APPLICATION OF GAUSSIAN MODEL AND DEEP LEARNING ENCODER-DECODER ALGORITHM FOR SINGLE-IMAGE REFLECTION REMOVAL

Ashraf Ishaq[1], Sumayyah Sophie Nandom[2],

Tsentob Joy Samson[3], and Maryam Suleiman[4]

[1]Department of Computer Science, Federal University Wukari.
Email: ishaqashraf@fuwukari.edu.ng

[2]Department of Computer Science, Federal University Wukari.
Email: ssnandom@gmail.com

[3]Department of Computer Science, Air Force Institute of Technology, Kaduna.
Email: tsentobsamson@gmail.com

[4]Department of Computer Science, Federal University Wukari.
Email: suleimammaryam554@gmail.com

Cite this article:

xxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxx

Manuscript History

Received: 13 May 2024

Accepted: 9 Jul 2024

Published: 22 Jul 2024

ABSTRACT: *Images of target scenes shot through clear, reflective materials like glass are frequently interfered by unwanted reflection scenes which often overlaid on top of the targeted scenes. This, has constantly degrades the quality of the captured images and affects their subsequent analyses. While cognitively, distinguishing a recognizable object from its reflection in a picture is not difficult for humans, it is highly difficult and more complex in computer vision due to the ill-posed nature of the problem. In this research an enhanced single-image reflection removal model was developed by combining Gaussian filter and deep learning encoder-decoder for effective performance. While the Gaussian filter denoises the reflection-contaminated image, the encoder-decoder network learns the features of the image to produce reflection-free image. The proposed network is an end-to-end trained network with three losses. The experimental findings showed that the proposed model out-performed several state-of-the-art methods both qualitatively and quantitatively on five different datasets.*

KEYWORDS: Gaussian filter, encoder-decoder, deep learning, single-image, reflection removal.

INTRODUCTION

When taking photographs through a transparent material such as glass, the photographs often contain unwanted reflection scenes. These reflection scenes not only reduce the image quality but can also negatively impact subsequent analysis of the image such as object detection, image segmentation, or classification (Wan et al., 2018). Hence, this makes image reflection removal an important task in the field of computer vision. Mathematically, single-image reflection could be modelled according Chi et al. (2018) as in equation 1:

$$I = \alpha I_B + \beta I_R + \mathrm{n} \qquad (1)$$

where I, I_B, and I_R are the mixture image, background image, reflection image respectively; α and β are the transmittance and reflective rate of the glass; and n is the noise term. The aim of reflection removal is to improve the visibility of the image behind the glass for instance while removing reflections. To address this challenge, photographers have traditionally used polarizers installed in front of the camera lens to reduce reflections. However, a polarizer can only remove the reflection components with an incident angle equal to the Brewster angle (Born & Wolf, 2013).

Although it is easier for humans to detect and distinguish the background scene from the reflection scene; however, it is highly difficult for computers to do so because of the severe ill-posed of the captured image. The task of reflection removal for computers becomes more complicated due to similar morphological features of the background and reflection scenes. In recent years, many approaches were proposed as a solution to this problem. It is worth noting that these approaches are categorized into two groups: conventional mathematical methods (non-learning methods) and data-driven methods (deep learning methods) (Amanlou et al., 2022). The non-learning methods used handcrafted priors and complex mathematical analysis as a solution for this problem (Fan et al., 2017). Many of the non-learning approaches only work under special conditions and also require stringent physical settings (Fan et al., 2017). Apart from the inefficiency and high-time consumption of these approaches, the assumption made in these approaches that the formation of images that contain reflections are linear does not capture real-life scenarios (Wen et al., 2019). These and many other shortfalls of the non-learning approaches inspired the need to adopt deep learning methods which are data-driven, highly efficient, and less time-consuming (Amanlou et al., 2022).

The first deep learning method for image reflection removal was proposed by Fan et al. (2017), and since then, deep learning methods have been gaining significant attention in finding solutions for image reflection removal (Wan et al., 2022). The deep learning methods which are data-driven rely heavily on synthesized data in training the models due to less availability of real-world datasets. These approaches simply mapped images to high-dimensional features using deep learning techniques. To this end, deep learning methods have shown great capability in removing reflection in images, however, there are still glitches needed to be handled in order to improve the performances of the existing models.

Reflection removal could also fall into two categories: single-image reflection removal and multiple-image reflection removal. Multiple image reflection removal is the process of removing reflections in multiple captured images at a slightly different angle. The multiple image reflection removal could be categorized into multiple polarized images, focus and defocus image pairs, flash and non-flash images, and video sequences (Amanlou et al., 2022).

Whereas single-image reflection removal is the process of removing reflection in an individual captured image. Amanlou et al. (2022), posited that while the multiple image reflection removal algorithms are still effective, they can be more complex and computationally demanding as it involves considering temporal or spatial information across multiple frames. Moreover, acquiring multiple image reflection pair is more challenging compared to single image. Therefore, single-image reflection removal methods have garnered more attention from researchers.

In view of the possible practicality of single-image reflection removal in real-life scenarios, and the advantages of the deep learning methods over the non-learning methods, researchers had proposed several solutions in tackling single-image reflection removal leveraging the deep learning approach. For instance, Fan et al. (2017) proposed a cascaded edge and image learning network that learn an intermediate edge map that guides separation of reflection and non-reflection image. Chi et al. (2018), also proposed a deep encoder-decoder network that recovers reflection-free image by simply learning an end-to-end mapping of image pairs with and without reflection. However, existing networks still fall short in removing reflection edges due to the ill-posedness of the problem and the complexities of real-world data with reflection. This research, hence, proposes to adopt the application of gaussian model and deep learning encoder-decoder to optimally remove reflection and enhances the quality of the reflection-free image. While the role of the gaussian model is to initially denoise the image, the encoder-decoder network specifically predicts and generate reflection-free image.

REVIEW OF RELATED LITERATURE

There is a wide range of literature on reflection removal, encompassing non-learning techniques as well as learning techniques, including both multiple-image reflection removal and single-image reflection removal. However, deep learning methods have gained significant popularity in recent times due to their effectiveness, data-driven nature, and efficient removal of reflections. It is important to highlight that within deep learning techniques, single-image reflection removal methods have garnered more attention due to their practicality in real-life situations (Amanlou et al., 2022). Therefore, this review of related literatures focuses more on the literature related to single-image removal using deep learning approaches.

