Dehazing of Images using Unpaired Image-to-Image Translation

Rohan Reddy Pathi
College Of Engineering
Northeastern University
pathi.r@northeastern.edu

Vikramadithya Pabba
College Of Engineering
Northeastern University
pabba.v@northeastern.edu

Chaitanya Prudvi Balusu
College Of Engineering
Northeastern University
balusu.c@northeastern.edu

Param Madan
College Of Engineering
Northeastern University
madan.pa@northeastern.edu

Abstract

Haze removal, also known as dehazing, is the primary focus, specifically concerning desmoking laparoscopic images. This process is critical for performing surgical operations safely and efficiently. Laparoscopic surgery often generates smoke from instruments like cauterizers, which can obscure the surgeon's view, thereby reducing procedural precision and safety. Haze significantly degrades image quality by diminishing contrast and introducing opacity. It is a vital preprocessing step for various computer vision applications, including object detection, image segmentation, and autonomous driving. Traditional haze removal methods depend on handcrafted features and assumptions, such as the dark channel prior (DCP) or physical models of image formation. However, these techniques frequently fail under challenging conditions where their underlying assumptions are invalid. This project aims to develop a combined model utilizing GANs paired with Diffusion models for image dehazing. The proposed model leverages Generative Adversarial Networks (GANs) to learn complex features directly from data, enabling it to effectively handle diverse and difficult scenarios where conventional methods might falter.

1 Introduction

1.1 Problem Definition

Haze removal, or dehazing, with a specific focus on desmoking laparoscopic images, is essential for ensuring safe and efficient surgical procedures. Smoke generated during laparoscopic surgery, often from cauterizing instruments, can impair the surgeon's visibility, consequently compromising the precision and safety of the operation. This phenomenon significantly lowers image quality by reducing contrast and adding opacity. Haze removal serves as a crucial preprocessing step in numerous computer vision tasks like object detection, image segmentation, and autonomous driving. Conventional methods for haze removal rely on manually designed features and assumptions like the dark channel prior (DCP) or physical models of image formation. These approaches often prove ineffective in challenging situations where their assumptions do not hold true. This project endeavors to create a hybrid model combining Diffusion models with GANs for image dehazing. The model will utilize the capability of Generative Adversarial Networks (GANs) to learn intricate features directly from the data, allowing it to manage varied and demanding scenarios where traditional methods may be inadequate.

1.2 Motivation

In contemporary medicine, laparoscopic surgery is increasingly preferred due to its lower associated mortality rate compared to traditional open surgery. Nonetheless, the efficacy of laparoscopic procedures can be diminished by reduced visibility resulting from smoke produced during the operation. Employing dehazing techniques on laparoscopic imagery is vital for improving surgical efficiency and safety. Applying dehazing algorithms to laparoscopic images not only improves visual clarity but also acts as an essential preprocessing step for various computer vision tasks that could further aid surgeons. By enhancing the quality of laparoscopic images, dehazing methods can lead to more accurate diagnoses, precise surgical interventions, and ultimately, safer procedures, potentially preventing complications such as excessive bleeding or organ failure.

2 Literature Survey

Image dehazing represents a persistent challenge in computer vision, holding significant importance for applications like remote sensing, autonomous driving, and surveillance systems. Over time, numerous methods have been proposed, evolving from traditional image processing techniques to sophisticated deep learning approaches. More recently, the advent of diffusion models and hybrid strategies has introduced new possibilities for tackling the dehazing task with enhanced performance and efficiency.

2.1 Traditional and Early Deep Learning Approaches

Initial dehazing methods primarily relied on physical models describing haze formation, like the atmospheric scattering model (ASM) introduced by Koschmieder (1924). He et al. (2011) proposed the dark channel prior, which became a foundational concept for many subsequent dehazing algorithms. As deep learning gained traction, researchers explored its application to image dehazing. Cai et al. (2016) introduced DehazeNet, an end-to-end system for single image haze removal, while Ren et al. (2016) presented a multi-scale convolutional neural network for the same purpose.

2.2 GAN-based Dehazing Methods

The development of Generative Adversarial Networks (GANs) introduced a new approach to image dehazing. Li et al. (2018) suggested a conditional GAN-based method for single image dehazing, showing better results than previous techniques. Expanding on this, Akhtar et al. (2023) introduced Mobile-UNet GAN, which uses a GAN based on the UNet architecture to effectively extract features from hazy images. This method achieved leading results while being computationally efficient, reportedly processing 0.1 images per second. CycleGAN, introduced by Zhu et al., pioneered unpaired image-to-image translation, applicable to tasks like dehazing. It employs a cycle-consistency loss, enabling training without paired data, which is particularly beneficial when paired datasets are unavailable or impractical.

