

DEEP LEARNING-BASED REFLECTION REMOVAL FOR
ENHANCED IMAGE QUALITY

KONG PINGFAN

SUBMITTED TO THE
FACULTY OF ENGINEERING
UNIVERSITI MALAYA,
IN PARTIAL FULFLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SYSTEMS ENGINEERING

2024

ABSTRACT

This report investigates the problem of single image reflection removal, crucial for improving image quality in photography, medical imaging, and autonomous systems. A comparison study was done on three deep learning frameworks: location-aware single image reflection removal (SIRR), Cooperative Reflection Removal Network (CoRRN), and Iterative Boost Convolutional LSTM Network (IBCLN). Extensive experiments on multiple datasets identified IBCLN as the most effective, excelling in metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). To further improve IBCLN's performance, Generalized Histogram Equalization (GHE) was adopted to enhance brightness and visual quality after reflection removal. Although GHE reduced PSNR and SSIM values slightly, it significantly improved image brightness, contrast, and visual appeal. Despite IBCLN's strengths, challenges remain in processing speed and handling complex scenes or varying reflection intensities in real-time applications. Future work will focus on optimizing the model to balance these trade-offs, possibly through dynamic adjustment of GHE parameters, while reducing artifacts and improving real-time processing. This report advances deep learning-based image reflection removal, highlighting strengths and areas for improvement.

ABSTRAK

Laporan ini menyiasat masalah penyingkiran pantulan imej tunggal, yang penting untuk meningkatkan kualiti imej dalam fotografi, pengimejan perubatan, dan sistem autonomi. Kajian perbandingan dijalankan ke atas tiga rangka kerja pembelajaran mendalam: penyingkiran pantulan imej tunggal yang peka lokasi (SIRR), Rangkaian Penyingkiran Pantulan Kerjasama (CoRRN), dan Rangkaian Konvolusi LSTM Dorongan Berulang (IBCLN). Eksperimen yang meluas pada pelbagai set data mengenal pasti IBCLN sebagai yang paling berkesan, cemerlang dalam metrik seperti Nisbah Isyarat-Ke-Bisingan Puncak (PSNR) dan Ukuran Indeks Kesamaan Struktur (SSIM). Untuk meningkatkan lagi prestasi IBCLN, Penyekataan Histogram Umum (GHE) digunakan untuk meningkatkan kecerahan dan kualiti visual selepas penyingkiran pantulan. Walaupun GHE mengurangkan nilai PSNR dan SSIM sedikit, ia meningkatkan kecerahan imej, kontras, dan daya tarikan visual dengan ketara. Walaupun IBCLN mempunyai kekuatan, cabaran kekal dalam kelajuan pemprosesan dan menangani pemandangan kompleks atau intensiti pantulan yang berubah-ubah dalam aplikasi masa nyata. Kerja masa depan akan memberi tumpuan kepada mengoptimumkan model untuk mengimbangi pertukaran ini, berkemungkinan melalui pelarasan dinamik parameter GHE, sambil mengurangkan artifak dan meningkatkan pemprosesan masa nyata. Laporan ini memajukan penyingkiran pantulan imej berasaskan pembelajaran mendalam, menyerlahkan kekuatan dan bidang untuk diperbaiki.

ACKNOWLEDGEMENTS

As time flies, my academic journey is coming to an end. Along the way, Professor Dr. Chow Chee Onn has provided me with invaluable guidance, and my classmates have offered great help. Here, I would like to sincerely thank every teacher at University of Malaya who has taught me and all the classmates who have accompanied me in my growth.

First, I would like to express my special thanks to my research project supervisor, Prof. Dr. Chow. Prof. Dr. Chow has a rigorous teaching attitude, a strong logical mind, extensive knowledge, and a responsible work ethic. His guidance has greatly benefited me in both learning and personal development. Throughout the process of defining and revising the research project, he meticulously reviewed my work, reading my thesis multiple times and patiently correcting my mistakes, providing numerous valuable suggestions. I will always remember his teachings, striving to improve and surpass myself in both this subject and in life.

Secondly, I want to thank my family. They have taken care of me for over twenty years, allowing me to grow up carefree, without worries about food and clothing, and in happiness and health. Their unconditional support and trust in my future endeavors mean everything to me. I will always cherish their love and continue to forge ahead, hoping that they remain healthy and happy.

Furthermore, I want to thank my friends. I am grateful for our encounters and the joy you've brought into my life. You made my academic life less monotonous. To my dear friends, we have worked hard together and made progress. Although we may eventually

go our separate ways, I hope you all have bright futures, follow your dreams, and cherish our precious youth. Even if we may not see each other often in the future, the bond we share will remain strong.

Lastly, I want to thank all the teachers at Universiti Malaya who have imparted their knowledge to me. Your hard work has expanded my knowledge and broadened my horizons, laying a solid foundation for my future work.

Once again, thank you to everyone who has supported and helped me. It is because of your support that I have been able to complete this thesis successfully.

TABLE OF CONTENTS

| | |
|--|------|
| ABSTRACT | iii |
| ABSTRAK | iv |
| ACKNOWLEDGEMENTS..... | v |
| TABLE OF CONTENTS..... | vii |
| TABLE OF CONTENTS..... | viii |
| LIST OF FIGURES | ix |
| LIST OF TABLES | x |
| 1. INTRODUCTION..... | 1 |
| 1.1. Research Background..... | 1 |
| 1.2. Research Gap..... | 1 |
| 1.3. Research Objectives | 2 |
| 1.4. Research Questions | 2 |
| 1.5. Research Significance | 2 |
| 2. LITERATURE REVIEW..... | 4 |
| 2.1. Learning-Based Methods | 5 |
| 2.1.1. Prior-Learning-Based Methods | 5 |
| 2.1.2. Neural Network | 9 |
| 2.1.2.1. CNN Network | 9 |
| 2.1.2.2. RNN Network | 26 |
| 2.1.2.3. GAN Network..... | 30 |
| 2.1.2.4. DNN Network | 34 |
| 2.1.2.5. CNN+GAN Network | 36 |
| 2.1.3. Other Learning-Based Methods | 38 |
| 2.2. Limitation and Trend | 42 |
| 2.2.1 Limitation | 42 |
| 2.2.2 Trend | 44 |
| 2.3. Datasets | 45 |
| 2.4. Evaluation Metrics..... | 46 |
| 2.4.1. PSNR | 46 |
| 2.4.2. SSIM | 47 |
| 2.4.3. AMBE..... | 48 |
| 3. RESEARCH METHODS | 49 |
| 3.1. Comparison Study..... | 49 |
| 3.2. Proposed Improvement | 52 |
| 3.2.1. Enhancement Method..... | 52 |
| 3.2.2. GHE Equations | 53 |
| 3.2.3. RGB to HSV Convert Equations | 54 |
| 3.2.4. HSV to RGB Convert Equations | 55 |

TABLE OF CONTENTS

| | |
|---|-----------|
| 4. RESULTS AND DISCUSSIONS | 57 |
| 4.1. Results on Comparison Study | 57 |
| 4.1.1. Quantitative comparisons..... | 57 |
| 4.1.2. Qualitative comparisons | 61 |
| 4.1.3. Summary of Comparison Study | 62 |
| 4.2. Results on Proposed Improvement..... | 64 |
| 4.2.1. Optimization of α | 64 |
| 4.2.2. Enhancement Comparison | 69 |
| 5. CONCLUSIONS AND FUTURE WORK | 72 |
| 5.1. Conclusion..... | 72 |
| 5.2. Challenges and opportunities..... | 72 |
| REFERENCES..... | 74 |

LIST OF FIGURES

| | |
|--|----|
| Figure 2.1 Classification of Learning-Based Methods | 5 |
| Figure 3.1 Average loss trend while training SIRR [16] model | 50 |
| Figure 3.2 Average loss trend while training CoRRN [26] model | 50 |
| Figure 3.3 Average loss trend while training IBCLN [21] model | 51 |
| Figure 3.4 Various Metrics vs α curve | 53 |
| Figure 4.1 PSNR comparison of SIRR [16] model on different datasets | 58 |
| Figure 4.2 SSIM results of SIRR [16] model on different dataset | 58 |
| Figure 4.3 PSNR results of CoRRN [26] model on different dataset | 59 |
| Figure 4.4 SSIM results of CoRRN [26] model on different datasets | 59 |
| Figure 4.5 PSNR results of IBCLN [21] model on different datasets | 60 |
| Figure 4.6 SSIM results of IBCLN [21] model on different datasets | 60 |
| Figure 4.7 Qualitative comparisons between the 3 methods | 61 |
| Figure 4.8 Average AMBE vs α curve | 64 |
| Figure 4.9 Average PSNR Enhanced vs α curve | 65 |
| Figure 4.10 Average SSIM Enhanced vs α curve | 65 |
| Figure 4.11 Various Metrics vs α curves when $\alpha \in [0, 2]$ | 66 |
| Figure 4.12 Qualitative comparison when $\alpha \in [0.6, 1.7]$ | 67 |
| Figure 4.13 Average EVM vs α curve when $\alpha \in [0, 12]$ | 69 |
| Figure 4.14 Qualitative comparison of enhancement on 5 datasets | 71 |

LIST OF TABLES

| | |
|---|----|
| Table 2.1 Datasets used for training, testing, and comparing | 45 |
| Table 4.1 Quantitative comparison to chosen algorithms | 57 |
| Table 4.2 Quantitative comparison when $\alpha \in [0.6, 1.7]$ | 68 |
| Table 4.3 Quantitative comparison on five datasets when $\alpha = 0.8$ | 70 |

1. INTRODUCTION

1.1. Research Background

Reflections in images often degrade their quality, particularly in scenarios such as photography through glass surfaces or in underwater imaging. Traditional methods for reflection removal often fall short in handling complex scenarios and may produce unsatisfactory results. Leveraging the capabilities of deep learning, this research aims to explore the development of a robust and efficient solution for effectively removing reflections from images while preserving image details and enhancing overall visual quality.

1.2. Research Gap

Existing methods for single image reflection removal, including both traditional and deep learning-based approaches, have limitations in handling complex reflection scenarios, such as strong reflections or large reflective areas. Traditional methods often struggle with these challenges, while deep learning models may require extensive training data and complex network architectures. There is a need for more robust and efficient methods that can accurately separate reflection and transmission layers under various conditions.

1.3. Research Objectives

The primary aim of this project is to develop a deep learning-based framework for reflection removal that surpasses the limitations of traditional methods. The research objectives are:

1. To investigate the limitations of the existing image reflection removal techniques through comprehensive literature review and comparison studies.
2. To develop a deep-learning framework to accurately detect and remove reflections while preserving image details and enhancing visual quality.

1.4. Research Questions

This study seeks to address the following key questions:

1. How can deep learning frameworks be optimized to improve the removal of reflections from single images?
2. What are the comparative advantages of using the IBCLN framework over other methods like SIRR and CoRRN in terms of performance metrics such as PSNR and SSIM?
3. How can the brightness and visual quality of images be enhanced post-reflection removal while minimizing artifacts?

1.5. Research Significance

The significance of this research lies in its potential to advance the state-of-the-art in image reflection removal. By improving the effectiveness and efficiency of deep

learning-based methods, particularly the IBCLN framework, this study aims to enhance the quality of images in practical applications. These applications include photography, medical imaging, and autonomous systems, where clear images are crucial. The outcomes of this research could lead to better real-time processing capabilities and more adaptable solutions for a wide range of imaging scenarios.

2. LITERATURE REVIEW

Reflections in images often degrade their quality, particularly in scenarios such as photography through glass surfaces or in underwater imaging. The aim of reflection removal is to improve the quality of images and visibility by reducing the reflections.

The concept of reflection removal was first proposed very early. Numerous studies have investigated various approaches for single image reflection removal, highlighting both classical and deep learning-based methods. At the very beginning, the methodologies of reflection removal, which we call classical methods, are normally physical-based. During the first few years, the researchers always used traditional image processing methods to remove reflections from images such as image sequence-based methods, reflective areas-based methods, etc. The more commonly used non-learning-based methods include image sequence-based methods (Yu Li & Michael S. Brown, 2013) [43], ghosting cues-based methods (Shih Y. et al., 2015) [42], depth of field (DoF) guided methods (Wan R. et al., 2016) [41], pixel compensation and detail reconstruction methods (Tsai G. C. et al., 2018) [38], panoramic images-based methods (Yun J. S. & Sim J. Y., 2019) [30], specular reflection removal in iris images (S. V. Mahesh Kumar et al., 2019) [32], background region marker map-based methods (Zhou L. et al., 2022) [15], and multiple polarized images with different exposure times (Aizu T. & Matsuoka R., 2022) [14].

2.1. Learning-Based Methods

However, in recent years, methods based on machine learning and deep learning are preferred by researchers. These Learning-Based Methods can be divided into prior learning-based methods, neural networks, and other learning-based methods. The neural networks can also be divided into single neural networks and multiple networks, as shown in **Figure 2.1**.

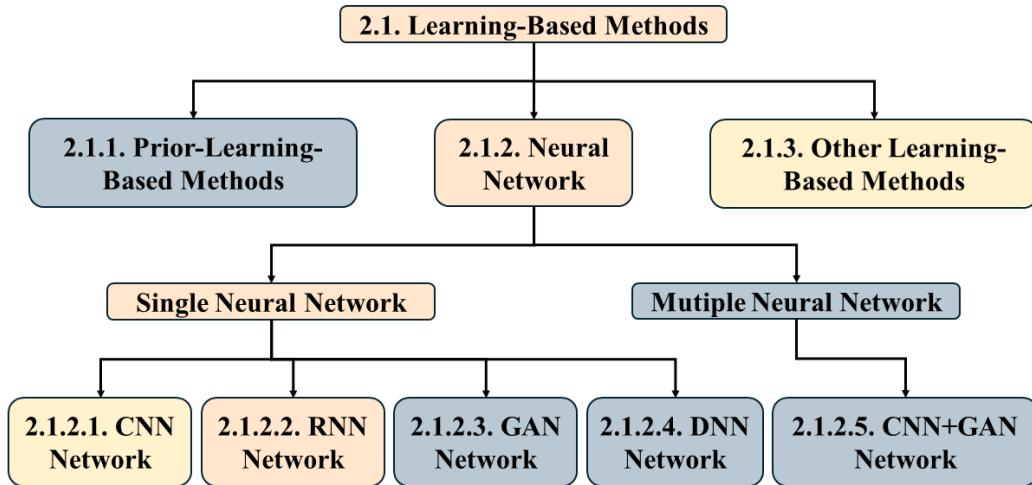


Figure 2.1 Classification of Learning-Based Methods

2.1.1. Prior-Learning-Based Methods

In 2017, Wan, R. *et al.* [40] introduced a reflection removal method based on sparsity prior and non-local image prior. They propose a unified optimization framework that combines sparse priors and non-local image priors to remove reflections in images by utilizing image patches retrieved from external data. The study points out that compared

with previous methods, their method does not require special requirements on the properties of the background layer and reflection layer, such as using two layers with different blur levels to assist separation. Instead, external image patches are used to generate more accurate sparse regularization terms by optimizing the sparsity coefficients. Experiments demonstrate that this method outperforms current state-of-the-art methods in both quantitative evaluation and visual quality.

However, this method also has some limitations in practical applications. First, the method relies on the construction and maintenance of an external database, which may limit its application effect for some scenarios where similar content is difficult to find. Second, the accuracy of image retrieval has a great impact on the effect of reflection removal. If inaccurate image blocks are retrieved, the final image quality may be reduced. In addition, building a large database of similar images requires a lot of work and resources, which increases the complexity of practical applications. Finally, the multi-step processing flow of this method, including image retrieval, registration, and matching, may require a long processing time and is not suitable for application scenarios that require real-time processing.

In the year of 2022, Chen, W. T. *et al.* [13] presented a novel algorithm centered around bi-channel priors, namely the dark channel prior (DCP) and the bright channel prior (BCP). It initiates by elucidating the critical role of these priors in reflecting image characteristics and their potential utility in mitigating reflections. Through meticulous statistical and mathematical analysis, the paper meticulously observes deviations in the values of dark and bright channel pixels when subjected to reflection scenarios. These

observations serve as pivotal insights, paving the way for leveraging bi-channel priors as constraints in the optimization process for reflection removal. The optimization problem formulated integrates the DCP and BCP as fundamental constraints, with the objective of effectively suppressing reflections. To enhance the optimization process, the paper introduces the application of L0-regularization, a technique aimed at promoting sparsity in the solution space. This optimization problem, enriched with bi-channel priors and L0-regularization, is tackled using the half-quadratic splitting method, known for its efficacy in solving non-convex optimization problems efficiently. A significant portion of the paper is dedicated to experimental validation, wherein the proposed method undergoes rigorous evaluation on benchmark datasets. Through meticulous experimentation and comparison with existing methods, the paper substantiates the superior performance of the proposed approach in effectively suppressing reflections in single images. This rigorous validation underscores the efficacy and practical utility of leveraging bi-channel priors and L0-regularization in the context of single-image reflection removal, thereby contributing to the advancement of the field.

Although this dual-channel prior-based method shows technical uniqueness, it also faces several limitations. First, this method relies heavily on the statistical properties of specific channel values in the image, which may lead to fluctuations in results when the scene or image content changes greatly. In addition, this method requires high computing resources when processing high-resolution images. Finally, this approach may not perform well when encountering new scenes or different types of reflections.

In the same year, Zhang, X. *et al.* [8] presented a novel approach that leverages the dark channel prior and sparsity regularization to effectively remove unwanted reflections from single images. The approach described in this article utilizes a combination of the dark channel prior and sparsity regularization techniques. The dark channel prior is employed to identify candidate reflection regions in the input image based on statistical properties of outdoor scenes. Subsequently, sparsity regularization is applied to these candidate regions to suppress the reflection while preserving the underlying scene details. The proposed method effectively separates the reflection layer from the original image, resulting in a reflection-free output that maintains the integrity of the scene content.

Although this method performs well in theory and experiments, it also has some limitations. First, this method relies on the assumptions of dark channel sparsity and gradient sparsity, which may not be fully applicable in some complex scenes, thus affecting the effect of reflection removal. Second, although this method introduces new data fidelity terms to deal with strong reflections, in practical applications, strong reflections may still be difficult to completely remove.