For instance, Fan et al. (2017) introduced the first deep learning network specifically designed for single image reflection removal called Cascaded Edge and Image Learning Network (CEILNet). In their approach, the authors formulated reflection removal as an edge simplification task and trained the network to learn an intermediate edge map that guides the separation of reflection and non-reflection layers. To train and evaluate CEILNet, the authors generated a dataset of 8,500 synthetic images from Flickr and PASCAL VOC datasets. They employed PSNR and SSIM as quantitative metrics and compared the results to those of Li and Brown (2013) on 100 synthetic images with reflections. The achieved values for PSNR and SSIM were 18.55 and 0.857, respectively. However, it should be noted that CEILNet was trained solely with a low-level loss function that combined differences in color space and gradients. This approach does not directly enable the model to learn high-level semantics, which are crucial for effective reflection removal.

Zhang et al. (2018) made improvements to the approach proposed by Fan et al. (2017) by incorporating perceptual information into their deep learning-based image reflection removal method. Their proposed method, called Perceptual Loss Network (PLNet), utilized a fully convolutional network that incorporated both low-level and high-level image features. To train the model, the authors collected a dataset consisting of 5,000 random synthetic images from Flickr, with one indoor and one outdoor image for each pair. They also gathered 110 real-world images with ground truths in natural environments for evaluation. The model performed well on the synthetic data, achieving SSIM and PSNR values of 0.853 and 22.63, respectively. For the real-world data, the model achieved SSIM and PSNR values of 0.821 and 21.30, respectively. When compared to the work of Fan et al. (2017) in terms of both quantitative and qualitative metrics, Zhang et al. (2018)'s model demonstrated superior performance in effectively removing reflections. However, it is important to note that Zhang et al. (2018) made the assumption that the reflection layer in real-world images is often blurry and less focused compared to the transmission layer. This assumption limited the efficiency of their model in cases where the reflection layer is as sharp as the transmission layer.

Wan et al. (2018) proposed a concurrent deep learning-based framework for effectively removing reflections from single images. Their framework unified gradient inference and image inference into a concurrent framework, integrating high-level image appearance information and multi-scale low-level features. To ensure general compatibility on real-world data, the authors constructed a large-scale Reflection Image Dataset (RID) consisting of 3,250 images. For quantitative evaluation, they used the SIR benchmark dataset. Wan et al. (2018) employed SSIM (Structural Similarity Index) and SI (Structure Index) as error metrics. The achieved values for SSIM and SI were 0.895 and 0.925, respectively, which outperformed the then state-of-the-art approaches. However, it is important to note that the performance of the concurrent deep learning framework may drop when the entire images are dominated by reflections. Additionally, the proposed approach was trained directly on images, which often suffer from the gradient vanishing problem. This issue can cause the convolutional neural network (CNN) to introduce color shifts to the estimated images.

To address the single image reflection removal problem, Yang et al. (2018) proposed a cascaded deep neural network that leveraged both the background image and the reflection image to estimate each other. Unlike training the network to estimate only the background image from the mixture image, this bidirectional network estimated both the background and reflection scenes. The network consisted of three subnetworks: the vanilla generator, the reflection estimator, and the refine background estimator. To train the model, the authors generated 50,000 training images from the training set of PASCAL VOC, which originally contained 5,717 images. The model was then evaluated using the SIR benchmark dataset to assess its performance quantitatively. For postcard images, the model achieved a PSNR of 20.4076 and an SSIM of 0.8548. For solid objects, the values were 22.7076 (PSNR) and 0.8627 (SSIM), and for wild scenes, the values were 22.1082 (PSNR) and 0.8327 (SSIM). Comparing their approach to the work of Fan et al. (2017), Yang et al. (2018) demonstrated significant improvements, particularly in wild scenes. However, it is important to note that this approach heavily relied on synthesized training data to train the model. This reliance on synthetic data may limit its effectiveness in real-world reflection removal scenarios, where the characteristics of reflections can vary significantly.

Chi et al. (2018), proposed a deep learning encoder-decoder network for the removal of reflection in single imaging. The network consists of 12 convolutional and deconvolutional

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

layers with one rectilinear unit between each layer. The convolutional layers were designed to extract and condense feature from the input image, while deconvolutional layers rebuild the details of reflection-free image from feature abstraction. Since reflection is often blurry, so to formulate synthetic data, Chi et al. (2018) used gaussian blur kernel of randomly selected variance of 1 to 5 with a transmittance rate between 0.75 to 0.80. Moreover, in order to train and test the model, the authors gathered 66327 synthesize data from Jokinen and Sampo (2016) and Quattoni and Torralba (2009); where 70% of the data goes to training and 30% for testing and validation. The authors added a benchmark dataset by Wan et al. (2018) for evaluation. Only PSNR was adopted as evaluation metric. The proposed model's PSNR value on synthetic data and the benchmark data is 29.08 and 18.70 respectively. When compared to the work of Arvanitopoulos et al. (2017) and Fan et al. (2017), the proposed model gave better performance. However, the network solely relied on synthetic data for training. This necessitated an improvement to the effectiveness of the model.

Since the introduction of deep learning methods in tackling the image reflection removal problems, approaches relied on synthetic dataset for training the models; which significantly affects the effectiveness of the models in real-life reflection removal. To better simplify the intrinsic ill-posedness and diminish ambiguity caused by reflection in imaging, Wei et al. (2019) proposed an enhanced network architecture that is sensitive to contextual features with an alignment-invariant loss function to help in maintaining the real-world data during training. Despite that this approach adopted the infusion of synthetic and real-world data as training dataset; the synthetic greatly outnumbered the real-world data. 7,643 images from PASCAL VOC were used as synthetic dataset and 90 real-world training images generated by Zhang et al. (2018) was used as real-world dataset. Wei et al. (2019) adopted PSNR, SSIM, NCC, and LMSE as evaluation metrics. The average PSNR, SSIM, NCC, and LMSE values are 23.59, 0.879, 0.956, 0.005, respectively. The authors compared their work to that of Fan et al. (2017), Zhang et al. (2018), and Yang et al. (2018) using 20 real-world images and SIR^2 benchmark dataset. The fact that Wei et al. (2019)'s method outperformed the other compared approaches; the effectiveness of the algorithm in the removal of reflection on wild images is still below the baseline 'Input'. The algorithms cannot robustly remove reflection on real-world images.

