February 2023

Master's Degree Thesis

A study on brightness enhancement of old images by estimating and combining curve maps and attention-guided illumination maps

Graduate School of Chosun University

Department of Information and Communication Engineering

Arshiana Shamir

A study on brightness enhancement of old images by estimating and combining curve maps and attention-guided illumination maps

인공지능 기반 곡선맵과 어텐션-가이드 밝기맵의 추정 및 조합을 이용한 고전 영상의 밝기 향상에 관한 연구

February 24, 2023

Graduate School of Chosun University

Department of Information and Communication Engineering

Arshiana Shamir

A study on brightness enhancement of old images by estimating and combining curve maps and attention-guided illumination maps

Advisor: Prof. Bumshik Lee

This thesis is submitted to Chosun University in partial fulfillment of the requirements for a Master's degree

October 2022

Graduate School of Chosun University

Department of Information and Communication Engineering

Arshiana Shamir

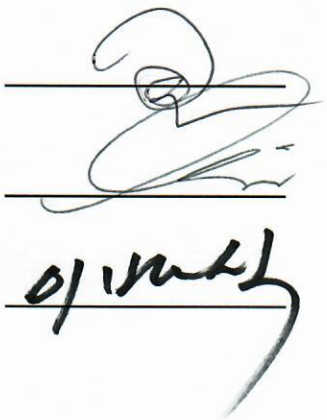
This is to certify that the master's thesis of Arshiana Shamir

has been approved by the examining committee for the thesis requirement for the master's degree in Engineering.

Committee Chairperson: Prof. Chanki Kim

Committee Member: Prof. Young-Sik Kim

Committee Member: Prof. Bumshik Lee



December 2022

Graduate School of Chosun University

Table of Contents

List of Figures

List of Tables

Abstract

A study on brightness enhancement of old images by estimating and combining curve maps and attention-guided illumination maps

Arshiana Shamir

Advisor: Prof. Bumshik Lee

Department of Information and Communication Engineering

Graduate School

Chosun University

In this paper, a novel deep-learning network for the brightness enhancement of old images is proposed. Although the task for brightness enhancement of old images is similar to the low-light enhancement problem of modern images, the cause of darkness and image characteristics are significantly different. The bright degradations in old images are mostly due to low-quality camera sensors and the harsh environment where the photos/films were stored. In contrast, the low-light condition occurs in modern images when the image is taken in dark places, or the camera's International Organization for Standardization (ISO) values are set very low. Even though the state-of-the-art low-light enhancement networks show good reconstruction capabilities in any light conditions, they produce overexposed, color distorted outputs if applied to old images and videos. A novel deep-learning network which is a combination of

image-specific curve estimation and attention-guided illumination network is introduced to overcome the above-mentioned problem. The image-specific curve estimation network estimates a curve map that automatically enhances the given input image by adjusting the dynamic range of the given input image. In addition to the curve map estimation, the illumination network with the attention module preserves the texture and color of the given input image and adjusts the given input's illumination. The attention module utilized in the illumination network helps to eliminate noises and restricts the image from getting over- or underexposed. Both networks perform in parallel to enhance the brightness of the image and preserve the inherent colors and texture details. Experimental results show that the proposed method outperforms state-of-the-art methods in terms of visual quality for brightness enhancement on old photo/video datasets.

한 글요 약

인공지능 기반 곡선맵과 어텐션-가이드 밝기맵의 추정 및 조합을 이용한 고전 영상의 밝기 향상에 관한 연구

아샤나 샤미르
지도교수: 이범식
정보통신공학과
조선대학교 대학원

본 논문에서는 오래된 이미지의 밝기 향상을 위한 새로운 딥 러닝 네트워크를 제안한다. 오래된 이미지의 밝기 향상을 위한 과제는 현대 이미지의 저조도 향상 문제와 유사하지만, 밝기 열화의 원인과 이미지 특성은 서로 크게 다르다. 오래된 이미지의 밝은 열화는 대부분 낮은 품질의 카메라 센서와 사진/필름이 저장되었던 가혹한 환경 때문이다. 반면, 현대 이미지의 밝기 저하는 어두운 곳에서 촬영되거나 카메라의 국제 표준화 기구(ISO) 값이 매우 낮게 설정될 때 주로 발생한다. 최근 발표되고 있는 최신 저조도 향상 딥러닝 네트워크는 어떤 조명 조건에서도 우수한 복원 성능을 보여주지만, 오래된 이미지 및 비디오에 적용할 경우 과도한 노출 노출과 색상이 왜곡된 출력을 발생시킨다. 위의 문제를 극복하기 위해 본 학위 논문에서는 이미지별 곡선 추정과 attention-guided illumination map 의 조합인 새로운 딥 러닝 네트워크를 제안한다. 이미지별 곡선

추정 네트워크는 주어진 입력 이미지의 동적 범위를 조정하여 주어진 입력 이미지를 자동으로 향상시키는 곡선 맵을 추정한다. 커브 맵 추정 외에도, attention-guided illumination map 은 주어진 입력 이미지의 텍스처와 색상을 보존하고, 주어진 입력의 조도를 조절하는 기능을 수행한다. attention-guided illumination netowrk 에서 사용되는 guided-module 노이즈를 제거하는 데 도움이 되며 이미지가 과노출되거나 과소노출되는 것을 제한한다. 두 네트워크는 이미지의 밝기를 향상시키고 고유한 색상과 텍스처 디테일을 보존하기 위해 병렬로 수행된다. 실험 결과는 제안된 방법이 오래된 사진/비디오 데이터 세트의 밝기 향상을 위한 시각적 품질 측면에서 기존 최신 방법을 능가한다는 것을 보여준다.

1. Introduction

1.1 Overview

Numerous studies have been done to improve the brightness of low-light images since this is a popular study area. Numerous supervised (RetinexNet [1], DeepUPE [2], KinD [3], etc.), unsupervised (EnlightenGAN [4], Zero-DCE [5], etc.), and semi-supervised (DRBN [6]) deep-learning methods have been used to enhance low-light images. Most state-of-the-art methods for improving dark images tend to enhance the image intensity values and dynamic range. Some also remove noise that appears after illumination adjustment. However, the main goal of these techniques is to enhance the brightness of modern, high-quality images that contain camera noise. Only a few methods have been suggested to increase the brightness of the old images while performing restoration or colorization of old images.

Unlike modern images, old images have smaller dynamic range, have less information concerning textures and intensity values and contain several types of degradation and noise [7, 8]. Figure. 1-1 displays a selection of the sample of images [9] and video frames from various decades. The majority of the photos in Figure. 1-1 are underexposed. The image quality of the video frames appears to be very poor. Due to the aforementioned issues, old photographs are highly sensitive to any type of enhancement [10]. The reasons for the low brightness and quality of old images are low camera performance, such as camera sensor lens and light conditions at the


1930


1940


1950


1960


1970


Rocky (1976)


Annie Hall (1977)


After hours (1985)

Howards End (1992)

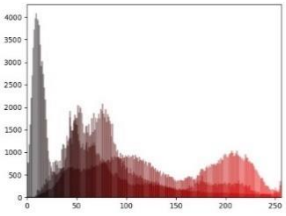

Clueless (1995)

Figure 1-1. Sample of images and video frames from different decades [9].

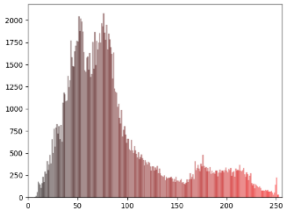
(a)Modern image

(b) Old image

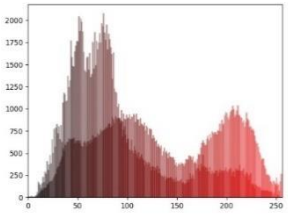
(c) Modern low-light image



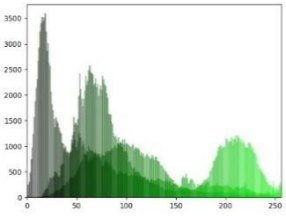
(d) Red channel histogram of (a)



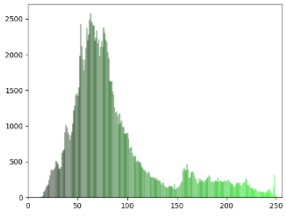
(e) Red channel histogram of (b)



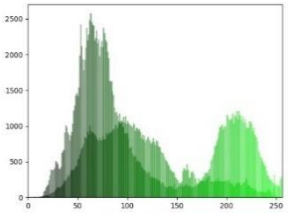
(f) Red channel histogram of (c)



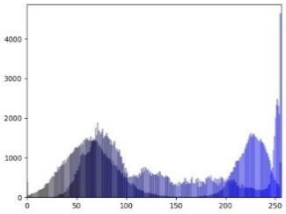
(g) Green channel histogram of (a)



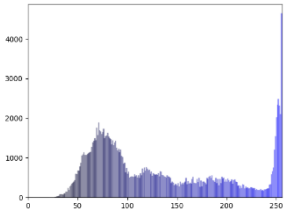
(h) Green channel histogram of (b)



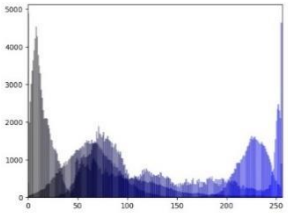
(i) Green channel histogram of (c)



(j) Blue channel histogram of (a)



(k) Blue channel histogram of (b)



(l) Blue channel histogram of (c)

Figure 1-2. Histogram comparison between old, modern and modern low-light images: (a) modern image, (b) old image and (c) modern low-light images. (d), (e), and (f) are their red channel histograms, (g), (h), and (i) are green channel histograms, and (j), (k), and (l) are blue channel histograms, respectively.

time of filming, and the environment in which films with images were stored. Figure. 1-2, shows examples of modern, old, and modern low-light images and their histograms. As shown in Figure. 1-2, the old images have smaller dynamic ranges, missing information, and intensity values compared to modern ones. Compared to old images, modern images intensity values are more densely distributed in the histogram. As old images contain less structural information, the intensity values are also less because of that. So, it is very important to give more importance to image structure and color information while doing brightness enhancement of old images.

