Article

Two Residual Attention Convolution Models to Recover Underexposed and Overexposed Images

Noorman Rinanto *,† and **Shun-Feng Su** †

Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan; sfsu@mail.ntust.edu.tw

* Correspondence: noorman.rinanto@ppns.ac.id
† These authors contributed equally to this work.

Abstract: Inconsistent lighting phenomena in digital images, such as underexposure and overexposure, pose challenges in computer vision. Many studies have developed to address these issues. However, most of these techniques cannot remedy both exposure problems simultaneously. Meanwhile, existing methods that claim to be capable of handling these cases have not yielded optimal results, especially for images with blur and noise distortions. Therefore, this study proposes a system to improve underexposed and overexposed photos, consisting of two different residual attention convolution networks with the CIELab color space as the input. The first model working on the L-channel (luminance) is responsible for recovering degraded image illumination by using residual memory block networks with self-attention layers. The next model based on dense residual attention networks aims to restore degraded image colors using ab-channels (chromatic). A properly exposed image is produced by fusing the output of these models and converting them to RGB color space. Experiments on degraded synthetic images from two public datasets and one real-life exposure dataset demonstrate that the proposed system outperforms the state-of-the-art algorithms in optimal illumination and color correction outcomes for underexposed and overexposed images.

Keywords: image enhancement; underexposed; overexposed; exposure correction; residual attention convolution network

1. Introduction

The symmetry of illumination in digital and photographic images plays a significant role in determining the quality of the resulting image. Generally, image exposure quality decreases due to the limited capabilities of the device (camera) and the influence of ambient light. Underexposure and overexposure are two phenomena that often arise due to lighting asymmetry. An underexposed image is produced when the camera sensor only receives a small amount of reflected light from an object, which causes the image to become dark and lose detailed information due to being immersed in its shadow. On the other hand, overexposure is caused by too much reflected light from objects hitting the camera so that detailed information is lost in the highlights, producing an overexposed image [1]. Therefore, exposure correction techniques are needed to improve the quality of the degraded image.

Over the last decade, a myriad of research has been conducted by scientists to deal with exposure correction issues in digital imaging ranging from conventional methods to deep learning algorithms. Exposure correction based on curve map estimations [2–5], the probabilistic method [6], histogram equalization [7–9], wavelet transform [10–12], and Retinex theory [13–17] are very popular conventional methods. Although these methods can overcome some light correction problems, their output still lacks image naturalness and detail.

Currently, deep learning algorithms application to overcome this exposure problem is an attractive option for some experts, along with the development of Graphics Processing

Unit (GPU) technology that can compute large amounts of data. Two deep learning architectural models are widely used in exposure correction: CNN-based models and generative-based models such as Generative Adversarial Networks (GANs). Many experts have successfully developed CNN-based models to restore underexposed images using either pure CNN networks [18–20] or by combining several other techniques such as discrete wavelet transform (DWT) [21], the Retinex model [22,23], and the estimation curve [24]. In addition, Gao et al. [25] constructed a convolution network to recover a single input image with overexposure defects. Several scientists have employed GAN-based models to improve underexposed images [26–28] and overexposed images [29]. Unfortunately, the aforementioned methods only recover one of the two types of light defects (underexposed or overexposed) in digital images. In recent years, several image enhancement methods have emerged that can deal with underexposed and overexposed images simultaneously. Some of them use traditional or estimation methods [30,31] or deep learning-based solutions [32–36]. However, their performance still needs to be optimized because sometimes unknown artifacts are still found (especially in the overexposed image), and some methods have not been able to recover image with contrast exposure defects.

In this study, we design a novel CNN-based exposure restoration method for underexposed and overexposed images by separating the illumination correction process and the color correction process through two different residual attention networks using the CIELab color space to achieve optimal outcomes compared to previous works. Our model uses a residual network due to its superior performance compared to the U-Net and fully convolutional network (FCN) models [37,38]. The CIELab color space was selected because it has an optimal output compared to the other color spaces in our experiments. The illumination correction task was performed using a model called the illumination correction attention network (ICANet), composed of residual memory blocks, each equipped with a self-attention layer. These residual memory blocks are inspired by MemNet [39], which has been successful in recovering image lighting defects [40]. Meanwhile, the color correction job is handled by another model named the color correction attention network (CCANet), which is constructed of a residual dense block (RDB) and a self-attention block. This RDB was adopted from the residual dense network (RDN), which has been able to solve several cases of color image restoration, such as image super-resolution, image artifact reduction, image denoising, and image deblurring [41]. The addition of a self-attention layer in both models is aimed at increasing the quality and accuracy of the output.

We conduct a series of comparative experiments with several advanced methods using degraded image synthesis data from the MIT-Adobe FiveK [42] and PASCAL VOC2012 [43] datasets, as well as the real-life exposure dataset used by Afifi et al. [36], to evaluate the effectiveness of our algorithm. The results of our experiments show that our approach is superior to those of existing works.

The contribution of our proposed method can be summarized as follows:

- We propose a novel illumination and color correction method, employing a dual convolution network based on dissimilar residual attention blocks to refine underexposed and overexposed images.
- Our model offers a solution to optimize image restoration results by separating the illumination and color correction processes through two convolution networks using the CIELab color space.
- We propose to add a self-attention layer to all residual blocks in our system to enhance system performance.
- We create a synthetic image dataset for underexposure and underexposure cases, along with related ground-truth images, based on two public datasets for the training process.

2. Related Works

Zhang et al. [30] describe the traditional methods for correcting underexposed and overexposed images by blending the forward illumination estimation of the input image

and the reverse illumination estimation of the inverted input image. Another traditional approach to normalizing exposure defect images is to use a fusion-based method that combines the reflectance and illuminance estimations of the input image developed by Fu et al. [31]. Non-traditional solutions applying deep learning models have been employed by several scientists, including Steffens et al. [32], who attempted to restore the contrast and global luminance of over- and underexposed images employing a U-Net architecture-based approach with a dilated convolution block. Another CNN-based method consisting of twenty convolution layers to correct the illumination of color painting on a reflective surface such as a notebook was constructed by Goswami et al. [33]. The application of illumination correction was not only for photographs and paintings but has also been used to improve the quality of scanned documents, as has been done by Li et al. [34] using a patch-based CNN model. In addition, the implementation of Retinex propagation theory in a dual-convolution network to address these exposure cases has also been proposed by Ma et al. [35]. Lastly, Afifi et al. [36] provided a solution to fix these exposure issues by applying a deep neural network with a Laplacian pyramid framework trained on 24,000 multi-exposure photos. All of the aforementioned algorithms have become state-of-the-art in addressing the exposure correction on under- and overexposed images.

The main differences between the proposed model and previous studies in the case of underexposed and overexposed image correction are shown in Table 1. ARPNet [35] and our system both have two convolution networks but the ARPNet networks are composed of identical structures and share weights between them, whereas our two networks are built from different residual attention networks, as described in the previous paragraph. Unlike IllNet [34] and ARPNet, our residual architecture is equipped with a self-attention layer that is widely used by scientists [44–46] to improve system performance. In addition, the number of residual blocks in our network is less than those in IllNet and ARPNet. For skip connections, the remaining blocks in our method use dense and recursive skip connections, whereas IllNet and ARPNet apply regular ones. Figure 1 illustrates the differences between these skip connections. Furthermore, data processing in our network is performed in the CIELab color space, unlike other models that typically use the RGB color space. The advantage of using this color space is that there is a special channel for setting the luminance (L) and a separate channel for adjusting the color (a and b). With the appropriate adjustment of both channels, an output image with optimal lighting and coloring will be obtained. A performance comparison of the proposed algorithm utilizing the CIELab color space with other color spaces that support luminance and color settings is discussed in the ablation study section.
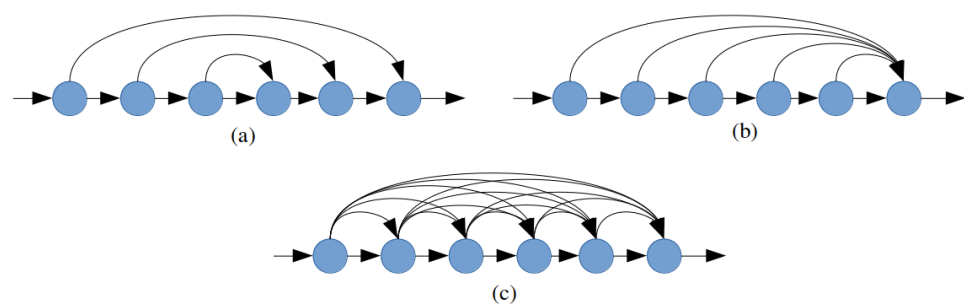

Figure 1. Illustration of various skip connection schemes. (**a**) Regular skip connection. (**b**) Recursive skip connection. (**c**) Dense skip connection.

