NeRF As A Denoiser

Boyu Zhang, and Shubo Yang

Abstract—Conventional denoising methods usually utilize the information of one single photo, which has limited information of the noise distribution and the content. One approach called burst denoising tackles this problem by aligning and merging the information in multiple images. However, this integration of images are implemented in 2D space, which is still limited to the lack of depth information in 2D RGB images. Therefore, inspired by burst denoising, we explore using Neural Radiance Field (NeRF) as an solution to denoise taking multiple images of the same scene. NeRF uses uses multilayer perceptron (MLP) to represent the 3D scene and then uses volume rendering to render images. When minimizing the loss between the rendered images the ground truth, the noise is reduced in the 3D space, achieving better denoising performance. In this work, we implement NeRF as a denoiser, and compare the performance with several baseline methods using diverse image quality metrics.

Index Terms—Neural Radiance Field, Denoising

1 Introduction

Nowadays, photo denoising is gaining more attention, especially due to the pursuit of capturing the enhanced photo quality from mobile devices, such as cellphones. However, but the physical dimensions of the mobile devices image sensor limit their ability to capture noise-free images, necessitating advanced denoising algorithms. Modern camera pipelines use the bursting denoising [1] [2] [3] technique to enhance signal-to-noise ratio (SNR) by aligning and merging a sequence of underexposed photographs, thus achieving superior image clarity. However, it is only using the 2D information in the photo, limiting their accurate of noise distribution in 3D space.

Meanwhile, Neural radiance field (NeRF) [4] was proposed for novel view synthesis, which naturally lifts the 2D images into a 3D space and store the scene in neural networks. NeRF takes multiple images captured from the same scene, and contains MLPs to represent the color and volume density of every voxel in the space. The network's weights are optimized to encode the representation of the scene so that the model can easily render novel views seen from any point in space.

Therefore, NeRF has the potential of contain the 3D content information as well as the noise distribution, which leads to better performance when rendering to 2D images. The potential of NeRF as a denoiser can be seen from two aspects. First, NeRF takes multiple images to optimize the weights for scene representation makes it suitable of merging information in images in the 3D space. Additionally, NeRF will first implement camera orientation estimation, which implicitly aligns different photos.

Therefore, in this work, we propose to use NeRF as a denoiser. We capture our own dataset in a low-lighting noisy environment, and subsequently derive the ground truth. We implement NeRF as a denoiser, merging multiple images. Additionally, we comprehensively evaluate NeRF method and other baseline methods qualitatively and quantatitively. Our metrics show that NeRF-based denoising method performs better visually and in structural similarity (SSIM), and shares similar quality as off-the-shelf AI-based denoising methods.

2 Related Work

2.1 Denoising Algorithms

The fundamental principle underpinning denoising algorithms is the aggregation of similar pixels to mitigate noise. Traditional methods such as Gaussian and median filters adopt a straightforward strategy, averaging local neighbourhood pixels. The bilateral filter [5], on the other hand, introduces a more nuanced technique, averaging pixels in the local neighborhood while selectively incorporating those with akin intensities, thereby preserving edge integrity. Moreover, non-Local Means (NLM) [6] approach extends its reach beyond immediate neighbors. By identifying and averaging pixels within similar patches throughout the image, NLM is able to incorporate distantly related pixels that share neighborhood characteristics, enhancing denoising effectiveness. However, these methodologies are inherently constrained by their usage of single-image information.

2.2 Burst Denoising

Burst denoising [1] [2] ingrates the information in multiple images, by averaging similar pixel values across many frames. This method takes a series of photographs in quick succession, to avoid the motion blur. Through sophisticated frame alignment processes, the algorithm matches corresponding pixels across different frames. By averaging these matched pixels, the algorithm effectively reduces noise in the resulting image. However, challenges such as exposure variation and subject movement can complicate the process of frame alignment. One approach is focusing on the alignment of a series of underexposed shots. Additionally, the cumulative effect of adding these aligned frames together mimics the effect of a longer exposure, thus providing the balance between noise reduction and image quality enhancement inherent in burst denoising strategies.

2.3 Neural Radiance Field

A neural radiance field (NeRF) [4] is originally proposed for novel view synthesis. It uses a multilayer perceptron

(MLP) network to convert 3D position and viewing angles to volume density and color. This MLP network transforms 3D spatial coordinates and viewing angles into volume density and color attributes, which are then integrated through volume rendering techniques to produce a 2D image from any viewpoint. The process takes with a set of images and their corresponding estimated camera poses, where the images also serve as ground truth. The accurate camera pose estimation together with volume rendering enable that rays traced from the camera in different images map back to the volxels.