As most state-of-the-art models for brightness enhancement are trained with modern low-light images, it is difficult to produce natural, normal bright outputs if the pre-trained supervised [1, 3, 11, 13] and unsupervised low-light enhancement networks [4, 5, 12] are applied to such old images. These networks are designed to enhance significant dark images, which is why over-exposed and color-distorted outputs are produced, as shown in Figure. 1-3. Figure. 1-3, shows the brightness enhancement results when conventional methods are applied to an old image. To obtain the results of Figure. 1-3, pre-trained conventional methods were used for testing, and the test images were taken from old films from different years, such as *A Time to Kill* (1996), *Butch Cassidy and the Sundance Kid* (1969), and *Taxi Driver* (1976). In Figure. 1-3, color distortion and overexposure were observed in most existing network outputs. Under-exposure or over-exposure was observed in the EnlightenGAN [4], Zero-DCE++ [12], and KinD++ [3] outputs. In contrast, Retinex-Net [1], RUAS [11], and

Zero-DCE [5] produce noisy artifacts and color distortion in their outputs. StableLLVE [13] exhibits an extra white-filtering effect in the output image.

The above-mentioned conventional methods are not aware of the characteristics and traits of old images. As for brightness or low-light enhancement, they mainly focus on low-light modern, high-quality images. Hence, they produce unnatural outputs for old images. The performance of the existing deoldifying methods for old images is also evaluated, as shown in Figure. 1-4. Figure. 1-4(b) shows that My Heritage [48], which offers enhancement, colorization, animation, and other functionalities for old photographs, brings a slight improvement in brightness when applied to an input image (http://www.myheritage.com). However, the other two methods (Bringing Old Photos Back to Life [10] and DeepRemaster [14]) do not show significant improvement because their networks give less attention to the low contrast of old images. Therefore, to improve the visual quality of old photographs, brightness enhancement is crucial.

Hence, enhancing the brightness of an old photograph is much more difficult than enhancing the brightness of a modern, low-light image. First, there are no paired benchmark datasets available for old image brightness enhancement. Thus, it is difficult to restore the brightness due to the lack of high-quality, low-light paired datasets to reflect the characteristics of old images. Therefore, generating old images with low brightness using generative networks such as generative adversarial

networks (GAN) is a challenging task. Second, the enhancement algorithms for old images and videos, such

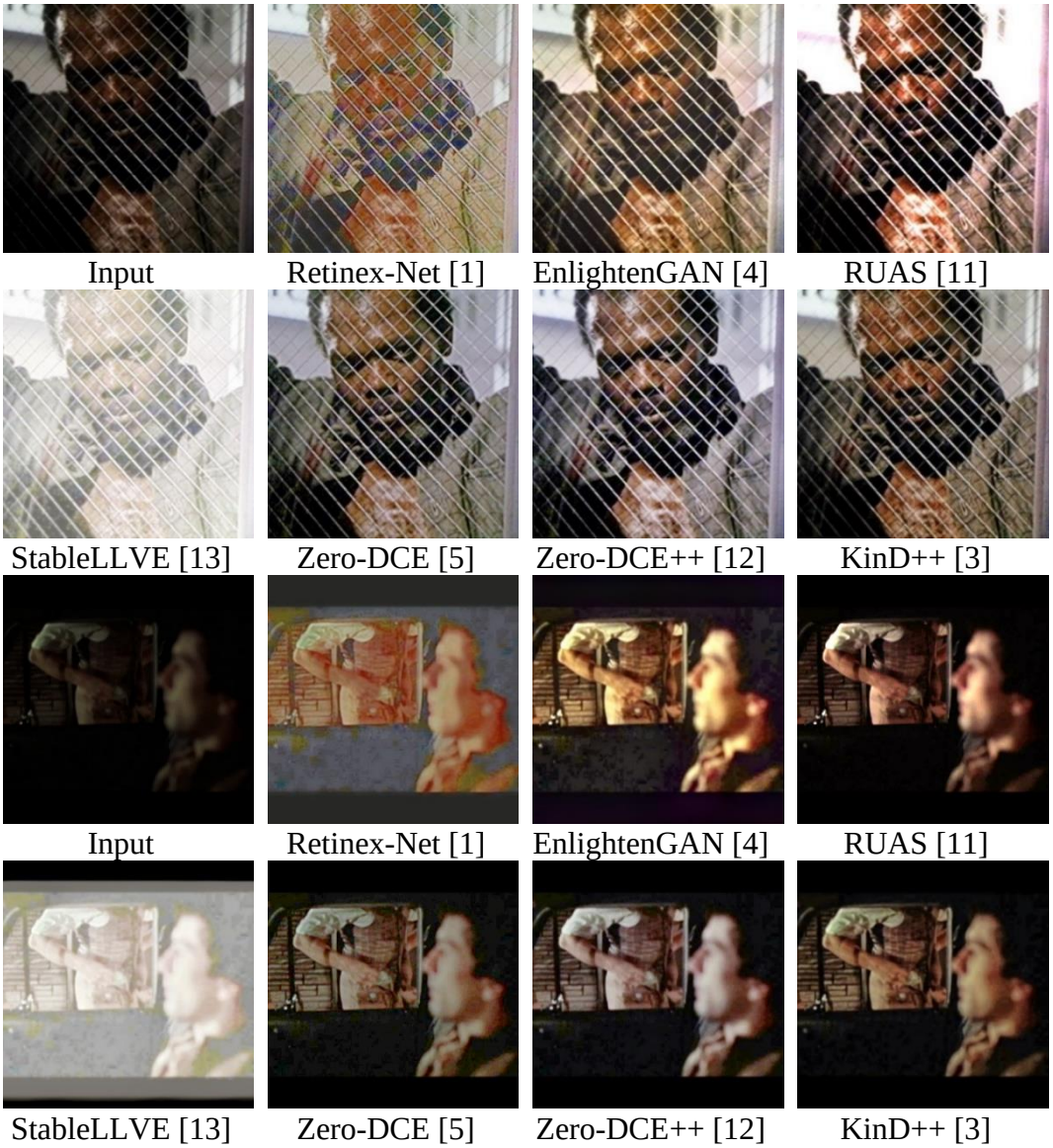


Figure 1-3. Visual comparisons of brightness enhancement for old video frames using state-of-the-art methods: Retinex-Net [1], EnlightenGAN [4], RUAS [11], StableLLVE [13], Zero-DCE [5], Zero-DCE++ [12], and KinD++ [3].

Figure 1-4. Visual comparisons of old image restoration using state-of-the-art models on video frames from *Highlander* (1986). (a) Original image, (b) My Heritage [48], (c) Bring old photos back to life [10], and (d) DeepRemaster [14].

as restoration, colorization, and brightness enhancement, may distort the color and noise of the original images because of the large amount of missing information in the old images. Thirdly, the state-of-the-art brightness enhancement methods attempt to enhance low-light (photographs obtained in low-light settings) and high-quality modern images from severe dark images to normal-light ones. However, old images do not contain the same darkness as modern low-light images. A few recent works that deal with old images have considered the low contrast of old images. DeepRemaster [14], which is a deoldifying network, considers the brightness, contrast, saturation, and other aspects of old images to generate old image or video datasets with old video and film noises. DeepRemaster achieves promising results in restoring and colorizing old images with generated datasets similar to authentic old images. However, it fails to achieve effective brightness enhancement because the proposed networks do not pay much attention to brightness enhancement rather than noise removal and restoration. In addition, datasets for brightness enhancement are not well constructed in their dataset. As shown in Figure. 1- 4, there is only a slight change in brightness when applied to an old image.

Moreover, various types of distortions and noises can be observed in old images compared to modern images. For example, old images are of low quality and missing details. Therefore, the network must support features that can preserve color and texture information while performing brightness enhancement on old images to prevent color distortion and degradation of output images. Unfortunately, conventional brightness enhancement networks have limitations because they do not

consider the above-mentioned points. Thus, most conventional brightness enhancement methods do not perform well on old images.

1.2 Research Objective

To resolve these problems, this study proposes a novel deep learning network to enhance the brightness of old images, consisting of several trainable components with a single end-to-end framework. It should be noted that the proposed algorithm is applicable to both old images and videos. For convenience, although the proposed method is targeted toward enhancing both old images and videos, here, old images are used as the target content in this study. The proposed method performs image-specific curve estimation and image-to-image mapping using an attention-guided illumination network for brightness enhancement and emerging noise and artifact removal. Both image-specific curve parameter maps and attention-guided illumination maps are obtained from the given input images, and the curve parameters and illumination maps are then used in the curve-mapping functions. The curve parameters and attention-based illumination maps are obtained so that the intensity range of the enhanced image is maintained. Those maps can also remove overexposed artifacts and produce a more natural output by increasing the dynamic range, brightness, and color distribution. In addition, RRDNet [15] is used to generate old dark and bright paired datasets for training the proposed network. For the objective of creating a dataset, old films and images from the 1960s to 2005 with various image properties, camera sensors, and storage systems were gathered. The proposed network with the generated new old dataset achieved significantly better

brightness enhancement results on old images than other conventional methods. The following is a summary of the contributions made by this study:

- This study proposes a novel deep network that combines image-specific curve estimation and attention-based image-to-image mapping to enhance the brightness of old images.
- An attention module with an illumination network is introduced to prevent noise and preserve the image information in the proposed network.
- A new curve equation that uses both the illumination and curve map values is proposed, resulting in brighter images.
- The old dark and bright paired datasets for brightness enhancement using a single image-specific state-of-the-art network were set up and used to train the proposed architecture.
- The proposed method achieves significantly better perceptual visual quality than existing state-of-the-art approaches for brightness enhancement.

1.3 Thesis Layout

There are five chapters that follow each other throughout this thesis. After the introduction, Chapter 2 provides an overview of numerous low-light image-enhancing approaches for recent images as well as some related works for old image restoration. In particular, the Deep curve estimation network and attention-guided illumination network are discussed in detail in Chapter 3 with the proposed technique and pipeline. In Chapter 4, the ablation investigation, the use of the suggested network, and a comparison of the experimental results with those of state-of-the-art networks are covered. Chapter 5 ends with a brief review of the research's findings.

