GAN Image Denoising

Tiansheng Wang

1. Summary

In this project, a GAN-based denoising network was trained on the SIDD dataset The network employed a combination of U-net structure denoiser and generator, as well as a CNN Discriminator and underwent training with data augmentation. Evaluation on the SIDD validation set showed improvements in PSNR and SSIM metrics. Further analysis is needed to assess the visual quality and performance of the denoised images compared to the original noisy images.

2. Introduction

In the field of digital image processing, noise poses a persistent challenge that significantly impacts the quality and clarity of images. A variety of sources, including limitations inherent to sensors, transmission errors, and environmental factors, introduce undesired noise that hinders the visual perception and interpretation of images. One prominent source of image noise is sensor limitations, whereby sensors rely on sufficient light to capture images effectively. Insufficient lighting conditions lead to heightened noise levels in sensor output, primarily determined by lens and sensor size. Additionally, environmental factors, such as dust on the lens or electromagnetic interference from other electronic devices, can introduce structured noise into images. Consequently, employing denoising methods becomes crucial in mitigating sensor limitations and enhancing image quality. The subsequent images presented here were captured using smartphones equipped with small CMOS sensors, and they demonstrate the tangible benefits of utilizing denoising techniques.

3. Background

In the realm of denoising methods, two distinct categories can be identified: traditional methods and neural network methods. Traditional methods, also known as conventional methods, rely on human-designed filters and make assumptions about noise characteristics. Some examples of these methods include the median filter, wavelet filter, and bilateral filter. These filters can be applied in various domains, such as spatial, temporal, and frequency domains. Among these methods, BM3D stands out as a powerful denoising technique, utilizing three dimensions of matched blocks to effectively remove noise. It was introduced in 2010.

On the other hand, neural network methods have gained significant attention in recent years. Neural networks have demonstrated remarkable capabilities in computer vision tasks, including image classification and image segmentation. Consequently, researchers began exploring the application of neural networks to image denoising tasks, which has yielded successful outcomes. These neural network methods encompass different architectures such as convolutional neural networks (CNNs), generative adversarial networks (GANs), and transformers, which have gained popularity in recent times. The U-net encoder-decoder structure has been widely adopted in the field of image denoising.

While neural network methods have showcased superior results, traditional methods still hold an important role, particularly in scenarios where computational efficiency is prioritized. For instance, many smartphone companies have incorporated AI-based denoising methods in photo mode. However, in video mode, due to limited computational resources on mobile devices, traditional methods remain the primary choice. Traditional methods can be implemented on chip, enabling real-time denoising with constrained computational requirements.

4. Approach

4.1. Network Structure

The overall network structure in my project, as described in the reference paper[5], consists of a GAN network with three components: the denoiser, the generator, and the discriminator. The denoiser and generator share the same U-net structure, while the discriminator utilizes an additional CNN structure. The network will be trained end-to-end.

4.2. Datasets

For data sourcing, I have chosen the SIDD dataset, which was published in 2018. This dataset comprises images captured by five different smartphones. One advantageous aspect of the SIDD dataset is that the noisy images are already aligned with their corresponding ground-truth clean images. The ground-truth clean images were obtained by using multiple image sequences.

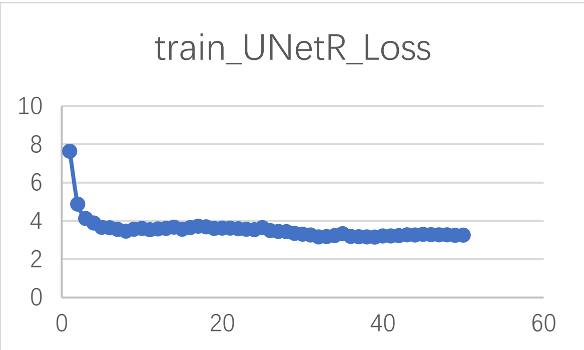
Although the SIDD dataset contains a total of 200 scenes, which might not be sufficient for network training, data augmentation techniques are needed. Data augmentation is a valuable technique for expanding the dataset. In this project, I have employed random crops, random flips, and random rotations (limited to 90/180/270 degrees). However, I have refrained from using arbitrary rotations or random light/color adjustments, as they led to unexpected results, such as a bright area appearing in the middle of the image. It is plausible that these transformations altered the noise characteristics in the image and introduced additional complexity to the denoising task.

