

GN-FR: Generalizable Neural Radiance Fields for Flare Removal

Gopi Raju Matta*
ee17d021@smail.iitm.ac.in
Brahul Siddartha
ee22m015@smail.iitm.ac.in

Rongali Simhachala Venkata Girish
girishrongali9@gmail.com
Sumit Sharma
ee20d042@smail.iitm.ac.in
Kaushik Mitra
kmitra@ee.iitm.ac.in

Computational Imaging Lab,
Dept. of Electrical Engineering,
Indian Institute of Technology Madras,
Chennai, India

Abstract

Flare, an optical phenomenon resulting from unwanted scattering and reflections within a lens system, presents a significant challenge in imaging. The diverse patterns of flares—such as halos, streaks, color bleeding, and haze—complicate the flare removal process. Existing traditional and learning-based methods have exhibited limited efficacy due to their reliance on single-image approaches, where flare removal is highly ill-posed. We address this by framing flare removal as a multi-view image problem, taking advantage of the view-dependent nature of flare artifacts. This approach leverages information from neighboring views to recover details obscured by flare in individual images. Our proposed framework, GN-FR (Generalizable Neural Radiance Fields for Flare Removal), can render flare-free views from a sparse set of input images affected by lens flare and generalizes across different scenes in an unsupervised manner. GN-FR incorporates several modules within the Generalizable NeRF Transformer (GNT) framework: Flare-occupancy Mask Generation (FMG), View Sampler (VS), and Point Sampler (PS). To overcome the impracticality of capturing both flare-corrupted and flare-free data, we introduce a masking loss function that utilizes mask information in an unsupervised setting. Additionally, we present the first-of-its-kind 3D multi-view flare dataset, comprising 17 real flare scenes with 782 images, 80 real flare patterns, and their corresponding annotated flare-occupancy masks. To our knowledge, this is the first work to address flare removal within a Neural Radiance Fields (NeRF) framework.

1 Introduction

The existence of flare artifacts can degrade image quality, potentially obscuring crucial object or scene details. Variations in lens characteristics, light source placement, and the camera's orientation towards the light source can result in diverse flare types. This diversity makes

flare removal an ill-posed problem. Recent works [10, 11, 12, 13] attempted to remove flare by utilizing the flare mask as a guiding cue in conjunction with paired data by using Unet and transformer-based architectures. Qiao *et al.*[12] utilized adversarial networks to learn from both flare and flare-free unpaired data by acknowledging the position of flare and light source. Lately, kotp *et al.*[13] leveraged depth information as a guide to remove flare. However, none of the aforementioned papers exploited the directional dependence of flare. We propose to utilize multi-view settings for flare removal, leveraging information lost due to flare in the current view from adjacent views. With the current framework, we can achieve novel view synthesis from images corrupted by flare and simultaneously eliminate the flare artifacts.



Figure 1: Our method: GN-FR is built on significant modifications to the GNT framework with flare masking loss, an effective view sampler and a Point Sampler. Our method consistently removes flare across the views on synthetic flare-imposed scenes and real scenes with flare.

Neural scene representations such as NeRF[10] can render realistic novel views, but they are limited by a tedious optimization process for each new scene and cannot generalize to new scenes. Recent works on generalizable neural rendering like IBRNet[11], PixelNerf[12], NeuRay[13] address these limitations by synthesizing a generalizable 3D representation through aggregation of image features extracted from seen views. We leverage GNT[14], a recent powerful method that utilizes transformers as aggregation functions to extract features from source views and to predict the target pixel color, eliminating the usage of the volumetric rendering equation. This makes GNT a reliable framework for flare removal alongside generating novel views. However, vanilla GNT cannot remove the flare as the network cannot learn to identify the flare region without sufficient cues. By leveraging flare occupancy masks as cues, we direct the network’s attention to the flare-affected regions, enabling it to learn how to remove flare and recover lost information from neighboring views effectively. This underscores our central concept that information obscured in one view can be extracted from adjacent views.