2. Related Works

2.1 Low-light Enhancement Using Classical Methods

Linear adjustment to the intensity value range or nonlinear gamma correction on the inputs is the conventional and most straightforward method for low-light enhancement. Techniques based on histogram equalization (HE) have been developed to increase the dynamic range of images both globally and locally [16]–[18]. However, because of over- and underexposed outputs as well as content blindness, the enhanced quality of these methods cannot be guaranteed. One of the most widely accepted theories for low-light image enhancement (LLIE) is the Retinex theory [19, 20], which claims that an image is mainly composed of illumination and reflection. Single-scale Retinex (SSR) [21] modifies the input image illumination using the estimated illumination after utilizing the Gaussian blurred input as its illumination map. Similar to SSR, Multi-scale Retinex (MSR) [22] mixes the output of numerous Gaussian blur algorithms with distinct variances. Later, alternative approaches were suggested by AMSR [23], NPE [24], and MF [25], including a weighting technique based on SSR, balancing the enhancement level to prevent over-exposure and processing the illumination map in a multi-scale manner. The initial illumination obtained using the Max-RGB assumption and the structure-preserving constraint is improved by LIME [26]. These techniques can help low-light photographs appear brighter to some extent. They can hardly fix the noise and color distortion issues. In order to accomplish reflection recovery, SRIE [27] and RRM [28] remove the noise from the enhanced images. The application of these

optimization-based methods has decreased because of the expensive calculation, sensitivity to hyperparameters, and insufficient enhancement quality. All of these conventional techniques have issues with color distortion, underexposure, and overexposure.

2.2 Learning-based Brightness Enhancement

The low-light photos have been improved using a variety of learning techniques, including supervised learning, reinforcement learning, semi-supervised learning, and unsupervised learning. End-to-end deep Retinex-based techniques and realistic data-driven methods are frequently utilized in supervised learning-based LLIE approaches. A variation of a stacked-sparse denoising autoencoder is used in the LLNet [29] end-to-end approach to carry out brightness enhancement and denoising together. However, the connection between noise and real-world illumination has not been discussed. Thus, residual noise and over-smoothing problems stand out. Later Lv *et al*. [30] proposed an end-to-end multi-branch enhancement network (MBLLEN) that improves performance with the help of feature extraction, enhancement, and fusion modules. The same authors later proposed three other subnetworks to enhance their performance further: Illumination-Net, Fusion-Net, and Restoration-Net. DUPE [2] and GLAD [31] learn illumination maps for image restoration. Despite reasonably better results, they are not sufficiently compelling to solve noise and color distortion issues. KinD [3] and KinD++ [3] use decomposition strategies, which are helpful for both illumination adjustment and reflectance refinement. DRBN [6] developed a deep recursive band network for semi-supervised

low-light enhancement. The DLN [32] introduces a light-darken tradeoff and feature aggregation blocks to restore the results. However, they frequently have problems with overexposure and color distortion. EnlightenGAN [4] attempts to utilize larger-scale unpaired training data in unsupervised settings employing the GAN mechanism, whereas Zero-DCE [5] alternatively learns a set of non-reference loss functions. Although they relieve the requirement of paired data, they focus on the light factor, thus making an insufficient ability to remedy other defects. To represent the conditional distribution of normally exposed images, LLFlow [45] uses a conditional normalizing flow. LLFlow adjusts lighting and reduces noise and artifacts via improved characterization of structural features and better estimation of visual distance. Self-calibrated Illumination (SCI) [46], a lightweight and effective framework for low-light image enhancement in a variety of real-world circumstances, was recently introduced. The SCI brings a different concept in method design by introducing a self-calibrated module to boost the model capability of the basic unit during the training stage to improve the practicability for additional low-level vision issues in real-world settings. Despite progress in addressing the low-light problem, most state-of-the-art methods cannot deal with the brightness problems of old images and videos.

2.3 Old Image Deoldifying

Old image deoldifying includes every enhancement task for old images to make them high-quality, similar to modern ones. Deoldifying includes classical mixed degradation problems such as restoration, colorization, brightness enhancement, and

inpainting. Earlier, most existing methods for old image restoration preferred to use inpainting [32]–[36] methods, which can detect defects such as scratches and blotches and then perform inpainting by taking the textures from the neighboring pixels. However, inpainting methods do not show satisfactory results for low-contrast old images because their performance for detecting defects is not remarkably high, resulting in marginal enhancement for old images.

DeepRemaster [14] is a well-known deoldifying network that performs the restoration and colorization of vintage films. It uses attention-based temporal CNNs trained on videos with data-driven deterioration simulations. For dataset generation, high-quality images were transformed into different brightness and contrast levels to generate old images. However, the brightness enhancement of old video frames in their network architecture shows promising results compared with restoration and colorization. The network did not perform brightness enhancement well on the input video frames, even though their generated training dataset contained dark and normal old, paired images. Two variational autoencoders (VAE) are employed in [10] to convert both old and new images into two latent spaces. With sythetic paired data, the model learns how to translate between these two latent spaces. Then, with the aid of a partial nonlocal block, the mapping that transforms the damaged pictures into clean ones in the latent space is learnt. Although the method in [10] achieves good restoration performance, it exhibits some limitations in brightness enhancement.

Even though there are many existing works on the restoration, inpainting, and colorization of old images, most of them do not focus on the brightness enhancement of old images. In this study, we propose a brightness enhancement network for old images, where we estimate image-specific curves and attention-based illumination maps and use them to enhance the brightness of old images. To this end, we used low-quality old images and video frames as our training dataset rather than modern, high-quality, low-light images.

3. Proposed Old Image Brightness Enhancement Network

3.1 Dataset Generation for Brightness Enhancement

Various low-light image-enhancement datasets (LOL [1], SID [37], ExDark [38], DICM [39], etc..) are available, all of which are constructed manually by capturing paired low-light and normal-light photographs, where the two basic methods for dataset generation are taking several shots with different camera settings or editing the obtained photographs. When these datasets are used for training the existing low-light enhancement models and tested on old images color-distorted, underexposed, overexposed, and degraded images are observed in the outputs, as shown in Figure. 2. Because conventional datasets for low-light image enhancement do not carefully reflect the actual qualities of old images, there is a limitation in enhancing the old image by producing an unnatural brightness enhancement output for old images. Thus, it is important to generate new datasets for the brightness enhancement of old images by fully reflecting the qualities of old images since there are significantly different characteristics between old and modern images concerning image capture and storage technologies. Datasets for the deoldifying methods are usually generated by synthesizing old videos for remastering old images. To generate the datasets in [15], modern images are degraded by adding historical image artifacts and noise, such as grain noise, dust, scratches, and fractal noise. As a result, software like Adobe After Effects is used to produce old image artifacts. To create dark photos, tone curves, brightness, and contrast are adjusted. However, synthesizing old images

may not produce actual old image characteristics, which results in a low brightness enhancement performance.

Unlike the datasets for conventional deoldifying methods, the proposed brightness-enhancement method utilizes actual old images that fully reflect the features of old images. Real old images allow the system to learn the actual characteristics of old images compared with synthesized images. Because it is difficult to obtain the ground truth in the task of bright enhancement for old images, the ground truths were generated using RRDNet [15] in our proposed method. RRDNet is an image-specific unsupervised CNN used to restore underexposed images, not requiring any prior image samples or training. Instead, internal image optimization is used to ensure generalization over a wide range of shooting scenarios and lighting conditions. It can perform restoration, reduce noise, and improve brightness. However, RRDNet cannot be utilized in real-time because it performs both training and testing simultaneously when a single image is provided. The above-mentioned network is utilized to create paired training datasets because it produces more realistic and bright outputs with only old input images. Old video frames were selected for dataset generation. Various old dramas or TV shows from different decades, such as *Brief Encounter* (1945), *Hogan Heroes* (1965), *Butch Cassidy and Sundance Kid* (1969), *Death Wish* (1974), *After Hours* (1985), and *My Girl* (1991) are involved in the training and testing dataset. The videos were converted into separate frames, and then frames with low contrast were selected. Each selected frame is trained and tested individually using RRDNet, and the final outputs of each individual frame are

considered the ground truth of the corresponding frames. To train the proposed network for old images (BOLD-Net), 1560 frame pairs with dark and bright images were used. Then, the network is tested using 200 frame pairs. Because each video has a different size (e.g., 1920×1080, 640×480, and 720×480), each frame was divided into small random patches of size 256×256 so that the datasets had the same resolution.

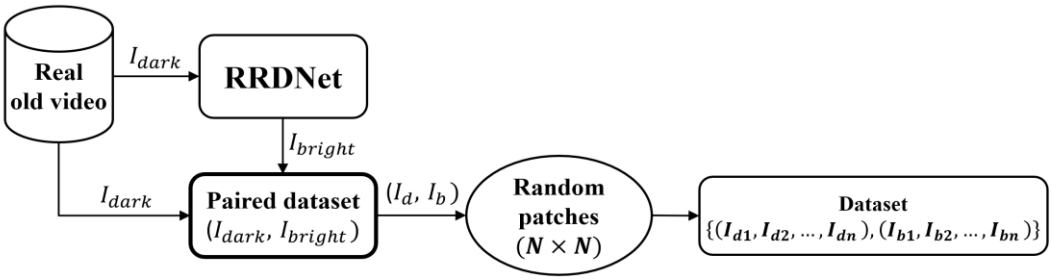
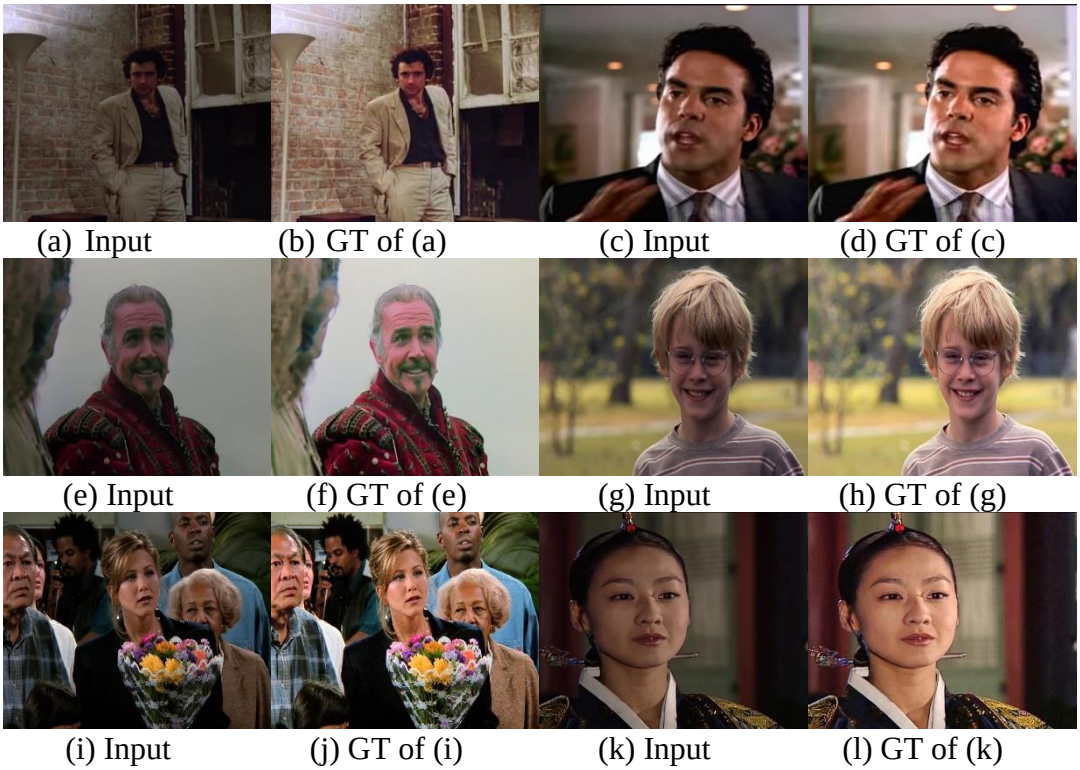


Figure 3-1. Process of generating datasets using RRDNet [15] for the proposed BOLD-Net.