NeRF operates under the premise that all input images collectively represent ground truth data. To minimize rendering loss, it 'averages' the values of corresponding pixels across the entire input image set for any given point in space. This averaging mechanism implies that NeRF possesses an inherent advantage over burst denoising techniques in terms of denoising capabilities. It can assimilate information from a broader array of photographs, leveraging its capability to identify image pixel correspondences beyond the limitations of conventional 2D frame alignment.

3 Dataset

We collect our own dataset, and the pipeline is shown in Figure 1.

To get a ground truth photo of a static scene, we took 100 photos of a static scene. These captures employ the same parameters. By aggregating these photographs, we applied an averaging process that effectively neutralized random noise present in individual images. This resultant image serves as the ground truth for the scene, as illustrated in Figure 2. One noisy photo is shown in Figure3. This reference photo serves as a singular representation of the scene, containing the noise, experimented by further denoising techniques.

In addition to this noisy image, we capture other 99 photographs, each from varying camera poses while maintaining identical camera settings. This approach ensured the consistency in exposure, aperture, ISO, and other relevant settings. The parameters of the camera are shown in Table 1. The only difference is the diversity in viewpoints, which will be inputted into and merged by NeRF. Examples of the inputs to NeRF are in Figure 4. This multi-view noisy dataset owns uniform technical parameters but is also diverse in spatial orientation, providing a robust foundation for NeRF algorithm.

TABLE 1
Camera settings.

Shutter Speed	1/60s
Aperture	F11
ISO	102400

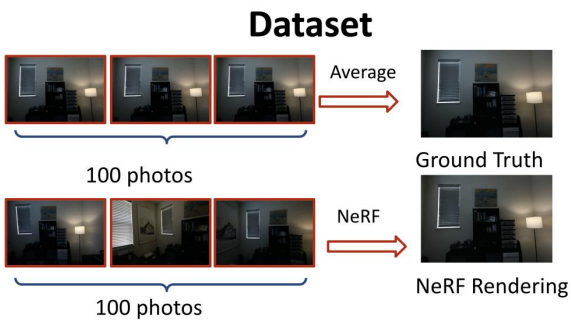

Fig. 1. The pipeline of deriving ground truth reference photo and photos for NeRF rendering.

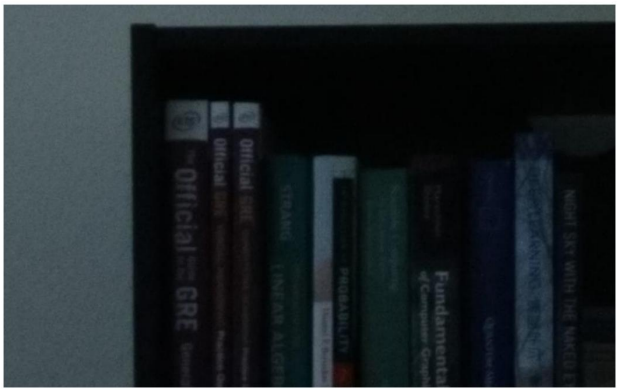

Fig. 2. Merged ground truth photo (cropped)

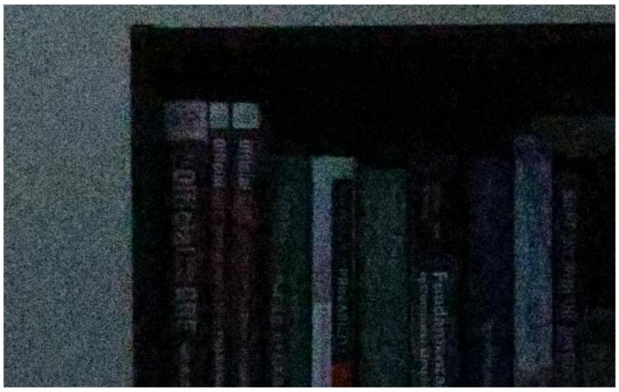

Fig. 3. The reference noisy photo (cropped)

Fig. 4. Sample images from NeRF Inputs

4 EXPERIMENTS

We implement the NeRF as a denoiser by inputting the noisy images into NeRF. The first design choice is the network. To experiment the NeRF process fast, we chose a NeRF model variant (nerfacto model in the nerfstudio framework) that does not need extensive training durations. Training this model approximately requires 5 minutes on an RTX 4090 GPU. This choice aligns with the objective of achieving a balance between computational efficiency and the advanced rendering capabilities of NeRF.

It can be observed that most NeRF models, including nerfacto, depend on Colmap for the estimation of camera poses based on the input photographs. The noise prevalent in these images posed a substantial hurdle, as Colmap struggled to accurately derive camera poses from such noisy data. To address this problem, baselines with a denoising preprocessing step was considered. Two denoising methods, the bilateral filter and Non-Local Means (NLM), were tested for their efficacy in enhancing photo clarity to a degree that would enable successful camera pose estimation by Colmap. The application of these denoising techniques allowed for the generation of viable camera poses. Thus, the camera poses utilized for the NeRF input were derived from the photos denoised using the bilateral filter. The outcomes of using both the bilateral filter and NLM methods on the input photos served as baseline comparisons. The investigation further extended to assessing the potential for improved results by denoising the photographs prior to the NeRF pipeline. This approach aimed to ascertain whether a sequential combination of initial denoising followed by NeRF processing could enhance the overall quality of the resultant imagery, thereby offering an baseline that leverages the strengths of both denoising techniques and the advanced rendering capabilities of NeRF.

