FIRE: Frequency Image Relighting Enhancement

Tianyang Hu, Jie Zhang, An Cao

Abstract—The image relighting task focuses on transferring a given input image to a desired new illumination condition. The relighting topic has been studied for years, but a new subdivision of the topic, single-image relight, has gotten more attention. Unlike sampling-based relighting which requires multiple input images of the same scene to perform relighting, the single-image relighting attempts to use only one input image. Although plenty of architectures have been proposed for this task, many of them suffer from incorrect or insufficient lighting condition predictions. For this project, we test if adding frequency domain feature analysis to the relighting architectures can help with its performance and result. As an early study on this subject, we find that adding specifically targeted frequency analysis can decrease the training time and increase the qualitative result of the architecture.

Index Terms—Computational Photography, Image Relighting, Deep Learning

1 Introduction

Relighting is the process of changing or simulating the lighting conditions of a scene or object in an image while maintaining other attributes such as geometry, texture, and reflectance. [Fig 1] This application has great potential in fields like photography, film production, gaming, and virtual reality, where realistic lighting effects are of significance.

Lighting condition in this context means the particular characteristics of light applied to a scene or object. It includes the direction, intensity, and spectral distribution of the source, as well as the resultant effects these properties create in the scene, such as patterns of overall illumination and shadowing.

In images, the overall illumination and shadow patterns correspond to the low-frequency components in the spatial domain, which varies slowly across a large area of the image in terms of intensity or color. Capturing and manipulating these low-frequency features is important for accurately simulating or transferring lighting in relighting tasks.

The Fourier transform provides a way of isolating the low-frequency information in the frequency domain efficiently. We want to find out if introducing Fourier-based techniques can force the network to focus more on the low-frequency details of the image and in the end improve the quality of output image and the efficiency of the network.

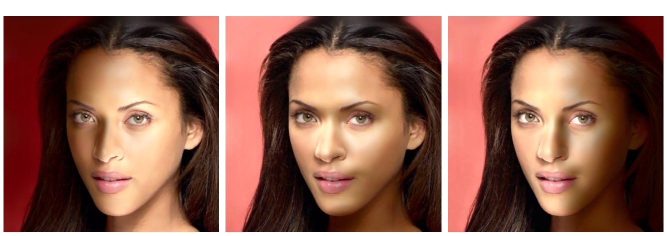

Fig. 1: Visual effects of relighting task.

2 Task Description

In this project, we address a specific fixed-to-fixed relighting task, where both the input and output images are captured under fixed external lighting conditions [Fig 2]. This constraint arises from the limited availability of training data. Although the external lighting conditions (such as sunlight or overcast skies, where the light source is not visible in the scene) remain consistent, variations may still occur within the scene itself due to light-emitting sources like fire or lamps. These internal lighting sources introduce localized differences in illumination while preserving the overall fixed lighting condition.

We use only RGB information of images for training, without auxiliary data such as depth maps, surface normals, or other scene descriptors. This is because many real-world applications, such as photography and image editing, have only RGB information.

3 Related Work

3.1 Image Relighting

The image relighting task is not a new topic in computer vision, and the approaches to address this task have been evolving through years of studies in the community. In the early stage, the task of image relighting takes a sampling approach [1], where multiple images of the same scene under different illumination conditions will be taken and the desired new illumination will be generated by interpolating those images. Later improvements [2] on this topic focus mainly on decreasing the image samples we need to generate the new desired image.

In recent years, the community continuous this focus on decreasing the sampling size with the inspiration from the image-to-image transformation [3]. There are various methods [4] afterwards use the idea of image-to-image transformation and successfully design a network to fulfill the relighting task using only images under 8 different illumination conditions as input.

(a) Input for Image 1 (b) Target for Image 1

(c) Input for Image 2 (d) Target for Image 2

Fig. 2: Fixed-to-fixed lighting conditions in training data

Currently, most of the work is focusing on the single-image relighting task where the input is now only one image. The most common approach is to use auxiliary information such as depth maps or image normals to support the network to determine the geometry of the scene [5] [6]; however, similarly to the sampling method, auxiliary information is not always accessible through real applications.

Several studies have been done to remove the need for auxiliary information. Wang et.al. [7] experiments on using a deep residual network for relighting (DRN). The algorithm had outperformed all other competitors at the time, but the network suffered significantly in determining the difference between material color and illumination condition.