These prior learning-based methods have some common limitations: First, most methods require high computing resources, especially when dealing with high-resolution images or complex network structures, which limits their application on resource-limited devices. Second, these methods usually involve multi-step processing flows and complex optimization processes, resulting in long processing times and are not suitable for application scenarios that require real-time processing. In addition, these methods have limited generalization capabilities and may perform poorly when faced with new scenes

or different types of reflections. Finally, whether it is training data or database construction, these methods have a high reliance on high-quality data, and insufficient data quality will directly affect the actual application effect. By understanding these common limitations, improvements and optimizations can be made more targeted.

2.1.2. Neural Network

As shown in **Figure 2.1**, the common use single neural network includes CNN Network, RNN Network, GAN Network, DNN Network, and the most common used multiple neural network is CNN+GAN Network.

2.1.2.1. CNN Network

In 2018, Chang, Y., & Jung, C. [34] proposed a method for single image reflection removal using convolutional neural networks (CNNs) that combines reflection physics and human visual system (HVS) perception to synthesize natural images for network training. The proposed method integrates reflection physics and human visual system (HVS) perception to generate synthetic images for training the CNN model. This involves synthesizing multiple images with reflections, considering parameters such as the thickness of glass and the angle of view. The CNN model is trained on these synthesized images to learn the mapping between input images containing reflections and their corresponding reflection-free outputs. The trained model is then applied to remove reflections from new single images, effectively restoring the underlying scene.

Although this approach shows significant theoretical and experimental potential, it also suffers from some limitations. First, there is the gap between synthetic data training and real-world applications, which may affect the effectiveness of the model in practical applications. Secondly, it is very difficult to design an appropriate training strategy to make the output close to the real situation. Simply using mean square error (MSE) or perceptual loss for regularization does not work well in terms of convergence. In addition, there is high computational complexity. Finally, the performance of processing different image types fluctuates.

In 2019, Wei, K. *et al.* [27] proposed a novel approach, termed ERRNet, which addresses the challenges by exploiting misaligned training data and incorporating network enhancements. ERRNet is designed to effectively remove reflections from single images by leveraging a fusion of synthetic and real data during training. This approach allows ERRNet to learn from both aligned and misaligned image pairs, enhancing its ability to handle real-world scenarios where perfect alignment is not guaranteed. The methods employed in this work involve the design of an alignment-invariant loss function, which measures the difference between feature representations extracted from aligned and predicted images. Additionally, ERRNet integrates channel-wise context modules and multi-scale spatial context modules into its architecture to capture contextual clues for more accurate reflection removal. The training process utilizes the Adam optimizer with a scheduled learning rate decay to fine-tune ERRNet on the combined dataset of synthetic and real images.

However, this method has several limitations: First, model performance is still highly dependent on the quality and coverage of these data. Second, since existing image synthesis procedures are heuristic, application of models on real images may suffer from accuracy issues due to inter-domain differences.

In 2020, Takahashi, T., Uruma, K., & Kobayashi, K [24] proposes a method for removing reflections from images. The authors employ a Deep Convolutional Neural Network (DeepCNN) as the basis for their reflection removal method. This method specifically uses the Denoising Convolutional Neural Network (DnCNN) architecture. They train the network using a dataset that includes images with human subjects in the reflection layer to enhance performance in human removal. The proposed method is compared with existing techniques including TV regularization, and the effectiveness of the proposed method in eliminating reflections while maintaining image quality is evaluated based on metrics such as PSNR and SSIM.

Although this method shows excellent potential in removing portraits from reflections, it also has several limitations. First, this method is highly dependent on a large amount of training data, especially images that must contain portrait reflections. Finally, this method may perform poorly when faced with new scenes that are significantly different from the training set because it is mainly optimized for a specific type of reflection.

Also in the year of 2020, Lei, C. *et al.* [20] focused on polarized reflection removal, aiming to overcome the limitations of existing methods by leveraging polarization information. Unlike conventional approaches that rely on assumptions like perfect

alignment and unpolarized light sources, the proposed method introduces a novel approach. It utilizes polarization data collected through the M-R method, which involves capturing paired images containing mixed, reflection, and transmission components. These data are then utilized to train a two-stage deep learning framework. In the first stage, the framework estimates the reflection component, followed by the estimation of the transmission component in the second stage. To effectively train the model, a combination of loss functions, including perceptual normalized cross-correlation (PNCC) loss and perceptual loss, is employed. Additionally, hypercolumn features extracted from the VGG-19 network are integrated to enhance the input and improve the overall performance of the framework.

However, this method also has some limitations. First, although this method does not rely on the assumption that the reflected image is out of focus, it is still based on the characteristics of polarized light, which may not be applicable in some specific scenarios. Second, although the new dataset contains more than 100 types of glass, in real applications, the complexity of glass types and environments may exceed the scope covered in the dataset. Finally, this method relies on polarization information. If it cannot be obtained or the polarization information is not obvious in some cases, the reflection removal effect may not be ideal.

Also in the year of 2020, Wan R. *et al.* [26] introduced CoRRN, a Cooperative Reflection Removal Network, which integrates both gradient and image inference into a cooperative framework. This integration is achieved through two key components: the Image Decomposition Network (IdecN) and the Gradient Decomposition Network

(GdecN). The IdecN extracts high-level context information and multi-scale low-level features from the input image, facilitating a comprehensive understanding of the image content. Meanwhile, the GdecN focuses on inferring the reflection gradients present in the image, aiding in the identification and removal of reflection artifacts. These two networks work collaboratively to generate a reflection-free output image by effectively removing reflections while preserving important image details. Moreover, CoRRN introduces a statistic loss based on gradient level statistics to further enhance the removal of locally strong reflections. Trained on real-world reflection image datasets, CoRRN demonstrates superior performance compared to existing methods in both quantitative evaluations and visual quality assessments. This underscores its efficacy in addressing the challenges associated with reflection removal from single images.

Although the CoRRN method demonstrates significant potential both theoretically and experimentally, it also suffers from some limitations. First, due to the loss of background information in the saturated reflection area, the reflection removal problem degenerates into an image repair problem. Secondly, this method may cause color shift in certain situations. Finally, the capture settings of reflection images and the diversity of scenes need to be further improved, and these issues may limit the generalization ability of the training dataset.

In 2021, Isana Funahashi *et al.* [19] introduced a novel approach to high reflection removal using CNNs with detection and estimation. Unlike previous methods, which focus on normal reflections, this method specifically targets high reflections, including those from intense light sources. It proposes a CNN model to detect high reflections based

on edge features, coupled with image inpainting to estimate background information. This method comprises several key steps. First, a CNN model is employed to detect high reflections in input images using RGB values, saturation masks, and edge information. Next, morphological operations are applied to refine the detection results. Finally, image inpainting is utilized to estimate background information in areas of high reflections, resulting in the removal of undesirable reflections. Experimental results show that although this approach demonstrates significant potential for processing high-intensity reflections, it also faces several significant limitations. First, the performance of this model is highly dependent on the accurate detection of edge features. If these features are not accurately identified, it may lead to poor reflection removal. Secondly, the training and performance of the model rely heavily on the quality and diversity of training data.

Also in the year of 2021, Lei, C. *et al.* [17] presented a novel approach to robust reflection removal in images, leveraging reflection-free flash-only cues. By recognizing that objects in reflection receive weaker light from the flash, the authors exploit this cue to enhance the accuracy and robustness of reflection removal algorithms. They construct a real-world dataset comprising raw flash and ambient image pairs, along with corresponding reflection images, facilitating the training of a dedicated deep learning model for reflection removal. The methodology involves dataset collection with consistent illumination conditions, generation of ground truth reflection images, training a deep learning model using the flash-only image as a cue, proposing a dedicated architecture to handle artifacts in flash-only images, and evaluating the model on both

real-world and synthetic datasets. This approach achieves state-of-the-art performance in removing reflections and improving overall image quality.

However, this method also has some limitations. First, the method is highly dependent on the quality of the flash image. In addition, compared with the single-image method, this method requires additional flash images, which increases the complexity of image capture and processing.

In the year of 2022, Peng, Y. T. *et al.* [12] proposed a novel method for SIRR based on knowledge-distilling content disentanglement. This approach innovatively combines knowledge distillation, a technique originally aimed at model compression in deep learning, with content disentanglement principles. By leveraging a Teacher-Student framework, the method effectively separates reflection features from the input image, enabling the Student network to disentangle the transmission and reflection layers. The method begins with the Reflection Teacher network, which is trained to extract reflection features from the input image using convolutional layers and residual blocks. These features are then transferred to the Content Disentangling Student network, which aims to decompose the input image into transmission and reflection layers. This decomposition process is achieved through a series of convolutional layers and Multi-kernel Strip pooling (MSP) based Content-aware Layers (CAL). These CALs attend to reflection-affected or transmission-dominated regions with varying degrees, facilitating the disentanglement of content. To effectively train both networks, various loss functions are employed, including reconstruction loss, content disentangling loss, representation mimicking loss, and fidelity loss. These loss functions ensure that the Student network

accurately learns to separate the transmission and reflection layers while maintaining structural fidelity and perceptual quality. Experimental evaluation on benchmark SIRR datasets demonstrates the effectiveness of the proposed method. Compared to existing state-of-the-art methods, the method achieves superior results in terms of perceptual quality and structural fidelity. By integrating knowledge distillation and content disentanglement techniques, the proposed approach offers a promising solution for the challenging task of single image reflection removal.

Although this method is technically innovative and performs well experimentally, it also has some limitations. First, the method has a high demand for high-quality training data, especially reflection and transmission image data that need to be accurately annotated. In addition, due to the need for multiple iterative calculations and long processing time, it is not suitable for real-time processing application scenarios. Finally, the complexity of the model also increases the difficulty of implementation.

In the year of 2023, Hong, Y. *et al.* [5] proposed an in-depth exploration into the development of PAR2Net, a cutting-edge solution for panoramic image reflection removal. At its core, PAR2Net harnesses the power of deep learning, specifically convolutional neural networks (CNNs), to tackle the complex task of removing reflections from panoramic scenes. The system is meticulously designed, comprising several integral components meticulously crafted to enhance its performance and effectiveness. A crucial aspect of the methodology involves the collection and preparation of a comprehensive dataset. This dataset consists of panoramic images paired with corresponding ground truth reflection-free images, providing the necessary training data

for the PAR2Net model. Leveraging this dataset, the PAR2Net architecture is meticulously crafted, incorporating key modules such as initial decomposition, reflection correspondence, and contextual information integration. These modules work synergistically to facilitate accurate and robust reflection removal across panoramic scenes. The training process of PAR2Net is a pivotal step in its development. Deep learning techniques are employed to train the model, allowing it to learn intricate patterns and relationships between reflection scenes and underlying content. Moreover, data synthesis techniques are utilized to bridge the domain gap between the training and testing data, ensuring the model's adaptability to various scenes and lighting conditions. Following the training phase, PAR2Net undergoes rigorous evaluation using both quantitative and qualitative metrics. Quantitative assessments involve measuring the model's performance against predefined error metrics, such as PSNR and SSIM, while qualitative evaluations entail visual inspections and comparisons against ground truth images. Additionally, PAR2Net's performance is benchmarked against existing methods for panoramic image reflection removal, providing insights into its comparative advantages. Through comprehensive analysis and discussion, the article delves into the strengths and limitations of PAR2Net. It identifies areas of success, highlighting the system's accuracy, robustness, and versatility across different scenes. Furthermore, the article discusses potential avenues for future research and improvement, paving the way for advancements in panoramic image reflection removal technology.

However, the PAR2Net method also has some limitations. First, if some areas in the reflection layer cannot be found in the reflection scene, the performance of the method

may decline. Second, although the method performs well in experiments, the training process is complicated and requires fine parameter adjustment and long training time to ensure the convergence of the model and stable performance.

Also in the year of 2023, Yan, T. *et al.* [7] proposed MRDNet, a novel approach for single-image reflection removal from glass surfaces. This method introduces a multi-scale reflection detection mechanism, allowing for the effective identification and removal of reflections of varying sizes and intensities from input images. Comprising three main modules—Reflection Extractor (R-Extractor), Transmission Extractor (T-Extractor), and Multi-scale Reflection-aware Transmission Recovery (MRTR) module—MRDNet synergistically extracts features of the foreground reflection layer and the background transmission layer, thereby achieving accurate reflection removal. Deep learning, specifically convolutional neural networks (CNNs), serves as the primary method in this work. The MRDNet architecture is tailored to learn from large-scale datasets, enabling the extraction of intricate features and patterns associated with reflections on glass surfaces. Furthermore, the incorporation of L0-regularization alongside bi-channel priors bolsters the network's capability to effectively suppress reflections. The proposed method undergoes training and evaluation using diverse datasets, including the introduction of the Physical Rendering-based Reflection Removal (PRGR) dataset.

Although MRDNet shows excellent potential in multi-scale reflection removal, the method also has some obvious limitations. First, this method is highly dependent on high-quality 3D virtual world generated data for training, which may affect its generalization ability and accuracy when processing real-world data. Secondly, the network may face

challenges in accurately detecting and segmenting reflective areas in scenes with large illumination changes or particularly high reflection intensity.

Also in the year of 2023, Daggubati, H. *et al.* [10] proposed method for image reflection removal using an Iterative Boost Convolutional LSTM Network through Feature Loss (IBCLN). It begins with an introduction highlighting the problem of unwanted reflections in images taken through glass windows and the need to improve image quality by removing these reflections. Then, they outlined the architecture of the IBCLN network, which includes the GT and GR sub-networks, each utilizing convolutional LSTM units and feature loss for training. The method utilizes pixel loss, adversarial loss, and feature loss functions to train the network and improve the quality of reflection removal. Additionally, it discusses the utilization of auxiliary information, such as predicted transmissions and residual reflections, to enhance the training process and achieve better results. Finally, it presents experimental analyses and quantitative evaluations comparing the proposed method with existing approaches, demonstrating the superiority of IBCLN in removing unwanted reflections from images.

Although this approach is technically significantly innovative, it also suffers from several significant limitations. First, the performance of the entire model is highly dependent on the quality of the initial estimate; if the initial estimate is inaccurate, more iterations may be required to achieve satisfactory results, or even little improvement may occur. Secondly, due to network complexity and iterative nature, the model is also at risk of overfitting, which may cause it to perform poorly on new, unseen images.

Also in the year of 2023, Xu Y. *et al.* [6] introduced a comprehensive methodology, the Superimposed Mask-Guided Contrastive Regularization (NM-DCR) approach, aimed at effectively separating multiple targets in Range-Doppler (RD) maps. The methodology begins with preprocessing steps, including the resizing of input RD maps to a standardized dimension of 128×128 . Subsequently, the NM-DCR model is trained using a batch size of 2 and the ADAM optimization algorithm over 300 epochs. During training, a combination of structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and separation accuracy (ACC) metrics is employed to evaluate separation performance rigorously. An essential aspect of the methodology involves conducting ablation experiments to dissect the individual contributions of different components within the NM-DCR framework. Specifically, the impact of the superimposed mask auxiliary branch and contrastive regularization is analyzed to understand their role in enhancing separation efficacy. Moreover, the paper meticulously compares the NM-DCR method with existing echo separation techniques, including DAD, G-DPS, and Pix2Pix. Through this comparative analysis, the superiority of the NM-DCR approach in terms of separation accuracy and preservation of structural information becomes evident. Overall, the methodology integrates deep learning principles, rigorous evaluation metrics, and comprehensive experimentation to establish the efficacy and superiority of the proposed NM-DCR method for echo separation on RD maps.

Although this method performs well both theoretically and experimentally, it has some limitations. First, although this method performs well in experiments, the training process is complex and requires fine parameter adjustment and long training time to

ensure model convergence and stable performance. Secondly, when processing the signals in the overlapping region, the signals may be significantly distorted, affecting the accuracy of subsequent tasks such as human activity classification. Finally, it may require further adjustments to the model structure and training strategy to ensure its robustness and effectiveness in various scenarios while processing different kinds of echo signals and targets in practical applications.

Also in 2023, Chen, W. T. *et al.* [11] presented a novel approach for single image reflection removal based on auxiliary prior learning. The work combines optimization techniques and deep learning methodologies to address single image reflection removal. Initially, an energy function is formulated, incorporating both data fidelity terms and constraint terms to effectively balance color fidelity and gradient preservation. Subsequently, Half-Quadratic Splitting (HQS) optimization is employed iteratively to minimize this energy function. HQS optimization utilizes auxiliary variables to update the reflection-free image and reflection layer separately, enhancing the accuracy of the removal process. Furthermore, Deep Convolutional Neural Networks (DCNNs) are integrated into the methodology to achieve more comprehensive results. Specifically, mapping functions based on DCNN architectures are utilized to predict edge components and the reflection layer, further improving the overall performance of the reflection removal task.

However, this method also has some limitations. First, although synthetic data is used for training, there is still a gap with the reflection scenes in the real world, which may affect the effect of the model in practical applications. In addition, the method

involves multi-step processing, including RRN and MRN, as well as complex optimization processes, which increases the calculation time and is not suitable for application scenarios that require real-time processing. Finally, the introduction of multiple loss functions and complex network architectures increases the difficulty of model implementation and debugging complexity.

In the same year, Chiang, C. C. *et al.* [4] proposed two preprocessing techniques: glare reflection removal and color distortion correction. These techniques utilize sophisticated image registration methods to establish pixel correspondences across multiple images, effectively mitigating the adverse effects of glare and color distortion. Following the preprocessing stage, deep learning models are trained on datasets comprising preprocessed samples. Two distinct training approaches are employed: uniform training and diverse training. In uniform training, each sample undergoes glare removal and color correction, resulting in a more standardized appearance. This approach enables the evaluation of how the preprocessing methods impact recognition rates under controlled and uniform conditions. Conversely, diverse training involves training models on samples without preprocessing, thereby exposing them to a wider range of appearance variations. This approach assesses the models' ability to handle glare reflections and color distortions through data-driven training on diverse samples. The methodology also introduces a novel ResDenseNet architecture, which amalgamates the strengths of ResNet and DenseNet networks. This architecture facilitates the integration of cross-level features through skip connections, enhancing the model's ability to distinguish between medications and improving recognition rates. Subsequently, the proposed methods are

rigorously evaluated through extensive experimentation. This evaluation includes medication detection and identification tasks, where the recognition rates of different models are compared and analyzed. Additionally, the authors assess the models' adaptability and reliability under varying imaging conditions. Overall, the methodology encompasses proposing and implementing preprocessing techniques, training deep learning models using diverse training approaches, introducing a novel architectural design, and conducting thorough experimentation to evaluate the effectiveness of the proposed methods in improving medication identification from transparent packaging.