2.3 Diffusion Models for Dehazing

The recent success of diffusion models in various image generation tasks prompted researchers to investigate their use for image dehazing. Yu et al. (2023) introduced DehazeDDPM, a novel framework merging diffusion models with physics-aware image dehazing. This method involves two stages: a physical modeling stage that aligns the hazy data distribution closer to the clear data distribution, and a diffusion stage that utilizes the powerful generation capabilities of Denoising Diffusion Probabilistic Models (DDPMs) to recover information lost due to haze. DehazeDDPM demonstrated state-of-the-art performance on several image dehazing benchmarks, showing particular strength in scenarios with dense haze.

2.4 Hybrid Approaches

Zhang et al. (2019) proposed an innovative unsupervised Image-to-Image Translation framework using Generative Adversarial Networks (GANs) to address both haze image synthesis and dehazing simultaneously. Inspired by this, we adopt a hybrid model that integrates CycleGAN for unpaired image translation with Diffusion Models for enhanced denoising and fine-detail recovery. This approach addresses challenges posed by imbalanced, unpaired, and limited-size training samples. By combining the structural consistency of CycleGAN with the generative fidelity of diffusion-based methods, our model ensures sharper, high-quality image restoration—crucial for applications such as laparoscopic surgery and autonomous systems.

3 Methods

3.1 Overview

This study investigated and compared several generative approaches for image dehazing, focusing on enhancing visual clarity by generating clear images from hazy inputs through different model architectures.

First, we implemented a standard CycleGAN architecture to serve as a baseline. This model leverages its established capability for unpaired image-to-image translation, primarily relying on adversarial and cycle consistency losses during training to learn the mapping between hazy and clear image domains.

Second, we explored augmenting the baseline by adding a diffusion model step after the standard CycleGAN process. In this approach, the initial dehazed output from the CycleGAN is further refined using a separate diffusion model, aiming specifically to reduce residual noise artifacts and improve the overall perceptual quality of the final image.

Third, and central to our contribution, we developed an integrated Diffusion-CycleGAN framework. This hybrid method embeds the diffusion process directly within the CycleGAN structure. It utilizes a UNet2DModel as part of the core architecture and incorporates an additional diffusion loss term alongside the standard adversarial, cycle consistency, and identity losses. This integration allows the model to leverage the denoising strengths of diffusion intrinsically during the adversarial training.

The overarching goal across these enhanced methods was to synergistically combine the image translation strengths of CycleGAN with the powerful denoising and detail-refinement capabilities of diffusion models. This comparison aimed to determine the most effective way to integrate these techniques to improve the perceptual quality and reduce noise compared to the baseline CycleGAN approach for image dehazing.

3.1.1 Intuition

The proposed model seeks to surpass the baseline CycleGAN method by incorporating a diffusion model, which adds an extra denoising capability to the image generation pipeline. While CycleGAN effectively generates realistic images through adversarial losses and cycle consistency, it can sometimes produce noisy or lower-quality outputs, particularly in tasks like image dehazing where noise is a concern.

The diffusion model within our hybrid approach functions by progressively adding noise to clear images and then learning to reverse this process. This effectively trains the model to denoise and recover the original image structure. This step-by-step refinement capability ensures higher and more consistent final image quality. Moreover, combining adversarial learning (from GANs) with denoising (from diffusion models) addresses both the visual appeal and structural integrity of the generated images, making this method more robust for complex tasks like image dehazing.

Additionally, while traditional GANs might solely focus on adversarial and cycle consistency losses, the diffusion component aids generalization by regularizing the model, leading to improved results concerning perceptual fidelity and noise reduction.

3.2 Models

Generator Network (G_{H2C} and G_{C2H}):

The generators are tasked with converting hazy images to clear ones (G_{H2C}) and vice versa (G_{C2H}). They employ a U-Net-like architecture featuring downsampling and upsampling layers interspersed with residual blocks. Residual blocks help retain fine details, while downsampling and upsampling layers manage spatial dimensions.

Discriminator Network (D_H and D_C):

These are convolutional neural networks designed to distinguish real images from generated (fake) ones. They operate on hazy (D_H) and clear (D_C) images respectively. The discriminators utilize LeakyReLU activations and instance normalization to stabilize training and mitigate mode collapse, a common issue in GAN training.