Figure. 3-1, shows the dataset generation process using the RRDNet for our proposed brightness enhancement method for the old images, where RRDNet was trained using the selected old video frames. Each dark frame (I_{dark}) is enhanced using the RRDNet to obtain the corresponding enhanced frame (I_{bright}). The N×N random patches are extracted from paired I_{dark} and I_{bright} frames at the same location, I_{di} and I_{bi}, creating the training batches {(I_{d1}, I_{d2}, I_{d3}, … , I_{dn}), (I_{b1}, I_{b2}, I_{b3}, … , I_{bn})}. Figure 3-2, shows examples of the datasets generated using the dataset generation process illustrated in Figure. 3-1. It is observed that our datasets contain realistic

dark-bright dataset pairs without overexposed, underexposed, or color-distorted outputs, as shown in Figure. 3-2.



(a) Input (b) GT of (a) (c) Input (d) GT of (c)

(e) Input (f) GT of (e) (g) Input (h) GT of (g)

(i) Input (j) GT of (i) (k) Input (l) GT of (k)

Figure 3-2. Examples of training datasets using RRDNet [15] for the proposed brightness enhancement network.

3.2 Network Architecture of BOLD-Net

Unlike low-light modern images for low-light enhancements, old images are not extremely dark but rather have lower contrast. As shown in Figure. 1-2, color distortion and information missing from old images are observed, even though some conventional methods can avoid overexposure, resulting in noise and artifacts. To overcome these problems, this study proposes a brightness enhancement network

that can brighten an image while preserving the colors and feature information of old images. The proposed network consists of two main subnets: one for the curve map estimation and the other for the illumination map estimation.

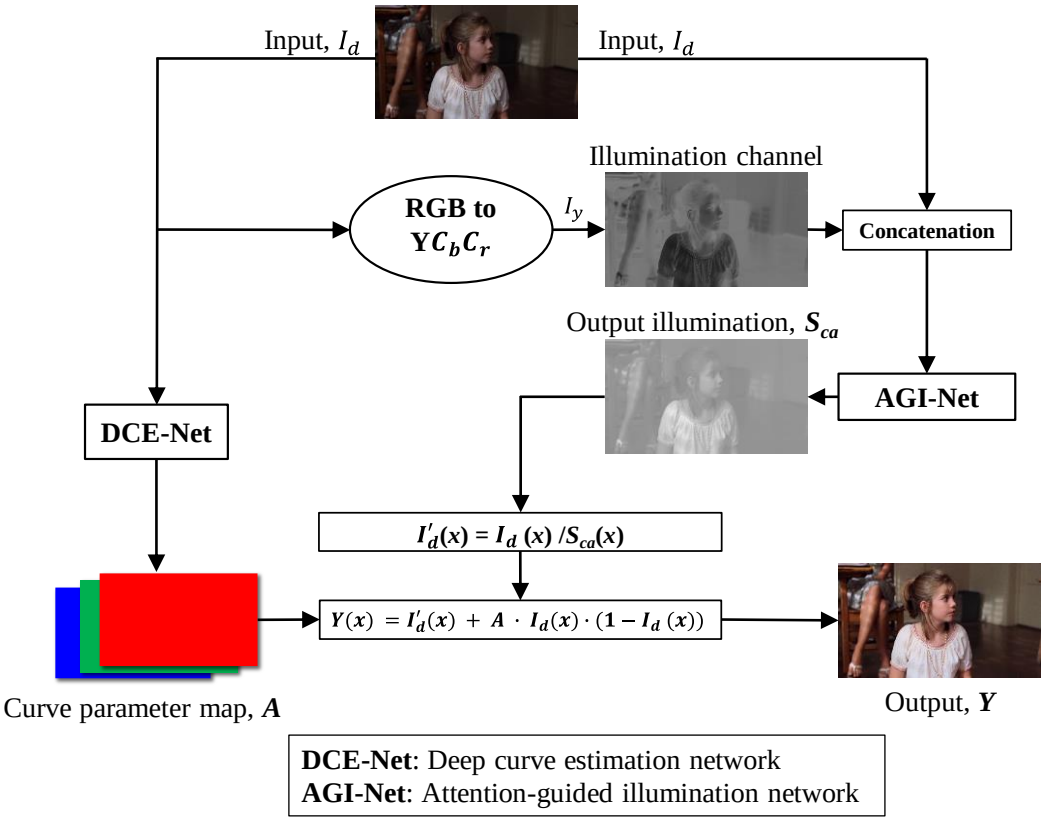


Figure 3-4. The overall architecture of the proposed brightness enhancement network.

Figure. 3-4 shows the overall network architecture of the proposed BOLD-Net. The DCE-Net maps all pixels of the RGB channels of the input by applying the curve equation to obtain the curve-adjusted image. An illumination map from an attention-

guided illumination network (AGI-Net) is used to adjust the input image. Both curve and illumination-adjusted images are added to obtain the final output image. This process helps in achieving a natural and bright image.

3.2.1 Self-adaptive Curve Equation for Brightness Enhancement

In general, color (RGB) image I of size $H \times W$ (e.g., 960×480) is a set of $H \times W \times 3$ pixels and defined as (1).

$$I = \{I(x,y,c) | I(x,y,c) \in [0,1], \\ x \in 0,\dots,H, y \in 0,..,W, c \in \{R,G,B\}\}, \tag{1}$$

The dynamic range adjustment of the image is performed by adjusting the values of each pixel $I(x,y,c)$. $F(x,y,c)$ is a function for optimizing the pixel values at (x,y,c), and $G(x,y,c)$ is a function for adjusting the illumination of a given input image pixel $I(x,y,c)$. This equation is expressed as (2).

$$I'(x,y,c) = G(x,y,c) + F(x,y,c), \\ x \in \{0,\dots,H\}, y \in \{0,\dots,W\}, c \in \{R,G,B\}, \tag{2}$$

where $I'(x,y,c)$ represents the final enhanced image. $F(x,y,c)$ is a self-adaptive curve that can translate a low-light image to its improved counterpart. The self-adaptive curve has to be designed in such a way that The enhanced image's pixel values should all fall between [0,1], and the curve should be monotonous to preserve neighboring pixel differences [5]. In addition, the curve should be differentiable to perform gradient backpropagation. Hence, to satisfy the aforementioned objectives, a quadratic curve was designed as (3) [5].

$$F(x,y,c) = A(x,y,c) \cdot \left(I(x,y,c) - I^2(x,y,c)\right)$$
$$where\ A(x,y,c): \mathbb{R}^{H\times W\times c} \rightarrow\ [-1,1], \tag{3}$$

where $F(x,y,c)$ is a curve in the form of a quadratic polynomial $(I(x,y,c) - I^2(x,y,c))$ and $A(x,y,c)$ is a trainable curve parameter with a constant value between −1 and 1, which can change the curve's magnitude and the exposure degree. All actions are performed pixel-by-pixel, with each pixel normalized to [0, 1].

Adding the input image with $F(x,y,c)$, results in only a slight improvement in brightness with noise suppression. Hence, to obtain a better brightness enhancement, an illumination adjustment to $F(x,y,c)$ was added. The illumination adjustment $G(x,y,c)$ provides an illumination-adjusted image. In Retinex theory, it is known [19, 20] that an underexposed image $I(x,y,c)$ can be decomposed into reflection and illumination. If the one-channel illumination map is $S(x,y,1)$, $G(x,y,c)$ can be expressed as (4).

$$G(x,y,c) = \frac{I(x,y,c)}{S(x,y,1)},$$
$$x \in 0,\dots,H, y \in 0,..,W, c \in \{R,G,B\}, \tag{4}$$

Note that $F(x,y,c)$ performs the adjustment of pixel values, whereas $G(x,y,c)$ helps to enhance the illumination of an image. The final proposed curve equation is expressed as (5).

$$I'(x,y,c) = \frac{I(x,y,c)}{S_{ca}(x,y,1)} + A(x,y,c) \cdot \left(I(x,y,c) - I^2(x,y,c)\right), \quad x \in 0,\ldots,H, y \in 0,..,W, c \in \{R,G,B\}, \tag{5}$$

where $I(x,y,c)$ is the input image, $S_{ca}(x,y,1)$ is the one-channel attention-guided illumination map, and $A(x,y,c)$is the three-channel curve-parameter map. The basic curve map parameter equation in (5) proposed in [5] was used with higher-order curves with a few iterations. However, the higher-order curve equation in [5] for old images results in critical drawbacks, such as an overexposed output with a whitening effect. To address these problems, a more straightforward curve estimation equation is proposed in our method, as shown in (5), which can suppress the noise while maintaining the brightness enhancement performance. For further brightness enhancement, an illumination adjustment $[I(x,y\,c)/S_{ca}(x,y,1)]$ with an attention-guided illumination map was performed, unlike in [5], where the input image was used without any adjustment. The attention-guided illumination map (S_{ca}) obtained from the Attention-Guided Illumination Network (AGI-Net) adjusts the illumination of the input image. The new curve equation yields more natural bright images.