Although the drug identification method based on ResDenseNet has demonstrated significant potential theoretically and experimentally, however, this method also has some limitations. The most typical one is: The denoising and color correction process may in some cases lead to a loss of image details, affecting the quality of the final generated image, especially when processing images with complex textures.

In the year of 2024, Lan, H., Zhang, E., & Jung, C. [2] introduced a novel method for face reflection removal that capitalizes on the multispectral data available from both RGB and NIR images. The core concept involves integrating the complementary information from these two modalities using a multi-spectral CNN architecture. Their proposed network architecture consists of four primary modules: the Context Enhancement Module (CEM), NIR Image Distortion Module (NIDM), Image Inverse Distortion Module (IIDM), and Reflection Confidence Generation Module (RCGM). Through the fusion of RGB and NIR data, the method aims to effectively eliminate reflections while retaining crucial facial details. The methodology unfolds through several

key steps. Firstly, the Context Enhancement Module (CEM) is employed to extract multi-scale features from both RGB and NIR images, with a focus on suppressing sparse reflection components. Subsequently, the NIR Image Distortion Module (NIDM) utilizes NIR images to suppress reflections, providing essential reflection-suppressed information for guiding subsequent processing steps. Following this, the Image Inverse Distortion Module (IIDM) estimates the face transmission layer, leveraging the reflection-suppressed features obtained from NIR images. Additionally, the Reflection Confidence Generation Module (RCGM) generates a reflection confidence map, facilitating the measurement of reflection dominance in various regions and aiding in accurate reflection removal. Finally, various loss functions, including perceptual loss, SSIM loss, and gradient-aware loss, are integrated into the methodology to optimize network parameters during training. Overall, this comprehensive approach offers a promising solution for reflection removal in face images, combining the strengths of RGB and NIR data to achieve enhanced visual quality and superior performance compared to existing methods.

However, this method also has some limitations. First, in some cases, although NIR images can provide more structural information, the denoising and multispectral fusion process may lead to the loss of RGB image details, affecting the quality of the final generated image. Second, although this method performs well in experiments, in practical applications, processing different kinds of reflections and images may require further adjustment of the model structure and training strategy to ensure its robustness and effectiveness in various scenarios.

Also in the year of 2024, Zhao, H., Zhou, R., Zhang, S., & Fu, Y. [3] proposed a novel single image reflection removal framework using feature difference enhancement. It highlights the limitations of existing methods and proposes a comprehensive approach that exploits the complementarity and distinction between the reflection and transmission layers in reflection-contaminated images. The proposed method consists of three primary components: a feature difference enhancement module (FDEM), an adaptive information exchange block, and a selective Instance Normalization strategy. These components synergistically enhance the efficiency and performance of reflection removal tasks, addressing challenges such as ill-posedness and error accumulation. The method employed in the work utilizes deep learning techniques within a single-stage framework for single image reflection removal. It leverages a pre-trained VGG-19 network for feature extraction and incorporates a shared encoder and two separate decoders for predicting the transmission and reflection images. The key techniques include the feature difference enhancement module, which enhances feature disparity between reflection and transmission layers, an adaptive information exchange block for inter-decoder communication, and a selective Instance Normalization strategy to improve feature calculation accuracy and speed. These methods collectively contribute to the efficiency and effectiveness of reflection removal tasks, offering advancements over existing approaches.

Although this method is technically significantly innovative and has shown superior performance in multiple experiments, it also suffers from several obvious limitations. First, the performance of the model is highly dependent on high-quality training data.

Secondly, its practicality in processing real-time data and adapting to changing environments still needs further verification. In addition, Finally, specific network designs may not perform well in the face of non-routine reflection situations.

These CNN Networks have some common limitations: First, these methods usually involve complex deep convolutional neural network structures, causing the training and inference processes to require extensive computing resources. Secondly, the performance of the method is highly dependent on high-quality training data. Lack of representativeness or low-quality data may significantly affect the generalization ability and actual effect of the model. Furthermore, in practical applications, these methods often exhibit performance fluctuations when processing different scenes and image types, indicating the need for further adjustments to model structures and training strategies. The training process is usually complex, requiring fine parameter adjustments and long training times to ensure model convergence and stable performance. Finally, many methods perform poorly when faced with specific reflection situations or extreme lighting conditions, and further research and improvements are needed to improve their adaptability and effectiveness in various scenarios.

2.1.2.2. RNN Network

In 2018, Zhang, X., Ng, R., & Chen [37] presented an approach to single image reflection separation using perceptual losses and a novel exclusion loss. The approach proposed in this paper employs a fully convolutional neural network trained with

perceptual losses for single image reflection separation. The network takes a single image as input and synthesizes two images: the reflection layer and the transmission layer. The method utilizes two perceptual losses: a feature loss from a visual perception network and an adversarial loss to refine the output transmission layer. Additionally, a novel exclusion loss is introduced to enforce pixel-level separation between the transmission and reflection layers. The authors create a dataset of real-world images with ground-truth transmission layers to facilitate model training and evaluation. Extensive experiments demonstrate the effectiveness of the proposed method compared to state-of-the-art approaches in terms of PSNR, SSIM, and perceptual user study.

Although this approach is technically promising, it faces several key limitations. First, due to the high reliance on perceptually oriented loss functions, this method may not work well in image scenarios where perceptual differences are not obvious. Secondly, although adversarial training can improve image quality, it may lead to instability in the training process, especially in the early training stages.

In 2020, Li, C. *et al.* [21] introduced an Iterative Boost Convolutional LSTM Network (IBCLN) designed for single image reflection removal through cascaded refinement. This method employs a cascaded architecture, utilizing two convolutional LSTM networks, each equipped with a convolutional LSTM unit. The aim is to iteratively refine transmission and reflection layers, leveraging LSTM units to enable effective training over multiple cascade steps. Additionally, the methodology incorporates a residual reconstruction loss for further training guidance, along with a multi-scale perceptual loss to capture contextual information from various scales. Implementation

details include the utilization of PyTorch and the Adam optimizer, with training conducted on a dataset containing both synthetic images and real-world image patches.

The IBCLN method has several advantages in single image reflection removal tasks. First, the method adopts a cascade network structure to gradually optimize the prediction of the transmission layer and the reflection layer through multi-stage iteration. The output of each step is based on the optimization result of the previous step, thereby gradually improving the accuracy of the overall prediction. In addition, the convolutional LSTM structure can effectively retain and transfer information, avoid the vanishing gradient problem in deep network training, and maintain the stability of long-term memory and information flow. The two sub-networks of IBCLN cooperate with each other and use each other's output as input to achieve complementary feedback, which significantly enhances the model's predictive ability. In addition, by introducing residual reconstruction loss and multi-scale perception loss, the model can reconstruct the input image more accurately and capture image features at multiple resolution levels, thereby improving the visual quality and overall perception effect of the transmission layer.

However, the IBCLN method also has some limitations. First, although the cascade and iteration strategies can improve the prediction accuracy, they increase the complexity of model training. Second, the model structure is complex and the processing speed may not be sufficient in real-time application scenarios.

In 2021, Dong, Z. *et al.* [16] proposed a location-aware single image reflection removal (SIRR) network as its main contribution, aimed at effectively identifying and removing reflections from images while preserving high-frequency details in the

transmission layer. The innovation lies in leveraging learned Laplacian features to emphasize strong reflection boundaries, such as reflected highlights. The methodology involves several key steps: First, the Reflection Detection Module (RDM) is used to roughly identify reflections by extracting multi-scale Laplacian features. Next, the Transmission Selection Module (TSM) selects the transmission layer based on the detected reflections. Contextual information is then integrated using Long Short-Term Memory (LSTM) to refine the reflection removal process. Additionally, a combination of loss functions, including pixel loss, SSIM loss, and adversarial loss, is employed for network training and optimization. Finally, location-aware processing is applied, exploiting learned Laplacian features to guide the removal process effectively. Overall, these steps enable the proposed method to achieve superior performance in various real-world scenarios.

However, this method also has some shortcomings: First, it cannot accurately detect or remove particularly strong reflections in some cases. Second, the training of the model relies on a composite loss function, including perceptual loss, pixel and structure similarity (SSIM) loss, and adversarial loss. The design and weight adjustment of these loss functions have a great impact on the performance of the model.

These RNN Networks have some common limitations: First of all, they are highly dependent on training data and require a large amount of high-quality and accurately labeled data to ensure the effectiveness of the model. Second, due to the complexity of the model, these methods require extensive computing resources during both training and inference, which may limit their applicability on resource-constrained devices.

Furthermore, these methods may suffer from insufficient generalization capabilities in new scenarios that are significantly different from the training data distribution. Finally, the design of the loss function and weight adjustment have a significant impact on model performance, and all methods require careful design and tuning of the loss function to achieve optimal performance.

2.1.2.3. GAN Network

In 2019, Heo, M., & Choe, Y. [33] introduced a deep neural network structure based on conditional GAN to render realistic images. They formulate the problem with a simple objective function that exploits gradient information to preserve low- and high-frequency details. The proposed network consists of a generator that aims to generate realistic images, and a discriminator that differentiates between real and generated images. The loss function combines the mean square error (MSE) of content loss and gradient loss to preserve high-frequency details. Synthetic datasets were created for training, and the network architecture was based on the pix2pix network and optimized using ADAM. The training process involves using synthetic images with reflections to train the network to effectively remove reflections from real images. By not relying on physical prior information, the proposed network effectively eliminates reflections from individual images. Experimental results are evaluated using metrics such as PSNR and SSIM, demonstrating the competitive performance of the proposed method with respect to previous methods.

Although conditional GANs have performed well in image generation and conversion tasks, this method still faces some technical challenges in single image reflection removal. First, the training process of GANs is inherently challenging and requires carefully designed training strategies to avoid mode collapse and ensure a good dynamic balance between the generator and the discriminator. Second, although this method may perform well on the training set, its generalization ability in new scenarios or situations that are significantly different from the training data is still an issue that needs attention.

Also in 2019, Chou, N. H., Chuang, L. C., & Lee, M. S. [29] proposed a novel intensity-aware generative adversarial network (GAN) for single image reflection removal. This method exploits the discriminative ability of reflective layers in low-intensity areas to directly estimate the function that converts a mixed image into a reflective image. The proposed architecture includes dual generators to handle different features of low-intensity and high-intensity regions respectively. Unlike optimization-based methods that rely on handcrafted priors, this method learns the transformation function directly from the data. Due to the lack of specific datasets for reflection removal, synthetic data are used for training. In addition, a reflection image synthesis method based on the screen blend model is also introduced. A screen hybrid model was chosen to simulate the interweaving effect between reflected and transmitted images, with a reflection intensity weighted L1 loss adapting to the reflection phenomenon. Overall, this method provides a novel approach to single-image reflection elimination, demonstrating promising results in real-world scenarios.

Although this intensity-aware GAN method is technically innovative and demonstrates good experimental results, it still faces several important limitations in practical applications. First, this method is highly dependent on large amounts of training data and mainly uses synthetic data to train the model, which may not fully capture the reflection complexity in the real world, which limits its generalization ability. Second, a careful balance is required between the generator and the discriminator in GAN training. Improper balance may lead to training instability or even mode collapse.

In the year of 2020, Zou Z. *et al.* [22] proposes a novel approach that leverages adversarial training within a deep learning framework. The proposed method primarily utilizes Generative Adversarial Networks (GANs), a powerful architecture consisting of a generator and discriminator network. The generator network aims to decompose the superimposed image into its constituent parts, while the discriminator network evaluates the quality of the generated outputs. Through adversarial training, wherein the generator network competes with the discriminator network, the method learns to produce increasingly accurate separations. The main steps of the proposed method involve inputting the superimposed image into the adversarial framework, where the generator network attempts to separate it into individual components. The discriminator network then provides feedback on the quality of the generated outputs, guiding the generator network to improve its performance iteratively. This iterative training process enables the method to achieve state-of-the-art results across various challenging tasks in computer vision and signal processing, including image deraining, reflection removal, and shadow removal. The proposed framework offers a unified solution for addressing the separation

of superimposed images, demonstrating its effectiveness and versatility in handling multiple tasks without the need for task-specific tuning.

However, this method also has some shortcomings: First, although this method can handle nonlinear mixing, the model's recovery ability is still limited for complex mixing situations. Second, since the generated output may be unordered, a specially designed "crossroad 11" loss function needs to be used to guide training, which increases model design complexity. Third, although local perception networks (such as PatchGAN) are introduced to improve the perceptual quality of separated images, this method may still face challenges when processing images with complex textures or details.

These GAN Networks have some common limitations: First, they are highly dependent on high-quality and large amounts of training data, which directly affects the model's generalization ability and performance in real-life scenarios. Secondly, due to the high computing resource requirements of GANs, the application of these methods on devices with limited computing capabilities is limited. In addition, the training process of GAN is inherently complex and unstable, requiring a carefully designed training strategy and a good balance between the generator and the discriminator to avoid mode collapse and training instability. Finally, these methods have limitations in their generalization capabilities in new scenarios or situations that are significantly different from the training data, and require further attention and improvement.

2.1.2.4. DNN Network

In 2017, Fan, Q. *et al.* [39] contributed to proposing a deep neural network structure called Cascaded Edge and Image Learning Network (CEILNet), a generic deep architecture designed for edge-sensitive image processing tasks. The work employs a combination of traditional image processing techniques and deep learning methodologies. Specifically, it addresses the challenges of single-image reflection removal and image smoothing. For reflection removal, it utilizes gradient projection, flash-exposure sampling, and sparsity priors, along with a novel reflection image synthesis method. This approach significantly advances the state-of-the-art by effectively separating reflection from the background image. Moreover, the document presents advancements in deep learning-based image smoothing, achieving superior results compared to traditional methods. The proposed CEILNet architecture serves as the backbone for implementing these methods effectively.

Although CEILNet shows potential for processing single-image reflection removal and image smoothing, it also faces several important limitations: First, this network architecture relies heavily on a large amount of training data, which may be difficult to obtain in practical applications. Furthermore, the generalization ability of the network may be limited, especially when dealing with new tasks that are significantly different from the training data. Finally, the training process of this method contains multi-stage optimization, which increases the complexity and challenge of training. Therefore, although CEILNet performs well in specific image processing tasks, further research and

development are required to address these challenges before it can be widely applied to other image processing problems.

In 2020, Yang, J. *et al.* [36] contributed to proposing a novel deep learning approach, termed the Bidirectional Network (BDN), for single image reflection removal. The BDN architecture is designed to estimate both the background and reflection layers simultaneously, leveraging bidirectional information exchange to enhance the removal process. The work presented in this paper employs a deep learning framework, implemented using PyTorch, to train the Bidirectional Network (BDN). The training process utilizes the Adam optimizer with default parameters and employs a combination of objective functions, including adversarial loss and mean squared error (MSE) loss, to optimize the network parameters. The BDN architecture consists of multiple components, including vanilla generators, reflection estimators, and background estimators, which are trained jointly to achieve robust performance in single image reflection removal tasks.

Both CEILNet [39] and Yang J. *et al.*'s [36] method have some common limitations. First, these methods have a high reliance on high-quality and diverse training data, which may be challenging to obtain in practical applications. Secondly, both methods have insufficient generalization capabilities when dealing with new tasks that are significantly different from the training data, which limits their widespread use in practical applications. In addition, complex network structures and high computing resource requirements also limit the application of these methods in resource-constrained environments. These shared limitations indicate that although these methods perform well in specific image

processing tasks, further research and optimization are required to broadly apply them to other image processing problems.

2.1.2.5. CNN+GAN Network

In the year of 2019, Li T., & Lun D. P. [25] proposed a novel two-stage reflection removal algorithm using deep neural networks. The approach aims to address the limitations of existing methods by effectively suppressing reflection residues, particularly those with strong gradient components. In the first stage, a CNN is trained with a modified loss function to effectively suppress reflection residues while preserving background features. The second stage involves refining the background estimate using a confidence map based on gradient analysis, followed by background reconstruction using a GAN. The training data for the networks are synthesized images with reflections using existing datasets, and the networks are trained sequentially to avoid overfitting. The performance of the proposed method is evaluated quantitatively and qualitatively against existing DNN-based methods using benchmark datasets, demonstrating its effectiveness in single-image reflection removal.

Nevertheless, this approach also faces some obvious limitations. The most typical one is: The feature suppression process may lead to the loss of background image details, making the background appear blurred in some cases.

In this year, Li T., & Lun D. P. [31] also proposed another novel deep learning-based method for reflection removal using Wasserstein Generative Adversarial Networks

(WGANS). The proposed method aims to overcome the limitations of traditional and existing learning-based approaches by leveraging the strengths of WGANS. This method involves capturing multi-view images of a scene and using a convolutional neural network (CNN) to extract depth information along the edges of the image. This depth information is then utilized in two WGANS to estimate the edges of the background and reconstruct the background image. By integrating depth-based information with the power of WGANS, the proposed method achieves state-of-the-art performance in reflection removal while significantly reducing computation time compared to traditional methods.

Although this method is obviously innovative in theory and shows high performance in experiments, the training process is still computationally intensive.

The limitations shared by both methods are mainly reflected in the following aspects: First, they both highly rely on high-quality and diverse training data. If the training data is insufficient or of low quality, it will seriously affect the output quality of the model. Secondly, these methods require a large amount of computing resources, and high computing requirements may limit their application on devices with limited computing capabilities. Finally, due to the complex network structure and significant pre- and post-processing steps, these methods are not suitable for application scenarios that require real-time or near-real-time response.

2.1.3. Other Learning-Based Methods

In addition to the prior-learning-based methods and neural network methods, there are some other learning-based methods that are also used for image reflection removal.

In 2018, Wan R. *et al.* [35] introduced the Concurrent Reflection Removal Network (CRRN). This novel method integrates image appearance information and multi-scale gradient information within a unified framework. By combining gradient inference and image inference stages into a single mechanism, the CRRN can remove reflections concurrently, overcoming the limitations of previous approaches. Additionally, the CRRN leverages a perceptually motivated loss function to suppress blurry artifacts and enhance result quality. The main steps involved in the proposed method include problem statement, shift to deep learning, introduction of CRRN, dataset creation, and experimental validation. The article introduces a new dataset containing a variety of real-world reflection images to facilitate the training and evaluation of the method. The effectiveness of CRRN is also verified through extensive experiments, showing that it outperforms existing methods in real scenes.