Abiko and Ikehara (2019) proposed a deep learning method for the single image reflection removal based on the Generative Adversarial Network (GAN) that leverage the gradient constraint to effectively separate the reflection layer from the background layer. The proposed Gradient Constraint Network (GCNet) made used of gradient loss in addition to the existing losses (pixel loss, feature loss, and adversarial loss) to robustly distinguish the reflection layer from the background layer. In order to train the model, the authors made use of the PASCAL VOC 2012 dataset and SIR^2 benchmark dataset was adopted for evaluation. PSNR and SSIM evaluation metrics were employed and the results were compared to the work of Fan et al. (2017), Yang et al. (2018), Zhang et al. (2018), and Wan et al. (2018). The PSNR and SSIM values for the proposed model are: 19.64 and 0.918 for Postcard; 23.87 and 0.928 for Solid; and 24.97 and 0.932 for Wild. The fact that Abiko and Ikehara (2019)'s model outperformed the rest, the model still relied completely on synthesis data for the training which its efficiency on real-world data is low.

Li et al. (2020), argued that recent deep learning approaches for single image reflection removal did not utilize the refinement method. Inspired by the iteration structure reduction approaches, the authors proposed an Iterative Boost Convolutional LSTM Network (IBCLN)

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

for transmission and reflection decomposition in a cascaded form – where the transmission and reflection were progressively refined during iteration. For training the model, the authors made used of 4000 images that comprised of 2800 synthetic data, 290 real-world images that created 1200 patche images, 90 real-world images from Zhang et al. (2018), and 200 generated captured images. For evaluation, the authors used the SIR^2 benchmark dataset, Zhang et al. (2018)'s real world dataset, and 200 real world captured dataset. Li et al. (2020) adopted PSNR and SSIM as evaluation metrics. The average PSNR and SSIM achieved by the proposed model on the five datasets were 24.87 and 0.893, 23.39 and 0.875, 24.71 and 0.886, 21.86 and 0.762, 23.57 and 0.783, respectively. And, finally, the authors compared the results to the work of Fan et al. (2017), Zhang et al. (2018), Wan et al. (2018), and Yang et al. (2018) and achieved better performances. However, the cascaded prediction refinement approach falls short to effectively decompose transmission and reflection layers on real-world images that has challenging characteristics like raindrop, flares, haze, etc. This has called for attention to better improve the effectiveness in real-life images.

Dong et al. (2021), proposed a location-aware single image reflection removal network to improve the effectiveness of reflection removal results. The network leveraged on Laplacian features that emphasized the strong reflections' areas and remove them; such as reflected highlights. The proposed network has a reflection detection module (RDM) that takes multi-scale Laplacian features as inputs to detect reflection roughly. The authors infused both synthetic and real-world data as training dataset. For synthetic data, 13700 image pairs was generated from the PASCAL VOC dataset; whereas for the real-world dataset, 200 pairs provided by Yang et al. (2018), 90 pairs provided by Zhang et al. (2018), and 200 provided by Wan et al. (2018). In order to evaluate the model, SIR^2 benchmark dataset was used and a combination of Zhang et al. (2018), and Yang et al. (2018) real-world dataset was used. The averaged PSNR and SSIM of the model were 24.179 and 0.893. Finally, the model was compared both qualitatively and quantitively to the work of Fan et al. (2017), Zhang et al. (2018), Yang et al. (2018), Wan et al. (2018), and Li et al. (2020). Although, the proposed network achieved better performances, its architecture is however too complex as much parameters of about 10.926 million were used, which greatly exceed that of selected state-of-the-arts models.

Chen et al. (2022) argued that existing single-image reflection removal approaches do not take into consideration the loss of information caused by intensity overflow when capturing an image affected by reflections. As a result, the contaminated reflection image may be clipped due to intensity overflow. Thus, the authors proposed a missing recovery strategy to compensate for the subsequent reflection removal process. To model the network, the authors combined deep CNNs and handcrafted priors to address the disadvantages of both strategies. Furthermore, the revisitation of the handcrafted priors aims to model a network based on auxiliary prior learning. The authors employed PSNR, SSIM, LMSE, and NCC as quantitative evaluation metrics to compare the effectiveness of the model to six state-of-the-art approaches (Wen et al. (2019), Zhang et al. (2018), Wei et al. (2018), Yang et al. (2018), Li et al. (2020)). The average values obtained were 31.91, 0.92, 0.004, and 0.987, respectively. While the approach has shown promising results, the model still contains a large number of parameters of about 7.7×10^7, which poses a challenge when deploying the model on edge devices.

Thungborg and Astrom (2022) argued that none of the existing reflection removal networks were trained on images with reflection domes. Consequently, the authors fine-tuned four

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

existing networks (DADNet, ERRNet, IBCLNet, and RAGNet) and trained them on the same dataset. To create a dataset with dome reflections, the authors generated their own datasets. The final dataset consisted of 4000 blended training images and 445 blended test images. The fine-tuned networks outperformed the pre-trained networks on the same dataset in both quantitative and qualitative assessments. Among the fine-tuned networks, ERRNet demonstrated superior performance on both synthetic and real-world datasets, achieving an average PSNR value of 32.926 and an SSIM value of 0.978. Although this approach exhibited promising results, however, there is still room for improvement when implementing the networks on real dome reflections as the networks were solely trained on synthetic data.

He et al. (2023) argued that existing deep learning approaches rely on decomposing the target scene into transmission layers and reflection layer, which neglect the physical formation of an image with reflection in real-life scenario; thus, resulting to unsatisfactory results especially on images that contained strong reflection. The authors thereby proposed a two-stage reflection intensity-guided network that in form of a divide-and-conquer for reflection removal and transmission recovery. The two parts of the network are jointly interconnected and mutually informative. The reflection intensity network provides essential information about the distribution and intensity of reflection in the input image, which act as guidance for the guided transmission recovery part of the network. The authors implored four different evaluation metrics: PSNR, SSIM, LMSE, and NCC and achieved the values 24.824, 0.906, 0.004, and 0.974, respectively on SIR2 benchmark dataset. While the network outperformed state-of-the-art models; the network was overwhelmed with large parameter count at about 29.07 million which is significantly higher than the compared state-of-the-arts models. This is so due to the lack of use of knowledge-distilling strategy in both transmission generation and transmission recovery modules.

