Real-time computer-aided detection of colorectal neoplasia during colonoscopy,” *Annals of Internal Medicine*, vol. 176, no. 9, pp. 1209–1220, 2023, doi: 10.7326/m22-3678.

[10] A. Oubaalla, H. El Moubtahij, and N. El Akkad, “Medical image segmentation using deep learning: a survey,” in *Lecture Notes in Networks and Systems*, vol. 669 LNNS, 2023, pp. 974–983.

[11] S. Tanwar, S. Vijayalakshmi, M. Sabharwal, M. Kaur, A. A. Al Zubi, and H.-N. Lee, “[Retracted] detection and classification of colorectal polyp using deep learning,” *BioMed Research International*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/2805607.

[12] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, “Imbalance problems in object detection: a review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3388–3415, Oct. 2021, doi: 10.1109/TPAMI.2020.2981890.

[13] S. Z. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, “End-to-end object detection with transformers,” in *European conference on computer vision*, 2013, pp. 213–229, [Online]. Available: https://github.com/facebookresearch/detr.

[14] A. Ravishankar, S. Anusha, H. K. Akshatha, A. Raj, S. Jahnavi, and J. Madhura, “A survey on noise reduction techniques in medical images,” in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, Apr. 2017, vol. 2017-January, pp. 385–389, doi: 10.1109/ICECA.2017.8203711.

[15] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[16] H. Chen *et al.*, “Low-dose CT denoising with convolutional neural network,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Apr. 2017, pp. 143–146, doi: 10.1109/ISBI.2017.7950488.

[17] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, “Medical image denoising using convolutional neural network: a residual learning approach,” *The Journal of Supercomputing*, vol. 75, no. 2, pp. 704–718, Feb. 2019, doi: 10.1007/s11227-017-2080-0.

[18] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, 2015, pp. 234–241.

[19] M. Tahmid, M. S. Alam, N. Rao, and K. M. A. Ashrafi, “Image-to-image translation with conditional adversarial networks,” in *2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, Nov. 2023, pp. 1–5, doi: 10.1109/WIECON-ECE60392.2023.10456447.

[20] J. Lehtinen *et al.*, “Noise2Noise: learning image restoration without clean data,” *35th International Conference on Machine Learning, ICML 2018*, vol. 7, pp. 4620–4631, 2018.

[21] K. He *et al.*, “Transformers in medical image analysis,” *Intelligent Medicine*, vol. 3, no. 1, pp. 59–78, Feb. 2023, doi: 10.1016/j.imed.2022.07.002.

[22] A. Alsaiari, R. Rustagi, A. Alhakamy, M. M. Thomas, and A. G. Forbes, “Image denoising using a generative adversarial network,” *2019 IEEE 2nd International Conference on Information and Computer Technologies, ICICT 2019*, pp. 126–132, 2019, doi: 10.1109/INFOCT.2019.8710893.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[24] Q. Dou *et al.*, “Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016, doi:10.1109/TMI.2016.2528129.

[25] K. D. Gupta, D. K. Sharma, S. Ahmed, H. Gupta, D. Gupta, and C. H. Hsu, “A novel lightweight deep learning-based histopathological image classification model for IoMT,” *Neural Processing Letters*, vol. 55, no. 1, pp. 205–228, 2023, doi: 10.1007/s11063-021-10555-1.

BIOGRAPHIES OF AUTHORS

Dr. Vikrant Shokeen works on the segmentation, feature extraction, classification of images. He uses various techniques like PCA, SIFT, and SURF for feature extraction and SVM, LDA, KNN for classification. He is particularly interested in hybrid approaches like SVM integrated with feed forward back propagation neural network to create a hybrid algorithm that further helps in reducing the computation complexity of the classification. His present research area is focused on the factual findings of the potential usage of the combinational feed-forward back propagation neural network as a judgement making for disease detection. He can be contacted at email: shokeen18@gmail.com.

Dr. Sandeep Kumar has over 19 years of teaching and research experience across various institutions, he holds the position of associate professor in the Department of Computer Science and Engineering at Maharaja Surajmal Institute of Technology, Delhi. His research interests span data mining, image processing, and related areas, where he extensively employs open-source technologies to tackle challenges. He integrates the technological aspects of computer engineering with real-life challenges in his work. In recognition of his contributions, he was honored with the national-level top mentor award for IBM TGMC 2013 during the IBM TGMC felicitation ceremony on October 15, 2014, at CMR Institute of Technology, Bangalore, India. Furthermore, he who has also been bestowed with the best paper award at IRCET-2021, is currently mentoring Ph.D. scholars and has published papers in many reputed SCI/SCIE/Scopus indexed journals. He also has national and international patents to his name. He can be contacted at email: san.jaglan@gmail.com.

Vidhu Mathur is a skilled machine learning engineer with experience in developing efficient, data-driven artificial intelligence systems. He has expertise in computer vision, NLP and generative ai, with the ability to confidently assess, analyze, and organize large amounts of data. He has experience in designing and developing machine learning algorithms and deep learning applications and is passionate about staying up to date with developments in the machine learning industry. He can be contacted at email: vidhumathur2002@gmail.com.

Dr. Amit Sharma an emerging educator and researcher, awarded Ph.D. in computer science in the year 2022 from Motherhood University Roorkee, India. He earned his M.Tech. from MDU University, India, in 2009. His expert academician is Associate Professor in the Department of Information Technology in IMS Engineering College, Ghaziabad, India. He has given over 20 invited presentations at events, FDPs, and training sessions nationwide. He teaches machine learning, deep learning, soft computing, and data analytics at undergraduate and postgraduate level. In pursuit of scholarly excellence, he has published 20 research papers in different international journals and conferences. His impact on academia is seen in his passion for teaching, research, and student counselling. He can be contacted at email: amit.faculty@gmail.com.

Dr. Indrajeet Gupta an emerging educator and researcher, received a 2019 Ph.D. in computer science and engineering from ISM Dhanbad, India. He earned his M.Tech. from NIT Rourkela, India, in 2012. His expert academic is a professor in the School of Computer Science and Artificial Intelligence at SR University, Warangal, India. Since 2019, he has been an AWS Cloud Practitioner and Educate Faculty Brand Ambassador. Microsoft's MS Learn Program qualified him as an Azure Educator. Cloud resource optimization, AI, and problem-based learning are his specialties. He has given over 20 invited presentations at events, FDPs, and training sessions nationwide. He studies process scheduling, cloud resource provisioning, data visualization, and problem-based learning. In pursuit of scholarly excellence, he has published in 10 science citation index (SCI) journals and reviews for Future Generation of Computer Systems (Elsevier), IEEE Transactions on Cloud Computing, IETE Journal of Research, IEEE Access, Journal of Grid Computing, and The Journal of Supercomputing. His impact on academia is seen in his passion for teaching, research, and student counselling. He can be contacted at email: indrajeet7830@gmail.com.

Dr. Parita Jain completed her Ph.D. degree from Amity University, Noida and M. Tech from CDAC, Noida. Having vast teaching experience of approx. 12 years. Currently, she is working as associate professor in Department of Computer Science and Engineering and Assistant Dean (Academics and Autonomous) at KIET Group of Institutions Delhi-NCR, India. Her area of research includes software engineering, artificial intelligence, and machine learning. She has published more than 25 research papers in reputed conferences and journals. He can be contacted at email: paritajain23@gmail.com.