Diffusion Model (UNet2DModel):

A key component of our method, the diffusion model acts as a denoising autoencoder. It introduces noise to the input image and learns the reverse process, denoising the image incrementally. Based on the U-Net architecture, it incorporates attention blocks to capture high-level features. The integration of a DDPMScheduler ensures the noise addition and reversal occur over multiple timesteps, enhancing the model's ability to produce clearer outputs.

Loss Funtions:

Adversarial Loss: Generator and discriminator networks are trained using mean squared error (MSE) loss for adversarial learning. The generator aims to deceive the discriminator by creating realistic images.

$$L_{GAN} = \mathbb{E}_{x\sim p_{data}}[logD(x)] + \mathbb{E}_{x\sim p_z}[log(1 - D(G(z)))]$$

where G is the generator, D is the discriminator, x is the real image, and z is input noise.

Cycle Consistency Loss: This loss ensures that applying transformations in both directions (e.g., hazy to clear and back to hazy) reconstructs the original image, maintaining consistency.

$$L_{cycle} = \mathbb{E}_{x\sim p_{data}}\left[\|G(F(x)) - x\|_1\right]$$

where F is the mapping from the dehazed domain back to the hazy domain.

Identity Loss: This loss encourages the generator to leave images from the target domain unchanged when passed through it (e.g., a clear image through G_{H2C} should remain clear). This helps prevent unwanted alterations and stabilizes training.

$$L_{identity} = \mathbb{E}_{x\sim p_{data}}\left[\|G(x) - x\|_1\right]$$

Diffusion Loss: This loss ensures effective denoising by minimizing the difference between the predicted noise and the actual noise added during the diffusion process.

$$\mathcal{L}_{diffusion} = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1)}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2\right]$$

4 Experiments

4.1 Testbed Description and Experiment Objectives

Experiments were designed to assess the proposed Diffusion+CycleGAN model's effectiveness against the baseline CycleGAN for dehazing laparoscopic surgery images. The following questions framed the experimental design:

- Does integrating diffusion models with CycleGAN enhance image dehazing performance?
- How does the proposed method influence metrics related to perceptual and structural quality?
- Can the proposed method demonstrate better generalization across diverse laparoscopic surgery images?

Model	Average FADE	Average JNBM	Average REA
Diffusion+CycleGAN	2.1674	24.984	9.8
CycleGAN	2.5761	36.1544	7.5609
Cycle Diffusion	5.4152	54.984	4.654

4.2 Experiment Details

Dataset:

- The dataset comprised unpaired hazy and clear laparoscopic images, reflecting real-world surgical conditions.
- Images were resized to 256×256 pixels for computational manageability, and standard data augmentation techniques were applied.

Evaluation Metrics:

To assess the quality of the dehazed images, the following metrics were used:

FADE (Fog Aware Density Evaluator): Measures image fog density. Lower values signify reduced haze and improved clarity.

$$FADE = \frac{1}{N} \sum_{i=1}^{N} f(I_i)$$

JNBM (Just Noticeable Blur Metric): Assesses the perceived sharpness of an image. Higher values indicate sharper images.

$$JNBM = \frac{1}{N} \sum_{i=1}^{N} (TotalEdges(I_i))(NoticeableEdges(I_i))$$

REA (Relative Edge Assessment): Measures edge enhancement or degradation. Lower values suggest better preservation or enhancement of image edges.

$$REA = \frac{1}{N} \sum_{i=1}^{N} \frac{\|E_{ref}\|}{\|E_{dehazed} - E_{ref}\|}$$

Baseline and Proposed Model Comparison:

The baseline CycleGAN model was trained using standard settings for image translation tasks. The proposed Diffusion+CycleGAN model was trained with a hybrid loss function combining CycleGAN's adversarial and cycle-consistency losses with the diffusion model's reconstruction loss.

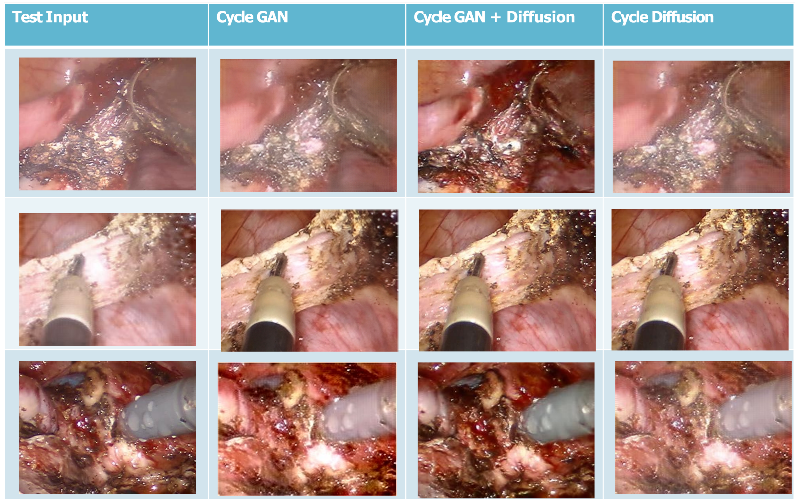
Experimental Setup:

Models were trained for 125 epochs on NVIDIA A100 GPUs using a batch size of 4. The Adam optimizer was employed with learning rates set at 0.0002 for both models.

Results:

Model performance based on the evaluation metrics is summarized below:

Here are the generated image results,



5 Conclusion

This study evaluated three image restoration models, Diffusion + CycleGAN, CycleGAN and Cycle Diffusion, for dehazing hazy images. Experimental results indicate that the Diffusion + CycleGAN model surpasses the traditional CycleGAN model across all evaluation metrics (FADE, JNBM, and REA). Images generated by the Diffusion + CycleGAN model showed superior clarity, reduced haze, and more detailed restoration compared to those from the CycleGAN model. This suggests that integrating diffusion processes with the CycleGAN architecture significantly boosts the model's capacity for restoring image quality.

The comparison highlights the potential of combining diffusion techniques with generative adversarial networks for tackling complex image restoration challenges. The Diffusion + CycleGAN model demonstrates promise not only in denoising images but also in preserving intricate details often lost in conventional approaches.

5.1 Future work

Despite promising results, considerable scope exists for improvement and further investigation. Future research could explore these areas:

1. **Enhancement of Diffusion Process:** The current diffusion loss function could be further optimized to improve denoising, especially for images with substantial haze or noise. Refining the model's handling of such complex data could lead to better restoration quality.
2. **Evaluation on Larger Datasets:** Assessing the models on larger, more diverse datasets, including real-world images, is essential to verify their robustness and generalization abilities. This would offer insights into performance under varied conditions and with complex data.
3. **Model Optimization:** Hyperparameter tuning and exploring alternative network architectures are key areas for potential improvement. Experimenting with different generator and discriminator designs might enhance performance, enabling the model to tackle more challenging dehazing tasks.
4. **Real-Time Applications:** Investigating the feasibility of these models for real-time image restoration could create new opportunities in applications like autonomous vehicles, surveillance, and satellite imaging. Optimizing inference speed and model efficiency will be crucial for practical deployment.
5. **Multi-Domain Dehazing:** Extending the models to handle dehazing across various domains, such as different weather conditions or haze types, could broaden their applicability. This would enable more robust and versatile image restoration across diverse environmental contexts.

Addressing these points could significantly enhance the performance, scalability, and usability of these models, making them suitable for a wider array of real-world image restoration applications.

References

[1] Koschmieder, H. (1924). Theorie der horizontalen Sichtweite. Beiträge zur Physik der freien Atmosphäre, 12, 33-53.

[2] He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12), 2341-2353.

[3] Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). DehazeNet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing, 25(11), 5187-5198.

[4] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. (2016). Single image dehazing via multi-scale convolutional neural networks. In European Conference on Computer Vision (pp. 154-169). Springer, Cham.

[5] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2018). Aod-net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4770-4778).

[6] Akhtar, N., Jalal, A., & Kim, K. (2023). Mobile-UNet GAN: An Efficient Approach for Image Dehazing Using Generative Adversarial Networks. Sensors, 23(3), 1610.

[7] Yu, T., Ma, J., & Li, C. (2023). DehazeDDPM: A Generative Dehazing Method with Denoising Diffusion Probabilistic Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 21804-21813).

[8] https://arxiv.org/html/2407.13132v1Wenhui Chang, Hongming Chen LSD3K: A Benchmark for Smoke Removal from Laparoscopic Surgery Images

[9] https://www.ijcai.org/proceedings/2021/0604.pdfA Comprehensive Survey on Image Dehazing Based on Deep Learning Jie Gui , Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang and Dacheng Tao

[10] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9110497/ DeSmoke-LAP: improved unpaired image-to-image translation for desmoking in laparoscopic surgery Yirou Pan, Sophia Bano, Francisco Vasconcelos, Hyun Park, Taikyeong Ted. Jeong, and Danail Stoyanov