Table 1. Comparison of the proposed method with previous studies.

Method	Number of Networks	Architecture	Number of Residual Blocks	Type of Connection	Color Spaces
DualIE [30]	-	Dual Illumination *	-	-	RGB
FBEI [31]	-	Reflectance and Illumination *	-	-	RGB
ReExposeNet [32]	1	UNet	-	-	RGB
FCN20 [33]	1	Fully Convolutional Network	-	-	RGB
IllNet [34]	1	Residual Network	5	Regular Skip	RGB
ARPNet [35]	2	Residual Network	16	Regular Skip	RGB
MSPEC [36]	1	UNet	-	-	RGB
Ours	2	Residual Attention Network	3 and 4	Recursive and Dense	CIELab

* Non-Deep Learning.

3. Proposed Method

This section discusses the mechanisms of the proposed system, including the system overview, ICANet architecture, self-attention mechanism, CCANet architecture, and loss functions.

3.1. System Overview

The main architecture of our system is presented in the pipeline diagram shown in Figure 2. Overall, this system can be divided into two sub-systems: the illumination correction attention network (ICANet) and the color correction attention network (CCANet). These two networks play different roles in recovering underexposed and overexposed images. The main task of ICANet is to correct the illumination of images that suffer from exposure defects by processing the luminance (L) channel of the input information. Meanwhile, CCANet is responsible for refining the color of the incorrectly exposed image by processing the incoming information using the chromatic channel (a, b). RGB input images must be converted to the CIELab color space first before being processed in these networks. Once the processing on this network is completed, the two network outputs are combined and converted to the RGB color space to obtain the final image.
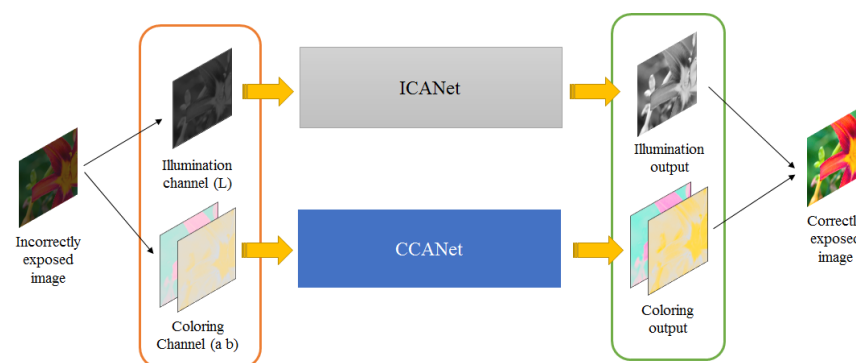

Figure 2. The overall pipeline diagram of the proposed model consists of an illumination correction attention network (ICANet) to improve illumination issues and a color correction attention network (CCANet) to handle color refinement in degraded underexposed and overexposed images.

3.2. ICANet Architecture

Figure 3 shows the ICANet architecture of the proposed method. The main structure of this network consists of a feature extraction convolution layer (f_e) with 32 filters ($n = 32$), a feature reconstruction convolution layer (f_r) at the end of the network, and three memory attention blocks (MABs). All convolution layers in this network use 2D convolution plus a

ReLU activation function with several parameters, such as the kernel size (k), stride size (s), and padding size (p), set to 3, 1, and 1, respectively.

This network was designed with two skip connections, which connect the system input to the output f_r and from the output f_e to the final MAB output through a summation (+) operation. The construction of this MAB, inspired by the memory block of MemNet [39], consists of three convolution layers triggered by the ReLU activation function using 32 filters; a self-attention layer with a block structure, as shown in Figure 4; and a gate unit that will concatenate the output of the previous layer. Unlike the memory block in MemNet, we removed batch normalization and added a self-attention layer to our memory block to maintain output quality. In addition, we used three recursive units in our proposed block, whereas the original MemNet block used five recursive units. This reduction was intended to minimize processing time and information loss from our network. In addition, according to research by Jin et al. [47], this recursive network can improve the performance of the model. The ICANet was trained using the ADAM optimization algorithm and several parameters such as epoch, batch size, learning rate, momentum, and weight decay with values of 100, 1, 10^{-4}, 0.9, and 0.9, respectively.
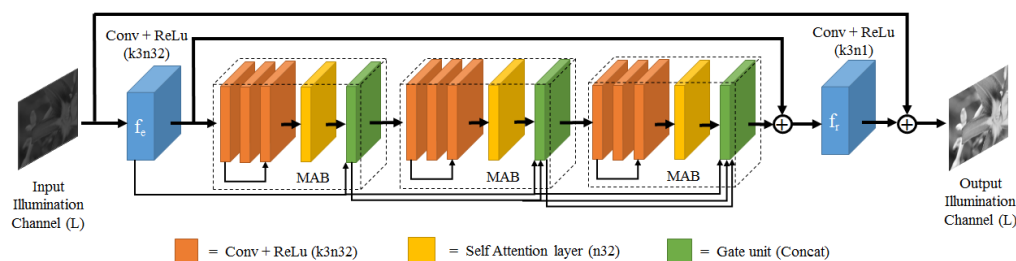

Figure 3. The ICANet architecture of our system consists of two feature convolution layers, known as feature extraction (f_e) and feature reconstruction (f_r), and three memory attention blocks (MABs).
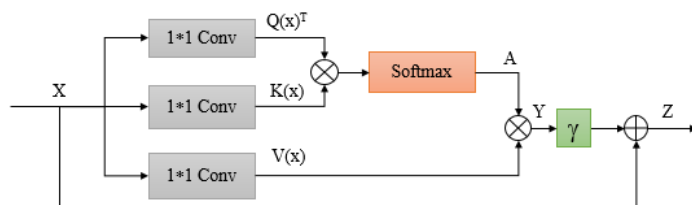

Figure 4. Block diagram of the self-attention layer applied to the proposed method.

3.3. Self-Attention Mechanism

Figure 4 shows the structure of the self-attention layer implemented in our method. The self-attention mechanism applied to the generator module in the GAN model was first introduced by Zhang et al. [44] and was used to help models drive long-distance dependencies and produce fine detail in drawing generation tasks. This layer has two output components: the feature map (Z_i) from the preceding convolution layer and the self-attention map (A). Z_i is obtained by adding up the attention matrix value (Y_i) multiplied by the gamma parameter (γ) for the optimization process with data input (X_i) as residual learning, which can be formulated as:

$$Z_i = \gamma Y_i + X_i, \tag{1}$$

In the initial stage, γ is set to zero and is increased for optimization during network training. The value of the attention matrix is the product of the multiplication between the self-attention map (A) and the value of the input matrix ($V(x)$), which can be expressed as (2), and by applying (3), we can obtain the value of A.

$$Y_i = AV(x), \tag{2}$$

$$A = softmax(Q(x)^T K(x)), \tag{3}$$

where $Q(x)^T$ is the transpose matrix of the query from input X, and the key matrix of X is denoted as $K(x)$.

