International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN(Online): 2347-5552, Volume-12, Issue-3, May 2024
https://doi.org/10.55524/ijircst.2024.12.3.22
Article ID IJIRD-1299, Pages 135-139
www.ijircst.org

Advancing Surgical Imaging with cGAN for Effective Defogging

Lingxi Xiao[1], Ruilin Xu[2], Yiru Cang[3], Yan Chen[4], and Yijing Wei[5]

[1] Georgia Institute of Technology , Atlanta, USA
[2] The University of Chicago, Chicago, USA
[3] Northeastern University, Boston, USA
[4] Stevens Institute of Technology, Hoboken, USA
[5] Northwestern University, Evanston, USA

Correspondence should be addressed to Lingxi Xiao ; Lingxi.xiao@gatech.edu

Received: 2 May 2024 Revised: 17 May 2024 Accepted: 31 May 2024

ABSTRACT- Image-based defogging technology can significantly enhance intraoperative image quality and shows great promise in various medical fields. A new image removal algorithm based on conditional generative adversarial networks (cGAN) has been developed. This algorithm employs the Tiramisu model instead of the conventional U-Net, thereby improving its computational accuracy. Additionally, the quality of the resulting images is enhanced by incorporating more textual data. A novel visual perception method is proposed, utilizing a contrast-based approach to improve the similarity between images with the same semantic content. Experiments demonstrate that this method not only excels at fog removal but also better preserves the key visual features of the images. Compared to existing image defogging technologies, this method offers superior qualitative analysis capabilities. This advancement can aid doctors in better visualizing intraoperative images. The effectiveness and robustness of the proposed method are validated through comparative analysis with several existing image noise reduction techniques.

KEYWORDS- Deep Learning; Image Elimination; U-Net; Generative Adversarial Networks; Image Removal Algorithm

I. INTRODUCTION

Due to the continuous development of abdominal surgical instruments, medical imaging and the requirement of minimally invasive surgery, the number of abdominal surgeries has been increasing. Under a laparoscope, the surgeon makes a small incision, puts in a needle, and injects CO2 to inflate the abdomen to provide enough room for other devices. Doctors also use special equipment such as endoscopes and ultrasound probes to examine the abdomen. In surgery, image is a very important information format, and high-definition image is the key to the doctor's operation and navigation. However, the presence of smoke, blood, dynamic light reflection, and other factors result in poor imaging effect. The smoke produced by laser ablation and electric burn will seriously affect the contrast and brightness of the scene. Such degraded images will inevitably impair the surgeon's visibility during operation. In addition, the current machine vision methods based on image navigation technology are mostly for clear images, especially in foggy areas, which has a greater limit on its effect[1]. Therefore, how to effectively remove fog can ensure that surgeons can obtain better imaging effects without affecting the performance of machine vision algorithms . The traditional method based on histogram equalization is used to de-fog the endoscopic imaging[2], but the existing methods mostly use the same transformation mapping, resulting in inconsistent results with the actual. Some scholars have adopted a linear conversion algorithm based on brightness [3]. This algorithm can eliminate the haze while maintaining the visual key information, but there is a large amount of computation in the process of solving. In this paper, a new image removal model based on conditional production adversarial network is designed, which can remove fog in endoscope images well. In addition, the new algorithm proposed in this project has certain reference value for improving the visual effect of medical images.

II. RELATED WORK

The use of deep learning for enhancing medical image clarity, as explored in this paper through a conditional generative adversarial network (cGAN) augmented by the Tiramisu model, aligns with significant advancements in the field of medical imaging. Research such as Xu et al. [4] and Zhang et al. [5] on multimodal and multi-scale deep learning underscores the potential of sophisticated neural architectures to refine the granularity and accuracy of image analysis, which is central to our methodology. The application of convolutional neural networks (CNNs) by Xiao et al. [6] for cytopathology image classification and Zhao et al. [7] for image segmentation demonstrates the broad utility of neural networks in extracting meaningful information from complex medical images, reinforcing the relevance of our cGAN approach in surgical settings.