Although CRRN is significantly innovative technically, it also faces some limitations. First, CRRN's performance may degrade when dealing with situations where the entire image is mainly occupied by reflection layers. Second, the network may introduce color shifts during image restoration, which requires further improvement in future work.

In 2019, Wen, Q. *et al.* [28] presents a novel approach to single image reflection removal that goes beyond the limitations of linear methods. The proposed method

comprises two key components: a reflection synthesis network and a multi-branch reflection removal network. The reflection synthesis network generates diverse reflection images by predicting a non-linear alpha blending mask. This mask is then used as a side output for supervision in the multi-branch reflection removal network. The network is trained using synthetic and real-world datasets, incorporating reconstruction loss, gradient loss, and pixel-wise loss to optimize the removal performance. Finally, quantitative and qualitative evaluations are conducted to assess the effectiveness of the proposed method in handling various reflection scenarios.

However, this method also has a series of limitations. First, compositing reflection data via alpha blending masks to correctly simulate various physical conditions in scenes with complex lighting and viewing angles remains a challenge. Secondly, adopting complex loss functions and multi-task learning strategies requires fine tuning and is very sensitive to various hyperparameters during training.

In 2021, Zheng, Q. *et al.* [18] introduced a novel approach that explicitly accounts for the absorption effect in the reflection removal process. This approach involves a two-step solution: first estimating the absorption effect and then utilizing this estimation to facilitate accurate recovery of transmission images. By incorporating absorption modeling into the reflection removal pipeline, the method aims to overcome the limitations of existing techniques and achieve state-of-the-art performance. To implement this approach, the method employs deep learning techniques, leveraging a sophisticated neural network architecture. This architecture consists of two branches: one responsible for estimating the reflection image, and the other for predicting the absorption effect. By

training this network using a carefully designed loss function that incorporates constraints to ensure accurate absorption estimation, the method aims to optimize the network parameters for improved generalization capacity. The proposed method's main steps involve first estimating the absorption effect from the input image data. Subsequently, the estimated absorption effect is utilized to guide the accurate recovery of transmission images. Throughout this process, the method emphasizes rigorous training and optimization of the neural network to minimize the loss function while effectively modeling the absorption effect.

However, this method also has some shortcomings: First, estimating the absorption effect e from the image I is an ill-posed problem, which may lead to inaccurate estimation results. Second, the simplification of modeling the absorption effect may make it difficult to directly validate the model with real data. Third, the estimation process is affected by the unknown scene and the camera's image signal processor (ISP).

In the year of 2023, Rosh G. *et al.* [9] proposed an innovative unsupervised reflection removal method utilizing diffusion models. Leveraging Denoising Diffusion Probabilistic Models (DDPM), the approach learns a distribution of reflection-free images autonomously. During inference, it employs a forward diffusion process conditioned on the input image corrupted by reflections to generate an initial estimate of the reflection-free output. Subsequently, a reverse diffusion process, facilitated by a cascade of denoisers, refines the initial estimate to produce the final reflection-free image. Noteworthy is the method's capability to dynamically control the strength of reflection removal by adjusting diffusion parameters during inference, providing flexibility in

balancing output quality and computational complexity. The proposed methodology revolves around utilizing Denoising Diffusion Probabilistic Models (DDPM) for unsupervised reflection removal. In the training phase, a DDPM is unconditionally trained solely on reflection-free images to learn a distribution of such images. During inference, the method employs a forward diffusion process conditioned on the input image corrupted by reflections to generate an initial estimate of the reflection-free output. Subsequently, a reverse diffusion process, implemented through a cascade of denoisers, refines the initial estimate to yield the final reflection-free image. This methodology enables dynamic control over the strength of reflection removal by adjusting diffusion parameters, allowing users to fine-tune the balance between output quality and computational complexity during inference.

However, this method also has some limitations. The most typical one is: In addition, this method may cause the loss of image details in some cases, especially when processing images with complex textures, the denoising process may affect the quality of the final generated image.

In 2024, Z. Chen *et al.* [1] proposed a new solution to the problem of single image reflection removal, namely SRNet. The article first introduces the challenges and importance of single-image reflection removal, and then describes the structure and working principle of SRNet in detail. SRNet mainly consists of two modules: reflection estimator and reflection removal module. The reflection estimator uses the visual encoder and reflection decoder to estimate the reflection part in the input image, and the reflection removal module uses the reflection estimation results and the mixed image to generate

the final reflection removal result. The article also introduces the training process and optimization goals of SRNet, including reconstruction loss, perceptual loss, mask loss and adversarial loss, etc. Finally, the article verifies the effectiveness and performance advantages of SRNet through comparison with other existing methods and experiments on multiple data sets.

This network also has some limitations: First, when SRNet handles areas with lower transmission layer intensity, the reflection layer covers most of the content. Second, domain adaptation methods are needed to improve the generalization ability of the model. Overall, this article proposes an innovative method to solve the problem of single image reflection removal and experimentally proves its effectiveness and superiority in this field.

2.2. Limitation and Trend

2.2.1 Limitation

There are some common limitations for learning-based methods, for non-learning-based methods, and for all methods.

The common limitations of non-learning methods include alignment issues, limitations in handling complex scenes, multi-image dependency, and computational resource requirements. These methods often rely on image alignment, such as SIFT-flow image alignment. If the alignment is inaccurate, it may lead to incomplete separation of reflections and background. In addition, non-learning methods often have difficulty in accurately separating images containing multiple layers of overlapping reflections or

blurred edges between background and reflection layers. These methods usually require images from multiple perspectives, limiting their applicability in single-image scenarios. Although non-learning methods may not require the training of deep learning models, complex image processing steps such as alignment and gradient processing still require high computational resources.

The limitations of learning-based methods are mainly reflected in the dependence on training data, alignment issues, challenges in complex scenes, computing resource requirements and real-time processing limitations. Deep learning methods are highly dependent on high-quality and diverse training data. Insufficient or low-quality training data will affect the performance of the model in practical applications. Although some learning-based methods are able to handle unaligned data, alignment issues may still affect the performance of the model, especially when dealing with moving objects or handheld images. The effectiveness of the model may decline when dealing with reflections with high reflection intensity or reflections that are highly similar to the background. Deep learning models usually require a lot of computing resources for training and inference, which limits their application on resource-constrained devices. In addition, multi-step deep learning methods are usually difficult to achieve real-time processing and cannot meet the rapid response requirements of some practical applications.

Common limitations of all methods include artifact issues, handling diverse exposure times, and the diversity of application scenarios. Whether it is non-learning or learning methods, dealing with artifact issues such as color distortion, non-uniform

lighting, and new shadows caused by occlusion remains a challenge. When the ambient image and the flash image require different exposure times, accurately calculating the reflection-free flash-only image requires additional processing steps, which increases the complexity of the method. Different methods perform differently in different types of reflective scenes and image content, and their universality still needs to be further improved.

2.2.2 Trend

Future research trends include improving alignment techniques, enhancing model robustness, optimizing computational resources, expanding data diversity, and developing real-time processing technologies. Further studies on more efficient image alignment techniques are necessary to improve alignment accuracy and robustness, addressing issues like handheld photography jitters. Developing methods capable of handling complex scenes and extreme conditions, such as exploring multispectral imaging technologies and various auxiliary information, is crucial. Researching more efficient algorithms and model compression techniques to reduce the demand for computational resources and storage space will make reflection removal methods applicable in resource-constrained environments like mobile devices. Expanding the diversity of training datasets to cover more types of reflections and scenarios will enhance model generalization and practical effectiveness. Lastly, developing technologies capable

of real-time reflection removal is essential to meet the rapid response needs of practical applications, particularly in video processing and real-time imaging systems.

2.3. Datasets

'CID' Dataset

Wan R. *et al.* [44] constructed a specific dataset for colored glass called ColoredDataset Wan R. et al. constructed a dataset called ‘ColoredDataset’ (‘CID’ for short). This dataset contains two parts: ‘Part A’ and ‘Part B’. ‘Part A’ includes 329 groups of images, and ‘Part B’ includes 165 groups of images. The dataset ‘CID Part A’ is used for model training, and ‘CID Part A’ is used for model testing.

'SIR²' Dataset

Wan R. *et al.* [44] constructed a dataset for postcard, solid object, and wild scenes called ‘SIR2’ Dataset’. This dataset includes 3 parts, 500 groups of images in total. Where, ‘Postcard’ includes 199 groups of images, ‘SolidObject’ includes groups of images, ‘Wildscene’ includes groups of images. These 3 Datasets are used for comparing.

Table 2.1 shows some basic information of these datasets.

Table 2.1 Datasets used for training, testing, and comparing.

| Dataset' | CID Part A | CID Part B | Postcard | SolidObject | Wildscene |
|----------|--------------------|--------------------|-------------------|-------------------|-------------------|
| Contents | Real-world data | Real-world data | Synthetic data | Synthetic data | Synthetic data |
| Used for | Model training | Model testing | Comparing | Comparing | Comparing |
| Size | 329 | 165 | 199 | 200 | 101 |

2.4. Evaluation Metrics

2.4.1. PSNR

PSNR stands for Peak Signal-to-Noise Ratio. It's a metric used to measure the quality of reconstructed or compressed images or videos compared to the original, uncompressed version. It's calculated as the ratio of the maximum possible power of a signal to the power of corrupting noise that affects the fidelity of its representation. PSNR is often expressed in decibels (dB), and a higher PSNR value indicates better quality.

Here's the formula for PSNR:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (1)$$

Where:

- MAX is the maximum possible pixel value of the image (for example, 255 for 8-bit grayscale images or 65535 for 16-bit grayscale images).
- MSE is the Mean Squared Error between the original and the compressed image.

A higher PSNR value typically indicates a higher quality image, meaning that the compressed image is closer to the original. However, it's essential to note that PSNR doesn't always perfectly correlate with perceived image quality, especially in cases where the compression artifacts are visually noticeable.

2.4.2. SSIM

SSIM stands for Structural Similarity Index Measure. It's a method used to measure the similarity between two images. Unlike PSNR, SSIM takes into account the perceived changes in structural information, luminance, and contrast, which makes it more closely aligned with human perception. The SSIM index is calculated based on three components: luminance similarity, contrast similarity, and structure similarity. The formula for SSIM is as follows:

$$SSIM(x, y) = \frac{(2 \cdot \mu_x \cdot \mu_y + C_1) \cdot (2 \cdot \sigma_{x,y} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) \cdot (\sigma_x^2 + \sigma_y^2 + C)} \quad (2)$$

Where:

- x and y are the compared images.
- μ_x and μ_y are the means of x and y , respectively.
- σ_x^2 and σ_y^2 are the variances of x and y , respectively.
- $\sigma_{x,y}$ is the covariance of x and y .
- C_1 and C_2 are constants to stabilize the division with weak denominator.

The SSIM index ranges from -1 to 1, where 1 indicates perfect similarity between the images. Higher SSIM values indicate greater similarity between the images, while lower values indicate more dissimilarity.

SSIM is widely used in image processing and compression applications as it correlates better with human visual perception compared to simpler metrics like PSNR.

2.4.3. AMBE

AMBE stands for Absolute Mean Brightness Error. It represents the absolute difference in average brightness between the enhanced image and the original image. The formula for AMBE is as follows:

$$AMBE = |\mu_{orig} - \mu_{enh}| = \left| \frac{1}{N} \sum_{i=1}^N I_{orig}(i) - \frac{1}{N} \sum_{i=1}^N I_{enh}(i) \right| \quad (3)$$

Where:

- N is the total number of pixels in the image
- $I_{orig}(i)$ represents the grayscale value of the i^{th} pixel.
- $I_{enh}(i)$ represents the grayscale value of the i^{th} pixel after enhancement.

The smaller the value of AMBE, the better the enhanced image is at maintaining the original average brightness, that is, the smaller the average brightness change. Ideally, AMBE should be close to zero, which means that the average brightness of the enhanced image is consistent with the original image.

3. RESEARCH METHODS

In reflection removal, an image I with reflections can be modeled as the weighted additive composition of a transmission layer T and a reflection layer R . Precisely, following the alpha blending model in [21][37][39], we express the composition procedure as:

$$I = W \circ T + R \quad (4)$$

Where: W here is an alpha blending mask and \circ indicates the element-wise multiplication.

3.1. Comparison Study

Implement SIRR [16], CORRN [26] and IBCLN [21] 3 algorithms using PyTorch on a Laptop with an Nvidia GeForce RTX 3070 Laptop GPU. Use dataset ‘CID Part A’ to retrain these 3 models. After training is complete, compare the model's performance when inferring the training set, test set, and comparison sets.

For the SIRR [16] model, following the training procedure outlined by Dong, Z *et al.*. The model is trained using the ADAM optimizer with the following settings: an initial learning rate of $2e^{-4}$ for the first 60 epochs, reduced to $1e^{-4}$ afterward, a batch size of 1, and ADAM parameters set to $\beta_1 = 0.5$ and $\beta_2 = 0.99$.

The training involves minimizing the loss function used by Dong, Z *et al.* [16], and the epoch average loss change curve is shown in **Figure 3.1**.

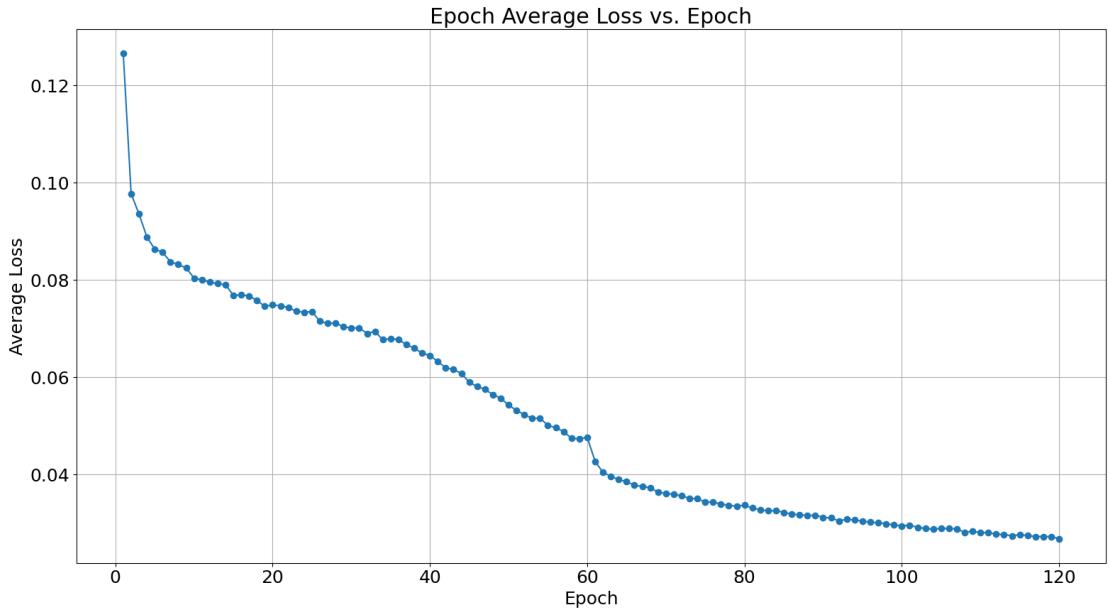


Figure 3.1 Average loss trend while training SIRR [16] model

For the CoRRN [26] model, following the training procedure outlined by Wan, R. *et al.*. The model is trained using the ADAM optimizer with the following settings: an initial learning rate of $1e^{-4}$ for the first 60 epochs, reduced to $1e^{-5}$ afterward, a batch size of 1. The training involves minimizing the loss function used by Wan, R. *et al.* [26], and the epoch average loss change curve is shown in **Figure 3.2**.

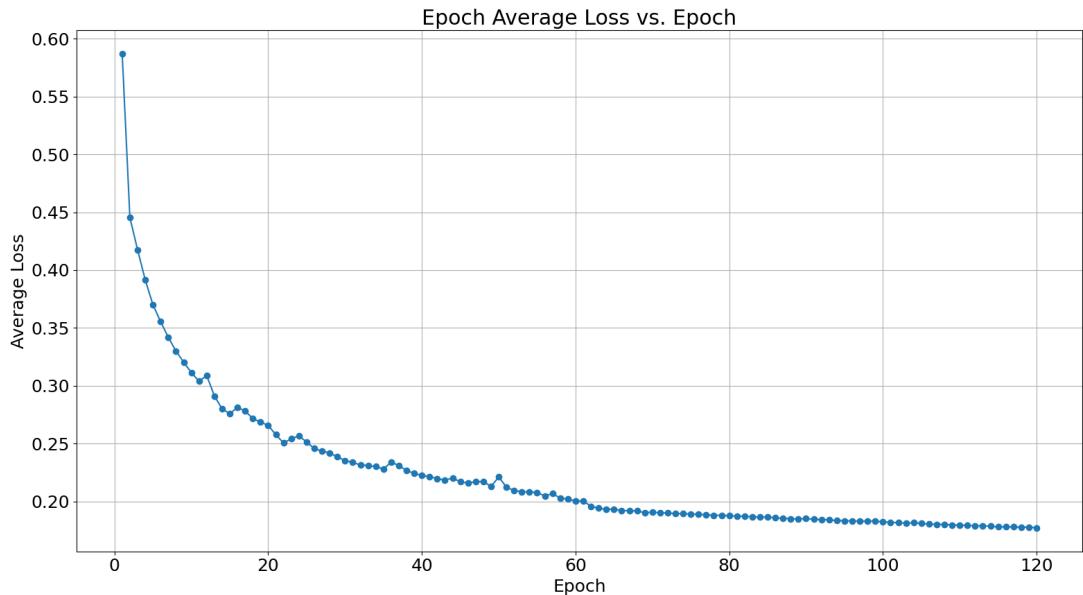


Figure 3.2 Average loss trend while training CoRRN [26] model

For the IBCLN [21] model, following the training procedure outlined by Li, C. *et al.*. The model is trained using the ADAM optimizer with the following settings: an initial learning rate of $2e^{-4}$ for the 120 epochs, a batch size of 1, and ADAM parameters set to $\beta_1 = 0.5$ and $\beta_2 = 0.99$.