Research Gap

Recently, data-driven methods using deep neural networks have become a trend due to their effectiveness and optimization, particularly in single-image reflection removal approaches, which are more practical than multiple-image reflection approaches (Amanlou et al. 2022). However, most existing single-image reflection removal approaches rely heavily on synthetic dataset for training models and also, simulating synthetic data with noise that depict real-life scenarios (like haze, raindrop, flare, etc.) is highly challenging; this has constantly degraded the quality of the training data. In an attempt by Chi et al. (2018) to model realistic synthetic data using variance gaussian blur kernel selected randomly from 1 to 5; it only incurred blurriness to the synthetic data which is often not the only challenges of real-world images. Also, the autoencoder architecture presented by Chi et al. (2018) was overparametrized with a large number of convolutional layers (6 convolutional layers for feature extraction; 6 convolutional and 6 deconvolutional layers for reflection recovery and removal; and another 6 deconvolutional layers for transmission restoration) which might lead to overfitting, especially with a limited amount of data. Despite this overparameterization the network falls short to captured high-level features of images as its number of convolutional filters is fixed at 32 throughout. In this light, this research aim to present a robust encoder-decoder model with gaussian filter that rely heavily on real-world data for training and could also learn low- and high-level features of images – having fewer convolutional layers and less computational cost.

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

METHODOLOGY

Data Collection

Deep learning networks rely on large training datasets to effectively learn and fit parameters. The data used in this research is in the form of an images. These images consist of two set: the groundruth image (reflection-free image) and the mixture image (reflection-contaminated image). While using real-world data to train model is presume to give optimal results, it is worth noting that obtaining a large dataset that contained real-world image with groundtruth and mixture image is labor-intensive and challenging. Therefore, most researches made use of synthetic data to train their model.

However, this research made use of five different real-world datasets to train, test, and evaluate the proposed model:

(a) **Zhang et al. (2018) Dataset:** This data, otherwise known as Real, is comprised of 110 real-world image pairs specifically for image reflection removal tasks from the UC Berkeley's Lab. These images represented diverse real-world scenarios and served as another source of real-world training data.

(b) **Wei et al. (2019) Dataset:** This is called the DSLR; and it is additional 250 real-world misaligned dataset captured from natural scenes.

(c) **Li et al. (2021) Dataset:** This dataset, otherwise named as NATURE, contain 200 real-world image pairs captured from natural scenes. It provided additional real-world samples for training and testing the proposed model.

(d) **Wan et al. (2022) Dataset:** This is the popularly known SIR^{2+} benchmark dataset that comprised of 1,700 images with diverse mixture of groundtruth, background, and reflection scenes.

(e) **Ash80 Dataset:** This is a real-world local dataset obtained using an Android smartphone, specifically the Oppo A93 model. It comprised of 1,500 diverse scenes of groundtruths and mixture images.

In total, this research gathered 3,760 real-world data for training, testing and evaluation. Where 70% of the total data goes for training and 30% of the data goes for testing and evaluation.

This research obtained the Ash80 dataset using a smartphone placed on a tripod to avoid shaking. To obtain a reflection-contaminated scene, a clear plane transmitted glass is placed between the camera and the target scene. And to clearly achieved this, the glass is at a distance of 10-30 cm from the camera and 10-15 cm from the target scene. Whereas the groundtruth is captured when the glass is removed. Each image pair was taken with the same exposure setting. Figure 1 give the real-world data collection setup whereas Table 1 present some samples of the captured image.

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

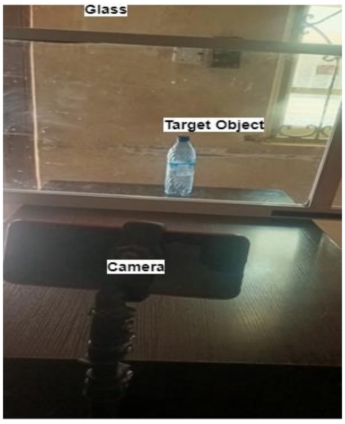


Figure 1: Data Collection Setup

Also, the local dataset was captured taken into consideration the following:

i. Environments: indoor and outdoor;

ii. Lighting conditions: skylight, sunlight, and incandescent;

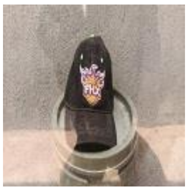
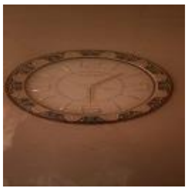
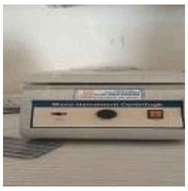
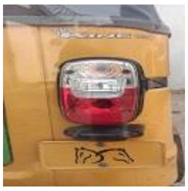
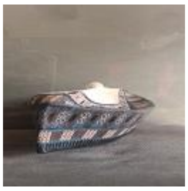
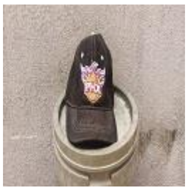
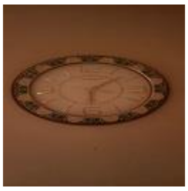
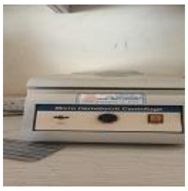
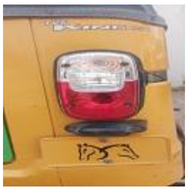
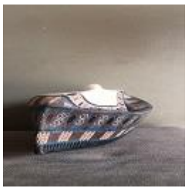
iii. Camera viewing angles: back-facing camera;

This research was implemented using Tensorflow and skimage as dependencies. And, to minimize the training loss, this research adopted Adam optimizer to train the network for 100 epochs with a learning rate of 0.0002 and batch size of 32. The network weights are initialized using a normal distribution (mean: 0, variance: 0.02), and the iteration number *N* is set to 2. The network neither overfit nor underfit; achieving a training accuracy of 95% and validation accuracy of 93%.

Data Preprocessing

The image dataset is normalized to 224 by 224 pixels for effective processing. These datasets contain a mixture of images and their corresponding ground truth. For easy recognition and training, each mixture image is named with the same name as its corresponding ground truth. Table 1 presents samples of the processed Ash80 dataset.