3.2.2 DCE-Net

As shown in Figure. 3-5, to understand the relationship between an input picture and its best-fitting curve parameter maps, a deep curve estimation network (DCE-Net) is presented. DCE-Net generates three-channel pixel-wise fitting curve parameter maps from a low-light input image. It's an image-specific network. The proposed DCE-Net contains nine convolutional blocks in total. The number of features increasing to 256 until the fifth convolutional blocks and then returning to three at the ninth convolutional block. Depth-wise separable convolutions [40] are employed in each convolutional block to minimize the amount of network parameters while ensuring performance. The input image is concatenated with the output of the ninth convolutional block to avoid information loss owing to many feature extractions. Using a basic convolution layer, the concatenated maps were reduced to a three-channel map. The tanh activation function was then used to form three curve parameter maps for the three input channels. Tanh activation helps adjust the curve magnitude and the exposure level. The output three-channel maps are of range [-1, 1].

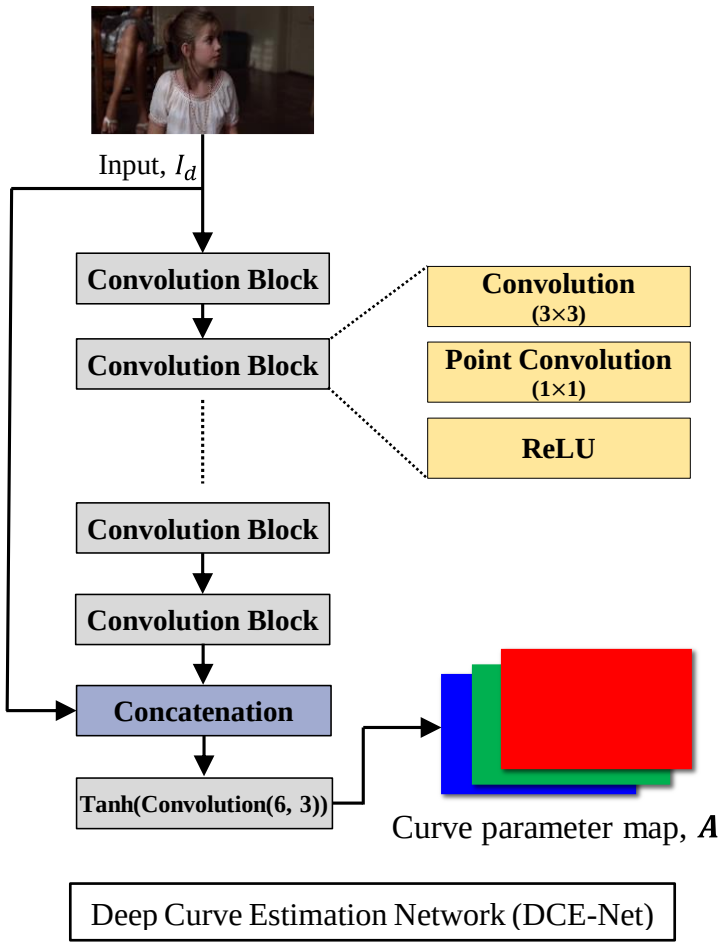


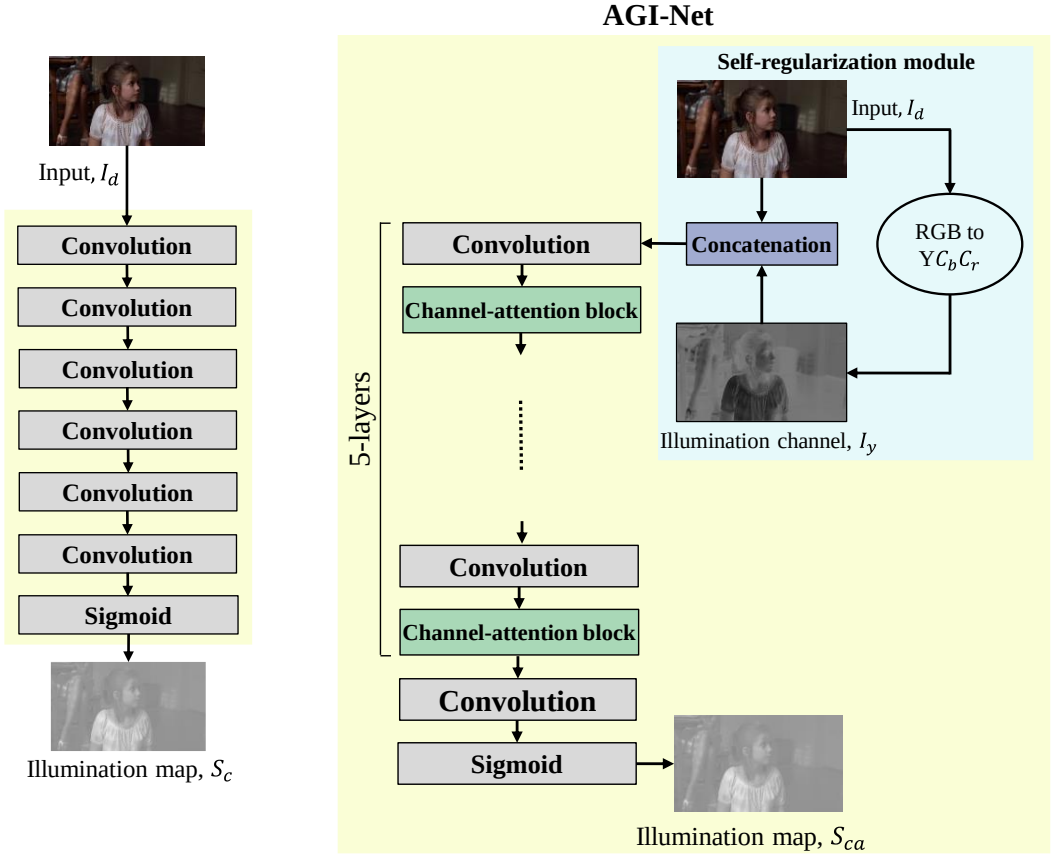
Figure 3-5. Curve estimation network with convolutional blocks and depth-wise separable convolutions.

3.2.3 AGI-Net

According to Retinex theory [19, 20], an image can be decomposed into a combination of illumination and reflection. Assuming that S is the source image, it can be expressed as in (6) according to the Retinex theory.

$$S = R \circ I \tag{6}$$

where R and I represent the reflectance and illumination, respectively, and $\circ$ is an element-wise multiplication operator. It is noted that reflectance (R) represents the inherent quality of an image regardless of luminance, and illumination (I) is the variable brightness of objects in an image. Darkness and unequal illumination distributions are common problems in low-contrast images. In general, to brighten the illumination (I), an illumination network is used with the constraint that low- and normal-light images have the same reflectance. Despite the good performance in brightness enhancement, conventional methods result in some drawbacks in that noise is also enhanced in the resulting images. To address this issue, in AGI-Net, implementation of an attention mechanism is done that enables feature recalibration. and learn to selectively highlight informative features while suppressing less relevant ones (such as noise) using global information.



(a) Simple Illumination Network

(b) Attention-guided Illumination Network (AGI-Net)

Figure 3-6. Illumination Networks: (a) Simple Illumination Network, (b) AGI-Net

Figure. 3-6-(b), shows the proposed AGI-Net. First, the RGB image $I_{RGB}(x, y, 3)$ was converted to $I_{YC_bC_r}(x, y, 3)$ to obtain the luminance channel. Second, to obtain a self-regularized attention map, only the luminance channel $I_Y(x, y, 1)$ was normalized to [0, 1], and the element-wise difference was performed with $(1 - I_Y(x, y, 1))$. Finally, the self-regularized attention map was concatenated with the

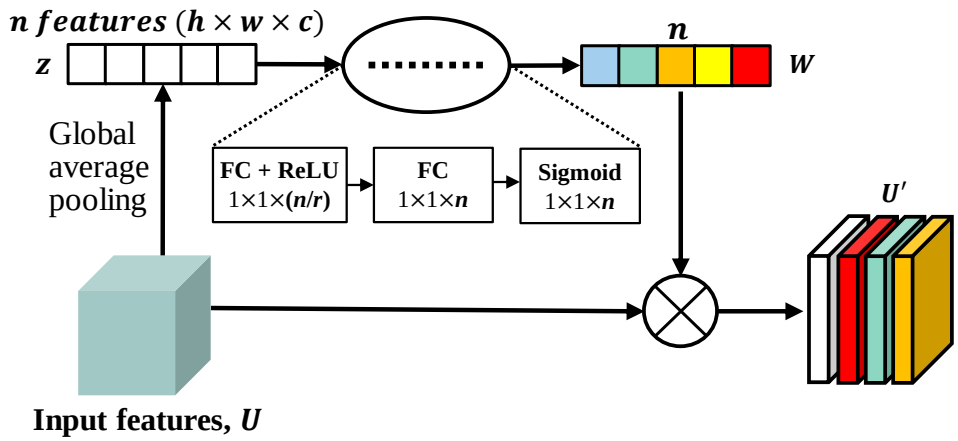


Figure 3-7. Channel-Attention Block used in AGI-Net.

input image $I_{RGB}(x, y, c)$ at the input stage of our AGI-Net. AGI-Net includes a series of five convolutions and channel-attention block pairs. Our goal is to enhance the dark areas of an old image more than the bright ones so that the resulting image is not over- or underexposed. To highlight valuable information (especially dark areas) from the input features [Figure. 3-7], we adopted a channel-wise attention block motivated by squeeze-and-excitation [41] that is placed after each convolutional layer, except the final layer. To obtain a channel-wise descriptor, global average pooling on the output of the convolutional layer is applied. Channel-wise descriptors are used to encode the global distribution of channel-wise feature responses, enabling all layers to access information from the network's global receptive field. If U is the output from the convolution layer, z_c, the channel-wise descriptor is generated by shrinking U through its spatial dimensions $H \times W$, such that the c-th element of z is calculated by

$$z_c = \frac{1}{(H \times W)} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} u_c(i,j) \tag{7}$$

where u_c is the c-th channel of U. H and W refer to the height and width of U, respectively. Furthermore, z_c passes through a fully connected (FC) layer with ReLU, followed by an FC layer without ReLU. Finally, the sigmoid activation function was used to obtain channel-wise weights W as in (8).

$$\begin{gathered} W = \sigma(W_2 \times \delta(W_1 \times z + b_1) + b_2), \\ W_1 \in R^{n/r \times n}, W_2 \in R^{n \times n/r}, b_1 \in R^{n/r}, b_2 \in R^n \end{gathered} \tag{8}$$

where σ refers to the sigmoid function and δ is the ReLU function. $W_1 \in R^{n/r \times n}$, $W_2 \in R^{n \times n/r}$, $b_1 \in R^{\frac{n}{r}}$, and $b_2 \in R^n$ are the weights and biases of the FC layers, respectively, and r is the reduction ratio. By rescaling u_c channel of U with activation w_c of W, the block's final output for each channel is achieved.

$$u'_c = u_c \cdot w_c \tag{9}$$

The output of the channel attention block U' is the input to the next convolutional layer. This process is repeated for five convolutional layers. After the sixth convolution layer, σ is used to ensure an intensity range of [0, 1]. Hence, the final output obtained from our illumination network is a one-channel illumination map $S_{ca}(x, y, 1)$: $\mathbb{R}^{H \times W \times 1} \rightarrow [0, 1]$. Despite the simplicity of the attention guidance module, it controls the noise that can emerge because of illumination map

adjustment. It also improves the visual quality consistently and the PSNR, SSIM, and LIPIP values. Using the proposed AGI-Net, three major problems in brightness enhancement were resolved with the help of the channel-wise attention module: 1) adjusting the illumination map to brighten the input old image, 2) enhancing the dark regions more than bright regions, and 3) preventing the emergence of noise while tweaking the illumination map.