The training involves minimizing the loss function used by Li, C. *et al.* [21], and the epoch average loss change curve is shown in **Figure 3.3**.

After completing the training process, the three retrained models (SIRR [16], CORRN [26], and IBCLN[21]) were applied to five different datasets to evaluate their performance. The evaluation included quantitative analyses, detailed in **Chapter 4.1.1**, and qualitative analyses, discussed in **Chapter 4.1.2**.

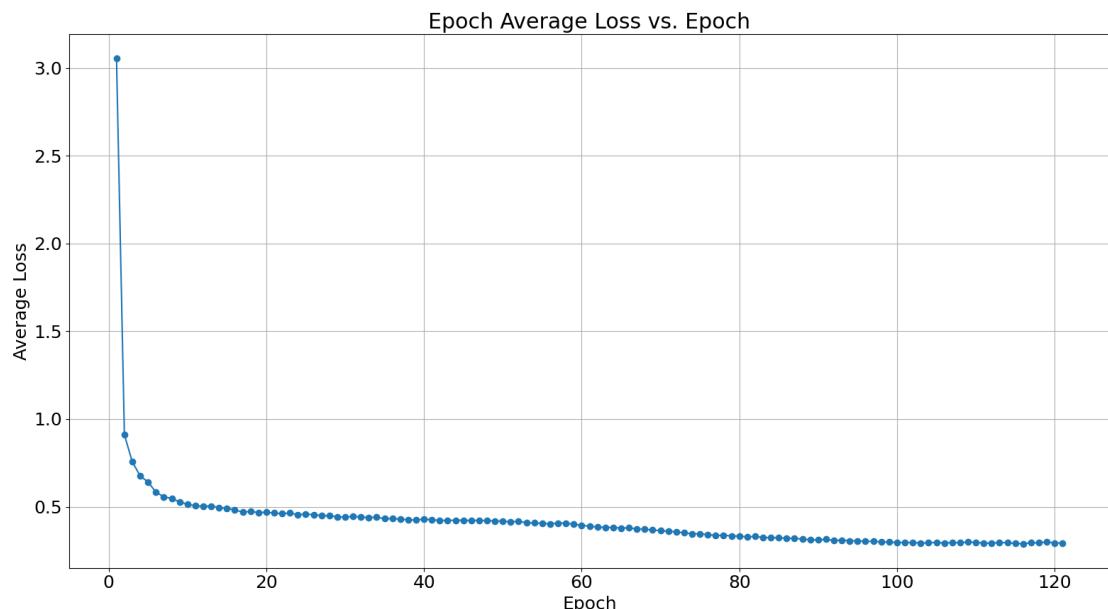


Figure 3.3 Average loss trend while training IBCLN [21] model

3.2. Proposed Improvement

Based on the findings of the comparative study, the IBCLN [21] model was selected for further enhancements. The primary focus of the improvement efforts was to enhance the brightness of the inferred images. Histogram equalization (HE) and differential grayscale histogram equalization (DHE) are both good brightness enhancement methods. Tanaka, H., & Taguchi, A. proposed a generalized histogram equalization (GHE) [23] method including the HE and the DHE in 2020. From Tanaka, H., & Taguchi, A.'s article,,

3.2.1. Enhancement Method

In Tanaka, H., & Taguchi, A.'s experiment, they set the value of parameter α in **Equation (7)** to 2.6, 1.5, and 6.8 for images of Couple , Airplane, and Girl respectively.

To optimize the parameter α for this work, increase the values of α within the range [0, 12] with a step size of 0.1. GHE [23] is applied with these α values to enhance the output images of IBCLN [21] on the training('CID' part A) dataset. The average AMBE, average SSIM, and average PSNR of the enhanced output images are calculated for each corresponding α value. **Figure 3.4** plots the changes in Average AMBE, Average PSNR, and Average SSIM after applying GHE [23] as alpha increases.

From the analysis of the optimization result detailed in **Chapter 4.2.1**, find the optimal α value which let average AMBE less and let average SSIM and average PSNR larger. Then, apply GHE [23] to the output of IBCLN [21] on other datasets with this find optimal α value.

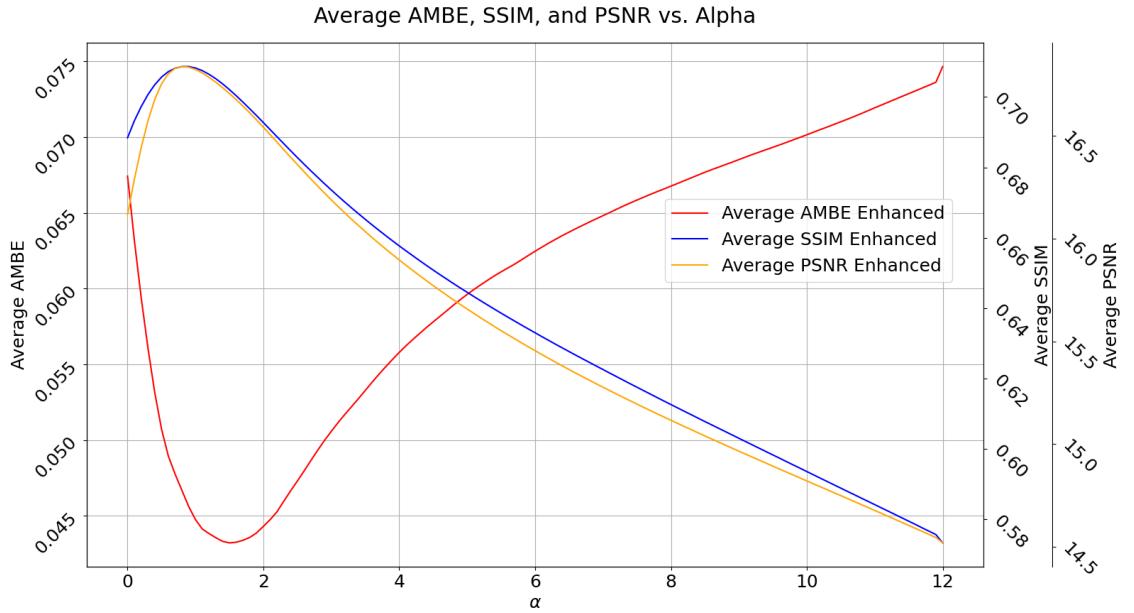


Figure 3.4 Various Metrics vs α curves

3.2.2. GHE Equations

The equations of GHE [23] are as follows:

The GHE [23] will map an input gray-level r into an output gray level t using the following transformation function.

$$t = (L - 1) \cdot c(r) \quad (5)$$

Where, L is the number of gray levels (for example, 256 for 8-bit images). $c(r)$ is the cumulative distribution function of the generalized histogram and is given by the following equation.

$$c(r) = \frac{\sum_{k=0}^r h_d^\alpha(k)}{\sum_{k=0}^{(L-1)} h_d^\alpha(k)} \quad (6)$$

Where, the histogram created by the power of α of the gradient value of gray-level is given by the following equation.

$$h_d^\alpha(r) = \sum_{(i,j) \in D_r} \{\Delta(i,j)\}^\alpha \quad (7)$$

Where, $int\{\}$ represents the truncated integer conversion processing. D_r is a region composed of pixels whose value is r . $\Delta(i,j)$ is the gradient value at the point (i,j) of the gray-level r of the input image given by the following equation.

$$\Delta(i,j) = int \left\{ \sqrt{\Delta_H(i,j)^2 + \Delta_V(i,j)^2} \right\} \quad (8)$$

Where:

$$\begin{aligned} \Delta_H(i,j) &= \{I(i+1,j+1) + 2I(i+1,j) + I(i+1,j-1)\} \\ &\quad - \{I(i+1,j+1) + 2I(i-1,j) + I(i-1,j-1)\} \\ &\quad - \{I(i+1,j+1) + 2I(i-1,j) + I(i-1,j-1)\} \end{aligned} \quad (9)$$

$$\begin{aligned} \Delta_V(i,j) &= \{I(i-1,j+1) + 2I(i,j+1) + I(i+1,j-1)\} \\ &\quad - \{I(i-1,j+1) + 2I(i,j-1) + I(i+1,j-1)\} \end{aligned} \quad (10)$$

3.2.3. RGB to HSV Convert Equations

The equations for converting RGB color space to HSV color space [45] are as follows:

$$h = \begin{cases} 0, & \text{if } max = min \\ \frac{g-b}{delta} + 0, & \text{if } max = r \text{ and } g \geq b \\ \frac{g-b}{delta} + 60, & \text{if } max = r \text{ and } g < b \\ \frac{b-r}{delta} + 2, & \text{if } max = g \\ \frac{r-g}{delta} + 4, & \text{if } max = b \end{cases} \quad (11)$$

$$s = \begin{cases} 0, & \text{if } max = 0 \\ \frac{delta}{max}, & \text{otherwise} \end{cases} \quad (12)$$

$$v = max \quad (13)$$

Where, (r, g, b) are the red, green, and blue coordinates of a color respectively, and their values are real numbers between 0 and 1. max is equal to the maximum of these values, min is equal to the minimum of these values, and $delta$ is equal to the difference between the maximum and minimum of these values, given by the following equation.

$$\begin{cases} max = max\{(r, g, b)\} \\ min = min\{(r, g, b)\} \\ delta = max - min \end{cases} \quad (14)$$

3.2.4. HSV to RGB Convert Equations

The equations for converting HSV color space to RGB color space [45] are as follows:

$$(r, g, b) = \begin{cases} (v, t, p), & \text{if } hi = 0 \\ (q, v, p), & \text{if } hi = 1 \\ (p, v, t), & \text{if } hi = 2 \\ (p, q, v), & \text{if } hi = 3 \\ (t, p, v), & \text{if } hi = 4 \\ (v, p, q), & \text{if } hi = 5 \end{cases} \quad (15)$$

Where hi is the color bin index, f is the fractional part of the hue, p is a dim version of a color, q is an intermediate color, and t is another intermediate color. The purpose of these intermediate variables is to determine the main components of the RGB color based on the color bin hi where the hue h is located, and use different combinations of p, q, t and v to calculate the RGB color components. The specific combination depends on the

value of hi , which determines which main color bin the current hue belongs to, and interpolates between adjacent colors. This results in the corresponding RGB color, representing the corresponding hue, saturation, and brightness, given by the following equation. [45]

$$\begin{cases} hi = \lfloor h \rfloor \\ f = h - hi \\ p = v * (1 - s) \\ q = v * (1 - f * s) \\ t = v * (1 - (1 - f) * s) \end{cases} \quad (16)$$

4. RESULTS AND DISCUSSIONS

4.1. Results on Comparison Study

4.1.1. Quantitative comparisons

Table 4.1 reports the performance comparisons on five datasets. It can be seen that the IBCLN [21] method ranked **top** on the ColoredDataset PartA, ColoredDataset PartB, Postcard, SolidObject, Wild datasets (in both PSNR and SSIM rankings). All three methods ranked **top** in the PSNR rankings on the ColoredDataset PartB dataset, and SIRR [16] and IBCLN [21] ranked **top** on the SolidObject dataset, while CoRRN [26] ranked **top** on the Postcard dataset (SSIM ranking).

Table 4.1 Quantitative comparison to chosen algorithms.

| Method Index | SIRR [16] | | CoRRN [26] | | IBCLN [21] | |
|----------------------------|-----------|--------|------------|--------|------------|--------|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ColoredDataset PartA (329) | 18.0359 | 0.7667 | 14.7393 | 0.5359 | 18.6705 | 0.7754 |
| ColoredDataset PartB (165) | 19.3358 | 0.6019 | 18.1488 | 0.5851 | 20.3224 | 0.6251 |
| Postcard (199) | 13.3401 | 0.6001 | 14.5883 | 0.6329 | 14.8523 | 0.6779 |
| SolidObject (200) | 17.2956 | 0.7678 | 13.1529 | 0.5096 | 19.2015 | 0.7941 |
| Wildscene (101) | 16.1946 | 0.7230 | 14.7979 | 0.5655 | 19.3970 | 0.7916 |

Figures 4.1 through **4.6** illustrate the comparison of PSNR and SSIM distributions for the SIRR [16], CoRRN [26], and IBCLN [21] models across the five datasets. Specifically, **Figure 4.1** shows the PSNR distribution for the SIRR [16] model, while **Figure 4.2** depicts its SSIM distribution. Similarly, **Figures 4.3** and **4.4** present the PSNR

and SSIM distributions, respectively, for the CoRRN [26] model. Lastly, **Figures 4.5** and **4.6** display the PSNR and SSIM distributions for the IBCLN [21] model. These figures collectively provide a comprehensive comparison of the models' performance across different datasets.

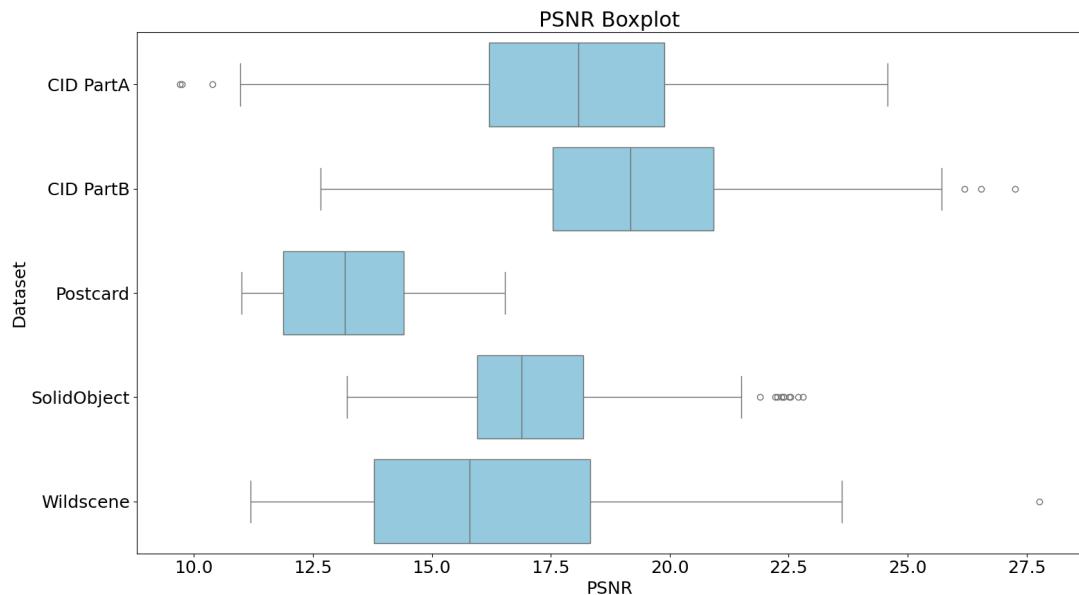


Figure 4.1 PSNR comparison of SIRR [16] model on different datasets.

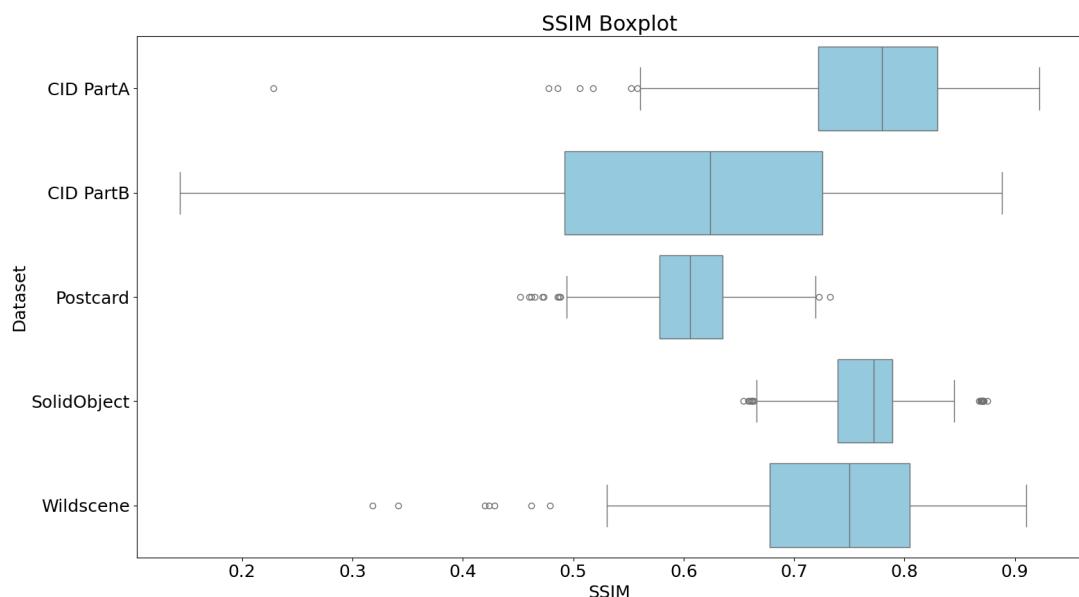


Figure 4.2 SSIM results of SIRR [16] model on different datasets.

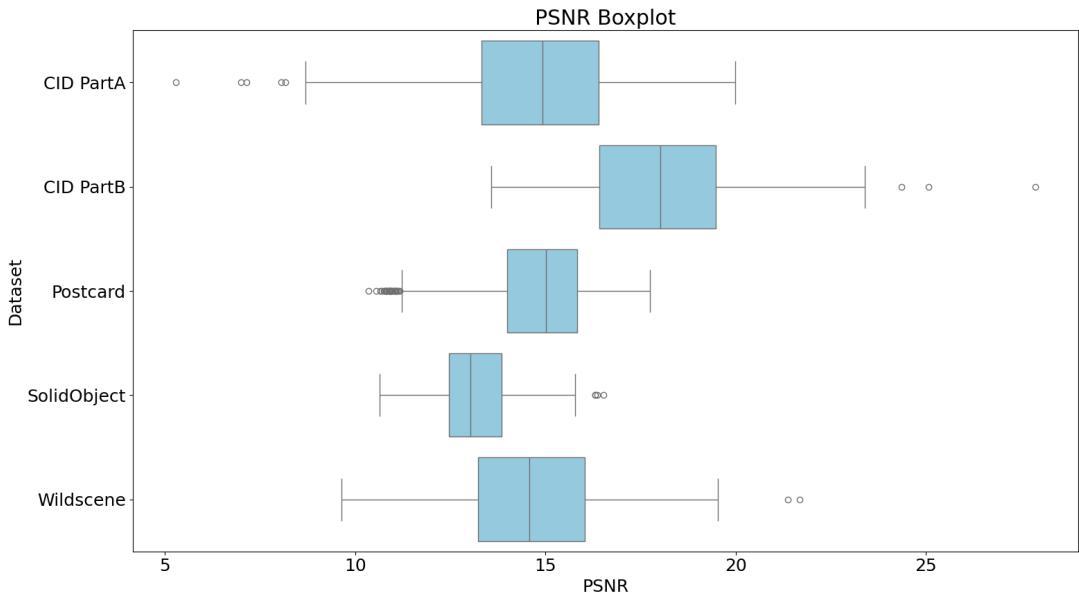


Figure 4.3 PSNR results of CoRRN [26] model on different datasets.