To this end, we propose our model GN-FR: Generalizable Neural Radiance Fields for Flare Removal, which is built on effective modifications to the GNT framework. Our model includes three sub-modules:

- 1) **Flare-occupancy Mask Generation network (FMG):** For each input view, it generates a binary mask to detect the regions w/ and w/o flare.
- 2) **View Sampler (VS):** Samples the views that are minimally affected by flare using the flare occupancy mask as a cue.
- 3) **Point Sampler (PS):** An effective attention masking-based sampling mechanism in the view transformer to guide the network to attend only to points that are valid in the flare occupancy mask.

As ground truth is not available for real scenes, we propose **Masking loss**, this will only attend to the pixels where there is no flare while calculating loss. Hence it ensures the rendering of the regions where the flare is prominent by extracting information from corresponding nearby views' regions where there is no flare.

In summary, our contributions include:

- A generalizable neural rendering network for flare removal that can render novel views and effectively remove the flare in complex scenarios.
- Masking loss, Flare-occupancy Mask Generator (FMG), View Sampler (VS) and Point Sampler (PS) modules to intelligently sample only flare-free regions in the source views for target view rendering.
- A first-of-its-kind 3D flare dataset of 17 scenes, together with real flare pattern dataset of 80 different flare patterns of real flare to enable our network to detect flare affected regions well and then to effectively de-flare them.

To the best of our knowledge, this is the first work that does flare removal and also produces novel views in a single framework.

2 Related Work

2.1 Flare Removal

Flare removal remained a challenging problem for a longer period due to its inherent complexity and diverse reasons causing flare. Despite these challenges, recent traditional [23] and deep learning [8, 9] based methods have shown decent progress. Wu *et al.*[15] proposed Unet-based flare removal using paired data and flare occupancy mask information, where the mask information is used to blend the input light source back into the prediction where the light source is removed during training. Qiao *et al.*[16] utilized the position of the light source and flare to remove the flare effectively. [8, 9] used Uformer[17] to remove flare. While the above methods primarily focus on single-image flare removal, they overlook the potential benefits of multi-view image flare removal. In multi-view approaches, information lost due to flare in one image can be potentially recovered from neighboring views, enhancing the overall quality of the reconstructed image.

2.2 Novel View Synthesis

NeRF introduced by [18] synthesizes consistent and photorealistic novel views by fitting each scene as a continuous 5D radiance field parameterized by an MLP(Multi-layer Perceptron). Several efforts have been made recently to improve efficiency and quality [1, 2, 7, 19] of NeRF. Some works have introduced lighting and reflection modeling [8, 12] to NeRF. Recently, Mildenhall *et al.* proposed RawNeRF[10], which can yield novel HDR views by training directly on linear RAW images.

2.3 Generalizable NeRF

The original NeRF[18] approach is limited to training a neural network for individual scenes, requiring optimization from scratch for each new scene without prior knowledge. Generalizable neural rendering methods overcome this constraint by training on multiple scenes, allowing the network to develop a broader understanding of synthesizing novel views by

leveraging source observations. Earlier methods acquire $\text{RGB}\sigma$ of each point by weighted summing of the image features of its 2D projections [19] and through cost volume induced by MVSNet [2]. To enhance generalization capabilities and rendering quality, recent approaches have incorporated transformer-based architectures [8] to aggregate features from source images [2, 15]. IBRNet [16] uses MLP conditioned on feature vectors extracted from the source images to predict color and radiance values, which are aggregated using volumetric rendering and a ray transformer to compute densities along the camera ray. GNT [13] uses transformers for the entire pipeline. However, the above methods cannot remove flare while generating novel views.

3 Preliminary

Generalizable Neural Radiance Fields: Vanilla NeRF [11] represents the novel views using volumetric rendering, which is time-consuming and produces artifacts while rendering reflective and metal surfaces. GNT [13] overcomes these issues by considering the problem of novel view synthesis as a two-stage information process: the multi-view image feature fusion, followed by the sampling-based rendering integration. It is composed of (i) *view transformer* to aggregate pixel-aligned image features from corresponding epipolar lines to predict coordinate-wise features, (ii) *ray transformer* to compose coordinate-wise point features along a traced ray via attention mechanism. More formally, the entire operation can be summarized as follows:

$$\mathcal{F}(\mathbf{x}, \theta) = \text{View-Transformer}(\mathbf{F}_1(\Pi_1(\mathbf{x}), \theta), \dots, \mathbf{F}_N(\Pi_N(\mathbf{x}), \theta)) \quad (1)$$

where $\text{View-Transformer}(\cdot)$ is a transformer encoder, $\Pi_i(\mathbf{x})$ projects position $\mathbf{x} \in \mathbb{R}^3$ onto the i^{th} image plane by applying projection matrix, Π onto \mathbf{x} and $\mathbf{F}_i(\mathbf{z}, \theta) \in \mathbb{R}^d$ computes the feature vector at position $\mathbf{z} \in \mathbb{R}^2$ via bi-linear interpolation on the feature grids. The multi-view aggregated point features are fed into the ray transformer, and the output from the ray transformer is fed into a view transformer and this process is repeated where the view transformer and ray transformer are stacked alternatively. Then the features from the last ray transformer are pooled to extract a single ray feature to predict the target pixel color.

$$\mathbf{C}(\mathbf{r}) = \text{MLP} \circ \text{Mean} \circ \text{Ray-Transformer}(\mathcal{F}(\mathbf{o} + t_1\mathbf{d}, \theta), \dots, \mathcal{F}(\mathbf{o} + t_M\mathbf{d}, \theta)) \quad (2)$$

where t_1, \dots, t_M are uniformly sampled between near and far plane, Ray-Transformer is a standard transformer encoder.

4 Method: Generalizable Neural Radiance Fields for Flare Removal

Overview: We introduce GN-FR to perform generalizable novel view synthesis on flare-corrupted images. Given a set of N input views with known camera parameters $\{(I_i \in \mathbb{R}^{H \times W \times 3}, P_i \in \mathbb{R}^{3 \times 4})\}_{i=1}^N$, corrupted with flare, our goal is to synthesize novel views from arbitrary camera positions and also generalize to new scenes.

Our pipeline has three stages, as shown in Fig. 2. Specifically, given a set of N flare input images, GN-FR utilizes a UNet with residual connection to encode each image to a feature map, expecting it to extract shading and local/global complex light transport information via its multi-scale architecture. We employ **FMG**(Flare-occupancy Mask Generator) (Sec. 4.1)

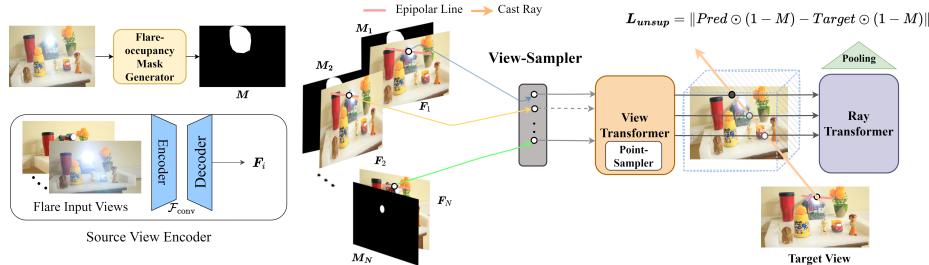


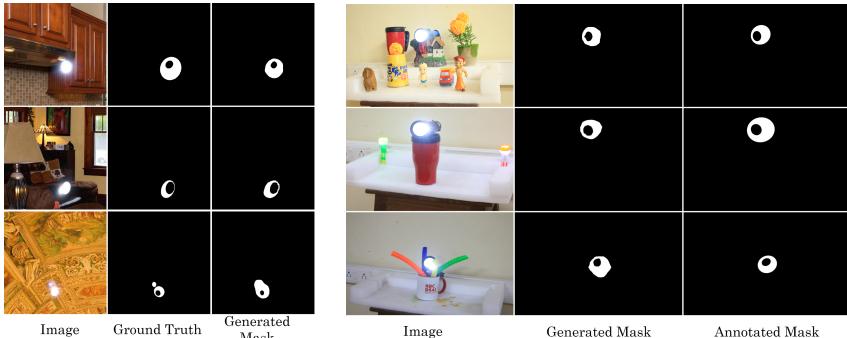
Figure 2: Overview of GN-FR: 1) FMG: Flare-occupancy Mask Generation network: It generates a binary mask (M) that contains zero values wherever there is no flare. Our FMG is capable of discerning flare regions effectively, excluding areas corresponding to the light source. This mask is used as a cue to guide the flare removal at regular stages of the network.

to generate a mask (M) that contains zero values wherever there is no flare. Our FMG is capable of discerning flare regions effectively, excluding areas corresponding to the light source. This mask is used as a cue to guide the flare removal at regular stages of the network.