Technological enhancements in training deep learning models, as discussed by Lu et al. [8] with federated learning, provide a framework for collaboratively improving model robustness without compromising data privacy, an aspect critical to deploying AI in healthcare. Mei et al. [9] focus on optimizing the efficiency of large-scale models, which is pertinent to processing the vast data involved in image defogging. Furthermore, the automation capabilities of deep learning, as shown in the automated medical report generation by Wang et al. [10] and survival prediction in oncology by Yan et al. [11],highlight the adaptability of these technologies to both generalize and specify tasks within medical practice.

The foundational research by Dai et al. [12] addresses the mitigation of unintended biases in AI applications, a concern that is paramount in ensuring the clinical utility of image defogging technologies. Similarly, the comparison of feature selection algorithms by Liu and Song[13]contributes to understanding how to optimize data preprocessing to enhance the performance of deep learning models in medical imaging. Additionally, the work by Wang et al.[14] on neuroinflammation imaging shows deep learning's expanding role in addressing complex medical conditions through advanced imaging techniques, further validating the depth and applicability of AI-driven methods in medical diagnostics.

Together, these studies not only provide a solid theoretical and empirical foundation for employing advanced neural networks in medical image enhancement but also illustrate a collective move towards more precise, efficient, and adaptable medical diagnostic tools. This body of work enriches the context of our proposed image defogging method, highlighting its potential impact on improving surgical outcomes through clearer intraoperative imaging.

III. TEXT GENERATION IMAGE METHOD BASED ON SINGLE-STAGE GENERATION ADVERSARIAL NETWORK

The proposed schema architecture of this study is shown in Figure 1 (image cited in A Tour of Generative Adversarial Network Models). The whole pattern has been improved following the single-level architecture of DF-GAN. 1) Affine transformation is performed on the first 3 text subareas sampled above, so that they can be combined with text information, and combine the text information that is most closely related to the subfield. 2) Dynamic convolution is used to replace traditional convolution, and multiple convolution cores are fused according to the input image characteristics.

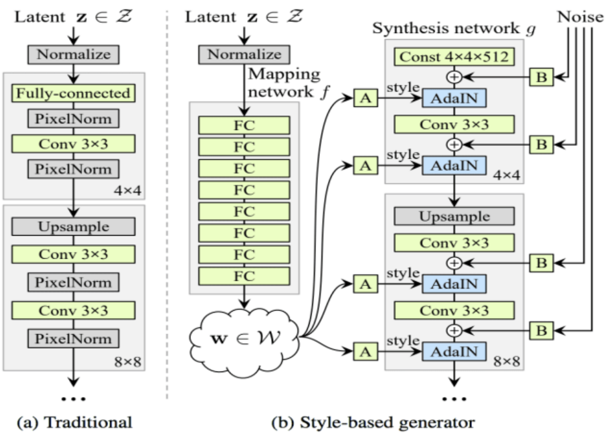

Figure 1: Frame diagram of text generation image method for generating adversarial network

A. Text Coding

Text encoding is to encode the text as a recognizable vector, so as to produce a picture consistent with the text. A multi-dimensional information fusion algorithm based on bidirectional short-time memory is proposed [15]. The 2-direction connection of a hidden layer is used as the encoding of a word, and the 2-direction connection of the last hidden layer is used as the encoding of the current sentence[16]. The algorithm is based on an in-depth study of a pre-trained Inception-V3 in ImageNet. The actual image is compressed according to a certain size, and then the coding $\overline{f} \in R^{2048}$ of the original image is obtained on the basis of the last level mean, and then the coding $f \in R^{768\times289}$ of the image is obtained in the mixed _6e. An image has 289 subareas, each of which has 768 dimensions. A video compression algorithm based on DAMSM is proposed. In this way, the text generated by the text encoder is recognizable. In the learning process of production adversarial network, two different parameters, text encoding and image encoder, are used to reduce the number of parameters required in the learning process.