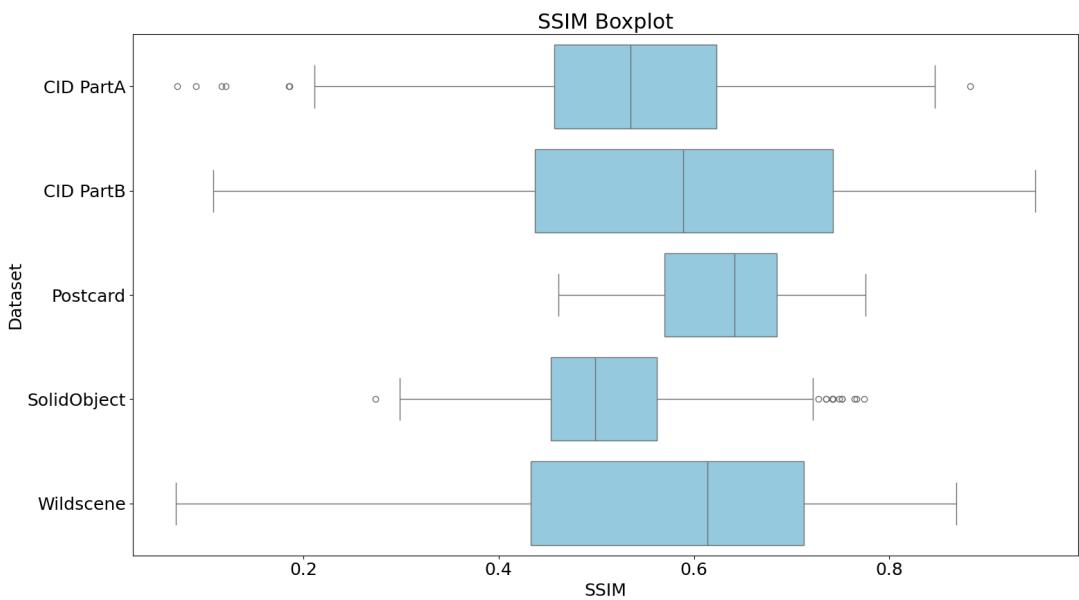


Figure 4.4 SSIM results of CoRRN [26] model on different datasets.

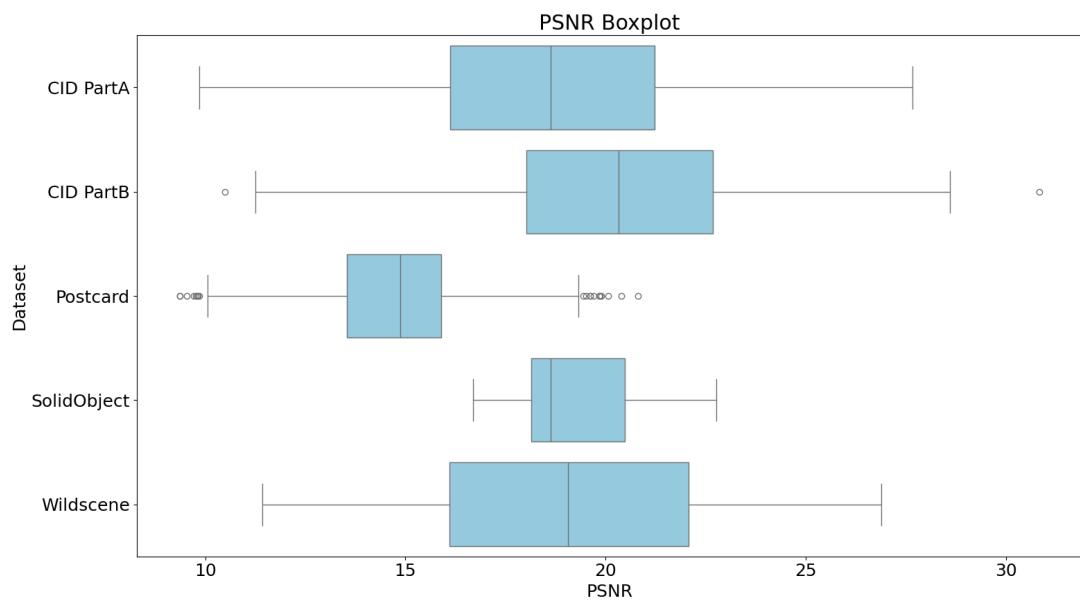


Figure 4.5 PSNR results of IBCLN [21] model on different datasets.

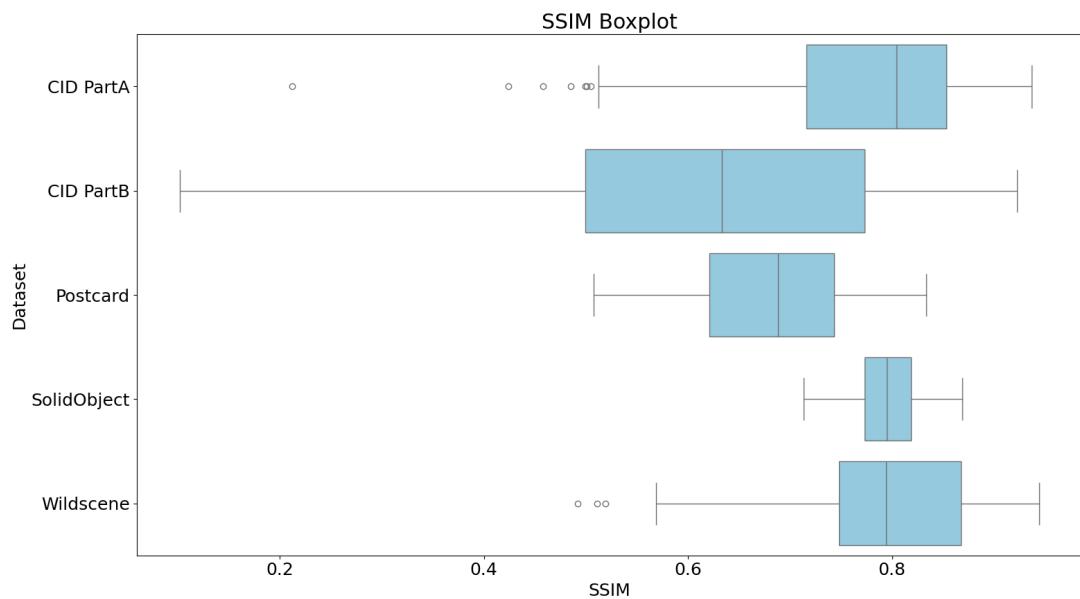


Figure 4.6 SSIM results of IBCLN [21] model on different datasets.

4.1.2. Qualitative comparisons

Figure 4.7 shows the reflection removal results of the 3 methods. These images are from the ‘training’ dataset (row 1), the ‘testing’ dataset (rows 2-3), the ‘Postcard’ dataset (row 4-5), the ‘SolidObject’ dataset (row 6) and the ‘Wildscene’ dataset (rows 7-8).

It can be seen that existing methods typically fail to remove large-area reflections and strong highlights. In contrast, IBCLN [21] can remove most undesirable reflections while preserving high frequency details in the transmission layer.

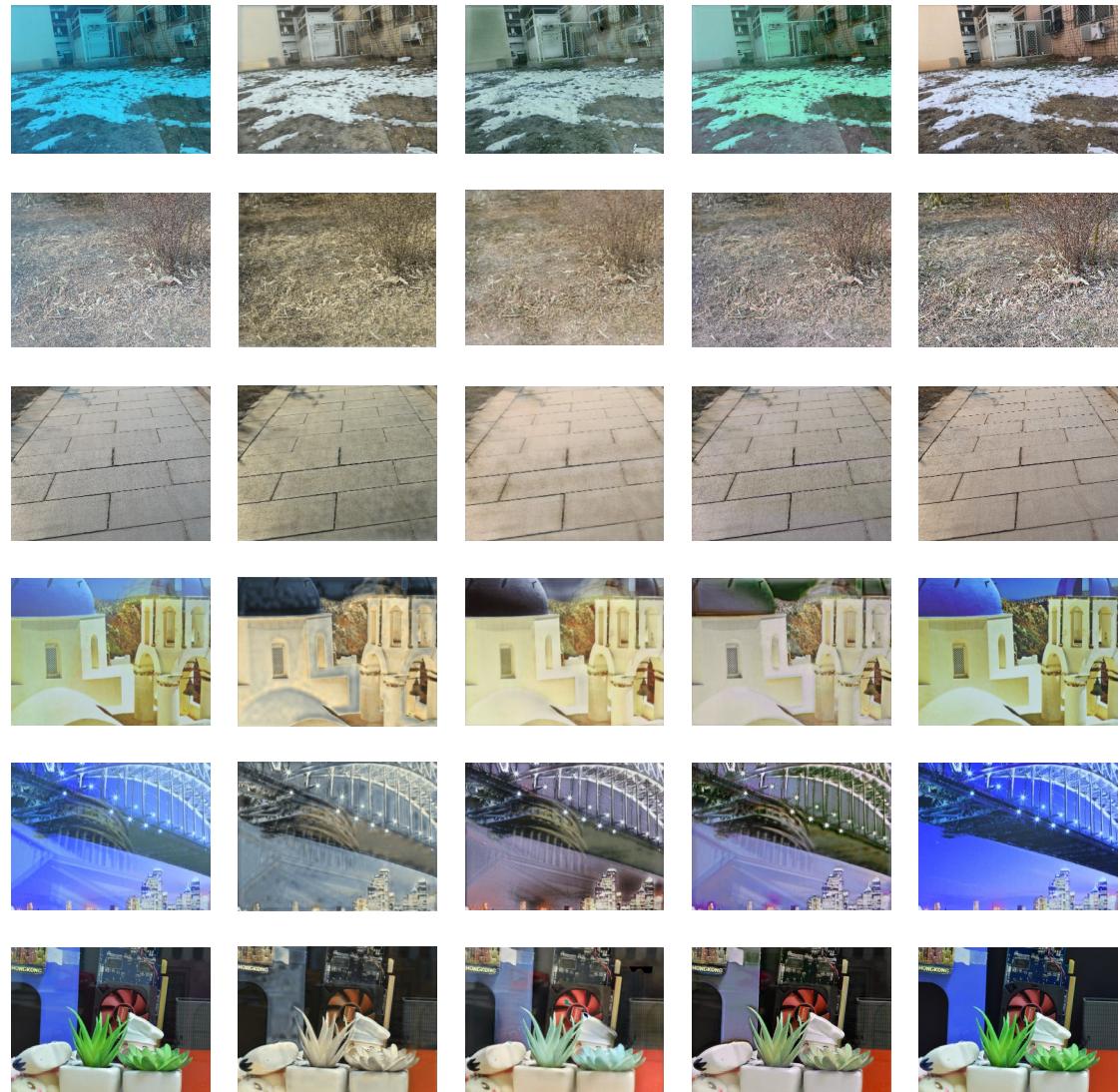


Figure 4.7, continued

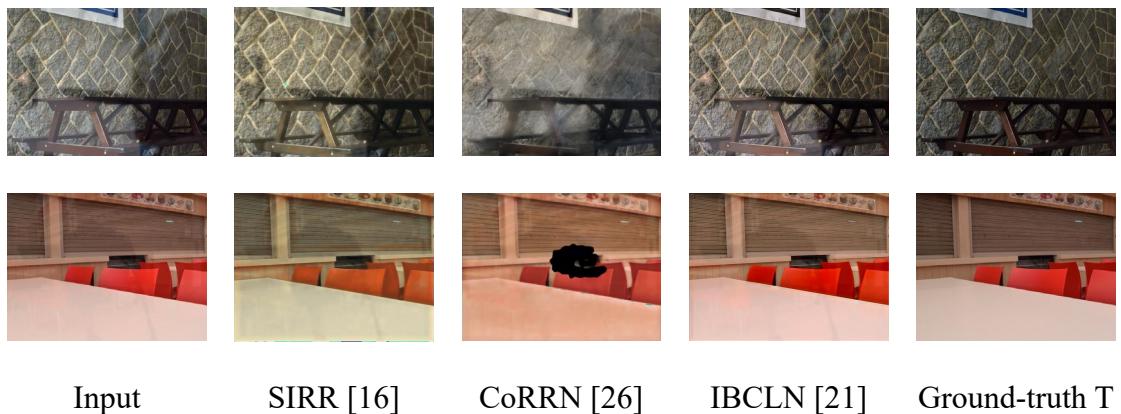


Figure 4.7 Qualitative comparisons between the 3 methods.

4.1.3. Summary of Comparison Study

In this study, the comprehensive quantitative and qualitative comparisons of three state-of-the-art reflection removal methods: SIRR [16], CoRRN [26], and IBCLN [21] is conducted. The findings can be summarized as follows:

In terms of quantitative performance, the IBCLN [21] method consistently achieved top rankings across all five datasets (**ColoredDataset PartA**, **ColoredDataset PartB**, **Postcard**, **SolidObject**, and **Wildscene**) in terms of both PSNR and SSIM metrics. Specifically, IBCLN [21] outperformed the other methods on the **ColoredDataset PartA**, **ColoredDataset PartB**, and **Wildscene** datasets. CoRRN [26] showed superior performance on the **Postcard** dataset, while both SIRR [16] and IBCLN [21] shared top rankings on the **SolidObject** dataset. **Figures 4.1 - 4.6** illustrated the PSNR and SSIM distributions for each model across the five datasets. These plots further confirmed the

dominance of IBCLN [21] in terms of higher PSNR and SSIM values, indicating better overall reflection removal performance and image quality.

Regarding qualitative performance, **Figure 4.7** presented the visual results from the CID and SIR2 benchmark datasets, showcasing the reflection removal capabilities of the three methods compared to the ground truth. The qualitative results highlighted that IBCLN [21] was superior in removing large-area reflections and strong highlights while preserving high-frequency details in the transmission layer. In contrast, SIRR [16] and CoRRN [26] struggled with these challenges, often leaving behind noticeable reflections and artifacts.

In conclusion, the IBCLN [21] method demonstrated superior performance both quantitatively and qualitatively in removing reflections from images, making it a more effective solution for the reflection removal task compared to the other evaluated methods. This comprehensive comparison provides valuable insights into the strengths and weaknesses of current reflection removal techniques and guides future research and development in this area. Despite its strengths, the IBCLN [21] method has several limitations. It can produce artifacts, and the resulting images may suffer from low brightness and low color saturation. These issues can affect the overall quality and visual appeal of the images.

To address these limitations, future work could focus on the following aspects: Reducing artifacts through advanced post-processing techniques or improved network architectures Improving brightness and color saturation to improve the visual quality of the output images. This could involve integrating brightness and color correction

mechanisms in the model or using additional training data with different lighting and color conditions.

4.2. Results on Proposed Improvement

4.2.1. Optimization of α

Figures 4.8, 4.9, and 4.10 analyze the impact of the parameter α on various metrics.

In **Figure 4.8**, the variation in the Average AMBE is depicted as α increases within the range [0, 12], highlighting the α value that results in the minimum Average AMBE.

Figure 4.9 presents the Average PSNR before applying GHE [23] and the subsequent change in Average PSNR as α varies within the same range, identifying the α value that maximizes the Average PSNR. Similarly, **Figure 4.10** illustrates the Average SSIM prior to GHE [23] and the corresponding change in Average SSIM post-GHE [23], pinpointing the α value that optimizes the Average SSIM.