Following GNT[2], our method projects x (position (x,y,z)) onto every source image and interpolates the feature vector on the image plane to obtain the feature representation at that position x . To render a target view, vanilla GNT[2] selects 8-10 nearby source views based on their relative distance. We modify this via our **View Sampler** (Sec. 4.2), which samples only those nearby views based on the flare occupancy mask information; typically, the source views where the percentage of flare is relatively less from other source views. We then use a **Point Sampler** (Sec. 4.2) imbued View transformer to aggregate all the feature vectors by masking the points affected by flare while computing attention. Such aggregated features are passed through the ray transformer (Eq.2) to render the final color. The network is trained end-to-end using the **Masking loss** (Sec. 4.3) in an unsupervised fashion.

4.1 Flare-occupancy Mask Generation

We use PSPNet [2], a semantic segmentation model, to generate the flare occupancy mask. We captured 80 different real flare patterns in a dark setting by showing torch light in multiple directions. We manually annotated these flare patterns to get binary flare-occupancy masks to highlight flare and flare-free regions. These flare patterns are imposed on the 24K Flickr image dataset[21] with a random affine transform on each flare image following the Flare7K methodology and the same affine transform is applied to annotated flare-occupancy mask to generate corresponding ground-truth flare-occupancy mask. This random affine transform simulates diverse flare patterns in multiple directions at various spatial locations. We used this dataset : flare images with their corresponding flare-occupancy masks to train PSPNet. We also observed that the inherent class (two classes : 1) regions w/ flare, 2) regions w/o flare) imbalance problem which further hinders the model's performance: regions affected by flare in an image is far less compared to flare-free regions, which creates class imbalance issues during the training process. So, we empirically assigned weights in the ratio 5:1 to the respective classes in calculating the binary cross-entropy loss during training with other settings unaltered. With these modifications, we are able to generate flare-occupancy mask to detect the flare regions considerably well on both synthetic images with quantitative metrics of 0.81/0.94 (mIoU/mAcc) and on real images, which is evident from the results of our proposed FMG module on both the synthetic and real datasets in Fig. 3.



(a) Synthetic Dataset

(b) Real Dataset

Figure 3: **Results of Flare-occupancy Mask Generation network (FMG):** on (a) synthetic datasets and (b) real datasets. Corresponding manually annotated masks for the real dataset in (b) match with the learned masks generated by our FMG.

4.2 View and Point Sampling Modules

We proposed two sampling modules to effectively sample source, target views and points on each ray while rendering each target view to accurately sample the information from only flare-free regions to enhance scene representation.

4.2.1 View Sampler

As seen from our captured real sample dataset, Fig. 3, some images are highly corrupted with flare. Utilizing any of these source images for rendering a target view introduces flare artifacts, hence compromising the scene representation. Therefore, we adopt a strategy of selecting nearby source views that are less impacted by flare using flare occupancy mask as a guide. Our sampling strategy 1) We exclude the image as a target view from training, if its flare occupancy(% of flare effected region) is beyond 10%, so this view can be rendered as a novel view during inference. 2) For rendering each target view, we only sample the nearby source views which are less effected by flare using flare occupancy information from flare occupancy mask. These two view sampling strategies ensure that network weights are not influenced by the largely effected flare views during optimization.