3.4. CCANet Architecture

The architecture of the CCANet, which is responsible for restoring the colors in the exposure-degraded images, is shown in Figure 5. This network was constructed by placing two feature extraction convolution layers (f_{e1}, f_{e2}) at the beginning of the network, two feature reconstruction convolution layers (f_{r1}, f_{r2}) at the end of the network, four residual dense attention blocks (RDABs) followed by a concatenated block (*Concat*) with a 1×1 convolution located in the middle of the network, and a self-attention block in the skip connection path, which has a multiplication operation at the end of the line. Here, all feature extractions and reconstructions were 2D convolution layers, with no activation function triggered. The structure of the RDABs can be seen in Figure 6. This residual block was adopted from RDN [41] and contains three convolution layers with a ReLU activation function and a concatenate layer with a 1×1 convolution. All convolution layers in this network used several parameters, as well as the filter (n), kernel (k), stride (s), and padding (p) with sizes of 32, 3, 1, and 1, respectively. We modified the residual block from RDN by adding a self-attention (SA) layer after the concatenate layer and multiplying its outcome with the output of the concatenate block in our network (see Figure 6). The construction and mechanism of all SA layers in this network were similar to those in ICANet. This network was also trained with the ADAM optimization algorithm and used training parameters identical to those in ICANet.
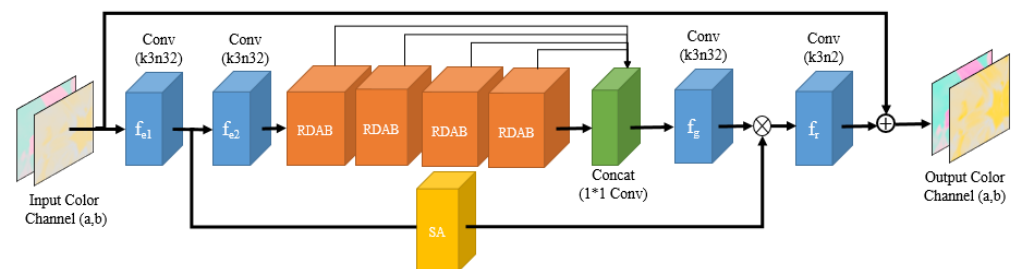

Figure 5. The CCANet architecture of our model consists of two feature extraction convolution layers (f_{e1}, f_{e2}), two feature reconstruction convolution layers (f_{r1}, f_{r2}), four residual dense attention blocks (RDABs), a merge block (Concat), and a self-attention (SA) block.
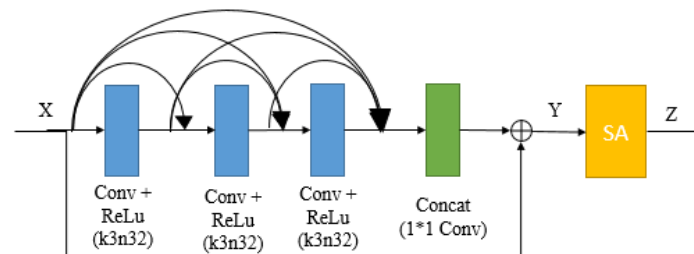

Figure 6. The structure of the residual dense attention blocks (RDABs) used in the CCANet architecture.

3.5. Loss Function

In this work, we used two total loss functions in our model during the training phase. The first total loss was for ICANet (L_{ICANet}) and was a combination of the L1 loss (L_{l1}) and structural similarity (SSIM) loss (L_{SSIM}). This total loss has been used in several systems for the illumination problem [48,49] and can be expressed as:

$$L_{ICANet} = \alpha_{l1} L_{l1} + \alpha_{SSIM} L_{SSIM}, \tag{4}$$

where α_{l1} and α_{SSIM} are the hyperparameters used to balance L_{l1} and L_{SSIM}, which have values of 0.2 and 0.8, respectively. L_{l1} or the mean absolute error (MAE), obtained from (5), and L_{SSIM} can be represented by (6).

$$L_{l1} = \frac{1}{N} \sum_{i=1}^{N} \left| y - y_{ref} \right|, \tag{5}$$

$$L_{SSIM} = 1 - SSIM(y, y_{ref}), \tag{6}$$

Here, y is the model output, the target output is denoted as y_{ref}, and N is the total number of data samples. SSIM(y, $yref$) is calculated with the following rules:

$$SSIM(y, y_{ref}) = \frac{(2\mu_y \mu_{y_{ref}} + c_1)(2\sigma_{yy_{ref}} + c_2)}{(\mu_y^2 + \mu_{y_{ref}}^2 + c_1)(\sigma_y^2 + \sigma_{y_{ref}}^2 + c_2)}, \tag{7}$$

where μ_y and $\mu_{y_{ref}}$ are the pixel samples of y and y_{ref}, respectively; σ_y^2 and $\sigma_{y_{ref}}^2$ are the variances of y and y_{ref}, respectively; the cross-correlation between y and y_{ref} is denoted as $\sigma_{yy_{ref}}$; and c_1 and c_2 represent divisional stabilizing constants with weak dominators. We obtain the values of c_1 and c_2 by applying $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$. Moreover, L is the dynamic range of the pixel values with default values of $k_1 = 0.01$ and $k_2 = 0.03$.

The other total loss used in the CCANet model (L_{CCANet}) consists of the mean square error (MSE) loss (L_{MSE}) and the Huber loss function (L_H). This total loss has been widely used in image color enhancement [40,50,51] and can be formulated as follows:

$$L_{CCANet} = \alpha_{MSE} L_{MSE} + \alpha_H L_H, \tag{8}$$

where α_{MSE} and α_H are hyperparameters that balance the sub-loss functions with values of 0.1 and 0.9, respectively. L_{MSE} is calculated using (9), and L_H can be obtained using (10).

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(y - y_{ref} \right)^2, \tag{9}$$

$$L_H = \begin{cases} 0.5(y - y_{ref})^2 & : if \left| y - y_{ref} \right| \leq \delta \\ \delta(\left| y - y_{ref} \right| - 0.5\delta) & : otherwise \end{cases} \tag{10}$$

where δ is the threshold value that controls the transition change between the *L1* loss and the *L2* loss function. We used the default value of $\delta = 1$ in this implementation.

4. Experiments

All topics related to our experiments are presented in this section, beginning with an explanation of the datasets and metrics, followed by an evaluation of the system's performance compared to other mainstream algorithms and an ablation study. As additional information, we developed our model using the Pytorch framework in a Linux environment, and we ran it on a computer with an Intel Core i7 processor utilizing an NVIDIA GeForce GTX 1070 GPU.

4.1. Datasets and Metrics

In this experiment, our proposed algorithm was trained using a synthetic exposure defect image dataset generated from the MIT Adobe FiveK dataset. This dataset is a paired dataset consisting of 5000 good-quality photos in sRGB format, along with associated images retouched by five experts [42]. This dataset has been widely used by scientists for image enhancement problems [28,30,32,36]. The procedure for creating our training synthesis dataset can be described as follows: First, a random sample of 800 images was

selected from the MIT Adobe FiveK dataset and resized to 256 × 256 pixels. This resizing was carried out due to hardware limitations and the large number of images used in the training phase. The next step was to generate a new synthetic image by adding lighting effects that varied with percentages of 30%, 50%, 70%, 100%, 130%, 150%, 170%, and 200%. This step produced 6400 images that represented both underexposed and overexposed images. Furthermore, we generated another new synthesized image representing low- and high-contrast defects by adding a contrast variation effect with percentages similar to those used in the previous step. So, we now had 12,800 degraded image data. To make our model more robust, we built a new synthetic image dataset from previously generated images by adding Gaussian noise and blur. Finally, the total synthesized image dataset resulting from the above process contained 38,400 degraded images. Figure 7 shows some exemplary synthetic images in our datasets.
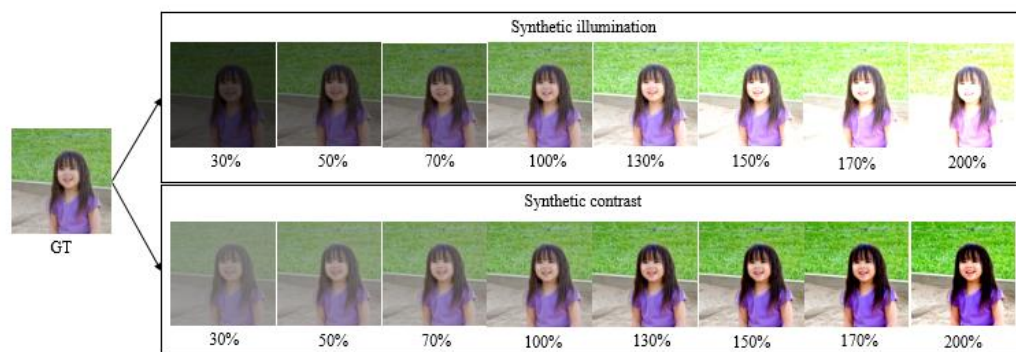

Figure 7. Examples of synthetic illumination and synthetic contrast images in our proposed dataset, with ground-truth (GT) image on the left.