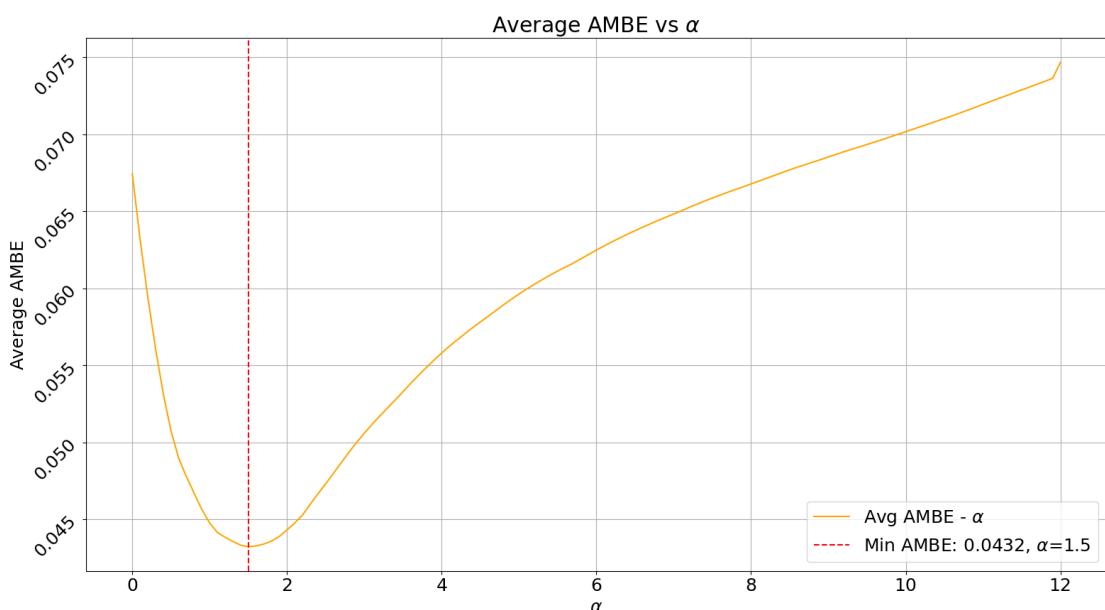


Figure 4.8 Average AMBE vs α curve

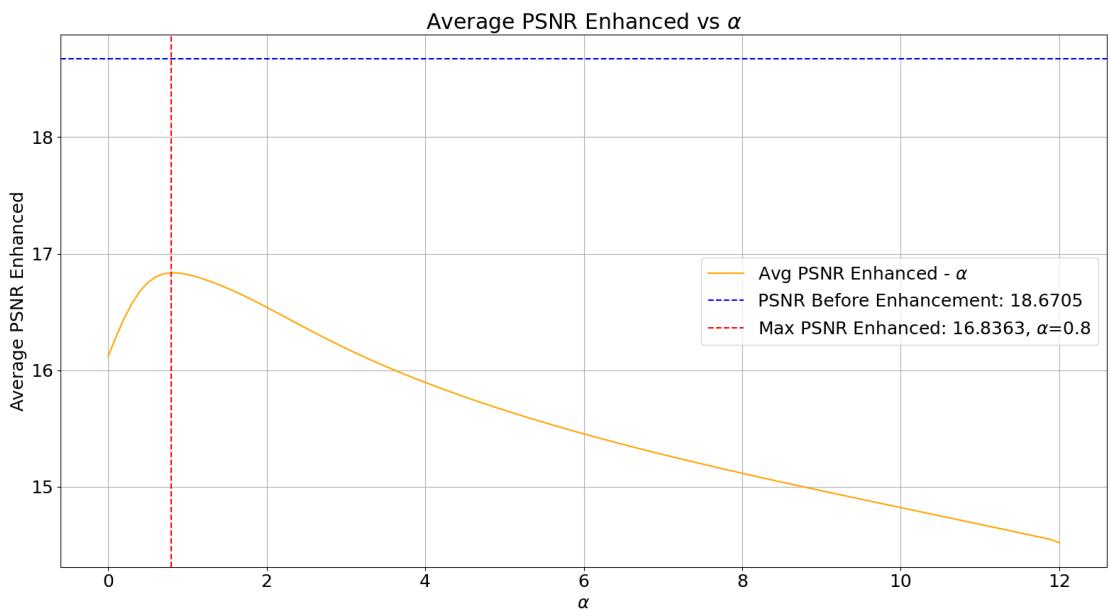


Figure 4.9 Average PSNR Enhanced vs α curve

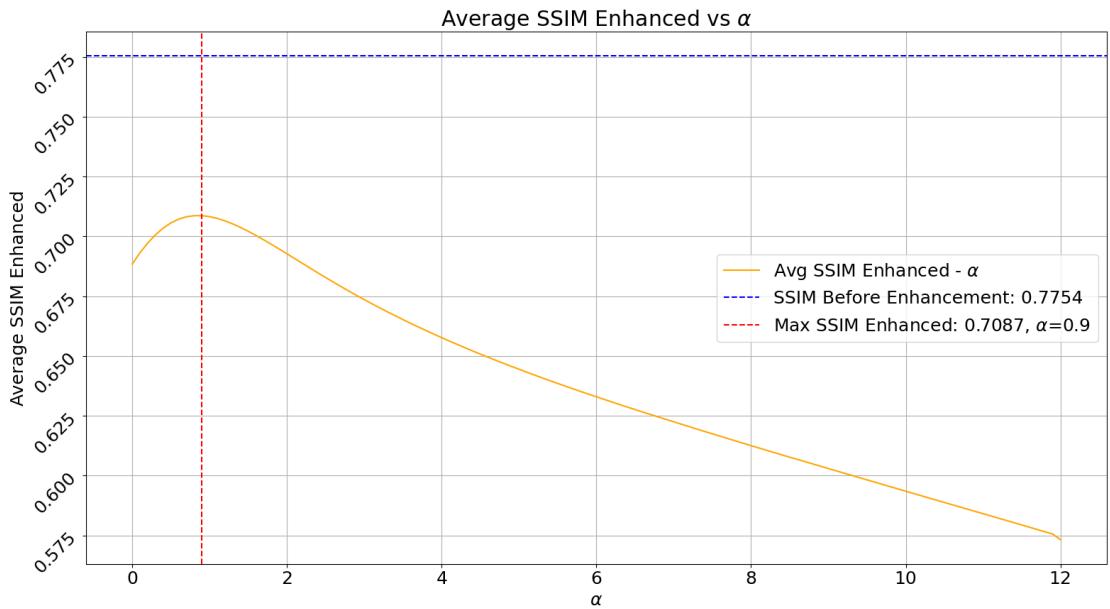


Figure 4.10 Average SSIM Enhanced vs α curve

As can be seen from **Figures 4.8-4.10**, it is obvious that the value of α that maximizes Average AMBE, Average PSNR, and Average SSIM is within the range of

$\alpha \in [0,2]$. For easier viewing, the changes in Average AMBE, Average PSNR, and Average SSIM after applying GHE [23] are plotted in a graph when $\alpha \in [0, 2]$ as α increases, as shown in **Figure 4.11**.

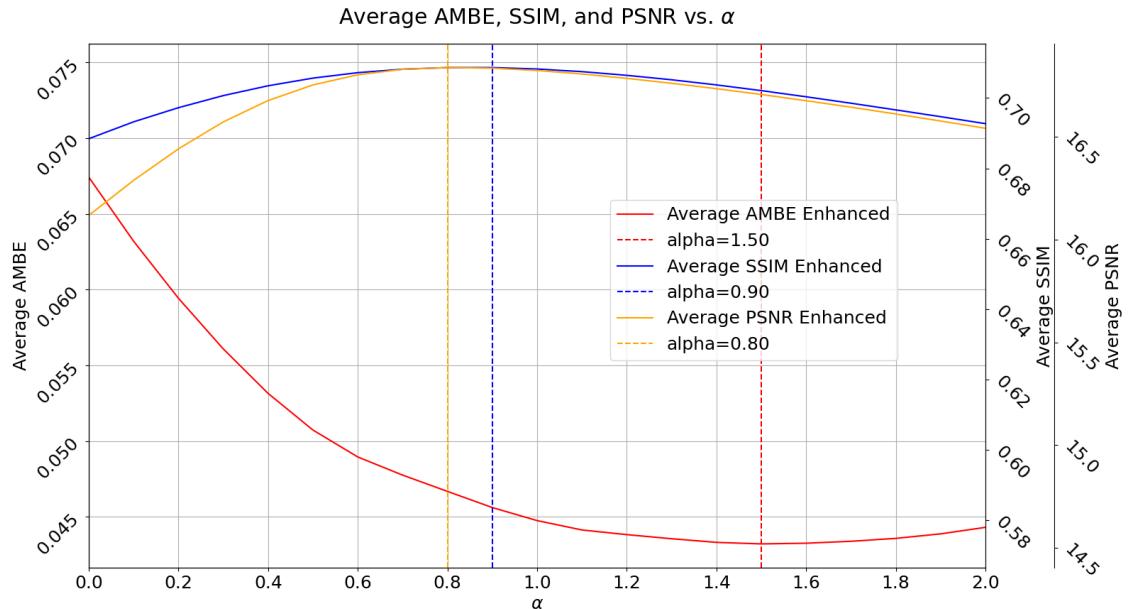


Figure 4.11 Various Metrics vs α curves when $\alpha \in [0, 2]$

From **Figure 4.11**, it can be seen that the values of alpha make metrics reach their maximum value are within the range of $\alpha \in [0.8, 1.5]$. To accurately compare these metrics, report the comparison within the range of $\alpha \in [0.6, 1.7]$, include qualitative comparison, as shown in **Figure 4.12**, and quantitative comparison, as shown in **Table 4.2**, where ‘Avg.’ is the abbreviation of ‘Average’. It can be seen that the Average PSNR Enhanced reaches its **maximum** value at $\alpha = 0.8$, Average SSIM Enhanced achieves its **maximum** value at $\alpha = 0.9$, and Average AMBE achieves its **minimum** value at $\alpha = 1.5$.

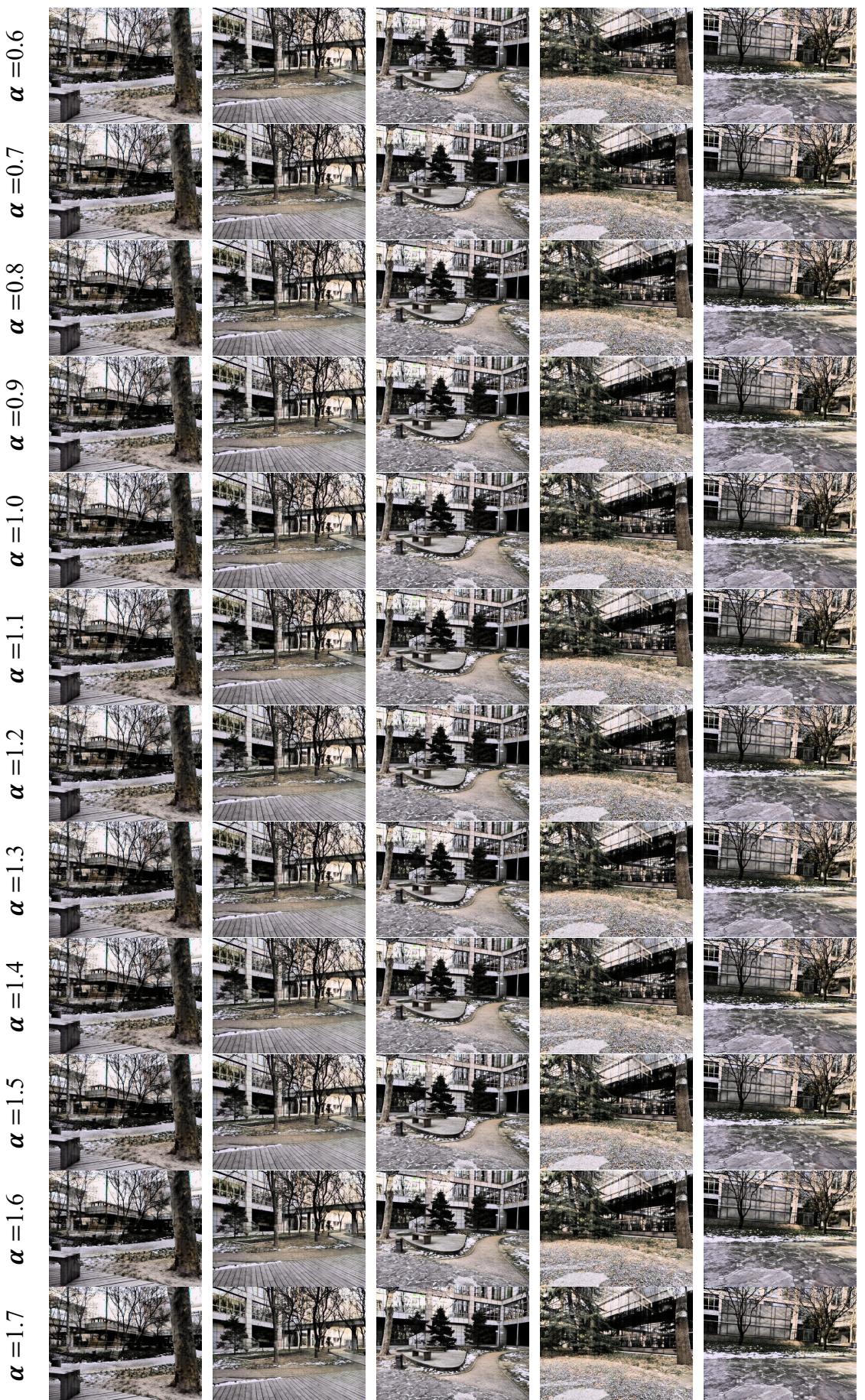


Figure 4.12 Qualitative comparison when $\alpha \in [0.6, 1.7]$

Table 4.2 Quantitative comparison when $\alpha \in [0.6, 1.7]$

| α | Avg. PSNR | Avg. PSNR Enhanced | Avg. SSIM | Avg. SSIM Enhanced | Avg. AMBE |
|----------|-----------|--------------------|-----------|--------------------|-----------|
| 0.6 | 18.670484 | 16.801096 | 0.775366 | 0.707205 | 0.048957 |
| 0.7 | 18.670484 | 16.827695 | 0.775366 | 0.708187 | 0.047760 |
| 0.8 | 18.670484 | 16.836328 | 0.775366 | 0.708645 | 0.046686 |
| 0.9 | 18.670484 | 16.833216 | 0.775366 | 0.708657 | 0.045620 |
| 1.0 | 18.670484 | 16.821941 | 0.775366 | 0.708243 | 0.044758 |
| 1.1 | 18.670484 | 16.805091 | 0.775366 | 0.707494 | 0.044139 |
| 1.2 | 18.670484 | 16.783767 | 0.775366 | 0.706443 | 0.043832 |
| 1.3 | 18.670484 | 16.759831 | 0.775366 | 0.705183 | 0.043561 |
| 1.4 | 18.670484 | 16.733507 | 0.775366 | 0.703720 | 0.043321 |
| 1.5 | 18.670484 | 16.705278 | 0.775366 | 0.702097 | 0.043220 |
| 1.6 | 18.670484 | 16.675162 | 0.775366 | 0.700346 | 0.043265 |
| 1.7 | 18.670484 | 16.643655 | 0.775366 | 0.698519 | 0.043393 |

Since the values of α that optimize these metrics are different, in order to find the optimal α , a new parameter Average Enhanced Various Metrics (Average EVM) is introduced to combine these metrics. EVM is defined by **Equation (16)**.

$$\text{Average EVM} = \frac{\frac{\text{Avg PSNR Enhanced}}{\text{Avg PSNR}} + \frac{\text{Avg SSIM Enhanced}}{\text{Avg SSIM}} - \text{Avg AMBE}}{2} \quad (16)$$

Calculate the value of the new indicator, Average EVM, and plot the variation in EVM as α increases within the range $[0, 2]$. Additionally, identify the α value at which the Average EVM reaches its maximum, as depicted in **Figure 4.12**. It can be observed

that when $\alpha = 0.8$, the Average EVM reaches its maximum value of 0.8550, indicating that the optimal α value is 0.8.

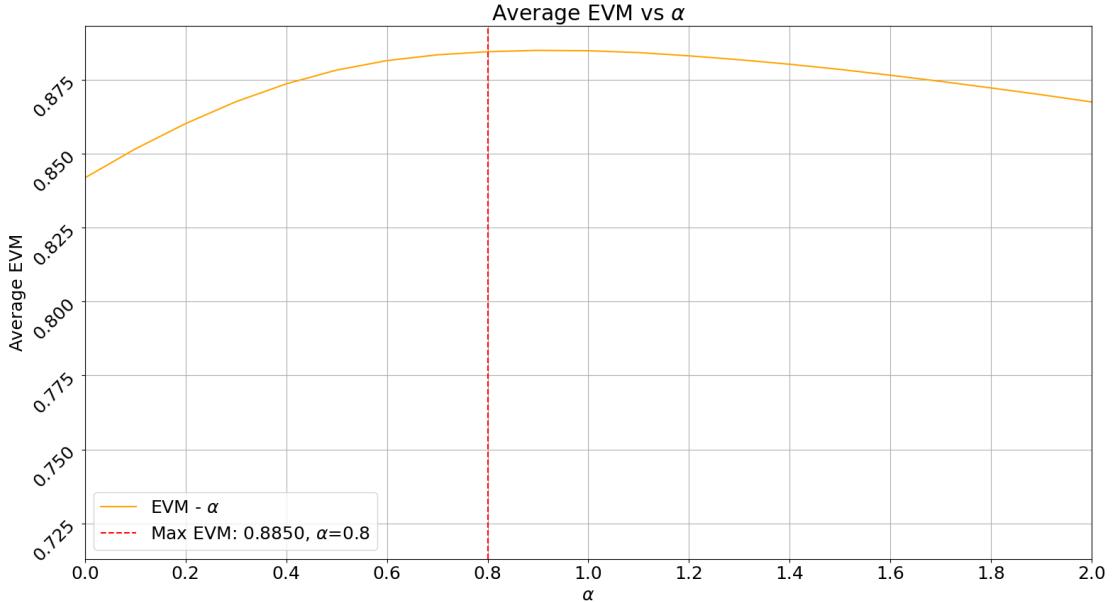


Figure 4.13 Average EVM vs α curve when $\alpha \in [0, 12]$

4.2.2. Enhancement Comparison

Table 4.3 reports the Enhancement performance comparisons on five datasets, where PSNR Reduce and SSIM Reduce are the first and second parts of the numerator in **Equation (16)**, respectively, as shown in **Equation (17)**.

$$\left\{ \begin{array}{l} \text{PSNR Reduce} = \frac{\text{Avg PSNR Enhanced}}{\text{Avg PSNR}} \\ \text{SSIM Reduce} = \frac{\text{Avg SSIM Enhanced}}{\text{Avg SSIM}} \end{array} \right. \quad (17)$$

It can be seen that the performance of enhancement is ranked as **top-1** on ColoredDataset PartA dataset (all metrics ranking), PSNR Reduce is **less than 10%** on the

ColoredDataset PartA dataset, SSIM Reduce is less than 10% on the ColoredDataset PartA and SolidObject datasets, AMBE is less than 10% on all 5 datasets, and EVM is greater than 80% on the ColoredDataset PartA, Postcard, and SolidObject datasets. As reported in **Table 4.1**, the original method ranks top-1 in PSNR on the ColoredDataset PartB dataset and top-1 in SSIM on the SolidObject dataset.