4.2.2 Point Sampler

While rendering ray, we sample points from corresponding epipolar lines of sampled source views. These sample points which fall on flare effected region will sample flare, hence it corrupts the target view while rendering. To nullify the effect of these flare sample points, we make the attention values due to these sample points zero in view transformer. This technique ensures that we sample almost no information from flare affected regions. Hence, scene representation by GN-FR is mostly unaffected by the flare. Our updated attention in view transformer is a simple multiplication with mask.

$$A' = A * (1 - M) \quad (3)$$

4.3 Flare Occupancy Mask Guided Loss

Capturing a real dataset with and without flare is highly impractical to train our GN-FR module, we leverage an unsupervised training strategy (we don't have ground-truth image

w/o flare) by constructing flare occupancy mask guided loss, where we multiply the flare-occupancy mask element wise to both the target image and its corresponding rendered image and then apply MSE loss; this ensures that the regions w/ flare in target image are rendered from corresponding regions w/o flare from the source views as like a novel region of the target view. Our idea follows as NeRF renders a novel views, it can also render novel regions in the views. This presupposes that at least one of the source views remains unaffected by flare in the target region while acknowledging flare presence in the specified area. As flare is view-dependent, this condition is mostly valid; a portion of the scene is unaffected by flare, at least in one of the images.

$$L_{unsup} = \|Pred \odot (1 - M) - Target \odot (1 - M)\| \quad (4)$$

5 Experiments

5.1 Settings

Following GNT, we construct a dictionary of each target view with its sampled source views for training by first selecting a target view, and then identifying a pool of $k \times N$ nearby views, from which N views are sampled as source views following the View Sampler strategy (Sec. 4.2.1). This sampling strategy ensures very minimal information from flare regions is passed through the network while rendering a target view. During training, the values for k and N are uniformly sampled at random from (1, 3) and (8, 12) respectively. From these views, N views are sampled at the end following the View Sampler strategy.

5.1.1 Our Proposed Flare Dataset



(a) Our Captured Flare Dataset



(b) Our Captured Flare Patterns

Figure 4: (a) Samples from our captured real flare-3D dataset of 17 scenes and (b) Samples from our captured real flare pattern dataset of 80 images, consisting of diverse flare patterns (both reflective and scattering).

We captured 80 different flare patterns with a light source in a dark setting as shown in Fig. 4 and then imposed these real flares on the 24K Flickr dataset with a random affine transform on each flare pattern image, this account for diverse flare patterns in multiple directions and orientations. Also, we collected a real flare 3D dataset of 17 scenes, comprising of 782 images to validate our method on real dataset. These are manually annotated with MATLAB using image segmentation application to generate flare occupancy mask. Please refer to the Fig. 3 to the samples from our dataset.

5.1.2 Training/Inference Details

We train the FMG following the default settings of PSPNet with modified weights of [5,1] (Sec. 4.1) to binary cross-entropy loss function. GN-FR is trained an end-to-end fashion on generated flare dataset with corresponding masks from 44 scenes of IBRNet dataset using the Adam optimizer to minimize the mean-squared error between predicted and target RGB pixel values using masking loss. For all our experiments, we train for 500,000 iterations with 256 rays sampled in each iteration with initialization from GNT. To test on real scenes, for better results our model is finetuned for 50k iterations. We run all our experiments on NVIDIA RTX 3090, 24GB. Unless otherwise stated, other settings of our model remains the same as that of GNT.

5.2 Quantitative Evaluations

For quantitative evaluation, as real dataset w/ flares and w/o flares is not available, we took IBRNet dataset, synthetically imposed the flares in random fashion and then send those to our network. As ours is the first work which does flare removal using NeRF, we compare our work with two methods, one is scene specific which is vanilla NeRF[] and the other is generalizable one, which is GNT. We first de-flare the images using Flare7K++ (existing SOTA flare removal technique) and then sent them to both NeRF and GNT for proper comparison. As we see from the Table 1, our method is producing superior results in all the metrics, surprisingly having 4dB PSNR improvement over the second best method.

Table 1: Quantitative results on real scenes with synthetically imposed flare. The **best** scores and **second best** scores are highlighted with their respective colors.