Table 1: Processed Data

Mixture Image					
Groundtruth					

Modeling of the System

In this reflection removal approach, two effective models are combined to optimally enhance reflection removal in images. The first model is the Gaussian filter which is responsible for denoising the captured input image through smoothing while preserving significant image features. The gaussian filter adopted in this network has a gaussian kernel of (5 × 5) and a standard deviation randomly chosen from 1 to 3. This is to ensure the filter significantly reduce reflection while preserving the essential information of the image. The choice of not using higher standard deviation from 4 and above is to avoid blur – further removing essential details in the input image. And, the second part of the model is the encoder-decoder network. The encoder part of the model is responsible for extracting meaningful and high-level features from the input image. It transforms the input image into a latent space representation where essential information about the image content is captured. Through multiple convolutional layers, the encoder builds a hierarchy of features, capturing both low-level details and high-level semantic information. The encoder consists of four convolutional blocks. In each of the blocks two convolutional layers with (32, 64, 128, 256) filters, each having a kernel size of (3, 3) and downsampling factor of (2,2) using the ReLu as activation function. The padding is set to ‘same’ meaning that the input image is padded with zeros so that the output has the same spatial dimensions as the inputs. The ‘skip_conn’ variable was utilized to store the output of the second convolutional network, which will be used later for upsampling phase. Whereas the decoder part of the model takes the compressed representation from the encoder and reconstructs an output image. It performs the opposite operation of the encoder, aiming to reconstruct the original image without the undesired reflections. The decoder consists of four deconvolutional blocks. Each blocks have two deconvolutional layers with (256, 128, 64, 32) filters respectively, and a kernel size of (3,3) and upsampling factor of (2,2). In-between the encoder and the decoder, a latent space (compressed represenation) where extracted features are compressed. This is to retain only the most relevant information while discarding unnecessary details. Figure 3.3 give the represenation of the model.

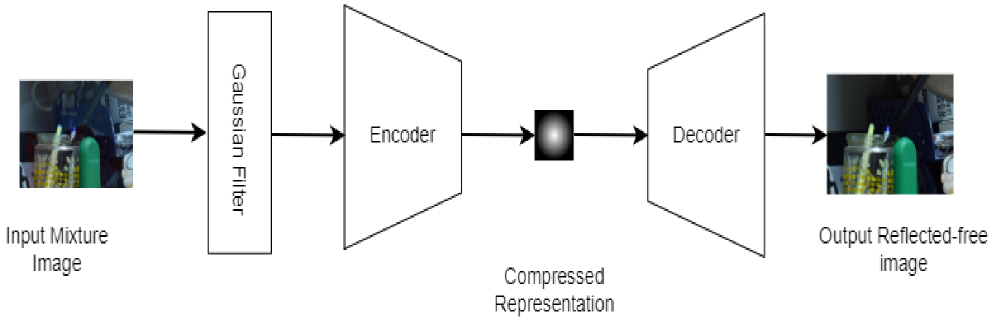


Figure 2: Modeling of the System

The Objective Loss Functions

This section describes the three loss functions employed in the cause of training the proposed network. For clarity, the groundtruth and reflection layers are denoted as T and R; whereas the predicted transmission and reflection layers at iteration *i* is denoted as $\widehat{T}_i$ and $\widehat{R}_i$ respectively.

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

Perceptual Loss

The perceptual loss involves using a pretrained CNN network to calculate the feature differences between the generated reflection-free image and the groundtruth. The perceptual loss helps improve visual quality and content preservation in the generated images. The loss function is described in equation 2 (Zhang et al., 2018).

$$\mathcal{L}_{perc} = \sum_{T,T^3,T^5 \epsilon D} \left(L_{CNN}(\mathrm{T}, \hat{\mathrm{T}}) + \gamma_3 L_{CNN}(T^3, \hat{T}^3) + \gamma_5 L_{CNN}(T^5, \hat{T}^5) \right) \qquad (2)$$

Where $T^3, \hat{T}^3$, and $\hat{T}^5$ corresponds to the output of Conv1_2, Conv3_2, and Conv5_2 for time step N in the CNN. And, lastly $\gamma_3 = 0.8$ and $\gamma_5 = 0.6$

Pixel Loss

The mean square error loss is often used when comparing pixel-wise difference between the generated reflection-free image and the groundtruth during training. In this research, in order to ensure the outputs, become as close to the groundtruth as possible, the L_{MSE} loss to measure the pixel-wise is employed. This pixel loss is defined in equation 3 (Li et al., 2020)

$$\mathcal{L}_{pixel} = \sum_{T \epsilon D} \sum_{t=1}^{N} \left[L_{MSE}(T, \hat{T}_t) + L_{MSE}(\tilde{R}, \hat{R}_t) \right] \qquad (3)$$

Where $\tilde{R}$ is the residual reflection, $\hat{T}_t$ and $\hat{R}_t$ are the outputs at the time step t.

Total Variation Loss

The total variation loss for short TV loss is used to promote smoothness in generated reflection-free images by penalizing rapid changes in the pixel values. In this research, the TV loss is defined as in equation 4 (Li et al., 2020).

$$\mathcal{L}_{TV} = \left\| \nabla \tilde{B}_t \right\|_1 + \left\| \nabla \tilde{R}_t \right\|_1 \qquad (4)$$

Where $\tilde{B}_t$ and $\tilde{R}_t$ are the predicted transmission and reflection layers, respectively.

Overall Loss

This research therefore, adopted the overall objective loss function for its proposed model as defined in equation 5.

$$\mathcal{L} = \lambda_1 \mathcal{L}_{perc} + \lambda_2 \mathcal{L}_{pixel} + \lambda_3 \mathcal{L}_{TV} \qquad (5)$$

Where this research empirically set the weights of the different losses as $\lambda_1 = 10$, $\lambda_2 = 100$, and $\lambda_3 = 1$.

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

RESULTS AND DISCUSSION

Figure 3 shows the loss/accuracy graph over 100 epochs. The graph demonstrates how intelligent the model is with minimal overfitting when faced with new sets of data that it has not been seen previously. The validation loss was a little bit lower than the training loss because of the regularization and augmentation applied during the training but not during validation in order to obtain higher validation accuracy and generalize better to data outside the validation data sets.