Table 4.3 Quantitative comparison on five datasets when $\alpha = 0.8$

| Metric | ColoredDataset PartA (329) | ColoredDataset PartB (165) | Postcard (199) | SolidObject (200) | Wildscene (101) |
|------------------|-------------------------------|-------------------------------|-------------------|----------------------|--------------------|
| PSNR | 18.670484 | 20.32239 | 14.852329 | 19.201505 | 19.397013 |
| PSNR Enhanced | 16.836328 | 14.301573 | 12.767821 | 17.131622 | 14.846086 |
| PSNR Reduce | 0.098238 | 0.296265 | 0.140349 | 0.107798 | 0.234620 |
| SSIM | 0.775366 | 0.62506 | 0.677853 | 0.794146 | 0.791561 |
| SSIM Enhanced | 0.708645 | 0.456299 | 0.545472 | 0.719527 | 0.644818 |
| SSIM Reduce | 0.086051 | 0.269992 | 0.195295 | 0.093961 | 0.185384 |
| AMBE | 0.046686 | 0.096798 | 0.054196 | 0.053082 | 0.09979 |
| EVM | 0.884513 | 0.668472 | 0.805081 | 0.87258 | 0.740103 |

Figure 4.14 shows the qualitative comparison of enhancement when on 5 datasets. These images are from the ‘training’ dataset (row 1), the ‘testing’ dataset (row 2), the ‘Postcard’ dataset (row 3), the ‘SolidObject’ dataset (row 4) and the ‘Wildscene’ dataset (row 5).

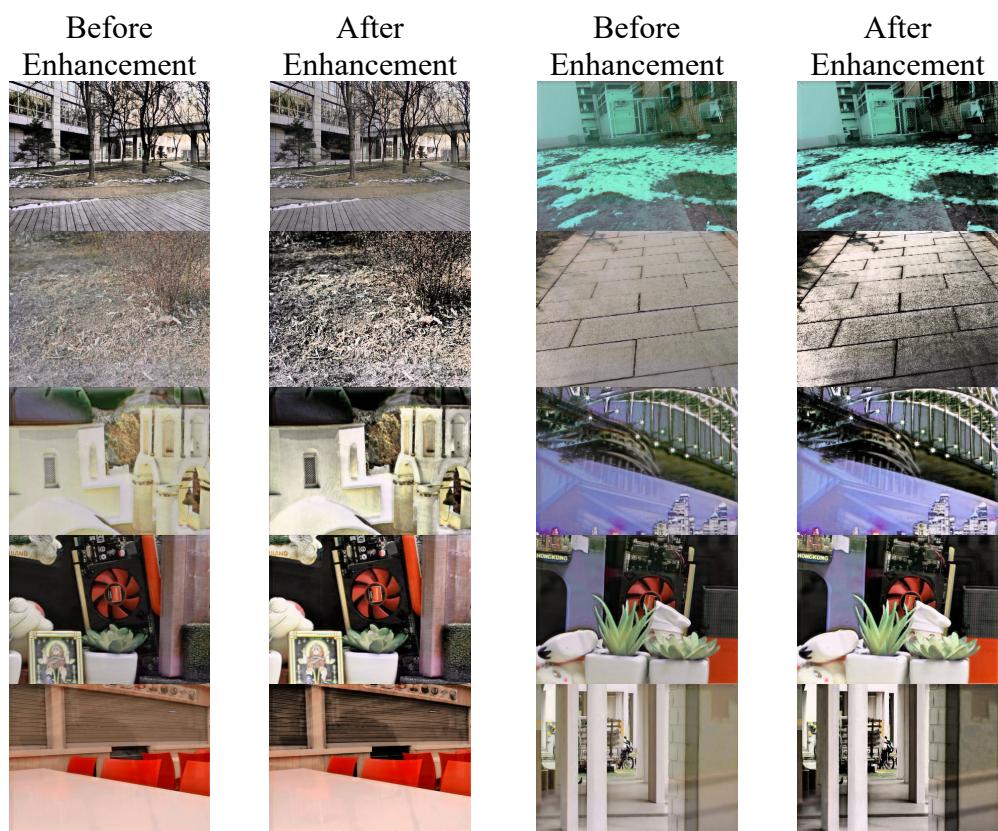


Figure 4.14 Qualitative comparison of enhancement on 5 datasets

5. CONCLUSIONS AND FUTURE WORK

5.1. Conclusion

The IBCLN [21] framework, selected based on a comparative analysis with other methods, has demonstrated superior performance in removing reflections from images, particularly in terms of PSNR and SSIM metrics. However, the complexity of the IBCLN [21] method can lead to slower processing times, making it less suitable for real-time applications. Additionally, artifacts may occur when handling reflections, resulting in lower brightness and color saturation in the output images, which can affect overall visual quality. To address these issues, the GHE [23] technique was adopted, with the parameter α optimized to enhance image brightness and visual quality. Although GHE [23] comes at the cost of some PSNR and SSIM metrics, it significantly improves the brightness and visual quality of the images, particularly in terms of contrast and visual appeal.

5.2. Challenges and opportunities

Despite the strong performance of the IBCLN [21] method in experiments, its practicality in handling real-time data and adapting to changing environments requires further validation. Additionally, the complex network design may not perform well in unconventional reflection scenarios. Future research should focus on improving the model's robustness and enhancing real-time processing capabilities. While the adoption of GHE [23] has improved image brightness and visual quality at the expense of some PSNR and SSIM metrics, finding the optimal parameter α across different scenes remains

a challenge. Further studies could explore how GHE [23] parameters can be dynamically adjusted to accommodate various image types and reflection intensities, while continuing to minimize artifacts and color distortions.

REFERENCES

- [1] Chen, Z., Long, F., Qiu, Z., Zhang, J., Zha, Z. J., Yao, T., & Luo, J. (2024). A Closer Look at the Reflection Formulation in Single Image Reflection Removal. *IEEE Transactions on Image Processing*. doi: [10.1109/TIP.2023.3347915](https://doi.org/10.1109/TIP.2023.3347915)
- [2] Lan, H., Zhang, E., & Jung, C. (2024). Face Reflection Removal Network Using Multispectral Fusion of RGB and NIR Images. *IEEE Open Journal of Signal Processing*. doi: [10.1109/OJSP.2024.3351472](https://doi.org/10.1109/OJSP.2024.3351472)
- [3] Zhao, H., Zhou, R., Zhang, S., & Fu, Y. (2024, April). Single Image Reflection removal Using Feature Difference Enhancement. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3090-3094). IEEE. doi: [10.1109/ICASSP48485.2024.1044675](https://doi.org/10.1109/ICASSP48485.2024.1044675)
- [4] Chiang, C. C., Yang, Y. Y., Liu, W. L., & Lin, Y. C. (2023). Improving Automatic Identification of Medications in Transparent Packaging by Glare Removal and Color Correction. *IEEE Access*, 11, 118812-118829. doi: [10.1109/ACCESS.2023.3327421](https://doi.org/10.1109/ACCESS.2023.3327421)
- [5] Hong, Y., Zheng, Q., Zhao, L., Jiang, X., Kot, A. C., & Shi, B. (2023). PAR 2 Net: End-to-End Panoramic Image Reflection Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10), 12192-12205. doi: [10.1109/TPAMI.2023.3286429](https://doi.org/10.1109/TPAMI.2023.3286429)
- [6] Xu, Y., Li, W., Yang, Y., Ji, H., & Lang, Y. (2023). Superimposed Mask-guided Contrastive Regularization for Multiple Targets Echo Separation on Range-Doppler Maps. *IEEE Transactions on Instrumentation and Measurement*. doi: [10.1109/TIM.2023.3320761](https://doi.org/10.1109/TIM.2023.3320761)
- [7] Yan, T., Li, H., Gao, J., Wu, Z., & Lau, R. W. (2023). Single image reflection removal from glass surfaces via multi-scale reflection detection. *IEEE Transactions on Consumer Electronics*. doi: [10.1109/TCE.2023.3303475](https://doi.org/10.1109/TCE.2023.3303475)
- [8] Zhang, X., Xing, K., Liu, Q., Chen, D., & Yin, Y. (2023). Single image reflection removal based on dark channel sparsity prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(11), 6431-6442. doi: [10.1109/TCSVT.2023.3270173](https://doi.org/10.1109/TCSVT.2023.3270173)
- [9] Rosh, G., Prasad, B. P., Boregowda, L. R., & Mitra, K. (2023, October). Deep unsupervised reflection removal using diffusion models. In *2023 IEEE International Conference on Image Processing (ICIP)* (pp. 2045-2049). IEEE. doi: [10.1109/ICIP49359.2023.10222641](https://doi.org/10.1109/ICIP49359.2023.10222641)
- [10] Daggubati, H., Chiluka, C., Reddy, M. D., & Kumar, D. P. (2023, January). Image reflection removal using iterative boost convolutional lstm network through feature loss. In *2023 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE. doi: [10.1109/ICCCI56745.2023.10128272](https://doi.org/10.1109/ICCCI56745.2023.10128272)
- [11] Chen, W. T., Chen, K. Y., Chen, I. H., Fang, H. Y., Ding, J. J., & Kuo, S. Y. (2022). Missing recovery: Single image reflection removal based on auxiliary prior learning. *IEEE Transactions on Image Processing*, 32, 643-656. doi: [10.1109/TIP.2022.3230544](https://doi.org/10.1109/TIP.2022.3230544)
- [12] Peng, Y. T., Cheng, K. H., Fang, I. S., Peng, W. Y., & Wu, J. S. (2022). Single image reflection removal based on knowledge-distilling content disentanglement. *IEEE Signal Processing Letters*, 29, 568-572. doi: [10.1109/LSP.2022.3148668](https://doi.org/10.1109/LSP.2022.3148668)

- [13] Chen, W. T., Chen, Y. W., Chen, K. Y., Ding, J. J., & Kuo, S. Y. (2022, October). Single image reflection removal based on bi-channels prior. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 2117-2121). IEEE. doi: [10.1109/ICIP46576.2022.9898067](https://doi.org/10.1109/ICIP46576.2022.9898067)
- [14] Aizu, T., & Matsuoka, R. (2022, August). Reflection removal using multiple polarized images with different exposure times. In *2022 30th European Signal Processing Conference (EUSIPCO)* (pp. 498-502). IEEE. doi: [10.23919/EUSIPCO55093.2022.9909712](https://doi.org/10.23919/EUSIPCO55093.2022.9909712)
- [15] Zhou, L., Qin, L., Zhang, Z., & Zheng, B. (2022, May). Single Image Reflection Removal Based on Reflection Region. In *2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 254-260). IEEE. doi: [10.1109/ICAIBD55127.2022.9820344](https://doi.org/10.1109/ICAIBD55127.2022.9820344)
- [16] Dong, Z., Xu, K., Yang, Y., Bao, H., Xu, W., & Lau, R. W. (2021). Location-aware single image reflection removal. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5017-5026). doi: [10.1109/ICCV48922.2021.00497](https://doi.org/10.1109/ICCV48922.2021.00497)
- [17] Lei, C., & Chen, Q. (2021). Robust reflection removal with reflection-free flash-only cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14811-14820). doi: [10.1109/CVPR46437.2021.01457](https://doi.org/10.1109/CVPR46437.2021.01457)
- [18] Zheng, Q., Shi, B., Chen, J., Jiang, X., Duan, L. Y., & Kot, A. C. (2021). Single image reflection removal with absorption effect. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13395-13404). doi: [10.1109/CVPR46437.2021.01319](https://doi.org/10.1109/CVPR46437.2021.01319)
- [19] Funahashi, I., Yamashita, N., Yoshida, T., & Ikebara, M. (2021, December). High Reflection Removal Using CNN with Detection and Estimation. In *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 1381-1385). IEEE. doi: <https://ieeexplore.ieee.org/abstract/document/9689615>
- [20] Lei, C., Huang, X., Zhang, M., Yan, Q., Sun, W., & Chen, Q. (2020). Polarized reflection removal with perfect alignment in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1750-1758). doi: [10.1109/CVPR42600.2020.00182](https://doi.org/10.1109/CVPR42600.2020.00182)
- [21] Li, C., Yang, Y., He, K., Lin, S., & Hopcroft, J. E. (2020). Single image reflection removal through cascaded refinement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3565-3574). doi: [10.1109/CVPR42600.2020.00362](https://doi.org/10.1109/CVPR42600.2020.00362)
- [22] Zou, Z., Lei, S., Shi, T., Shi, Z., & Ye, J. (2020). Deep adversarial decomposition: A unified framework for separating superimposed images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12806-12816). doi: [10.1109/CVPR42600.2020.01282](https://doi.org/10.1109/CVPR42600.2020.01282)
- [23] Tanaka, H., & Taguchi, A. (2020, November). Brightness preserving generalized histogram equalization. In *2020 IEEE REGION 10 CONFERENCE (TENCON)* (pp. 1397-1400). IEEE. doi: [10.1109/TENCON50793.2020.9293837](https://doi.org/10.1109/TENCON50793.2020.9293837)

- [24] Takahashi, T., Uruma, K., & Kobayashi, K. (2020, August). Single image reflection removal for privacy protection using deep CNN. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1028-1031). IEEE. doi: [10.1109/MWSCAS48704.2020.9184486](https://doi.org/10.1109/MWSCAS48704.2020.9184486)
- [25] Li, T., & Lun, D. P. (2019). Single-image reflection removal via a two-stage background recovery process. *IEEE Signal Processing Letters*, 26(8), 1237-1241. doi: [10.1109/LSP.2019.2926828](https://doi.org/10.1109/LSP.2019.2926828)
- [26] Wan, R., Shi, B., Li, H., Duan, L. Y., Tan, A. H., & Kot, A. C. (2019). CoRRN: Cooperative reflection removal network. *IEEE transactions on pattern analysis and machine intelligence*, 42(12), 2969-2982. doi: [10.1109/TPAMI.2019.2921574](https://doi.org/10.1109/TPAMI.2019.2921574)
- [27] Wei, K., Yang, J., Fu, Y., Wipf, D., & Huang, H. (2019). Single image reflection removal exploiting misaligned training data and network enhancements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8178-8187). doi: [10.1109/CVPR.2019.00837](https://doi.org/10.1109/CVPR.2019.00837)
- [28] Wen, Q., Tan, Y., Qin, J., Liu, W., Han, G., & He, S. (2019). Single image reflection removal beyond linearity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3771-3779). doi: [10.1109/CVPR.2019.00389](https://doi.org/10.1109/CVPR.2019.00389)
- [29] Chou, N. H., Chuang, L. C., & Lee, M. S. (2019, November). Intensity-aware GAN for single image reflection removal. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 590-594). IEEE. doi: [10.1109/APSIPAASC47483.2019.9023087](https://doi.org/10.1109/APSIPAASC47483.2019.9023087)
- [30] Yun, J. S., & Sim, J. Y. (2019, September). Cluster-wise removal of reflection artifacts in large-scale 3D point clouds using superpixel-based glass region estimation. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 1780-1784). IEEE. doi: [10.1109/ICIP.2019.8803126](https://doi.org/10.1109/ICIP.2019.8803126)
- [31] Li, T., & Lun, D. P. (2019, May). Image reflection removal using the wasserstein generative adversarial network. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. doi: [10.1109/ICASSP.2019.8683786](https://doi.org/10.1109/ICASSP.2019.8683786)
- [32] Kumar, S. M., Nishanth, R., Sani, N., Joseph, A. J., & Martin, A. (2019, March). Specular reflection removal using morphological filtering for accurate iris recognition. In *2019 International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE. doi: [10.1109/ICSSS.2019.8882863](https://doi.org/10.1109/ICSSS.2019.8882863)
- [33] Heo, M., & Choe, Y. (2019, January). Single-image reflection removal using conditional GANs. In *2019 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-4). IEEE. doi: [10.23919/ELINFOCOM.2019.8706433](https://doi.org/10.23919/ELINFOCOM.2019.8706433)
- [34] Chang, Y., & Jung, C. (2018). Single image reflection removal using convolutional neural networks. *IEEE Transactions on Image Processing*, 28(4), 1954-1966. doi: [10.1109/TIP.2018.2880088](https://doi.org/10.1109/TIP.2018.2880088)
- [35] Wan, R., Shi, B., Duan, L. Y., Tan, A. H., & Kot, A. C. (2018). CRRN: Multi-scale guided concurrent reflection removal network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4777-4785). doi: [10.1109/CVPR.2018.00502](https://doi.org/10.1109/CVPR.2018.00502)

- [36] Yang, J., Gong, D., Liu, L., & Shi, Q. (2018). Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 654-669). doi: https://doi.org/10.1007/978-3-030-01219-9_40
- [37] Zhang, X., Ng, R., & Chen, Q. (2018). Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4786-4794). doi: [10.1109/CVPR.2018.00503](https://doi.org/10.1109/CVPR.2018.00503)
- [38] Tsai, G. C., Chen, W. T., Yuan, S. Y., & Kuo, S. Y. (2018, November). Efficient reflection removal algorithm for single image by pixel compensation and detail reconstruction. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)* (pp. 1-5). IEEE. doi: [10.1109/ICDSP.2018.8631670](https://doi.org/10.1109/ICDSP.2018.8631670)
- [39] Fan, Q., Yang, J., Hua, G., Chen, B., & Wipf, D. (2017). A generic deep architecture for single image reflection removal and image smoothing. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3238-3247). doi: [10.1109/ICCV.2017.351](https://doi.org/10.1109/ICCV.2017.351)
- [40] Wan, R., Shi, B., Tan, A. H., & Kot, A. C. (2017, July). Sparsity based reflection removal using external patch search. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1500-1505). IEEE. doi: [10.1109/ICME.2017.8019527](https://doi.org/10.1109/ICME.2017.8019527)
- [41] Wan, R., Shi, B., Hwee, T. A., & Kot, A. C. (2016, September). Depth of field guided reflection removal. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 21-25). IEEE. doi: [10.1109/ICIP.2016.7532311](https://doi.org/10.1109/ICIP.2016.7532311)
- [42] Shih, Y., Krishnan, D., Durand, F., & Freeman, W. T. (2015). Reflection removal using ghosting cues. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3193-3201). doi: [10.1109/CVPR.2015.7298939](https://doi.org/10.1109/CVPR.2015.7298939)
- [43] Li, Y., & Brown, M. S. (2013). Exploiting reflection change for automatic reflection removal. In *Proceedings of the IEEE international conference on computer vision* (pp. 2432-2439). doi: [10.1109/ICCV.2013.302](https://doi.org/10.1109/ICCV.2013.302)
- [44] Wan R., Shi B., Li, H., Hong Y., Duan, L. Y., & Kot, A. C., *Benchmarking Single-Image Reflection Removal Algorithms* <https://sir2data.github.io/>
- [45] Wikipedia contributors. (2024, July 22). HSL and HSV. In *Wikipedia, The Free Encyclopedia*. Retrieved 11:33, August 2, 2024, from https://en.wikipedia.org/w/index.php?title=HSL_and_HSV&oldid=1236077814