Models	Cross-scene Generalization	PSNR↑	SSIM↑	LPIPS↓
NeRF[]	✗	21.92	0.827	0.045
GNT[]	✓	22.44	0.867	0.0386
(Flare7K++)[]+(NeRF)[]	✗	18.63	0.769	0.052
(Flare7K++)[]+(GNT)[]	✓	18.85	0.812	0.056
Ours	✓	26.18	0.8815	0.034

5.3 Qualitative Results

For qualitative evaluation, we tested our method on both synthetic and captured real datasets with methods (Flare7K++)+(GNT) and (Flare7K++)+(NeRF). As we see from the Figs. 5 and 6, our method is working well both on synthetically imposed flare scenes as well on real scenes. Whereas other methods are having difficulty in removing flare consistently, either they leave artifacts or introduces artifacts in flare-free regions. As reported in Fig. 7 our method is able to remove the flare artifacts and restore the lost scene content effected due to flare very well on the captured real dataset. For more results on synthetic as well real datasets, please refer supplementary material.

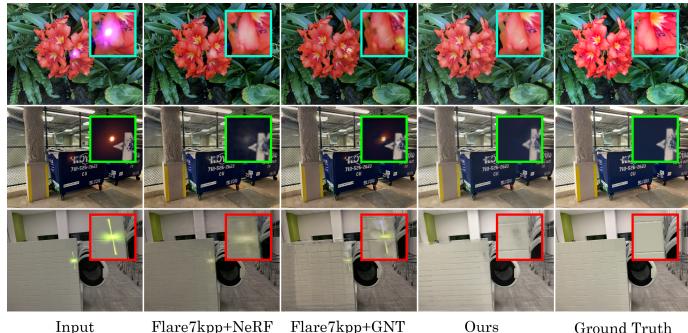


Figure 5: Qualitative results on Synthetic dataset: GN-FR framework is able to consistently remove the diverse flares present in the synthetically imposed flare images. As the insets highlight, GN-FR surpasses other methods in removing flare artifacts.

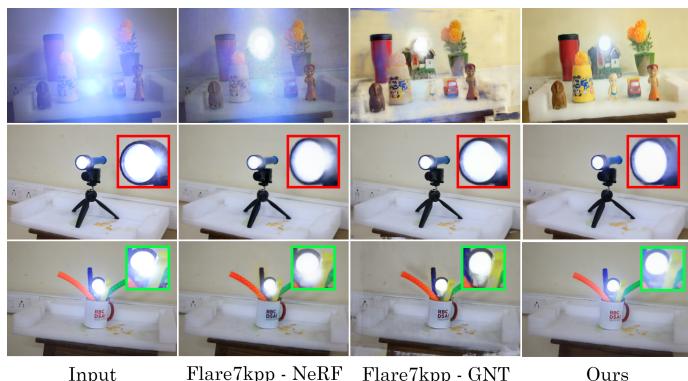


Figure 6: Qualitative results on Real dataset: Our GN-FR framework is able to remove the flare artifacts from flare-corrupted images effectively. It is able to render more visually appealing results consistently compared to other methods.

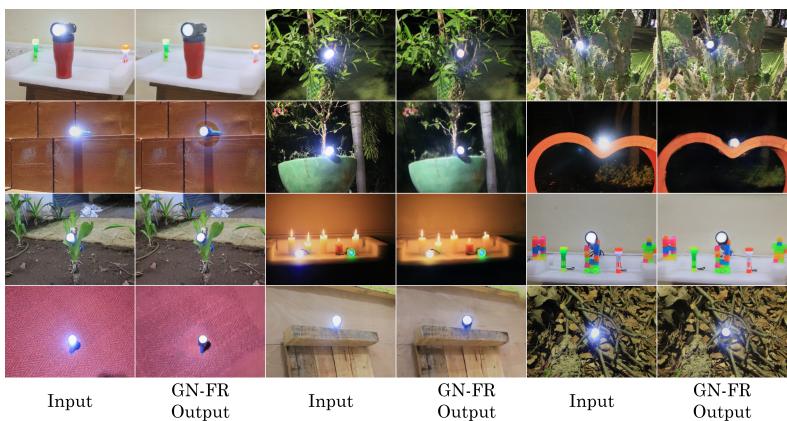


Figure 7: Flare removal results on captured Real dataset: GN-FR is able to consistently remove the flare-artifacts on different scenes of our real flare-3D dataset and also reproduce the lost scene information covered up due to flare effects(Zoom-in for better visualization).