B. Model structure

The image features of 256x256*32 are obtained by merging 6 upper samples of the image. The generated image, the actual image, and the statement vector are fed into the discriminator for recognition. An improved affine transformation for adding text information is shown in Figure 2 (image cited in Frontiers in Plant Science, 2024, 15:1327237). Using 1-by-1 convolution, the dimension of the word $a \in R^{L\times K}$ is converted to the same dimension as the current image characteristics:

$$a' = Va, V \in R^{\hat{L}\times L} \tag{1}$$

Multiply a' with the current image property $s \in R^{M\times\hat{L}}$ and standardize it to get. Since the pixel size of each pixel is different, M is the dimension of the current subregion. δ_{ji} is the weighting of the j region of the current image characteristics of the i word,

$$\delta_{ji} = \frac{\exp(s_j^K a_i')}{\sum_{i=1}^{K-1}\exp(s_j^K a_i')} \tag{2}$$

Multiplying δ_{ji} by a word produces a dynamic expression for all words related to that word:

$$z_j = \sum_{i=0}^{K-1} \delta_{ji} a_i' \tag{3}$$

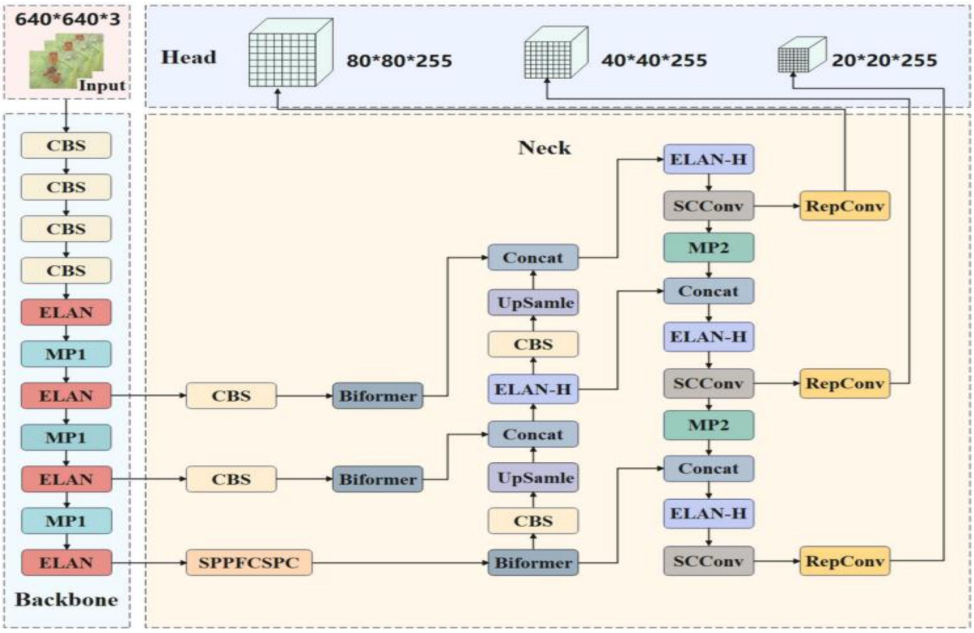

Figure 2: Improved affine transformation structure diagram

Concatenate the statement vector λ, the word content vector z_j, and the noise vector u and use them as the text state of the current subdomain. The scale and displacement coefficients of each subregion are extracted from the text state by using multi-level perceptron, so that each subregion has its own scale coefficient γ_j and displacement coefficien σ_j.

$$\gamma_j = MLP_1(concat(u, \lambda, z_j)) \tag{4}$$

$$\sigma_j = MLP_2(concat(u, \lambda, z_j)) \tag{5}$$

$$AFF(s_j \mid concat(u, \lambda, z_j)) = \gamma_j s_j + \sigma_j \tag{6}$$

This project intends to affine a number of existing texts, so that the subdomain not only contains the description of sentences, but also organically combines with the most important words in the subdomain, so as to solve the defect of traditional methods that only use sentences and ignore words. This allows for fine image generation.