Our work will be based on Illumination-Aware Network (IAN) [8] which is improving the original DRN network by having the network focus on the illumination condition. In this paper, we will try to use frequency domain features to improve IAN. Details about the IAN and how we will improve it will be discussed in Section 3.

3.2 Frequency Domain Information

Frequency domain analysis has been widely used in image-related areas. Various methods for different tasks use frequency features to enhance their performance, especially in illumination-evolving tasks. For example, Vasluianu *et al.* [9] uses a frequency-based network to perform ambient light normalization and shadow removal. Hu *et al.* [10] also uses the frequency features to support generating the HDR image from LDR images. Last but not least, Wang *et al.* [11] uses the frequency domain analysis to boost the low-light image enhancement task.

The relighting task we will improve on involves direct manipulation of the illumination condition of the image; however, this task is still a new topic in computational imaging with a few studies on whether adding specific frequency domain features extraction can enhance its performance. In the rest of the paper, we will perform some early studies on this subject.

4 Proposed Method

We use the Illumination-Aware Network (IAN) [8] as a baseline model and add the frequency domain feature onto its architecture. The pipeline of the IAN can be seen in Figure 3.

The encoder and decoder here follow the basic U-Net [12] structure where CNN and ReLu layers are applied with a filter size of 24. The basic U-Net structure is a very general structure that can be used in various different visual tasks; however, due to the ill-posed nature of the relighting task, we would like to modify the encoder to also include frequency domain features and test if this can help determine the illumination condition (see section 3.1 and 3.2).

As described in the original paper, the Illumination Aware Residual Block (IARB) is used to capture the illumination condition by simulating a physic-based rendering process using dilate blocks and spatial attention blocks. The IARB will be trained to capture an illumination descriptor that can be used to offset the input illumination to the output illumination. The detail of the implementation on why spatial attention blocks are used here to capture illumination is not specifically described in the paper, so we would like to see whether switching to a frequency attention block can improve the ability of IARB to capture illumination conditions (see section 3.3).

The IAN uses a pyramid-shaped structure where the Encoder-IARB-Decoder section has been applied to the original input, a half-size down-sampled input, and a quarter-size down-sampled input. As described in the original paper, this structure is used to better detect the illumination condition in different scales. Note that we separate the Encoder-IARB-Decoder section for each input size in the figure for better visualization, but the same Encoder-IARB-Decoder section will be used for all input sizes in the actual implementation.

4.1 Fourier Channel

Fourier Channel utilizes the Fourier transform to enhance the representation of input image data for lighting transformation. As is illustrated in Figure 4, the method begins with the applying Fourier Transform to each of the three RGB channels of the input images, projecting them into Fourier space. This transformation allows the model to capture frequency information within images that is also important for relighting tasks. The output Fourier space tensors are then concatenated with the RGB tensors to form six-channel tensors, which integrate both spatial and frequency-domain features. These enhanced tensors serve as the input to the original model, enriching its capacity to analyze and process lighting-related features [13]. To ensure compatibility with these modified inputs, the model's pre-processing module is adjusted accordingly.