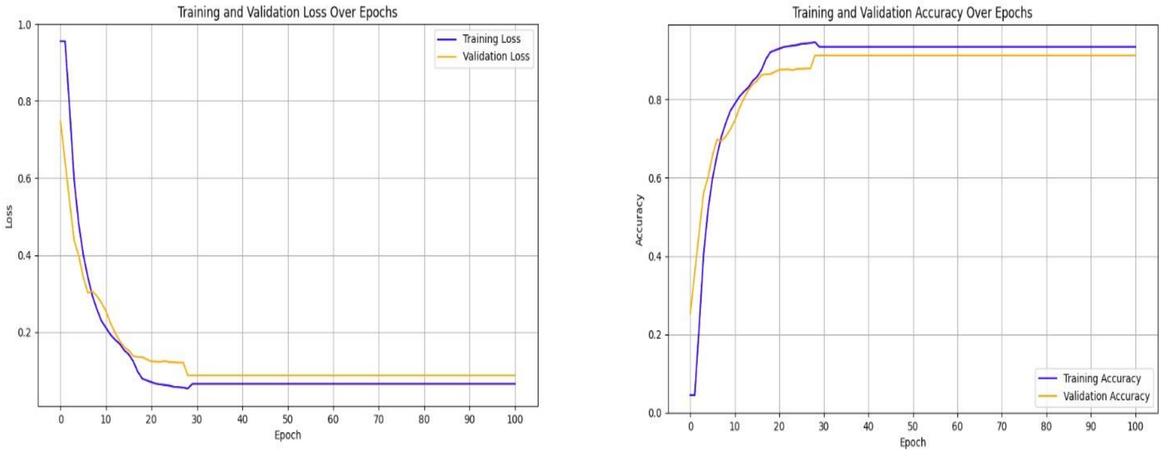


Figure 3: Training and validation loss and accuracy graph

Figure 4 showed the experimental results on some selected individual images. The input image which is the reflection-contaminated image is at the beginning whereas the generated reflection-free image by the model is at the center with its corresponding value of PSNR and SSIM; and the last image is the groudtruth.

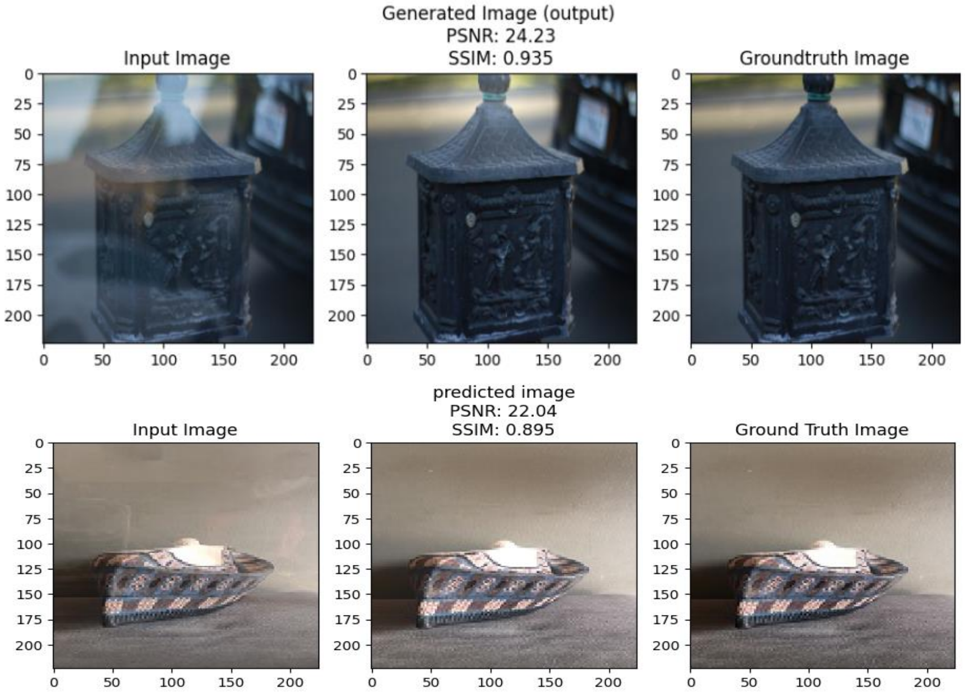
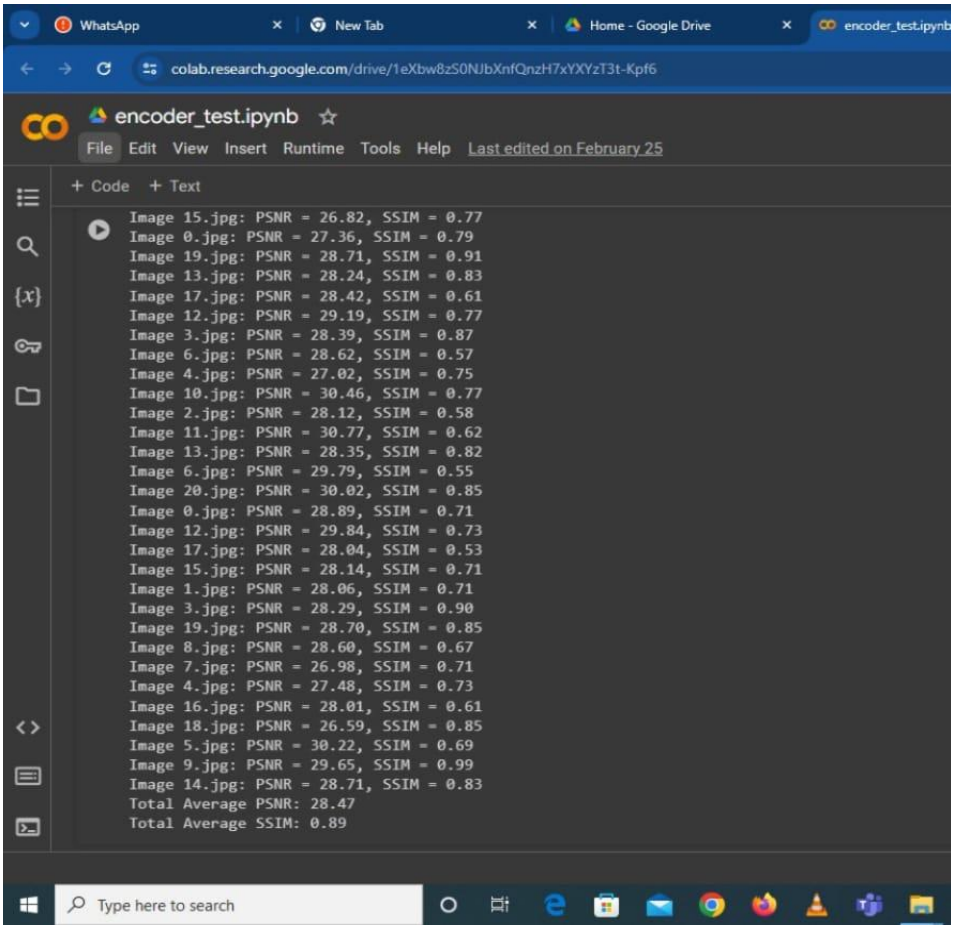