Additionally, we compared AI-denoised image and traditional pipelines as baselines, including lateral filter and non-local means methods.

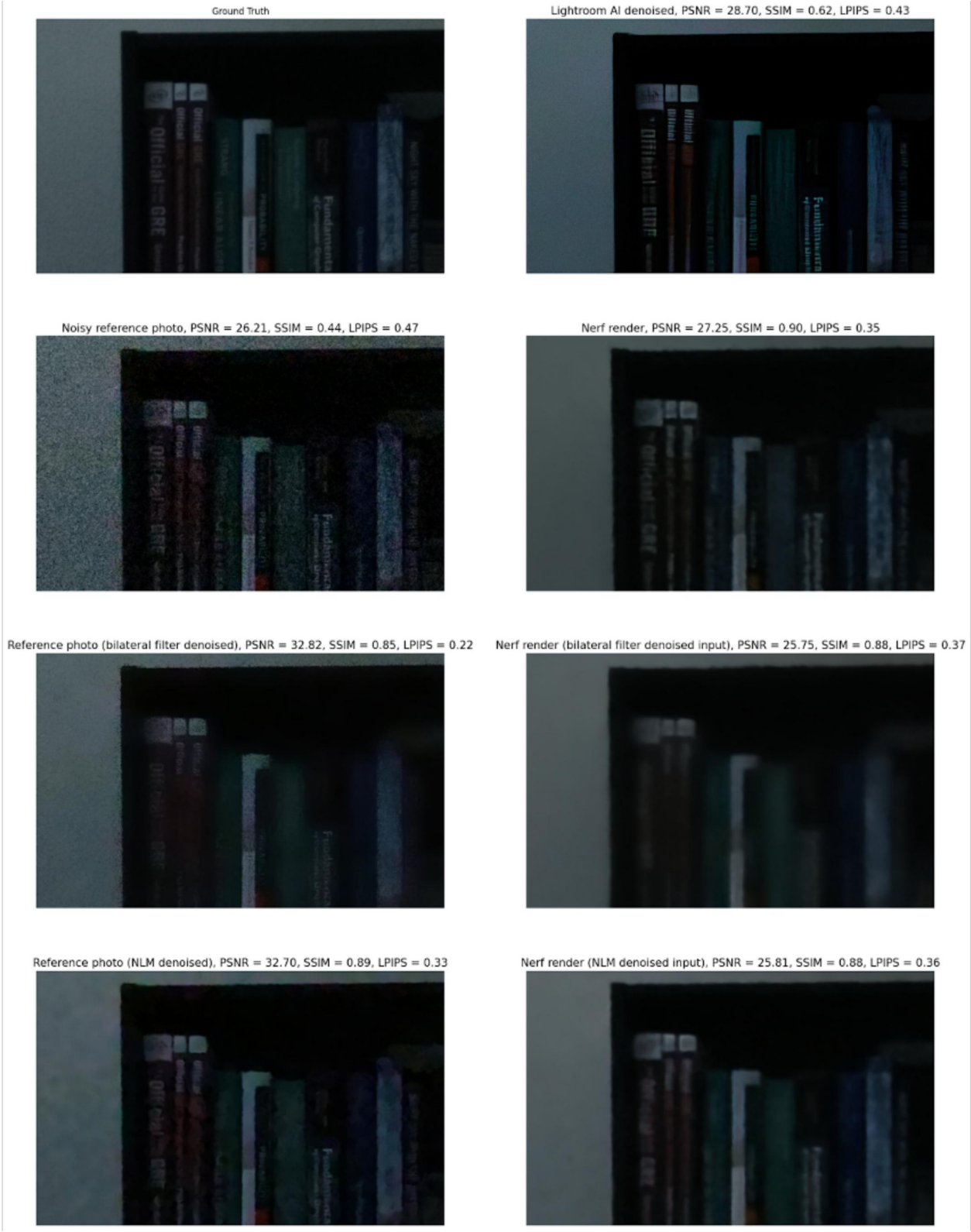

Fig. 5. Non-local Means denoised photo

TABLE 2
Qualitative metrics of different denoised images.