3.3 Loss Functions

3.3.1 $\mathcal{L}_2$ Loss

Both $\mathcal{L}_1$ and $\mathcal{L}_2$ loss functions can be used to minimize the difference between the output and ground-truth images in our proposed network. We empirically observed that the slope of the prediction line changes significantly when $\mathcal{L}_1$ loss function is used. Furthermore, the network starts to provide an unsatisfactory output with a large amount of noise and artifacts for $\mathcal{L}_1$ loss. Consequently, we selected $\mathcal{L}_2$ loss, which does not lose image information and degrades the output image after a few epochs. The loss function $\mathcal{L}_2$ is defined in (10).

$$\mathcal{L}_2 = \frac{1}{H \times W \times c} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \sum_{c \in [R,G,B]} |I'(x,,y,c) - Y(x,y,c)\,|^2 \qquad (10)$$

where $I'(x,,y,c)$ is the predicted image, and $Y(x,y,c)$ is the ground-truth image of the input image $I(x,,y,c)$.

3.3.2 Illumination Smoothness Loss

To maintain the monotonicity correlations between adjacent pixels, the loss of illumination smoothness was estimated. [5]. The illumination smoothness loss $\mathcal{L}_{\text{tv}\mathbb{A}}$ (total variation in $\mathbb{A}$ map) is defined as in (11).

$$\mathcal{L}_{tv\mathbb{A}} = \sum_{c\in x,y,c} (|\nabla_x \mathbb{A}^c| + |\nabla_y \mathbb{A}^c|)^2 \tag{11}$$

where, ∇_x and ∇_y respectively, stand for the horizontal and vertical gradient operations.

3.3.3 Total Loss

The total loss in our proposed architecture is expressed as (12).

$$\mathcal{L}_{total} = \mathcal{L}_2 + w_{tv_{\mathbb{A}}} \mathcal{L}_{tv\mathbb{A}} \tag{12}$$

where $w_{tv_{\mathbb{A}}}$ is the weights of the loss.

4. Experimental Results

4.1 Implementation Details

Training and testing datasets: A total of 11 distinct video clips from different decades for training, such as *Brief Encounter* (1945), *Hogan Heroes* (1965), *My Girl* (1991), *After Hours* (1985), *Death Wish* (1974), *Highlander* (1986), and so on, were selected for the datasets. In addition, eight films from various decades were selected for testing, including *Butch Cassidy* and *Sundance Kid* (1969), *Taxi Driver* (1976), *The Color of Money* (1986), *Howards End* (1992), and *Clueless* (1995). First, original videos were collected and converted into separate frames. Diverse scenes from these video frames were chosen; 1560+ frames were used for training and 200 frames for testing. As previously mentioned, the RRDNet [15] was used to generate the ground truth for the experiments. We randomly cropped the training dataset frames into small patches of size 256×256 to make the training faster and easier.

Training details: PyCharm was used to implement the suggested framework, and an NVIDIA GeForce RTX 3090 (max. clock speed 1.400 MHz) was used for training. A batch size of 4 was applied. Each layer's filter weights were initially set up with a Gaussian function with a standard zero mean and 0.02 standard deviation. The bias was initialized as a constant. An The network optimization was carried out using an ADAM [42] optimizer with default settings and a fixed learning rate of 10-4. The weight $W_{tv_{\mathbb{A}}}$ was set to 1000 to balance the scale of the losses. In the channel-wise attention module, a reduction ratio of 4 was maintained.

4.2 Performances and Comparisons

Most state-of-the-art networks are trained on high-resolution, low-light images. Due to the differences in the characteristics of old and new images, pre-trained brightness enhancement models with modern low-light images for enhancing old images can lead to distorted, overexposed, and underexposed outputs, as shown in Figure. 1-2. Here, the generated dataset is used to train some of the most popular state-of-the-art networks accessible with training codes with suggested parameter settings and then evaluated the final trained model with testing dataset to make a fair quantitative and qualitative comparison, except for KinD++. For KinD++ the pre-trained model is used as it was built to conduct brightness enhancements of images with different levels of darkness. The lighting adjustment ratio was set as 3. For comparison, the following conventional approaches were included:

Histogram Equalization (HE): One of the conventional strategies for brightness enhancement is histogram equalization.

Retinex-Net [1]**:** In Retinex-Net, the illumination map is estimated using the Retinex theory and CNN, and the low-light image is enhanced by modifying the illumination map.

Zero-DCE [5]**:** Using a deep curve estimating network with zero references, Zero-DCE estimates distinct curves and uses those curves to improve low-light images.

Zero-DCE++ [12]**:** Zero-DCE++ is an upgraded version of Zero-DCE that attempts to improve the complexity of the method.

KinD++ [3]**:** KinD++ is identical to Retinex-Net, except during the decomposition stage it links feature-level illumination with reflectance.

LLFlow [46]: LLFlow models the conditional distribution of normally exposed pictures using a conditional normalizing flow.

SCI [47]: SCI uses a self-calibrated module for illumination learning with weight sharing to ensure fast, flexible, and robust output.

Quantitative comparisons were performed using the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM) [43], and the learned perceptual image patch similarity (LPIPS) [44] image quality matrix. First, an evaluation was performed on the generated dataset for qualitative and quantitative comparisons. Second, a visual comparison between state-of-the-art methods and the suggested methods was performed.

4.2.1 Quantitative Comparison

Different models were validated on old video frames using three evaluation metrics for comparison. First, the low-level differences between the predicted output and ground truth were analyzed using the PSNR and SSIM. Second, the LPIPS metric was also measured, which evaluates the distance between the image patches. Higher LPIPs values indicate further differences, and lower LPIPS values indicate more similarities. Thus, smaller LPIPS values indicate better performance. Table 1 shows a quantitative comparison between the proposed BOLD-Net and state-of-the-art networks in terms of PSNR, SSIM, and LPIPS. From Table 1, the proposed BOLD-

Net achieves the best values for PSNR, SSIM, and LPIPS. LLFlow shows the second-best PSNR, SSIM, and LPIPS values for the testing dataset. In addition, both Zero-DCE [5] and Zero-DCE++ [12] yield the lowest PSNR, SSIM, and LPIPS values.

PSNR: The PSNR determines the PSNR ratio between two images in decibels. This ratio is frequently used to assess the quality between the Ground Truth and predicted images. The quality of the predicted image will be higher when PSNR is higher.

SSIM: Structural Similarity Index Measure, sometimes known as SSIM, is an abbreviation for a method of measuring the similarity of two images using luminance, contrast, and. structure. These characteristics are used to compare images as they correspond to how the actual human visual system sees the images.

LPIPS: Using the Learned Perceptual Image Patch Similarity(LPIPS), two images' perceptual similarity is measured. In essence, LPIPS determines how comparable two image patches' activations are for a given network. It's been demonstrated that this metric closely matches human perception. Image patches with a low LPIPS score are perceptually similar.

Table 1. Quantitative comparison between the proposed BOLD-Net and state-of-the-art networks in PSNR, SSIM, and LPIPS

Methods	PSNR	SSIM	LPIPS
HE [45]	15.00	0.7011	0.1855
Retinex-Net [1]	27.45	0.9389	0.1299
Zero-DCE [5]	12.74	0.6676	0.1630
Zero-DCE++ [12]	12.17	0.6607	0.1724
KinD++ [3]	25.34	0.9222	0.1418
LLFlow [46]	31.90	0.9863	0.0252
SCI [47]	22.69	0.9495	0.0590
Proposed BOLD-Net	**35.24**	**0.9911**	**0.0218**

4.2.2 Visual Comparison

To validate and compare the performance of the proposed BOLD-Net and other methods from a visual perspective, experiments were performed on classic videos from different decades. Figure 10 displays the visual outcomes and a comparison with latest low-light enhancement techniques.

In Figure. 4-1, the video frames were taken from *Clueless* (1995), *Taxi Driver* (1976), *The Color of Money* (1986), *A Time to Kill* (1996), and *Howards End* (1992). The state-of-the-art networks, except for HE and KinD++ [3] were trained with the generated old dark and bright datasets. As shown in the figure, the results for HE

shows an overexposed output, and the whitening effect is observed in both Zero-DCE [5] and Zero-DCE++ [10] outputs, resulting in low PSNR, SSIM, and LPIPS values. While Kind++ [3] produces relatively better-enhanced results, noise artifacts are observed in the output, SCI [47] produces over-exposed output for white color. Although Retinex-Net [1] shows a better result in removing noise and preserving color than the others, it is overly smooth and produces less bright images than the proposed BOLD-Net. LLFlow [46] shows the comparable output of the proposed network yet a slightly darker result than the proposed one. The proposed BOLD-Net achieves a more natural and bright output.

Furthermore, noise after illumination enhancement is not observed. In summary, the proposed network outperformed the other methods in terms of quality and quantity metrics.