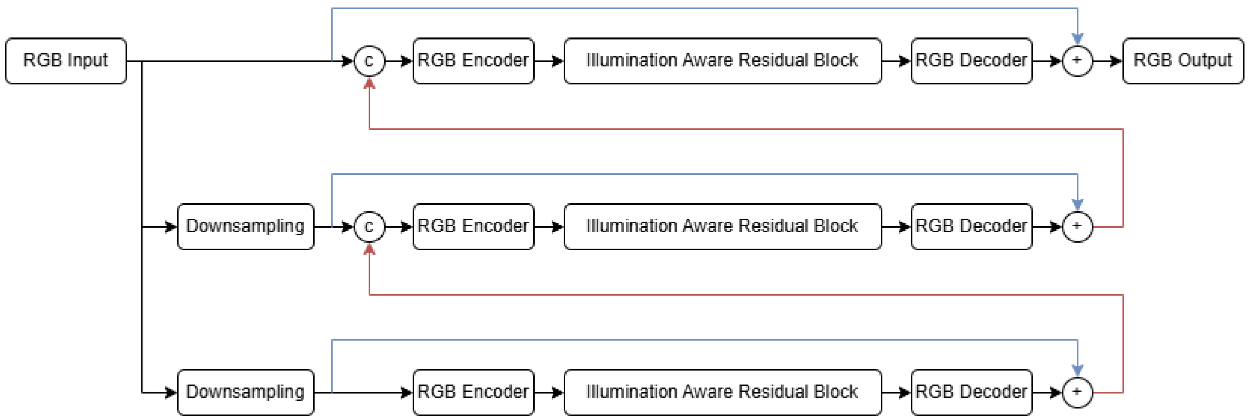

Fig. 3: Structure of the Illumination-Aware Network

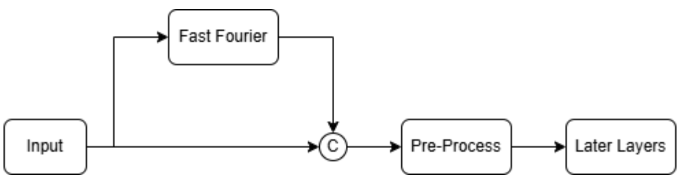

Fig. 4: Fourier Channel's Pipeline

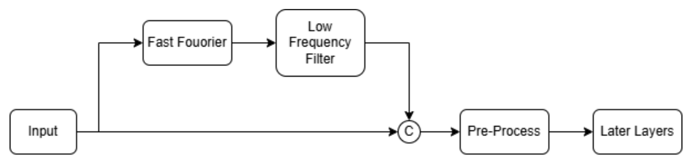

Fig. 5: Fourier Filter's Pipeline

4.2 Fourier Filter

Fourier Filter is another Fourier-based relighting method we propose, which is built based on Fourier Channel. We designed this method in order to enhance the performance by making the model focus on low-frequency information, which is more significant than high-frequency representations in relighting tasks. The pipeline of this method is presented in Figure 5. In the beginning, this method also applies Fourier Transformation to each RGB channel of input images, projecting them from the spatial space to the frequency space. However, not all information in the frequency contributes identically to relighting tasks. Since low-frequency information carries crucial information related to global lighting conditions and shadows, and high-frequency components typically just represent details of images, low-frequency are more relevant and crucial for relighting tasks [14], [15]. To address this issue, a low-frequency filter is introduced to the Fourier-transformed tensors. This filter selectively removes high-frequency components and preserves the low-frequency parts, which are important for relighting. The transformed and filtered tensors, which contain the image's low frequency information, is then concatenated with the original RGB tensors, creating a tensor that contains both spatial and frequency information. The filtered tensors, rich in essential lighting information, are then passed through a modified pre-processing module. Since its' shape different from that of the original input, the pre-processing module modifiy the input's shape into an expected shape for later layers, ensuring the integration of Fourier Filter with the original model's architecture. Once processed, the concatenated tensors containing the spatial and low-frequency information, are fed into the following layers of the model for further processing and feature extraction.

4.3 Frequency Encoder

Another modification we will be testing is the usage of a frequency encoder. The idea of the frequency encoder originates from the Low-Light Image Enhancement (LLIE) task which is also an illumination-related task. Among the recent studies on LLIE, FourLLIE [11], which takes advantage of using frequency domain features, has shown a promising result when compared with other State-Of-Art methods; therefore, we choose to apply a similar approach in the relighting task here.

Similar to FourLLIE we will use a multi-branch encoder structure to modify the original method (see Figure 6), where the frequency encoder will operate in parallel with the original spatial RGB encoder. Note that the frequency encoder will also be used for all input sizes in the pyramid-shaped structure of our original method to better address the frequency information in different scales.

The design of the frequency encoder is similar to the FourLLIE but with some modifications. The frequency domain information can be separated into amplitude and phase, where amplitude corresponds to lighting intensity and phase represents more about lighting structures [16]; thus, in the FourLLIE the frequency encoder has been divided into amplitude and phase branches that are convolving separately since the low-lighting situation needs to focus more on the lighting intensity.

In the relighting task, the amplitude and the phase are both relevant to the illumination condition and have a certain correlation with each other. We therefore design our frequency encoder to represent this correlation by convolving the frequency domain image using a three-by-three filter.