Method	PSNR	SSIM	LPIPS
Lightroom AI denoised	28.70	0.62	0.43
Noisy reference	26.21	0.44	0.47
Bilateral Filter	32.83	0.85	0.22
NLM	32.70	0.89	0.33
NeRF (Noisy Input)	27.25	0.90	0.35
NeRF (Bilateral Filter Denoised Input)	25.75	0.88	0.37
NeRF (NLM Denoised Input)	25.81	0.88	0.36

5 Experiments and Evaluation

5.1 Qualitative Evaluation

The qualitative result is shown in Figure 5. It can be seen that qualitatively, Lightroom AI-based denoising achieves similar performance as NeRF. Additionally, bilateral filter and NLM tend to produce blurry images.

Furthermore, NeRFs take denoised images perform worse than directly taking original inputs. This might due to the images denoised by traditional methods, such as bilateral filter, are distorted during the process. This makes NeRF harder to construct the scene. The orginal images with no denoising techniques may contain more information for NeRF to denoise.

5.2 Quantitative Evaluation

The quantitative evaluation is described in Table 2.

NeRF's denoising application excelled in the Structural Similarity Index Measure (SSIM), a metric designed to assess the visual impact of three characteristics of an image: luminance, contrast, and structure. SSIM's favorable results suggest that NeRF is particularly adept at preserving structural information in the denoising process, maintaining visual similarity in ways that align closely with human perception. This also aligns with our human qualitative evaluation.

Conversely, the performance of NeRF as a denoiser demonstrated a drop compared to traditional methods. This observation points to a potential distortion introduced by preliminary denoising algorithms, which might affect the quality or characteristics of the noise in a manner that diminishes the effectiveness of subsequent NeRF processing. Such pre-processed inputs could mislead the NeRF model, emphasizing the importance of raw data fidelity for optimal denoising outcomes.

In comparison, conventional denoising techniques outperformed NeRF in both the Peak Signal-to-Noise Ratio (PSNR) and the Learned Perceptual Image Patch Similarity (LPIPS) metrics. PSNR's sensitivity to luminance changes can skew its reliability as a comprehensive performance indicator, particularly in scenarios where luminance distortion is not the primary concern. LPIPS, on the other hand, offers a more nuanced reflection of perceptual similarity, suggesting that traditional methods may better capture the essence of the original image in certain respects.

The divergent performances across these metrics underscore the ongoing challenges in evaluating denoised images. The quest for the ideal balance between quantitative accuracy and qualitative perceptual fidelity remains at the forefront of image processing research. This complexity necessitates a multi-faceted approach to metric selection and emphasizes the need for further advancements in denoising technology. Future research could focus on developing or refining metrics that more accurately reflect human visual assessment, alongside exploring innovative denoising techniques that can seamlessly integrate with or enhance NeRF's capabilities.

5.3 Limitations

There are several limitations of this work. First, NeRF only works well in static scenes, without dynamic objects. This leads to huge limitation of applications of NeRF denoising. Second, this project is using jpeg as the NeRF input. Jpeg is a file format with compression. The compression may also lead to distortion of the noise distribution that can worsen the performance. Mildenhall et al. [7] proposed to use raw images as input to NeRF, which avoids the noise distortion in data compression. This is also a future work of us. Third, the metrics we used may not capture the metrics of our eyes. This is because the images that achieves the vest by visually evaluating are different from the by quantitative metrics.

6 Conclusion

In this work, we explore using Neural Radiance Fields (NeRF) as a novel denoising technique. Our approach was grounded in the creation of a manually captured dataset, from which we established a noise-free ground truth image. This foundation allowed us to apply NeRF technology specifically for the purpose of denoising, diverging from its traditional use in novel view synthesis. Through comprehensive assessments encompassing a range of metrics and comparative analyses against established baselines, we observed that NeRF-generated images exhibit commendable visual quality, showcasing the method's efficacy in enhancing image aesthetics and clarity. We also find a disparity in quantitative performance evaluations. Specifically, NeRF's application as a denoiser underperformed in certain statistical measures. This divergence underscores a critical insight into the complex nature of image denoising challenges, where the success of a technique can be multifaceted. Our work can serve as a start for future investigations to refine NeRF's application as a denoiser, potentially involving adjustments to its training or processing framework to better cater to the nuances of noise reduction. Moreover, this work sets the stage for further research into developing more metrics that can more accurately capture the essence of visual fidelity and noise suppression.

REFERENCES

[1] C. Godard, K. Matzen, and M. Uyttendaele, "Deep burst denoising," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 538–554.

[2] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun, "Fast burst images denoising," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–9, 2014.

[3] B. Mildenhall, J. T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll, "Burst denoising with kernel prediction networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2502–2510.

[4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[5] M. Zhang and B. K. Gunturk, "Multiresolution bilateral filtering for image denoising," *IEEE Transactions on image processing*, vol. 17, no. 12, pp. 2324–2333, 2008.

[6] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.

[7] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, "Nerf in the dark: High dynamic range view synthesis from noisy raw images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 190–16 199.