Input

Retinex-Net [1](21.24dB/0.7443)

Zero-DCE [5] (11.99dB/0.4840)

Zero-DCE++ [12] (11.80dB/0.5038)

KinD++ [3] (20.57dB/0.7063)

LLFlow [46] (32.10dB/0.9847)

SCI [47] (20.89dB/0.9817)

Proposed (40.35dB/0.9953)

Reference GT

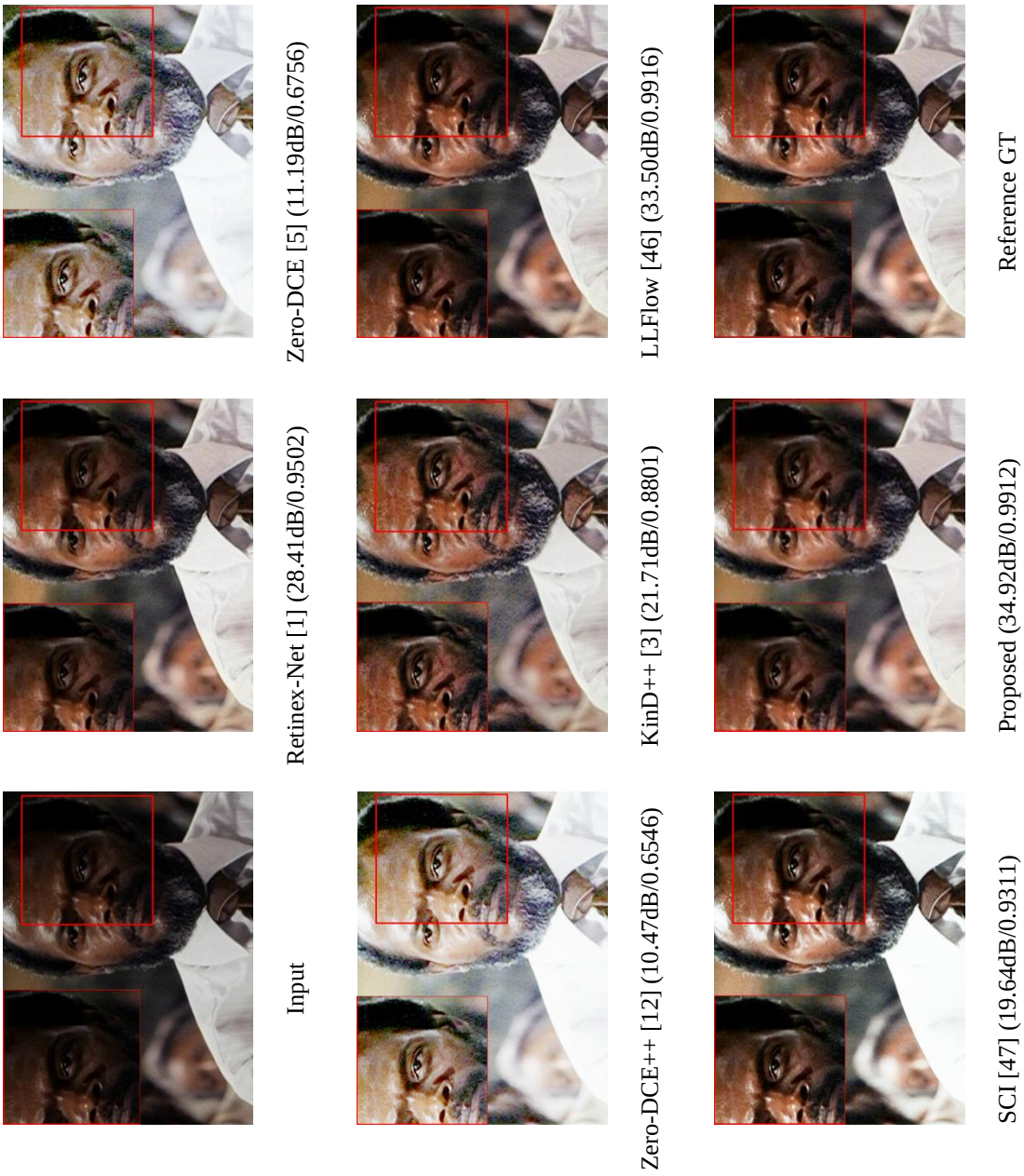


Figure 4-1: Qualitative comparison between our proposed network with state-of-the-art networks trained with our dataset.

4.3 Ablation Study

To investigate the efficiency of particular aspects of the proposed architecture, an ablation study, as in Table II was performed. Four versions by varying the modules and loss functions were tested on test dataset. The baseline in Table II indicates a direct comparison between the input image and ground truth. Method (a) used only DCE-Net without AGI-Net, and $\mathcal{L}_2$ loss and illumination smoothness loss were used. Method (b) was set up to compare the results between DCE-Net and Illumination-Net without the Attention Module (AM). In method (b), the illumination smoothness loss was not used because it was used for the DCE-Net. Method (c) was evaluated to investigate the performance of the combination of DCE-Net and Illumination-Net without AM. Finally, method (d) includes every component. Once all four networks completed 1000 epochs, the PSNR, SSIM, and LPIPS values were obtained for each network in the testing dataset. Figure. 4-2 shows the resulting images (*The Color of Money* (1986), *Dead Poets Society* (1989)) for all the methods in Table 2. Because method (a) includes only a curve estimation network without an illumination network, the final output is obtained using (13).

$$I'(x,y,c) = I(x,y,c) + A(x,y,c) \cdot \left(I(x,y,c) - I^2(x,y,c)\right) \\ where\ A(x,y,c): \mathbb{R}^{H\times W\times c} \rightarrow [-1,1] \tag{13}$$

Table 1. Ablation study comparison: (a) Baseline (only input image), (b) Only DCE-Net, (c) Only illumination network without any attention module, (d) DCE-Net with illumination network without any attention module, and (e) Curve estimation network with attention-based illumination network (proposed).

Methods	Proposed BOLD-Net					PSNR [dB]	SSIM	LPIPS
	Network			Loss functions				
	DCE-Net	Illumination-Net w/o AM	AGI-Net	$\mathcal{L}_2$ loss	ISL			
Baseline	✖	✖	✖	✖	✖	17.03	0.7595	0.1126
(a)	✓	✖	✖	✓	✓	30.49	0.9670	0.0366
(b)	✖	✓	✖	✓	✖	29.52	0.9740	0.0477
(c)	✓	✓	✖	✓	✓	29.78	0.9752	0.0432
(d)	✓	✖	✓	✓	✓	35.24	0.9911	0.0218
AM: Attention module, ISL: Illumination smoothness loss								

In method (a), the network adjusts the curves pixel-wise and produces an enhanced output with high PSNR, SSIM, and LPIPS values. Although method (a) could effectively control evolving artifacts, the effectiveness of the brightness enhancement was not significant, as shown in Figure. 4-2. For method (b), the network was trained to adjust the illumination of the input image through a simple illumination network [Figure. 3-6-(a)], where the AM was not used in this combination, as in (14).

$$I'(x,y,c) = \frac{I(x,y,c)}{S_c(x,y,1)} \qquad (14)$$
$$x \in 0,\dots,H, y \in 0,..,W, c \in \{R,G,B\}$$

where $S_c(x,y,1)$ is the illumination map obtained by plain convolution operations. Although the illumination map without AM can lead to bright outputs, it also generates output with noise and new artifacts. The results of methods (a) and (b) show that the illumination network produces a higher brightness with slightly higher noise than DCE-Net. Without AM, method (b) could produce a similar bright output but with artifacts. In method (c), a DCE-Net and simple illumination network [Figure. 3-6-(a)] combination without an AM was tested. The final input image is obtained using (15).

$$I'(x,y,c) = \frac{I(x,y,c)}{S_c(x,y,1)} + A(x,y,c) \cdot \left(I(x,y,c) - I^2(x,y,c)\right) \qquad (15)$$

(a) (b) (c)

(d) (e)

(f) (g) (h)

(i) (j)

Figure 4-2. Ablation study: (a) Input Image, (b) Only DCE-Net, (c) Only illumination Network without any attention module, (d) DCE-Net with illumination network without any attention module, (e) Curve estimation network with attention-based illumination network (proposed) and the same for (f), (g), (h), (i), (j), respectively.

Combining DCE-Net and a simple illumination network shows an enhanced output with more brightness and fewer artifacts than method (b). However, the PSNR value of the method (c) is still less than that of method (a) because it still contains staircase artifacts. It can be inferred that only the combination of DCE-Net and the illumination network does not significantly affect the quality of the output image. Finally, all the components were included in method (d). Method (d) outperformed the other combinations in all aspects of image quality assessments. It produces significantly high PSNR and SSIM values while maintaining the LPIPS values. The network is also capable of controlling artifacts, preserving structures, and enhancing colors better than method (a) because of the use of AM in the illumination network, as shown in Figure. 4-2. Usually, an AM module learns to highlight important features and suppresses unwanted perturbations, thereby resulting in the efficient enhancement of intermediate features. In the proposed work, the AM module efficiently enhances the intermediate features by emphasizing dark regions and suppressing evolving artifacts. Consequently, a final enhanced output containing natural colors and brightness without any evolving artifacts could be obtained.

To summarize, although simple DCE-Net has no distortion or artifacts, such as simple illumination network and curve estimation network with illumination network output, it produces a darker result than the other two. However, the BOLD-Net (method(d)) output does not contain distortions or artifacts when using an attention-guided illumination network with a curve estimation network, and the brightness is improved.

4.4 Application

The suggested framework can be applied as a pre- or post-processing step in order to improve the effectiveness of deoldifying techniques. To test the proposed method with the deoldifying method as pre- or post-processing, Bringing Old Photos Back to Life [10] as a photo deoldifier is used. Bring Old Photos Back to Life is an old image restorer. Figure. 4-3, shows examples of performances for brightness enhancement before and after old image restoration. The video frames are taken from *My Girl* (1991), *Clueless* (1995), *Taxi Driver* (1976), and *The Color of Money* (1986). As shown in Figure. 4-3-(b) and -(c), the proposed BOLD-Net shows superior brightness enhancement performance than the deoldifying method. When the proposed BOLD-Net is used as pre-processing, or postprocessing, the quality of the result images can be further enhanced. The effectiveness of the proposed method gets higher when it is used as preprocessing, as shown in Figure. 4-3-(d) and -(e). In conclusion, the proposed brightness enhancement framework is also crucially beneficial for restoring old photos and films when it is used for pre- or post-processing.