4.4 Fourier Attention mechanism

In the original method, a spatial attention mechanism is used to force the network to focus on important spatial

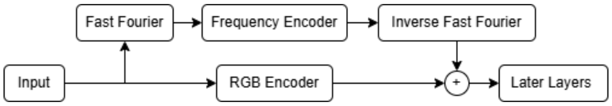

Fig. 6: Fourier Encoder Pipline

Model	Parameters (M)	FLOPs (G)
Original [?]	0.6070	53.89
Fourier Channel	0.6076	53.93
Fourier Filter	0.6076	53.93
Fourier Attention Map	0.5992	54.06
Fourier Encoder	0.6707	65.02

TABLE 1: Sizes of Models

patterns of the input. This mechanism uses measures of means and standard deviations across spatial dimensions as input, combined with lightweight convolutional layers to modulate the importance of specific pixel regions. While this is effective for capturing spatial patterns, it tends to focus more on localized spatial features and may not be effective in capturing global or repetitive structures.

To improve that, we add a Fourier-based attention mechanism that transforms the input features into the Fourier domain, enabling the model to focus on important frequencies. [17] The Fourier attention mechanism computes an attention map directly from the magnitude of the frequency components, modulates the importance of the frequency representation, and then reconstructs the enhanced features using the inverse Fourier Transform. This modification allows the model to capture both global structural information and localized textures, making it well-suited for relighting tasks.

However, due to the limitation of computational resources, we finally replaced the combination of spatial attention mechanism and frequency attention mechanism with the pure frequency attention mechanism. [Fig 7]

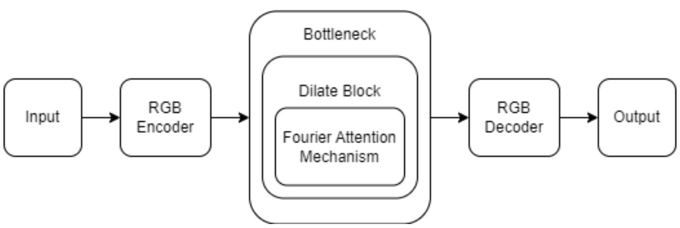

Fig. 7: Structure for Fourier Attention Mechanism.

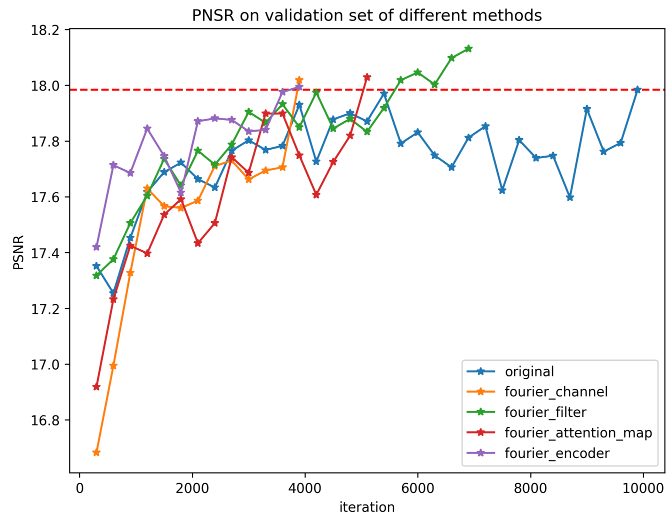

Fig. 8: PSNR vs Iteration

Model	PSNR	Iteration
Original [?]	17.9837	9900
Fourier Channel	18.0184	3900
Fourier Filter	18.1311	6900
Fourier Attention Map	18.0287	5100
Fourier Encoder	17.9945	3900

TABLE 2: Performance of Models

5 Experimental Results

5.1 Dataset

In order to highlight the relighting performance of four modified Fourier-based methods we propose and compare them with the original model's performance, we train Fourier Channel, Fourier Filter, Fourier Attention Map, Fourier Encoder as well as the original model on the training set of VIDIT ECCV 2020 AIM dataset [18], [19]. Since ground truths of the test set of VIDIT ECCV 2020 AIM dataset are not open to the public, we choose to test the performance of all these models on the validation set.

The training set of VIDIT ECCV 2020 AIM includes 300 items, and the validation set includes 45 items. Each of these items consists of one input and one ground truth, which are images of the same scene. All inputs in the dataset share the light situation, and all ground truth share another light situation.

In our experiment, we will pass the input of the dataset into the model, and then compare the obtained results with the ground truth to find the loss value required for model training or to calculate the PSNR of the model performance.