Three different datasets were used in the evaluation phase, including two synthetic exposure-degraded image datasets generated from the MIT Adobe FiveK [42] and PASCAL VOC2012 [43] datasets and a test dataset used by Afifi et al. [36]. The generation of the synthetic evaluation dataset based on the MIT Adobe Fivek and PASCAL VOC2012 datasets used the same procedure as the generation of the training dataset. Each dataset was generated from 200 random samples of images and at the end of the process, each dataset produced 9600 degraded images. VOC2012 contains images that differ in quality, especially in terms of illumination and color, making it a suitable test dataset for our algorithm. Meanwhile, the dataset by Afifi et al. is a collection of images that have been rendered, imitating real exposure from digital cameras with different exposure values (EV) of −1.5, −1, 0, +1, and +1.5. Thus, this dataset can simulate images with natural exposure defects.

In this evaluation, we used several quantitative metrics based on image quality assessment (IQA) with full-reference metrics, including the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index [52], and visual saliency with color display and gradient similarity (VCGS) [53]. The PSNR is used to measure the level of compression in decibels (dB) between the reference image and the predicted image by comparing the pixel values. It is calculated using the following rule:

$$PSNR = 10 * log_{10}\left(\frac{MAX^2}{MSE}\right),$$ (11)

where MAX is the maximum pixel value in the grayscale image and MSE is the mean square error between the reference image and the predicted image, which can be calculated using (9).

SSIM is a full-reference metric that compares the luminance, contrast, and similarity structure of the reference image and the degraded image. We used (7) to calculate this metric.

The PSNR and SSIM index metrics are commonly used in IQA in image enhancement models, including exposure correction cases [32,35,36].

To assess the color correction quality of the output image, this study used the reliable VCGS metric as a full-reference IQA model for color images. This metric combines the similarity of visual saliency with the color display (VC), gradient, and chrominance and is suitable for measuring the color quality in degraded images. The VCGS metric can be expressed as follows:

$$VCGS = \frac{\sum_\Omega S_{vc} \cdot (S_g)^\alpha \cdot (S_c)^\beta \cdot VC_{max}}{\sum_\Omega VC_{max}}, \tag{12}$$

where Ω is the spatial domain; S_{vc} is the similarity of the visual saliency map with a color appearance, which can be formulated using (13); the gradient similarity is denoted by S_g and obtained using (14); S_c is chrominance elements' similarity in the CIELab color space (*a-b* channel), which can be expressed as in (15); α and β are the important relative features among visual salience, structure, and chrominance; and VC_{max} is the maximum values of two visual saliency maps (VC_1 and VC_2) that represent the reference image and the predicted image in this case.

$$S_{vc} = \frac{2VC_1.VC_2 + K_{vc}}{VC_1^2 + VC_2^2 + K_{vc}}, \tag{13}$$

$$S_g = \frac{2g_1 \cdot g_2 + K_g}{g_1^2 + g_2^2 + K_g}, \tag{14}$$

$$S_c = \frac{2a_1.a_2 + K_c}{a_1^2 + a_2^2 + K_c} \cdot \frac{2b_1.b_2 + K_c}{b_1^2 + b_2^2 + K_c}, \tag{15}$$

where K_{vc}, K_g, and K_c are numerical stability control parameters [52]; g_1 and g_2 are the gradient magnitudes of the luminance channel (*L*) in CIELab color spaces, which are calculated by $g_{1,2} = \sqrt{g_x^2 + g_y^2}$, where g_x and g_y are the horizontal and vertical gradient operators of an input image (*X*), respectively. The equations g_x and g_y can be written as follows:

$$g_x = \frac{1}{16} \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} * X, \tag{16}$$

$$g_y = \frac{1}{16} \begin{bmatrix} 3 & 10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & -3 \end{bmatrix} * X, \tag{17}$$

4.2. Performance Evaluation

To evaluate the reliability of the proposed approach, we performed benchmarking experiments against other advanced methods that support both underexposed and overexposed image problems, including DualIllum [30], FBEI [31], ReExposeNet [32], FCN20 [33], IllNet [34], ARPNet [35], and MSPEC [36]. These quantitative evaluations are summarized in Table 2. All of these algorithms were tested using the MIT Adobe FiveK-based synthesis dataset, the VOC2012-based synthesis dataset, and the Afifi et al. [36] test dataset. It should be noted that the ground-truth images in the Afifi dataset are images corrected by an expert photographer C (refer to [42]). In the results, it can be seen that the proposed method outperformed other state-of-the-art algorithms, achieving the highest PSNR, SSIM, and VCGS scores on all of the testing datasets.

A visual comparison of the output for underexposed input images with or without noise and blur can be seen in Figure 8. Meanwhile, Figure 9 demonstrates a qualitative comparison of all outputs for the overexposed input images. Based on these images, it

appears that our algorithm produced predictive outputs that were close to the ground-truth images and achieved the highest scores across all IQA metrics.

Table 2. Quantitative comparison of the proposed system and other methods on different datasets.

| Method	MIT-Adobe FiveK-Based			PASCAL VOC2012-Based			Afifi et al. [36]		
	PSNR	SSIM	VCGS	PSNR	SSIM	VCGS	PSNR	SSIM	VCGS
DualIE [30]	17.83	0.686	0.913	17.81	0.687	0.912	19.16	0.855	0.967
FBEI [31]	16.84	0.681	0.913	16.34	0.671	0.911	15.82	0.800	0.959
ReExposeNet [32]	13.44	0.544	0.892	13.05	0.537	0.896	15.11	0.596	0.909
FCN20 [33]	18.64	0.655	0.916	18.08	0.647	0.914	16.81	0.755	0.946
IllNet [34]	18.77	0.680	0.931	18.56	0.690	0.931	17.45	0.790	0.954
ARPNet [35]	18.67	0.673	0.926	18.34	0.675	0.925	17.35	0.785	0.954
MSPEC [36]	19.43	0.730	0.935	19.33	0.727	0.936	21.23	0.874	0.971
Ours	**22.38**	**0.828**	**0.963**	**22.23**	**0.836**	**0.961**	**22.52**	**0.888**	**0.974**
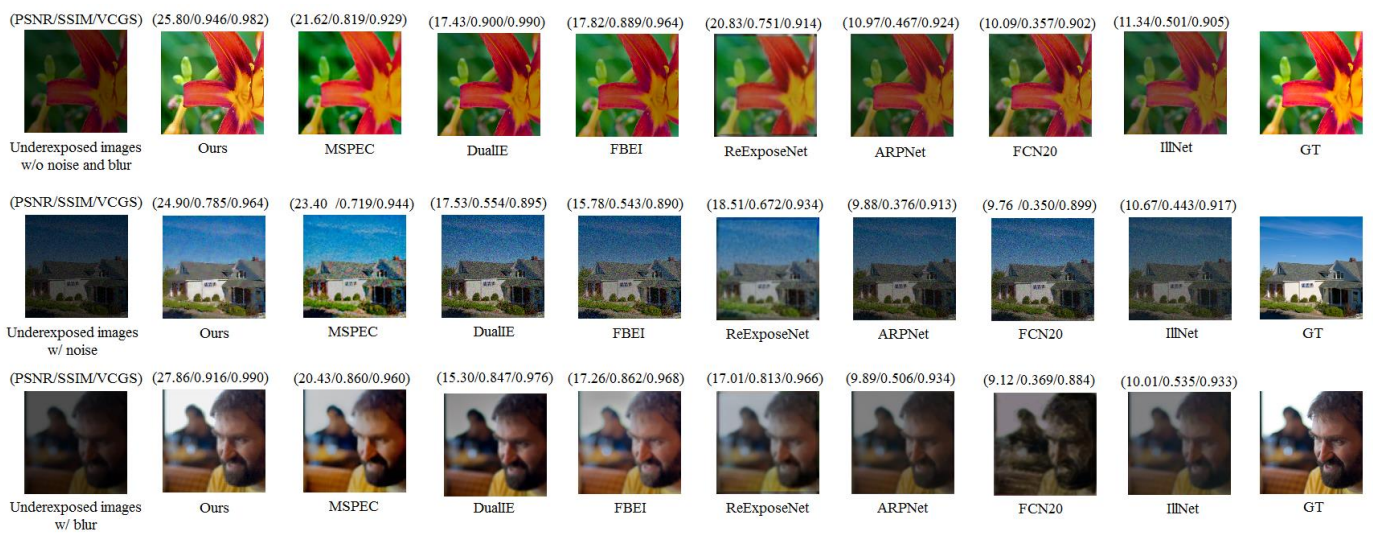
Figure 8. Qualitative comparison of all methods on underexposed images without noise and blur defects (**top**), with noise defects (**middle**), and with blur defects (**bottom**).
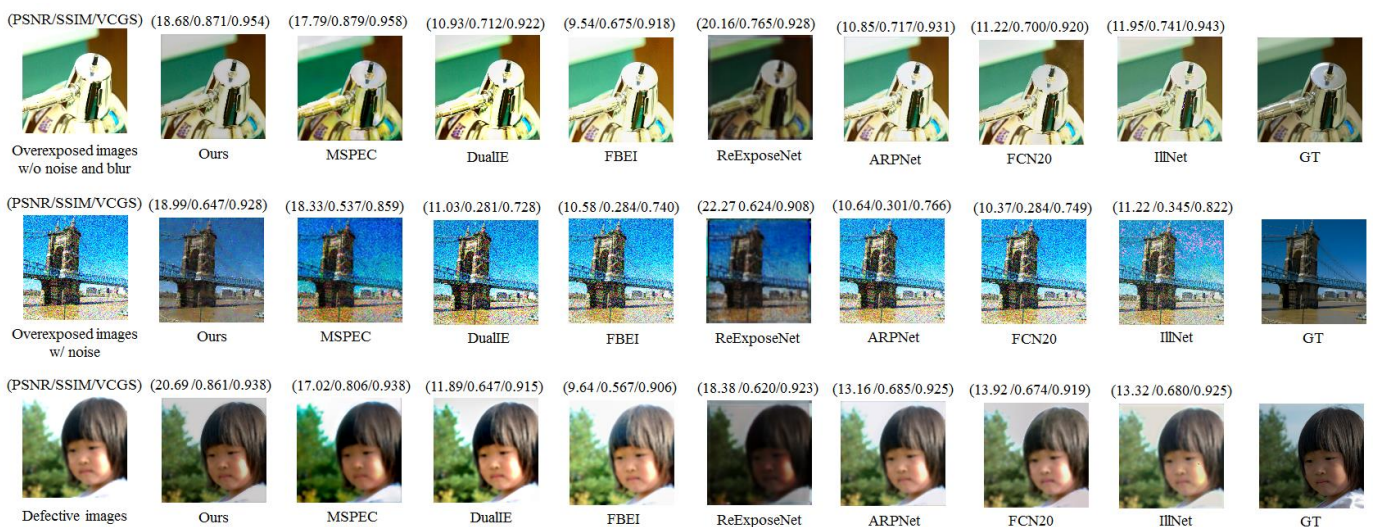