Figure 4: Selected Individual Results

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

In order to evaluate the performance of the proposed model, this research computed the extensive experimental results of the combined datasets (SIR^{2+}, Nature, Real, DSLR and Ash80). The total average of PSNR and SSIM achieved were 28.47 and 0.89, respectively. Figure 5 shows the evaluation metrics of the overall model.



```
Image 15.jpg: PSNR = 26.82, SSIM = 0.77
Image 0.jpg: PSNR = 27.36, SSIM = 0.79
Image 19.jpg: PSNR = 28.71, SSIM = 0.91
Image 13.jpg: PSNR = 28.24, SSIM = 0.83
Image 17.jpg: PSNR = 28.42, SSIM = 0.61
Image 12.jpg: PSNR = 29.19, SSIM = 0.77
Image 3.jpg: PSNR = 28.39, SSIM = 0.87
Image 6.jpg: PSNR = 28.62, SSIM = 0.57
Image 4.jpg: PSNR = 27.02, SSIM = 0.75
Image 10.jpg: PSNR = 30.46, SSIM = 0.77
Image 2.jpg: PSNR = 28.12, SSIM = 0.58
Image 11.jpg: PSNR = 30.77, SSIM = 0.62
Image 13.jpg: PSNR = 28.35, SSIM = 0.82
Image 6.jpg: PSNR = 29.79, SSIM = 0.55
Image 20.jpg: PSNR = 30.02, SSIM = 0.85
Image 0.jpg: PSNR = 28.89, SSIM = 0.71
Image 12.jpg: PSNR = 29.84, SSIM = 0.73
Image 17.jpg: PSNR = 28.04, SSIM = 0.53
Image 15.jpg: PSNR = 28.14, SSIM = 0.71
Image 1.jpg: PSNR = 28.06, SSIM = 0.71
Image 3.jpg: PSNR = 28.29, SSIM = 0.90
Image 19.jpg: PSNR = 28.70, SSIM = 0.85
Image 8.jpg: PSNR = 28.60, SSIM = 0.67
Image 7.jpg: PSNR = 26.98, SSIM = 0.71
Image 4.jpg: PSNR = 27.48, SSIM = 0.73
Image 16.jpg: PSNR = 28.01, SSIM = 0.61
Image 18.jpg: PSNR = 26.59, SSIM = 0.85
Image 5.jpg: PSNR = 30.22, SSIM = 0.69
Image 9.jpg: PSNR = 29.65, SSIM = 0.99
Image 14.jpg: PSNR = 28.71, SSIM = 0.83
Total Average PSNR: 28.47
Total Average SSIM: 0.89
```

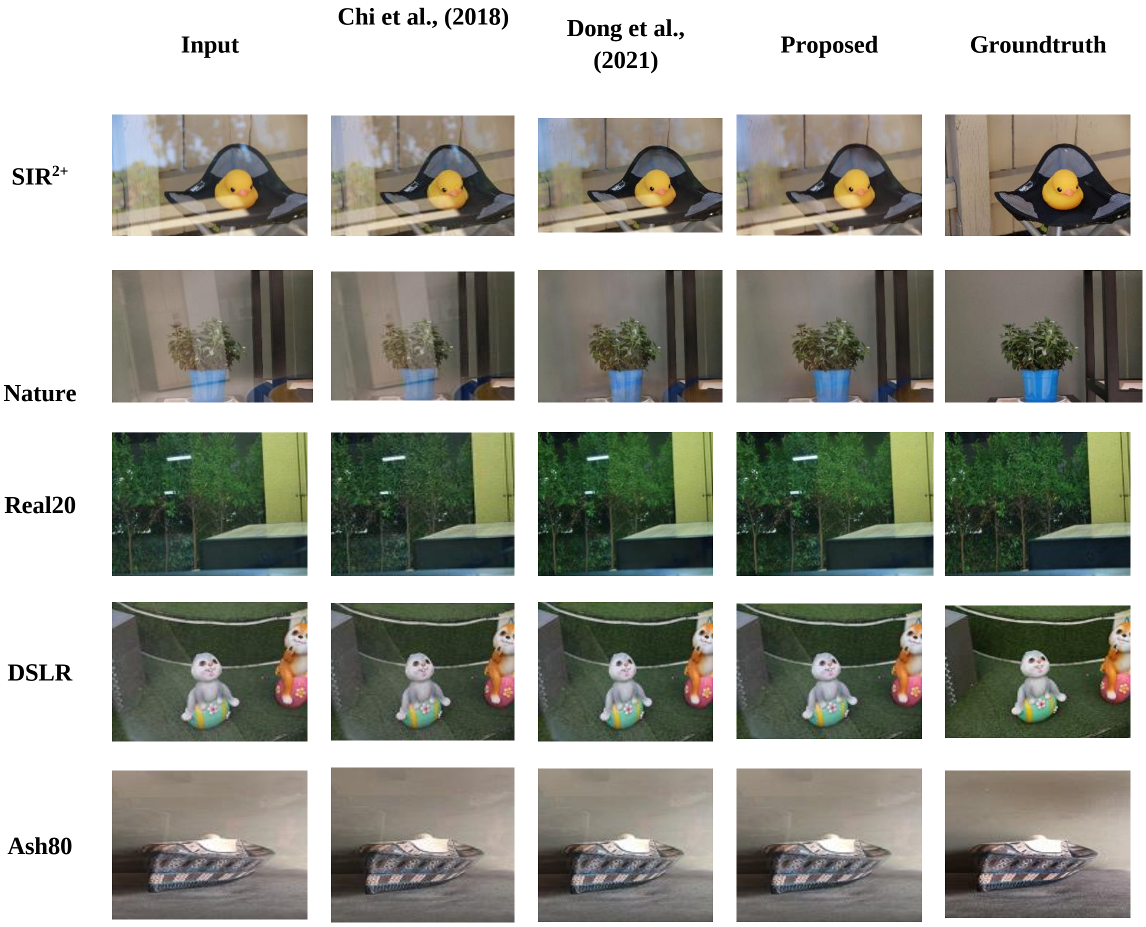
Figure 5: Evaluation metrics of the overall model

Moreover, this research also compares the extensive experimental results of the proposed model to two selected state-of-the-arts models both qualitatively and quantitatively. To testthe two selected state-of-the-art models, this research downloaded the pre-trained models without fine-tuning the models and test the models on the same datasets

Qualitative Evaluation

The qualitative evaluation of the proposed model was done by comparing the visual appealing scenes of the reflection-free scenes of the proposed model and two state-of-the-art models on five different datasets. It can be seen that existing models fail to remove concentrated reflections while the proposed model produced high-quality reflection-free images close to the grountruths. Table 2, give the reflection-contaminated images as inputs, the generated reflection-free images by each model and the groundtruths of the corresponding inputs.