5.2 Quantitative Result

In order to keep the fairness of the experiment, we used the same hyperparameters for all the five models. In this way, the parameters as well as the flops of these models are at the same level (TABLE 1), ensuring that any observed performance differences can be attributed to our modifications rather than the influence of model size.

Peak Signal-to-Noise Ratio (PSNR) measures the similarity between the relit image and the ground truth by quantifying the reconstruction quality, A higher PSNR indicates that the relit image closely matches the ground truth, reflecting better performance in preserving details and minimizing errors introduced during relighting [19]. In order to quantify the performance of each model, we calculate the PSNR between model outputs and ground truths as the experiment matrics.

The quantitative analysis of Fourier-based relighting methods compared to a baseline provides valuable insights into the efficiency and effectiveness of these methods. According to the Figure 8 and TABLE 2, the original model [8] was trained for 10,000 iterations as the baseline and achieved a PSNR of 17.9837. In contrast, all Fourier-based methods demonstrated better performance in PSNR than the original model, while requiring significantly fewer iterations. Therefore, we are convinced that Fourier-based methods obtain superior computational efficiency. Among our methods, the Fourier Filter achieves the best performance, achieving a PSNR of 18.1311, which is the highest among all approaches. This improvement was realized with only 6,900 iterations – roughly 30% fewer iterations than the baseline. Such a reduction in computational efficiency as well as the improved performance indicates the remarkable advantage of integrating Fourier-based techniques into relighting models.

Other Fourier-based methods, including the Fourier Channel, Fourier Attention Map, and Fourier Encoder, also obtained substantial improvements over the baseline. While requiring significantly fewer iterations, these methods consistently achieved comparable or better PSNR than the baseline. This consistent performance advantage across all the Fourier-based methods indicates their robustness and versatility in addressing relighting tasks.

Since low-frequency information of images captures the broad, global characteristics of an image, such as illumination, shading, and smooth transitions in light intensity, it has significant effects on achieving realistic and consistent lighting effects when relighting an image [14], [15]. From this analysis, we could obtain a critical result that Fourier Transformation can significantly improve models' abilities to focus on low-frequency information, which is crucial for relighting tasks, since accurately capturing and manipulating low-frequency details can significantly impact the final output of relighting. By implementing Fourier-based modifications, the models not only improved the fidelity of outputs and their PSNR, but also reduce training iterations, minimizing computational power and time usage.

The quantitative results of Fourier-based relighting methods demonstrate their significant advantages over the baseline. Our Fourier-based modifications not only improve the PSNR performance and overall fidelity of output images, but also achieve these better results with a reduced computational usage. These two benefits—enhanced performance and reduced computational power usage—highlight the efficiency provided by Fourier Transformation in addressing the complexities of relighting tasks. With Fourier transformation, all the methods we propose achieve better performance with fewer iterations, highlighting their computational efficiency. Their focus on low-frequency information of images allows capturing and manipulating global image characteristics accurately, achieving more realistic and consistent lighting effects.

5.3 Qualitative Result

The qualitative result comparisons of our relighting models, as shown in the Figure 9, validate the advantages of our models, which is proved by our quantitative results. These comparisons highlight the effectiveness of our Fourier-based methods in relighting tasks while achieving better fidelity and PSNR. For instance, the bottom left corner of the sample image in the first column showcases an area that is supposed to remain dark under the target lighting condition. Our Fourier-based methods outperform the original approach [8] by demonstrating superior luminance suppression in this region, closely adhering to the intended lighting dynamics. This ability to control and minimize excess illumination ensures the scene retains its intended atmosphere and avoids overexposure.

Although halos, a common and challenging artifact is a difficult problem remaining unsolved in relighting tasks [20], this is another aspect where our models are proficient. Halos often occur when bright areas blur into darker regions, affecting the visual coherence of the scene. In the lower-right part of the example images in the first column, our models effectively alleviate this issue. By reducing the prominence of halos, our Fourier-based methods achieve a smoother transition between bright and dark areas, enhancing the visual clarity and the quality of the scene. This improvement demonstrates the robustness of our methods in reducing artifacts, which often affects negatively the fidelity of relighting outputs.