Figure 9. Qualitative comparison of all methods on overexposed images without noise and blur defects (**top**), with noise defects (**middle**), and with blur defects (**bottom**).

In addition, an evaluation using low-contrast and high-contrast input images was also applied to the models. This test aimed to determine the performance of the proposed system against input images that have inappropriate exposure contrast. The qualitative comparison of this evaluation is presented in Figure 10 (low-contrast exposed images) and Figure 11 (high-contrast exposed images). In the visualization, it can be seen that the proposed method has better corrected image quality compared to other advanced methods, especially for images with low-contrast exposure. Moreover, the proposed method has promising performance when recovering improperly contrast-exposed images with noise and blur defects.
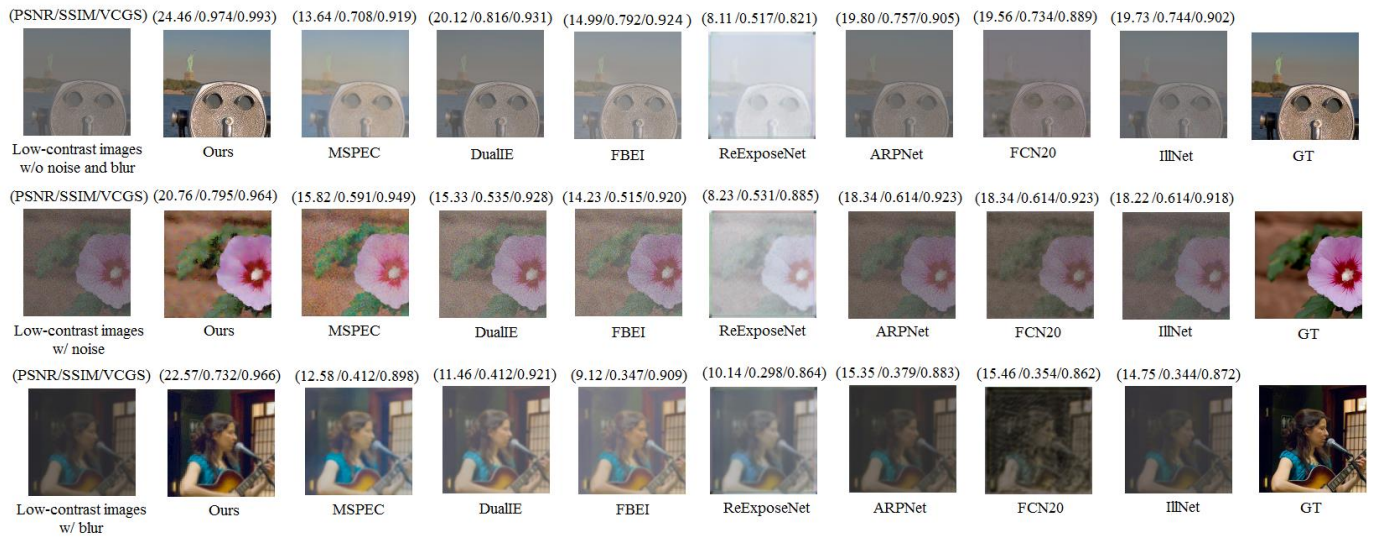

Figure 10. Qualitative comparison of all methods on low-contrast exposed images without noise and blur defects (**top**), with noise defects (**middle**), and with blur defects (**bottom**).
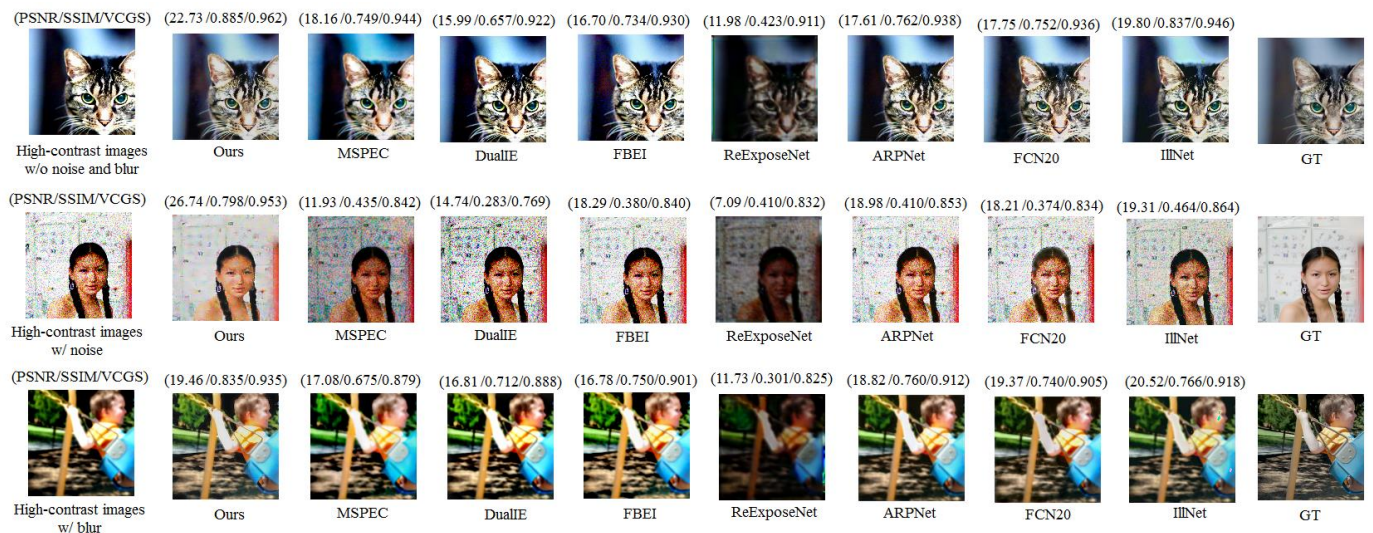

Figure 11. Qualitative comparison of all methods on high-contrast exposed images without noise and blur defects (**top**), with noise defects (**middle**), and with blur defects (**bottom**).