Table 2: Qualitative Comparisons



Quantitative Evaluation

A quantitative evaluations and comparisons of the proposed model is made using PSNR and SSIM respectively to two state-of-the-art models on five different datasets. The model with highest value(s) on a specific dataset is highlighted in dark color, and also compared the performance of that model to others. It can be noted that the proposed model achieved the highest values on all the different datasets. This verifies that the proposed model can achieved superior performance than other models. Table 3 shows the quantitative comparison and it can also be verified in the appendix of this work.

Table 3: Quantitative comparisons

Dataset	Metrics	Chi et al. (2018)	Dong et al. (2021)	Proposed
SIR^{2+}	PSNR	16.16	16.53	**20.92**
	SSIM	0.765	0.786	**0.895**
Nature	PSNR	20.25	19.54	**23.75**
	SSIM	0.846	0.773	**0.928**
Real20	PSNR	24.34	23.82	**26.27**
	SSIM	0.888	0.859	**0.959**

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

DSLR	PSNR	16.28	16.18	**19.20**
	SSIM	0.514	0.349	**0.699**
Ash80	PSNR	21.17	20.99	**22.04**
	SSIM	0.751	0.646	**0.895**

CONCLUSION

In conclusion, the task of removing reflection interference from an image is extremely difficult due to the close morphological features between the background image and the reflection image. This research designed and implemented an enhanced single-image reflection removal model, which combined deep learning encoder-decoder and a Gaussian filter to optimally remove reflections and enhance the quality of the generated reflection-free image. The proposed developed model was found to have achieved higher results both quantitatively and qualitatively when compared to two state-of-the-art approaches, especially on images that contained noise. For the quantitative evaluation, the proposed model achieved PSNR and SSIM values on five different datasets as 20.92 and 0.895; 23.75 and 0.928; 26.27 and 0.959; 19.20 and 0.699; 22.04 and 0.895; respectively. And, for the qualitative evaluation, the proposed developed produced highly appealing visual output close to the groundtruth.

Contribution to Knowledge

The main contributions of this research are:

i. This study proposed an enhanced approach for single-image reflection removal that combined encoder-decoder and Gaussian filter.

ii. Extensive experiment shows that the proposed model out-performed two state-of-the-art models in removing reflection on reflection-contaminated image.

iii. Thus, this study also proposed a new dataset of real-world images for reflection removal with their corresponding ground truths.

RECOMMENDATION FOR FURTHER STUDIES

This study hopes to inspire subsequent work on single-image reflection removal, as reflection removal is one ill-posed problem that has been challenging the research community. Several approaches have been deployed to address this issue. However, this research proposed an enhanced deep learning encoder-decoder network that has a Gaussian filter to effectively and robustly remove reflections while enhancing the quality of the generated reflection-free image. However, despite the proposed approach outperforming two state-of-the-art approaches on five different datasets, there is still room for improvement of the model, especially on images that contain strong reflections or on images that are dominated by reflections.

REFERENCES

Abiko, R., & Ikehara, M. (2019). Single image reflection removal based on GAN with gradient constraint. *IEEE Access*, *7*, 148790-148799.

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, *1*(1), 1003-1014.

Amanlou, A., Suratgar, A. A., Tavoosi, J., Mohammadzadeh, A., & Mosavi, A. (2022). Single-Image Reflection Removal Using Deep Learning: A Systematic Review. *IEEE Access*.

Arvanitopoulos, N., Achanta, R., & Susstrunk, S. (2017). Single image reflection suppression. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4498-4506).

Born, M., & Wolf, E. (2013). *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier.

Chang, Y., & Jung, C. (2018). Single image reflection removal using convolutional neural networks. *IEEE Transactions on Image Processing*, *28*(4), 1954-1966.

Chang, Y., Jung, C., & Sun, J. (2020). Joint reflection removal and depth estimation from a single image. *IEEE Transactions on Cybernetics*, *51*(12), 5836-5849.

Chang, Y. C., Lu, C. N., Cheng, C. C., & Chiu, W. C. (2021). Single image reflection removal with edge guidance, reflection classifier, and recurrent decomposition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 2033-2042).

Cheng, K., Song, J., Du, J., Rong, S., & Zhou, H. (2020). Single image reflection removal via attention model and SN-GAN. *IEEE Access*, *8*, 96046-96054.

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.

Chi, Z., Wu, X., Shu, X., & Gu, J. (2018). Single image reflection removal using deep encoder-decoder network. *arXiv preprint arXiv:1802.00094*.

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).

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).

He, L., Li, F., Cong, R., & Zhao, Y. (2023). Reflection Intensity Guided Single Image Reflection Removal and Transmission Recovery. *IEEE Transactions on Multimedia*.

Jokinen, L and Sampo, K. (2017, October 19). *Hel Looks*. https://www. hel-looks.com/

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).

Li, Y., Yan, Q., Zhang, K., & Xu, H. (2021). Image reflection removal via contextual feature fusion pyramid and task-driven regularization. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(2), 553-565.

Quattoni, A., & Torralba, A. (2009, June). Recognizing indoor scenes. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 413-420). IEEE.

Article DOI: 10.52589/AJSTE-GWXJPEN4
DOI URL: https://doi.org/10.52589/AJSTE-GWXJPEN4

Thunborg, T., & Åström, E. (2022). Single Image Dome Reflection Removal Using Neural Networks.

Wan, R., Shi, B., Duan, L. Y., Tan, A. H., & Kot, A. C. (2017). Benchmarking single-image reflection removal algorithms. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3922-3930).

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).

Wan, R., Shi, B., Li, H., Hong, Y., Duan, L., & Chichung, A. K. (2022). Benchmarking single-image reflection removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wang, Z., She, Q., & Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)*, *54*(2), 1-38.

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).

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).

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).

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).

Zhang, H., Xu, X., He, H., He, S., Han, G., Qin, J., & Wu, D. (2019). Fast user-guided single image reflection removal via edge-aware cascaded networks. *IEEE Transactions on Multimedia, 22*(8), 2012-2023

Zheng, Q., Shi, B., Jiang, X., Duan, L. Y., & Kot, A. C. (2019, September). Denoising adversarial networks for rain removal and reflection removal. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 2766-2770). IEEE

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. 13395-12404).