Furthermore, the reflection of objects' surfaces under specific lighting conditions is significant for creating fidelity in relighting tasks. In the dark region of the target image in the first column, a reflective spot is produced under the target lighting condition. Among our Fourier-based models, the Fourier Encoder performs best in reproducing this reflective spot with remarkable accuracy. This accuracy ensures that reflective surfaces behave realistically, contributing to the fidelity of the relighting scene. By faithfully replicating the reflective behavior, the Fourier Encoder captures subtle but essential details that elevate the fidelity of the final output.

The qualitative evaluation of our models demonstrates their abilities to produce visual coherence and generate realistic results. The better luminance suppression observed in the darker areas, the mitigation of halos, and the accurate replication of reflective spots collectively validate the effectiveness of our Fourier-based methods. Outperforming in these areas, our models demonstrate clear advantages over the original model. The results highlight their abilities to handle complex lighting interactions and generate outputs which precisely align with the target lighting conditions.

6 Conclusion

6.1 Summary

In this paper, we propose several methods of integrating the Fourier transform into the Illumination Aware Network (IAN) for a relighting task. In our experiment, our Fourier Transformation-based method has a significant advantage over the original method in terms of training efficiency as well as PSNR of the output results. This is also verified by the images presented in the qualitative results. The results show the promising potential of the Fourier transform in improving both the quality of the relighted images and computational efficiency.

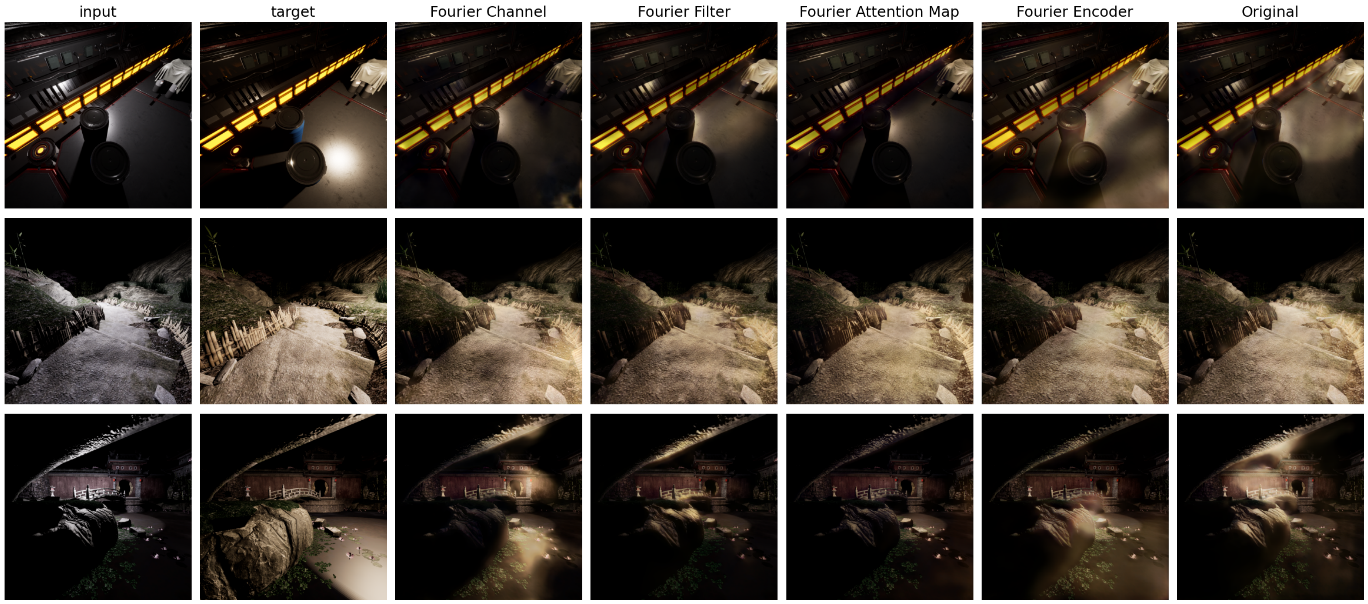

Fig. 9: Output Comparison

6.2 Limitation and Future Work

While we propose different methods of integrating the Fourier transform, they are applied to only one Illumination Aware Network. Future research should extend this integration with other network architectures, transformer-based, or hybrid methods to learn its adaptability and overall impact on image relighting tasks.