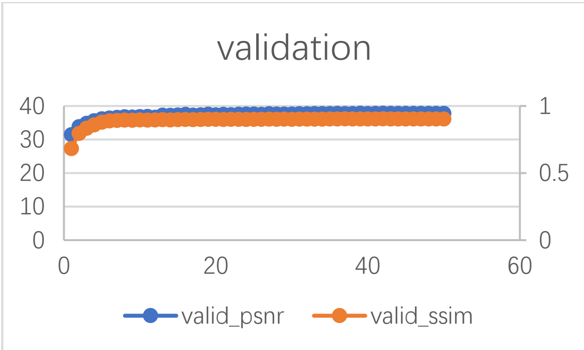
4.3. Model Training

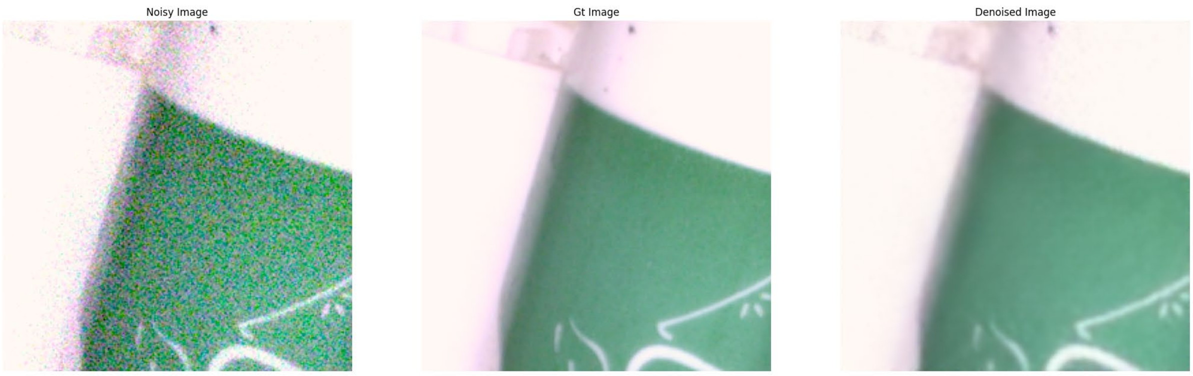
The network training was conducted on my local desktop with GPU acceleration. Several hyperparameters were employed in the project, including a relatively small batch size of 16. The Adam optimizer was used, along with weight decay for normalization purposes. The total number of epochs for training was set to 50, and the training process took approximately 6.5 hours to complete.

5. Results

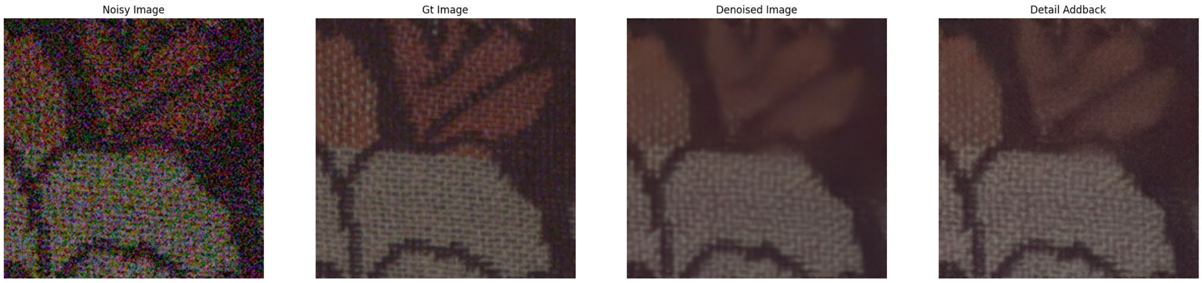
We can see that with the training progress goes on, the PSNR and SSIM on the validation set is getting higher, which means the network's denoising ability is getting stronger.







For perspective evaluation, I evaluate the network on the SIDD validation set. The patches have shown that the network does remove all the details from the given noisy patches. However, the output image would lose details compared to the ground truth. This may be caused by two reasons: 1) we use the L1 loss for the denoiser, which will make the network remove more noise and give the result as clean as possible. 2) the SNR on the given noisy patch is so low that even humans cannot recognize any textures. In that case, we could try to add some grain noise by blend the output and the input noisy patches to get better perspective quality.



6. Discussion

The denoising task raises important questions about the evaluation of network performance and the subjective perception of image quality. While objective evaluation metrics like PSNR and SSIM are commonly used in denoising projects, they may not always capture the true quality or aesthetic appeal of an image. Merely achieving a cleaner image does not necessarily mean it is visually superior or more desirable.

In some cases, introducing a controlled amount of grain noise can enhance the natural appearance of an image, avoiding an overly processed or artificial look. Denoising methods often excel in reducing noise in flat areas, but they can sometimes exhibit undesirable side effects in textured regions. As image scientists, it is crucial to delve deeper into developing better evaluation methods that accurately assess the quality of an image in terms of noise and texture balance. The exploration of new metrics that capture the nuanced interplay between noise reduction and preserving essential textures could lead to improved assessments of denoising effectiveness.

Furthermore, an intriguing consideration is the incorporation of prior knowledge about the noise profile in denoising tasks. Smartphone manufacturers, like Apple, employ calibration techniques on CMOS sensors to obtain precise noise profiles. With this prior knowledge, they can design denoising networks that are highly optimized and tailored to the specific noise characteristics of their devices. Such advanced calibration-driven approaches can potentially outperform conventional denoising networks trained on generic datasets, including those used in my project.

7. Reference

[1] Abdelhamed, Abdelrahman, Stephen Lin, and Michael S. Brown. "A high-quality denoising dataset for smartphone cameras." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1692-1700. 2018.

[2] Heinrich, Mattias P., Maik Stille, and Thorsten M. Buzug. "Residual U-net convolutional neural network architecture for low-dose CT denoising." Current Directions in Biomedical Engineering 4, no. 1 (2018): 297-300.

[3] Xue, Haolan, Chang Liu, Fang Wan, Jianbin Jiao, Xiangyang Ji, and Qixiang Ye. "Danet: Divergent activation for weakly supervised object localization." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6589-6598. 2019.

[4] Liu, Yang, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. "Invertible denoising network: A light solution for real noise removal." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 13365-13374. 2021.

[5] Yue, Zongsheng, Qian Zhao, Lei Zhang, and Deyu Meng. "Dual adversarial network: Toward real-world noise removal and noise generation." In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16, pp. 41-58. Springer International Publishing, 2020.

[6] Murali, Vineeth, and P. V. Sudeep. "Image denoising using DnCNN: an exploration study." In Advances in Communication Systems and Networks: Select Proceedings of ComNet 2019, pp. 847-859. Springer Singapore, 2020.

[7] Lehtinen, Jaakko, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. "Noise2Noise: Learning image restoration without clean data." arXiv preprint arXiv:1803.04189 (2018).