5.4 Ablation Study

Without View Sampler (VS), Point Sampler (PS) modules and generated masks: Our contributions include masking loss, VS and PS modules and flare occupancy masking network. So we did the ablations without VS and PS modules, as we can see from Fig. 8, inclusion of PS and VS modules contribute to effective removal of flare artifacts, where flare is significant. As we see from the results, our method including all gives the SOTA results. Also, render results with generated flare occupancy masks are on par with annotated masks, Fig. 8.



Figure 8: **Effectiveness of Learned mask together with PS and VS modules:** Proposed modules(PS & VS) contribute positively to render the target views with very minimal flare artifacts. Rendered target views with learned masks are almost on par with annotated masks.

6 Conclusion

We present GN-FR, an innovative unsupervised neural rendering framework designed to effectively remove flares by leveraging multi-view flare-free information from neighboring source views. Our approach minimizes flare artifacts in the target view, achieving superior performance. We collected real flare pattern dataset to generate flare occupancy mask which will attend effectively to the flare-free regions while rendering. Also, we contribute a 3D flare dataset with ground-truth flare occupancy mask annotations. Our model demonstrates significant effectiveness in flare removal across both synthetic and real-world scenarios, setting a new state-of-the-art in both quantitative and qualitative assessments. To our knowledge, GN-FR is the first framework to utilize multi-view information for enhanced flare removal, opening new avenues for research in this field. Future work will aim to extend our framework to handle diverse flare types and explore zero-shot generalization on real-world scenes.

Acknowledgement: This work was supported in part by IITM Pravartak Technologies Foundation.

References

- [1] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864, October 2021.

- [2] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14124–14133, 2021.
- [3] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser-Nam Lim, and Abhinav Shrivastava. NeRV: Neural representations for videos. In *NeurIPS*, 2021.
- [4] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yihang Luo, and Chen Change Loy. Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. *arXiv preprint arXiv:2306.04236*, 2023.
- [5] Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20783–20791, 2023.
- [6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [7] Tao Hu, Shu Liu, Yilun Chen, Tiancheng Shen, and Jiaya Jia. Efficientnerf efficient neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12902–12911, 2022.
- [8] Yousef Kotp and Marwan Torki. Flare-free vision: Empowering uformer with depth insights. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2565–2569. IEEE, 2024.
- [9] Yuan Liu, Sida Peng, Lingjie Liu, Qianqian Wang, Peng Wang, Christian Theobalt, Xiaowei Zhou, and Wenping Wang. Neural rays for occlusion-aware image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7824–7833, 2022.
- [10] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.
- [11] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan 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*, pages 16190–16199, 2022.
- [12] Xiaotian Qiao, Gerhard P Hancke, and Rynson WH Lau. Light source guided single-image flare removal from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4177–4185, 2021.
- [13] Mukund Varma T, Peihao Wang, Xuxi Chen, Tianlong Chen, Subhashini Venugopalan, and Zhangyang Wang. Is attention all that neRF needs? In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=xE-LtsE-xx>.

-
- [14] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022.
 - [15] Dan Wang, Xinrui Cui, Septimiu Salcudean, and Z Jane Wang. Generalizable neural radiance fields for novel view synthesis with transformer. *arXiv preprint arXiv:2206.05375*, 2022.
 - [16] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2021.
 - [17] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.
 - [18] Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T Barron. How to train neural networks for flare removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2239–2247, 2021.
 - [19] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021.
 - [20] Dafeng Zhang, Jia Ouyang, Guanqun Liu, Xiaobing Wang, Xiangyu Kong, and Zhezhu Jin. Ff-former: Swin fourier transformer for nighttime flare removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2824–2832, 2023.
 - [21] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4786–4794, 2018.
 - [22] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.
 - [23] Yuyan Zhou, Dong Liang, Songcan Chen, Sheng-Jun Huang, Shuo Yang, and Chongyi Li. Improving lens flare removal with general-purpose pipeline and multiple light sources recovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12969–12979, 2023.