IV. EXPERIMENTAL PROCESS

A. Data Collection

Obtaining massive samples not only consumes huge medical resources, but also puts forward higher requirements for accuracy and scale . However, in the existing methods, there are some problems such as shielding smoke and concentration, which makes the work of haze elimination more complicated. It is difficult to obtain such image pairs and concentration templates by hand. The paper uses the graphic rendering technology of PS to continuously draw the smoke of the endoscope image, and finally gets a composite image. This method has the following two advantages: 1) Since the smoke in surgery is generated locally and related to distance, it is notnecessary to use the conventional smog modeling method. 2) The image painter is able to generate realistic differences in fog morphology and concentration through carefully constructed internal modeling. This paper selects 1250 pictures without haze as the actual scene based on manual selection, and adjusts the percentage parameters of haze transparency and concentration in the drawing interface to realize real-time rendering of the scene. The transparency and concentration are set to 10,25,40%,55%,70%, respectively, and the proportion of each percentage is larger, indicating that the generated fog concentration is higher, so that 6250 groups of different effects of the smoke combination. At the same time, to ensure that the network does not overly fit a particular form of smoke, the image is reversed in 90 degrees, 180 degrees, 270 degrees in three directions to enhance the data. The result is 25,000 simulation results. The training set, test set and test set are divided at the ratio of 8:1:1.

B. Details of Data Training

Wgan is set to 2, WL1 is set to 100, and the learning rate of the network is fixed at 0.001. In this article, Adam optimizer is used to make a gradient drop. In the TensorFlow architecture, programming is done in Python. The testers installed it on a GTX1080 (8 GB) graphics card. Each step in this configuration takes only 1.6 seconds.

V. COMPARISON AND ANALYSIS OF EXPERIMENTAL RESULTS

This project intends to carry out theoretical and empirical research by combining theoretical simulation with practical application, and compare and verify it with four commonly used fog removal methods in literature [17].

A. Simulated Image Control Test

This paper first tests and evaluates the network in a simulated data set. Figure 3 shows one of these comparison scenarios.

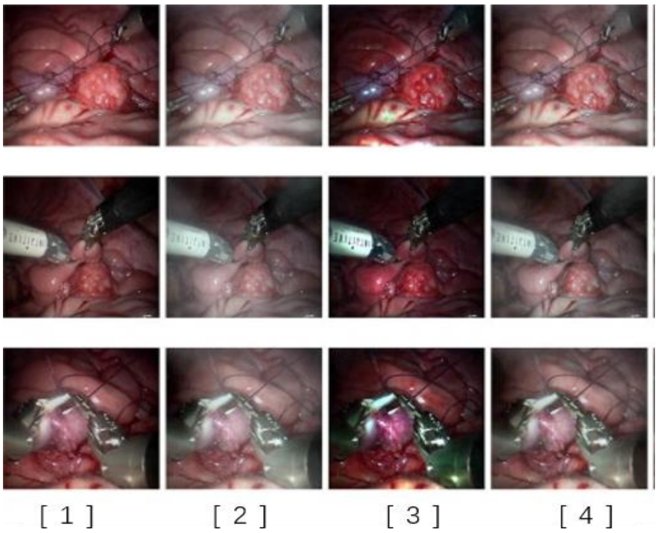

Figure 3: Comparison of results of different algorithms on synthetic fog images

In addition, the image sharpness can be evaluated from many aspects, and PSNR and SSIM are two widely used indicators [18]. The higher the PSNR, the better the performance when measuring the ability to remove fog. The structural similarity model (SSIM) is used to measure the similarity between two identical images, and the SSIM value is 1. From both qualitative and quantitative aspects, the method has obvious advantages in synthesizing test samples.

B. True Controlled Trial

Finally, taking the smoke in actual surgery as an example, the mathematical model is studied experimentally. Figure 4 shows the different schemes, and from this comparison, the robustness of the method can be seen directly in this article.