(a) Input (b) Restoration [10] (c) BOLD-Net d) Restoration [10] + BOLD-Net e) BOLD-Net + Restoration [10]

Figure 4-3. Images in column (a) Input images, images in column (b) Result images deoldified by Bringing Old Photos Back to Life [8], images in column (c) Results of the BOLD-Net, images in column (d) Results of Deoldified [8] + BOLD-Net, and images in column (e) Results utputs of BOLD-Net + Deoldified [8], respectively.

5. Conclusion

In this thesis, a multi-branch brightness enhancement network to enhance the brightness of old images while preventing the network from creating artifacts and distorting colors was proposed. The proposed brightness enhancement scheme for old images introduces curve estimation and an attention-guided illumination map to guide enhancement in a feature-adaptive manner. The curve estimation network estimates a curve parameter map that is used to adjust the intensity values of the input image. The attention-guided illumination network estimates an illumination map that adjusts the illumination channel of the input in such a way that both color and textures are preserved. The experimental findings show that, for the brightness enhancement of old images, the suggested framework greatly outperforms state-of-the-art low-light enhancement approaches because it produces output that is more bright and natural. In addition, this approach can be utilized as a pre- or post-processing strategy for restoring old images. The suggested network can act as a pre-processor, enabling older image restoration networks to produce outputs with higher visual quality. The proposed method also performs well when restoring the brightness of videos from previous decades. The future research direction includes extending the proposed method to personalized old image brightness.

References

[1] Wei, Chen, Wenjing Wang, Wenhan Yang, and Jiaying Liu. "Deep retinex decomposition for low-light enhancement." British Machine Vision Conference 2018, BMVC 2018, Aug. 2018.

[2] Wang, Ruixing, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. "Underexposed photo enhancement using deep illumination estimation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6849-6857. 2019.

[3] Zhang, Yonghua, Jiawan Zhang, and Xiaojie Guo. "Kindling the darkness: A practical low-light image enhancer." In Proceedings of the 27th ACM international conference on multimedia, pp. 1632-1640. 2019.

[4] Jiang, Yifan, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. "Enlightengan: Deep light enhancement without paired supervision." IEEE Transactions on Image Processing 30 (2021): 2340-2349.

[5] Guo, Chunle, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. "Zero-reference deep curve estimation for low-light image enhancement." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1780-1789. 2020.

[6] Yang, Wenhan, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3063-3072. 2020.

[7] Rosenthaler, Lukas, and Rudolf Gschwind. “Restoration of movie films by digital image processing,” IEE Colloquium (Digest), no. 49, pp. 55–59, Jan. 2001.

[8] Gschwind, Rudolf. "Restoration of movie films by digital image processing." Nissen, Dan (Hg.): Preserve Then Show. Copenhagen: https://www. google. ch/url (2002).

[9] Palermo, Frank, James Hays, and Alexei A. Efros. "Dating historical color images." In European Conference on Computer Vision, pp. 499-512. Springer, Berlin, Heidelberg, 2012.

[10] Wan, Ziyu, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, and Fang Wen. "Bringing old photos back to life." In proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2747-2757. 2020.

[11] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, “Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 10556–10565, 2021.

[12] Li, Chongyi, Chunle Guo, and Chen Change Loy. "Learning to enhance low-light image via zero-reference deep curve estimation." arXiv preprint arXiv:2103.00860 (2021).

[13] Zhang, Fan, Yu Li, Shaodi You, and Ying Fu., “Learning temporal consistency for low-light video enhancement from single images,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 4965–4974, 2021.

[14] S. Iizuka and E. Simo-Serra, “DeepRemaster: Temporal source-reference attention networks for comprehensive video enhancement,” ACM Transactions on Graphics, vol. 38, no. 6, 2019.

[15] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou, “Zero-shot restoration of underexposed images via robust retinex decomposition,” Proceedings - IEEE International Conference on Multimedia and Expo, vol. 2020-July, Jul. 2020.

[16] H. Ibrahim and N. S. P. Kong, “Brightness preserving dynamic histogram equalization for image contrast enhancement,” IEEE Transactions on Consumer Electronics, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[17] T. Arici, S. Dikbas, and A. Altunbasak, “A histogram modification framework and its application for image contrast enhancement,” IEEE Transactions on Image Processing, vol. 18, no. 9, pp. 1921–1935, 2009.

[18] K. Nakai, Y. Hoshi, and A. Taguchi, “Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms,” ISPACS 2013 - 2013 International Symposium on Intelligent Signal Processing and Communication Systems, pp. 445–449, 2013.

[19] E. H. Land, “The retinex theory of color vision.,” Sci Am, vol. 237, no. 6, pp. 108–128, 1977.

[20] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. “Retinex processing for automatic image enhancement.” JEI, 13(1):100–111, 2004.

[21] D. J. Jobson, Z. U. Rahman, and G. A. Woodell, “Properties and performance of a center/surround retinex,” IEEE Transactions on Image Processing, vol. 6, no. 3, pp. 451–462, 1997.

[22] D. J. Jobson, Z. U. Rahman, and G. A. Woodell, “A multi-scale retinex for bridging the gap between color images and the human observation of scenes,” IEEE Transactions on Image Processing, vol. 6, no. 7, pp. 965–976, 1997.

[23] C. H. Lee, J. L. Shih, C. C. Lien, and C. C. Han, “Adaptive multi-scale retinex for image contrast enhancement,” Proceedings - 2013 International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2013, pp. 43–50, 2013.

[24] S. Wang, J. Zheng, H. M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” IEEE Trans Image Process, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

[25] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, “A fusion-based enhancing method for weakly illuminated images,” Signal Processing, vol. 129, pp. 82–96, Dec. 2016.

[26] X. Guo, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[27] X. Fu, D. Zeng, Y. Huang, X. P. Zhang, and X. Ding, “A weighted variational model for simultaneous reflectance and illumination estimation,” Proceedings of the IEEE Computer pSociety Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 2782–2790, Dec. 2016.

[28] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, “Structure-revealing low-light image enhancement via robust retinex model,” IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[29] K. G. Lore, A. Akintayo, and S. Sarkar, “LLNet: A deep autoencoder approach to natural low-light image enhancement,” Pattern Recognition, vol. 61, pp. 650–662, Jan. 2017.

[30] F. Lv, F. Lu, J. Wu, and C. Lim, “MBLLEN: Low-light image/video enhancement using CNNs” In BMVC, vol. 220, no. 1, p. 4. 2018.

[31] W. Wang, C. Wei, W. Yang, and J. Liu, “GLADNet: Low-light enhancement network with global awareness,” Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, pp. 751–755, Jun. 2018.

[32] L. W. Wang, Z. S. Liu, W. C. Siu, and D. P. K. Lun, “Lightening network for low-light image enhancement,” IEEE Transactions on Image Processing, vol. 29, pp. 7984–7996, 2020.

[33] F. Stanco, G. Ramponi, and A. de Polo, “Towards the automated restoration of old photographic prints: A survey,” IEEE Region 8 EUROCON 2003: Computer as a Tool - Proceedings, vol. B, pp. 370–374, 2003.

[34] I. Giakoumis, N. Nikolaidis, and I. Pitas, “Digital image processing techniques for the detection and removal of cracks in digitized paintings,” IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 15, no. 1, 2006.

[35] R.-C. Chang, Y.-L. Sie, S.-M. Chou, and T. K. Shih, “Photo defect detection for image inpainting,” In Seventh IEEE International Symposium on Multimedia (ISM'05), pp. 5-pp. IEEE, 2005.

[36] V. Bruni and D. Vitulano, “A generalized model for scratch detection,” IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 13, no. 1, 2004.

[37] C. Chen, U. Q. Chen, I. Labs, and J. Xu, “Learning to See in the Dark.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3291-3300. 2018.

[38] Y. Peng Loh and C. Seng Chan, “Getting to know low-light images with the Exclusively Dark dataset,” Computer Vision and Image Understanding, vol. 178, pp. 30–42, 2019.

[39] C. Lee, C. Lee, and C.-S. Kim, “Contrast enhancement based on layered difference representation,” IEEE transactions on image processing 22, no. 12 (2013): 5372-5384.

[40] Chollet, François. "Xception: Deep learning with depth-wise separable convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258. 2017.

[41] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, Dec. 2018.

[42] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[43] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600–612.

[44] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586–595, Jan. 2018.

[45] R. C. Gonzalez, R. E. Woods, and P. Prentice Hall, “Digital image processing third edition Pearson international edition prepared by Pearson Education”.

[46] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. C. Kot, “Low-light image enhancement with normalizing flow,” In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 3, pp. 2604-2612. 2022.

[47] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, “Toward fast, flexible, and robust low-light image enhancement,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5637-5646. 2022.

[48] My Heritage, http://www.myheritage.com

Acknowledgment

First, all praises and thanks to the Almighty - ALLAH, for His blessings that helped me to successfully complete my research. In addition, I want to thank my parents for their unconditional support and for allowing me to pursue my Master's degree overseas.

I want to convey my gratitude to everyone who helped me finish my Master's degree and research. First of all, I sincerely want to take this chance to thank my supervisor, Prof. Bumshik Lee, for letting me pursue my Master's degree at Chosun University. I have been guided and motivated throughout my studies and research by his never-ending inspiration, encouragement, and wise suggestions. His regular supervision and guidance have helped me achieve higher-quality research. He taught me the importance of professionalism, efficiency, and focus, for which I will always be thankful. In addition, I appreciate the chance to work with faculties especially Dr. Shivarama Holla and Dr. Murali Krishna in the Department of Information and Communication Engineering at Chosun University. I want to express my gratitude to the Multimedia Information Processing Lab for providing me with such a great atmosphere and chance for academic growth. I've received both moral and academic support from my colleagues. I also want to thank my international friends at Chosun University, especially Manasa Movvar, Villiana Jainih, Hok Chamroeun, Ghafoor Muhammad Afnan, Aradhana Mishra and Hamza Shafiq for their kindness and support in making my time in South Korea enjoyable and simple. I couldn't have

accomplished anything without their inspiration and guidance. Last but not least, I want to thank my family, close friends, and relatives for their support over the course of my challenging times.

Publications

- ## International Journal Papers

1. Arshiana Shamir, Nokap Park and Bumshik Lee, “BOLD-Net: Brightness Enhancement for Old Images using Deep Curve Estimation and Attention Modules”, **Engineering Application of Artificial Intelligence. (Under review)**

- ## International Conference Papers

1. Arshiana Shamir and Bumshik Lee, "Estimating Deep Curve and Illumination Maps for Old Image Brightness Enhancement," 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), Oct., 2022. **(Presented)**