Figure 12 shows a qualitative comparison of the models tested against the sample of the Afifi dataset representing natural exposure-degraded images, with overexposed images (+1.5 EV) in the top row and underexposed images (−1.5 EV) in the bottom row. This figure demonstrates that our algorithm's output was the best, achieving the highest PSNR, SSIM, and VCGS values across the real exposure-degraded images. In underexposed images, our method was the best, with PSNR, SSIM, and VCGS values of 33.98, 0.981, and 0.993, respectively, whereas in overexposed images, our proposed method achieved slightly

higher metric values compared to the MSPEC method as its closest competitor, with PSNR, SSIM, and VCGS values of 28.88 dB, 0.929, and 0.987, respectively.
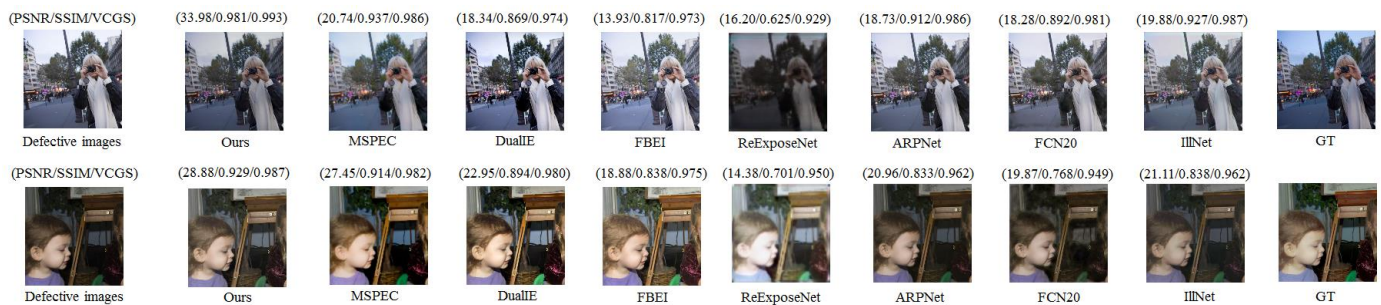

Figure 12. Qualitative comparison of all methods on overexposed images (+1.5 EV) in the top row and underexposed images (−1.5 EV) in the bottom row.

4.3. Ablation Study

The ablation study aimed to measure the effectiveness of the proposed method by changing or eliminating several parameters such as the use of different color spaces in the model, the use of self-attention layers in the network, and the use of mixed datasets (using noise and blur samples) during model training.

Table 3 shows the comparative performance of our model with different types of color spaces applied. The color space compared in this experiment is a color space that has lighting channels (illumination) and chromatic color channels such as hue-saturation-value (HSV), luminance-chroma blue-chroma red (YCbCr), luminance-green/red-blue/yellow (Luv), and luminance-green/red-blue/yellow (Lab).

Table 3. Quantitative comparison of the proposed model using different color spaces on the synthesis datasets.

Color Spaces	MIT Adobe FiveK-Based			VOC2012-Based		
	PSNR	SSIM	VCGS	PSNR	SSIM	VCGS
HSV	21.58	0.724	0.923	21.12	0.738	0.927
YCbCr	21.98	0.734	0.930	21.76	0.752	0.934
Luv	21.41	0.735	0.934	20.75	0.744	0.935
CIELab	**22.38**	**0.828**	**0.963**	**22.23**	**0.836**	**0.961**

In this test, the illumination channel was applied to the ICANet model, and the chromatic color channel was used in the CCANet model. As for the HSV color space, the V channel is the intensity or illumination channel, while the rest are the chromatic colors. In the YCbCr color space, the Y channel is the illumination channel or the light channel, whereas the other channels are the color channels. As for Luv and Lab, the illumination channel is the L channel (luminosity) and the other channels are the chromatic colors. The entire training model used ADAM optimizations and the same configuration of parameters such as epoch, batch size, learning rate, momentum, and weight decay with values of 100, 1, 10^{-4}, 0.9, and 0.9, respectively. All of these color space tests were applied to two exposure-degraded image synthesis datasets, including those based on MIT Adobe FiveK and PASCAL VOC2012 with the number of test samples for each dataset of 9600 images. It can be seen in Table 3 that the model with the Lab color space performed the best compared to the other models that applied the HSV, YCbCr, and Luv color spaces by achieving the highest PSNR, SSIM, and VCGS scores across all test datasets (marked with values in bold). In addition, the model with the HSV color space did not perform well in the similarity index and coloration assessment, as it achieved the lowest SSIM and VCGS values across all the degraded image synthetic datasets.

The model's performance with various combinations of self-attention layers is presented in Table 4. In this experiment, the effectiveness of the self-attention (SA) layer of the two residual models (ICANet and CCAnet) was tested by installing an SA layer (+SA) or removing an SA layer (−SA) in each network. In this table, it can be seen that the best configuration of the two residual networks is the model using the self-attention layer with the highest assessment metric score (marked by the value in bold) for both the MIT Adobe FiveK-based test dataset and the VOC2012-based dataset. In the table, it can also be seen that the CANNet model with the SA (+SA) layer achieved better SSIM and VCGS values compared to without using the SA (−SA) layer in both test datasets. In addition, the use of the SA layer in the ICANet model increased the PSNR score. This can be seen from the PSNR value in the MIT Adobe FiveK-based synthesis dataset, which obtained a score of 20.19 dB, and the VOC2012-based synthesis dataset, which obtained a score of 20.77 dB.

Table 4. Ablation study of the proposed model using different combinations of self-attention layers.

ICANet	CCANet	MIT Adobe FiveK-Based			VOC2012-Based		
		PSNR	SSIM	VCGS	PSNR	SSIM	VCGS
−SA	−SA	19.35	0.712	0.923	19.53	0.720	0.922
+SA	−SA	20.19	0.732	0.926	20.77	0.738	0.926
−SA	+SA	19.48	0.770	0.951	19.19	0.773	0.950
+SA	**+SA**	**22.38**	**0.828**	**0.963**	**22.23**	**0.836**	**0.961**

−SA = without self-attention; +SA = with self-attention.

Figure 13 shows a comparison of the training performance of the two proposed network models (CCANet and ICANet) using degraded image datasets with noise defects and blurred images (w/mixed dataset) and without using noise and blurred images (w/o mixed dataset). Both models were trained using the same dataset (FiveK-based) and the same parameter configuration. In the CCANet training performance chart, the model using a mixed dataset (red line graph) achieved a higher average score of about 1.18 dB compared to the model without a mixed dataset (blue line graph). Meanwhile, the ICANet model trained on a mixed dataset (red line graph) outperformed the ICANet model trained using an unmixed dataset (blue line graph), with a difference in the average PSNR value of 3.17 dB. A comparison of the loss graphs of our method both with and without mixed datasets is shown in Figure 14. In the CCANet model, it appears that the model trained on mixed datasets exhibited a slightly faster loss reduction (epoch ≤ 10) compared to other models without mixed datasets, whereas in the ICANet model, the loss graph model trained on mixed datasets was superior to the model not trained on mixed datasets. The loss value dropped at epoch 8 for models with mixed datasets, whereas for models with non-mixed datasets, the loss value dropped when the epoch reached 18. Based on these facts, it is clear that the proposed model trained on a mixed dataset demonstrated better performance compared to the model trained on an unmixed dataset.
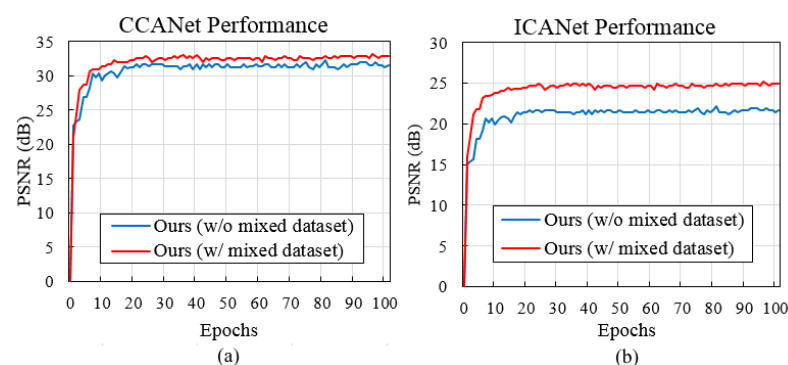