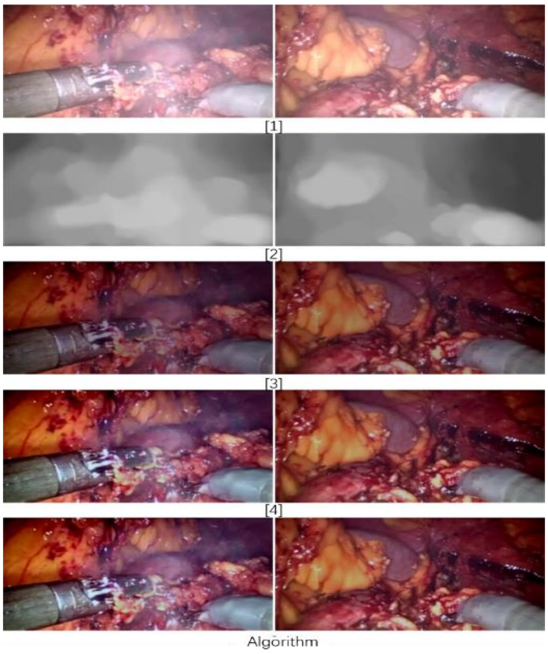

Figure 4: Comparison of results of different algorithms on real fog images

Fog removal through the network of this article is the most thorough. The research shows that the conventional algorithm is not effective in medical image, not only cannot effectively restore the real color in haze area, but also lead to color distortion in non-haze area. At present, the traditional defogging algorithms mostly adopt parametric modeling rather than medical image oriented. At present, the fog removal method based on deep neural network still has some residual smoke that has not been completely eliminated, which is largely due to the insufficient light intensity during training. However, this new algorithm can solve the above problems well . The algorithm can not only focus on the fog area, but also keep the haze free area, but also use the learned characteristics to accurately organize the color restoration.

C. Robustness Test Experiment

The loss of image data is irreversible based on the scale of the fog. The robustness of the method is verified by the comparative analysis of 5 fog degrees. Their concentration values correspond to the five percentage rendering parameters set in the rendering board, namely 10%, 25%, 40%, 55% and 70%. A test set containing 25 photos of each concentration was selected in this study (fig.5 and fig.6).

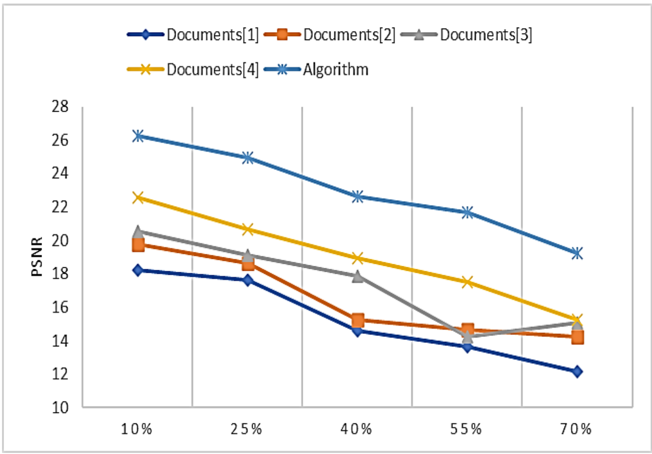

Figure 5: PSNR result curve for robustness test of different algorithms

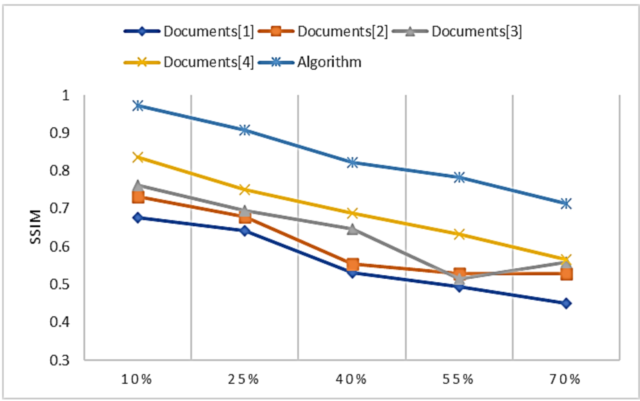

Figure 6.SSIM results of robustness test for different algorithms

With the increase in disturbance concentration, the exponential curve gradually decreases, indicating that the fog is more difficult to cope with. The proposed method can effectively deal with various types of disturbances, thus improving the robustness of the method.