Another limitation of the current modified network is that it only takes RGB image data without auxiliary information as inputs. This is due to a lack of access to diverse training datasets. In follow-up work, networks should be tested by using richer inputs to see precisely how the contribution of the Fourier transform works for more complicated scenarios of relighting.

Finally, the methods in this paper integrate Fourier transform into IAN in the input, encoder, and attention stages. Future research should extend the exploration to other stages of the network.

References

[1] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," vol. 2000, 07 2000, pp. 145–156.

[2] Z. Xu, K. Sunkavalli, S. Hadap, and R. Ramamoorthi, "Deep image-based relighting from optimal sparse samples," *ACM Transactions on Graphics*, vol. 37, pp. 1–13, 07 2018.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018. [Online]. Available: https://arxiv.org/abs/1611.07004

[4] P. Gafton and E. Maraz, "2d image relighting with image-to-image translation," 2020. [Online]. Available: https://arxiv.org/abs/2006.07816

[5] H.-H. Yang, W.-T. Chen, H.-L. Luo, and S.-Y. Kuo, "Multi-modal bifurcated network for depth guided image relighting," 2021. [Online]. Available: https://arxiv.org/abs/2105.00690

[6] A. Yazdani, T. Guo, and V. Monga, "Physically inspired dense fusion networks for relighting," 2021. [Online]. Available: https://arxiv.org/abs/2105.02209

[7] L.-W. Wang, W.-C. Siu, Z.-S. Liu, C.-T. Li, and D. P. K. Lun, "Deep relighting networks for image light source manipulation," 2020. [Online]. Available: https://arxiv.org/abs/2008.08298

[8] Z.-L. Zhu, Z. Li, R.-X. Zhang, C.-L. Guo, and M.-M. Cheng, "Designing an illumination-aware network for deep image relighting," *IEEE Transactions on Image Processing*, vol. 31, pp. 5396–5411, 2022.

[9] F.-A. Vasluianu, T. Seizinger, Z. Wu, R. Ranjan, and R. Timofte, "Towards image ambient lighting normalization," 2024. [Online]. Available: https://arxiv.org/abs/2403.18730

[10] T. Hu, Q. Yan, Y. Qi, and Y. Zhang, "Generating content for hdr deghosting from frequency view," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 25 732–25 741.

[11] C. Wang, H. Wu, and Z. Jin, "Fourllie: Boosting low-light image enhancement by fourier frequency information," 10 2023, pp. 7459–7469.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[13] X. Lv, S. Zhang, C. Wang, Y. Zheng, B. Zhong, C. Li, and L. Nie, "Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25 378–25 388, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:272723365

[14] C. Li, C. Guo, M. Zhou, Z. Liang, S. Zhou, R. Feng, and C. C. Loy, "Embedding fourier for ultra-high-definition low-light image enhancement," *ArXiv*, vol. abs/2302.11831, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257102476

[15] C. Wang, H. Wu, and Z. Jin, "Fourllie: Boosting low-light image enhancement by fourier frequency information," in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 7459–7469. [Online]. Available: https://doi.org/10.1145/3581783.3611909

[16] J. Huang, Y. Liu, F. Zhao, K. Yan, J. Zhang, Y. Huang, M. Zhou, and Z. Xiong, "Deep fourier-based exposure correction network withnbsp;spatial-frequency interaction," in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*. Berlin, Heidelberg: Springer-Verlag, 2022, p. 163–180. [Online]. Available: https://doi.org/10.1007/978-3-031-19800-7_10

[17] Y. Chaudhuri, A. Kumar, A. B. Buduru, and A. Alshamrani, "Fga: Fourier-guided attention network for crowd count estimation," 2024. [Online]. Available: https://arxiv.org/abs/2407.06110

[18] M. E. Helou, R. Zhou, J. Barthas, and S. Süsstrunk, "Vidit: Virtual image dataset for illumination transfer," 2020. [Online]. Available: https://arxiv.org/abs/2005.05460

[19] M. El Helou, R. Zhou, S. Süsstrunk, R. Timofte *et al.*, "AIM

2020: Scene relighting and illumination estimation challenge," in *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*, 2020.

[20] Y. Yang, H. A. Sial, R. Baldrich, and M. Vanrell, "Relighting from a single image: Datasets and deep intrinsic-based architecture," 2024. [Online]. Available: https://arxiv.org/abs/2409.18770