Figure 13. The training performance graph of the proposed method with and without the mixed dataset. (**a**) On CCANet training. (**b**) On ICANet training.
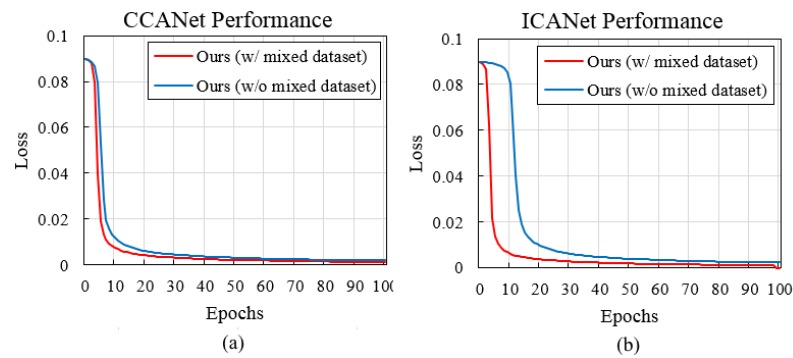
Figure 14. The loss graph comparison of the proposed method with and without the mixed dataset. (**a**) On CCANet training. (**b**) On ICANet training.

Furthermore, the performance of the proposed model was also evaluated against other types of imagery such as satellite (remote sensing) [54], industrial applications (visual inspection) [55], and biomedical imagery (human lung X-ray photos) [56]. These experiments aimed to determine the reliability of the system when applied outside of photographic image restoration. Figure 15 demonstrates the restoration results of all models in other computer vision applications using the same parameters and trained models as previous experiments (without training on new datasets). For satellite image restoration, our model achieved the highest measurement metric values with a PSNR = 26.89 dB, an SSIM = 0.933, and a VCGS = 0.969. Our system also outperformed other benchmark models on PCB visual inspection image recovery with PSNR, SSIM, and VCGS values of 23.71 dB, 0.979, and 0.990, respectively. In biomedical image applications, the proposed method achieved image quality assessment results above the average of the state-of-the-art model. These evaluation results indicate that our approach is superior compared to competing models for restoring exposure-degraded images and show that our proposed system can be applied to restoration tasks for various types of images.
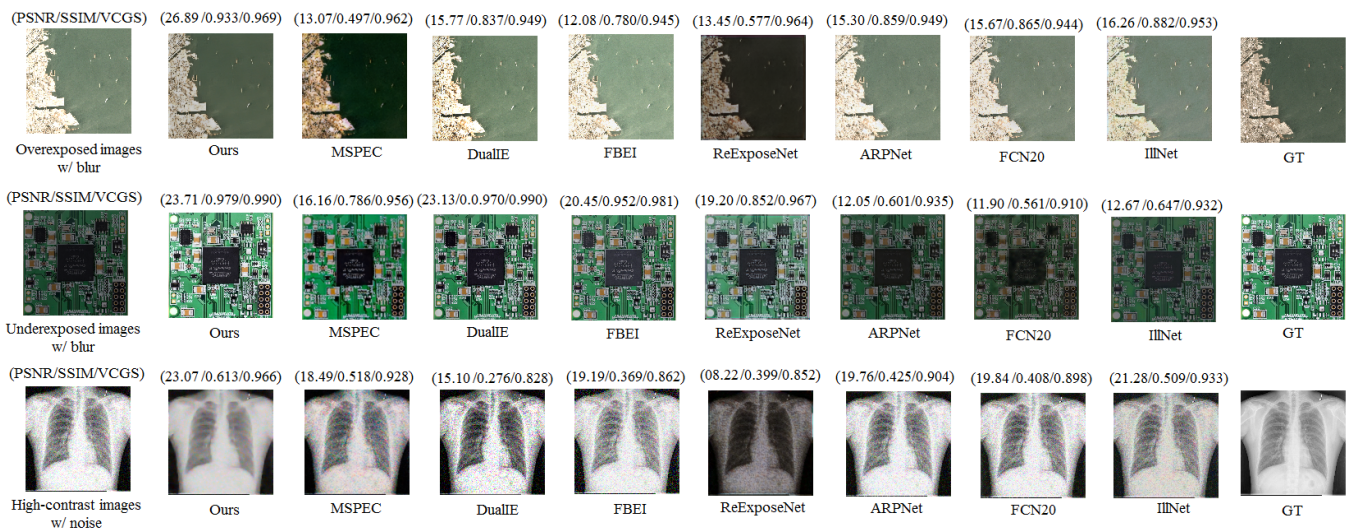

Figure 15. Qualitative comparison of all methods on degraded image samples for different application datasets. The (**top row**) shows a satellite image [54], the (**middle row**) shows an image from an industrial PCB inspection [55], and the (**bottom row**) shows an X-ray photo of human lungs in a biomedical imaging application [56].

5. Conclusions

In this study, we propose a model to improve lighting and color problems in under-exposed and overexposed images by separating the illumination correction process and the color correction process using two residual attention-based convolution models imple-

mented in the CIELab color space. The quantitative and qualitative evaluations on two synthetic degraded image datasets and one real-life exposure image dataset demonstrated the superiority of the proposed model over state-of-the-art methods. The ablation study proved the effectiveness of our model using the CIELab color space as a color channel for the input model, the utilization of self-attention layers within the network, and improved model performance when using mixed image datasets (with image samples that have noise defects and blur defects) during the training phase. In addition, testing on various types of images, such as satellite images, industrial inspection images, and biomedical images, showed that the proposed system could be applied to exposure-degraded images beyond photographic images, even without retraining the model. However, the results of this experiment could be optimized by retraining the model using additional datasets appropriate to the desired restoration application object. In the future, we plan to improve the generalizability of our method to various other image enhancement applications by adding domain adaptation or domain generalization methods.

Author Contributions: Conceptualization, N.R. and S.-F.S.; Methodology, N.R.; Project administration, S.-F.S.; Software, N.R.; Supervision, S.-F.S.; Visualization, N.R.; Writing—original draft, N.R.; Writing—review and editing, S.-F.S. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: MIT Adobe FiveK dataset can be downloaded at https://data.csail.mit.edu/graphics/fivek/, VOC2012 dataset can be downloaded at http://host.robots.ox.ac.uk/pascal/VOC/voc2012/, Afifi dataset can be downloaded at https://github.com/mahmoudnafifi/Exposure_Correction.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Payne, T. *Another Photography Book*; Adobe Education Exchange: San Jose, CA, USA, 2018; Volume 2020-7, pp. 121–122. Available online: https://edex.adobe.com/teaching-resources/another-photography-book (accessed on 12 January 2023)
2. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [CrossRef]
3. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef] [PubMed]
4. Song, Q.; Cosman, P.C. Luminance enhancement and detail preservation of images and videos adapted to ambient illumination. *IEEE Trans. Image Process.* **2018**, *27*, 4901–4915. [CrossRef] [PubMed]
5. Reinhard, E.; Stark, M.; Shirley, P.; Ferwerda, J. Photographic tone reproduction for digital images. In *Seminal Graphics Papers: Pushing the Boundaries*; Association for Computing Machinery: New York, NY, USA, 2023; Volume 2, pp. 661–670.
6. Fu, X.; Liao, Y.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Trans. Image Process.* **2015**, *24*, 4965–4977. [CrossRef] [PubMed]
7. Abdullah-Al-Wadud, M.; Kabir, M.H.; Dewan, M.A.A.; Chae, O. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [CrossRef]
8. Veluchamy, M.; Subramani, B. Image contrast and color enhancement using adaptive gamma correction and histogram equalization. *Optik* **2019**, *183*, 329–337. [CrossRef]
9. Li, C.; Tang, S.; Yan, J.; Zhou, T. Low-light image enhancement based on quasi-symmetric correction functions by fusion. *Symmetry* **2020**, *12*, 1561. [CrossRef]
10. Zhang, W.; Liu, X.; Wang, W.; Zeng, Y. Multi-exposure image fusion based on wavelet transform. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1729881418768939. [CrossRef]
11. Jung, C.; Yang, Q.; Sun, T.; Fu, Q.; Song, H. Low light image enhancement with dual-tree complex wavelet transform. *J. Vis. Commun. Image Represent.* **2017**, *42*, 28–36. [CrossRef]
12. Demirel, H.; Ozcinar, C.; Anbarjafari, G. Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition. *IEEE Geosci. Remote Sens. Lett.* **2009**, *7*, 333–337. [CrossRef]
13. Jobson, D.J.; Rahman, Z.u.; Woodell, G.A. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [CrossRef]