VI. CONCLUSION

In this study, we introduced a novel image defogging algorithm based on conditional generative adversarial networks (cGANs) and the Tiramisu model, specifically designed to enhance intraoperative imaging. This method not only effectively removes fog from endoscopic images but also preserves key visual features essential for surgical precision. Our contrast-based visual perception approach ensures the semantic consistency of images, which is vital for accurate interpretation during surgeries. The comparative analysis with existing image noise reduction techniques confirms the superior performance of our method in both fog removal and feature preservation. These advancements in medical imaging are poised to significantly improve surgical outcomes by offering clearer and more accurate visualizations. As we look to the future, we plan to further refine this technology, focusing on optimizing its application in real-time surgical settings and extending its use to other medical imaging modalities. This will broaden the impact of our work, potentially transforming various aspects of medical practice through improved visual accuracy.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest

REFERENCES

[1] M. R. Pavan Kumar and P. Jayagopal, "Generative adversarial networks: a survey on applications and challenges," International Journal of Multimedia Information Retrieval, vol. 10, pp. 1-24, Jan. 2021.

[2] W. L. Jun and Z. H. U. Rong, "Image defogging algorithm of single color image based on wavelet transform and histogram equalization," Applied Mathematical Sciences, vol. 7, no. 79, pp. 3913-3921, 2013.

[3] S. Suresh and S. Lal, "Modified differential evolution algorithm for contrast and brightness enhancement of satellite images," Applied Soft Computing, vol. 61, pp. 622-641, 2017.

[4] T. Xu, I. Li, Q. Zhan, Y. Hu, and H. Yang, "Research on Intelligent System of Multimodal Deep Learning in Image Recognition," Journal of Computing and Electronic Information Management, vol. 12, no. 3, pp. 79-83, 2024.

[5] H. Zhang, S. Diao, Y. Yang, J. Zhong, and Y. Yan, "Multi-scale image recognition strategy based on convolutional neural network," Journal of Computing and Electronic Information Management, vol. 12, no. 3, pp. 107-113, 2024.

[6] M. Xiao, Y. Li, X. Yan, M. Gao, and W. Wang, "Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example," arXiv preprint arXiv:2404.08279, 2024.

[7] B. Zhao, Z. Cao, and S. Wang, "Lung vessel segmentation based on random forests," Electronics Letters, vol. 53, no. 4, pp. 220-222, 2017.

[8] S. Lu, Z. Liu, T. Liu, and W. Zhou, "Scaling-up medical vision-and-language representation learning with federated learning," Engineering Applications of Artificial Intelligence, vol. 126, p. 107037, 2023.

[9] T. Mei, Y. Zi, X. Cheng, Z. Gao, Q. Wang, and H. Yang, "Efficiency optimization of large-scale language models based on deep learning in natural language processing tasks," arXiv preprint arXiv:2405.11704, 2024.

[10] S. Wang, Z. Liu, and B. Peng, "A Self-training Framework for Automated Medical Report Generation," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Dec. 2023, pp. 16443-16449.

[11] X. Yan, W. Wang, M. Xiao, Y. Li, and M. Gao, "Survival Prediction Across Diverse Cancer Types Using Neural Networks," arXiv preprint arXiv:2404.08713, 2024.

[12] W. Dai, J. Tao, X. Yan, Z. Feng, and J. Chen, "Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach," in 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA), Nov. 2023, pp. 375-379.

[13] Z. Liu and J. Song, "Comparison of Tree-based Feature Selection Algorithms on Biological Omics Dataset," in Proceedings of the 5th International Conference on Advances in Artificial Intelligence, Nov. 2021, pp. 165-169.

[14] Q. Wang, S. E. Schindler, G. Chen, N. S. McKay, A. McCullough, S. Flores, ... and T. L. Benzinger, "Investigating White Matter Neuroinflammation in Alzheimer Disease Using Diffusion-Based Neuroinflammation Imaging," Neurology, vol. 102, no. 4, p. e208013, 2024.

[15] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," arXiv preprint arXiv:1508.01991, 2015.

[16] J. Yao, T. Wu, and X. Zhang, "Improving depth gradient continuity in transformers: A comparative study on monocular depth estimation with cnn," arXiv preprint arXiv:2308.08333, 2023.

[17] B. Das, J. P. Ebenezer, and S. Mukhopadhyay, "A comparative study of single image fog removal methods," The Visual Computer, vol. 38, no. 1, pp. 179-195, 2022.

[18] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in 2010 20th International Conference on Pattern Recognition, Aug. 2010, pp. 2366-2369.