14. Jobson, D.J.; Rahman, Z.u.; Woodell, G.A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [CrossRef] [PubMed]

15. Li, J. Application of image enhancement method for digital images based on Retinex theory. *Optik* **2013**, *124*, 5986–5988. [CrossRef]

16. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef]

17. Ren, X.; Yang, W.; Cheng, W.H.; Liu, J. LR3M: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Trans. Image Process.* **2020**, *29*, 5862–5876. [CrossRef] [PubMed]

18. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]

19. Jiang, L.; Jing, Y.; Hu, S.; Ge, B.; Xiao, W. Deep refinement network for natural low-light image enhancement in symmetric pathways. *Symmetry* **2018**, *10*, 491. [CrossRef]

20. Li, Q.; Wu, H.; Xu, L.; Wang, L.; Lv, Y.; Kang, X. Low-light image enhancement based on deep symmetric encoder-decoder convolutional networks. *Symmetry* **2020**, *12*, 446. [CrossRef]

21. Guo, Y.; Ke, X.; Ma, J.; Zhang, J. A pipeline neural network for low-light image enhancement. *IEEE Access* **2019**, *7*, 13737–13744. [CrossRef]

22. Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognit. Lett.* **2018**, *104*, 15–22. [CrossRef]

23. Wang, R.; Zhang, Q.; Fu, C.W.; Shen, X.; Zheng, W.S.; Jia, J. Underexposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6849–6857.

24. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4225–4238. [CrossRef]

25. Gao, Z.; Edirisinghe, E.; Chesnokov, S. OEC-cnn: A simple method for over-exposure correction in photographs. *Electron. Imaging* **2020**, *32*, 1–8. [CrossRef]

26. Wang, J.; Tan, W.; Niu, X.; Yan, B. RDGAN: Retinex decomposition based adversarial learning for low-light enhancement. In Proceedings of the 2019 IEEE international conference on multimedia and expo (ICME), Shanghai, China, 8–12 July 2019; pp. 1186–1191.

27. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [CrossRef] [PubMed]

28. Ma, T.; Guo, M.; Yu, Z.; Chen, Y.; Ren, X.; Xi, R.; Li, Y.; Zhou, X. RetinexGAN: Unsupervised low-light enhancement with two-layer convolutional decomposition networks. *IEEE Access* **2021**, *9*, 56539–56550. [CrossRef]

29. Cao, Y.; Ren, Y.; Li, T.H.; Li, G. Over-exposure correction via exposure and scene information disentanglement. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.

30. Zhang, Q.; Nie, Y.; Zheng, W.S. Dual illumination estimation for robust exposure correction. *Comput. Graph. Forum* **2019**, *38*, 243–252. [CrossRef]

31. Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* **2016**, *129*, 82–96. [CrossRef]

32. Steffens, C.R.; Messias, L.R.V.; Drews, P., Jr.; Botelho, S.S.d.C. Contrast enhancement and image completion: A cnn based model to restore ill exposed images. In Proceedings of the 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki, Finland, 22–25 July 2019; Volume 1, pp. 226–232.

33. Goswami, S.; Singh, S.K. A simple deep learning based image illumination correction method for paintings. *Pattern Recognit. Lett.* **2020**, *138*, 392–396. [CrossRef]

34. Li, X.; Zhang, B.; Liao, J.; Sander, P.V. Document rectification and illumination correction using a patch-based CNN. *ACM Trans. Graph. (TOG)* **2019**, *38*, 1–11. [CrossRef]

35. Ma, L.; Jin, D.; Liu, R.; Fan, X.; Luo, Z. Joint over and under exposures correction by aggregated retinex propagation for image enhancement. *IEEE Signal Process. Lett.* **2020**, *27*, 1210–1214. [CrossRef]

36. Afifi, M.; Derpanis, K.G.; Ommer, B.; Brown, M.S. Learning multi-scale photo exposure correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 9157–9167.

37. Shen, Y.; Sheng, V.S.; Wang, L.; Duan, J.; Xi, X.; Zhang, D.; Cui, Z. Empirical comparisons of deep learning networks on liver segmentation. *Comput. Mater. Contin.* **2020**, *62*, 1233–1247. [CrossRef]

38. Cao, Y.; Liu, S.; Peng, Y.; Li, J. DenseUNet: Densely connected UNet for electron microscopy image segmentation. *IET Image Process.* **2020**, *14*, 2682–2689. [CrossRef]

39. Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4539–4547.

40. Atoum, Y.; Ye, M.; Ren, L.; Tai, Y.; Liu, X. Color-wise attention network for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 506–507.

41. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2480–2495. [CrossRef] [PubMed]

42. Bychkovsky, V.; Paris, S.; Chan, E.; Durand, F. Learning photographic global tonal adjustment with a database of input/output image pairs. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 97–104.

43. Everingham, M.; Winn, J. The PASCAL visual object classes challenge 2012 (VOC2012) development kit. *Pattern Anal. Stat. Model. Comput. Learn. Tech. Rep* **2012**, *2007*, 5.

44. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 10–15 June 2019; pp. 7354–7363.

45. Huang, Z.; Chen, Z.; Zhang, Q.; Quan, G.; Ji, M.; Zhang, C.; Yang, Y.; Liu, X.; Liang, D.; Zheng, H.; et al. CaGAN: A cycle-consistent generative adversarial network with attention for low-dose CT imaging. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1203–1218. [CrossRef]

46. Guo, M.; Lan, H.; Yang, C.; Liu, J.; Gao, F. AS-Net: Fast photoacoustic reconstruction with multi-feature fusion from sparse data. *IEEE Trans. Comput. Imaging* **2022**, *8*, 215–223. [CrossRef]

47. Jin, Z.; Iqbal, M.Z.; Bobkov, D.; Zou, W.; Li, X.; Steinbach, E. A flexible deep CNN framework for image restoration. *IEEE Trans. Multimed.* **2019**, *22*, 1055–1068. [CrossRef]

48. Wang, J.; Wang, X.; Zhang, P.; Xie, S.; Fu, S.; Li, Y.; Han, H. Correction of uneven illumination in color microscopic image based on fully convolutional network. *Opt. Express* **2021**, *29*, 28503–28520. [CrossRef]

49. Kim, B.; Jung, H.; Sohn, K. Multi-Exposure Image Fusion Using Cross-Attention Mechanism. In Proceedings of the 2022 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 7–9 January 2022; pp. 1–6.

50. Yoo, S.; Bahng, H.; Chung, S.; Lee, J.; Chang, J.; Choo, J. Coloring with limited data: Few-shot colorization via memory augmented networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11283–11292.

51. Zhang, R.; Zhu, J.Y.; Isola, P.; Geng, X.; Lin, A.S.; Yu, T.; Efros, A.A. Real-time user-guided image colorization with learned deep priors. *arXiv* **2017**, arXiv:1705.02999.

52. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

53. Shi, C.; Lin, Y. Full reference image quality assessment based on visual salience with color appearance and gradient similarity. *IEEE Access* **2020**, *8*, 97310–97320. [CrossRef]

54. Hammell, R. Ships in Satellite Imagery. 2018. Available online: https://www.kaggle.com/datasets/rhammell/ships-in-satellite-imagery/ (accessed on 20 May 2023).

55. Kuo, C.W.; Ashmore, J.; Huggins, D.; Kira, Z. Data-Efficient Graph Embedding Learning for PCB Component Detection. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019.

56. Candemir, S.; Jaeger, S.; Palaniappan, K.; Musco, J.P.; Singh, R.K.; Xue, Z.; Karargyris, A.; Antani, S.; Thoma, G.; McDonald, C.J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans. Med. Imaging* **2013**, *33*, 577–590. [CrossRef] [PubMed]