A Color Vision Approach Based on the Autoencoder Technique and Deep Neural Networks for ReconstructingColor Images under Various Lighting Conditions

by

Asm Faisal
(16201049)
Ashhab Ahmad
(17301162)
Asif Tazwar
(21301732)

A thesis submitted to the Department of Computer Science and Engineeringin partial fulfillment of the requirements for the degree of
B.Sc in Computer Science

Department of Computer Science and Engineering
BRAC University
January 2022

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Asm Faisal
(16201049)

Ashhab Ahmad(17301162)

Asif Tazwar
(21301732)

Approval

The thesis titled "A Color Vision Approach Based on the Autoencoder Technique and Deep Neural Networks for Reconstructing Color Images under Various Lighting Conditions" submitted by

1. Asm Faisal (16201049)
2. Ashhab Ahmad (17301162)
3. Asif Tazwar (21301732)

Of Fall, 2021 has been accepted as satisfactory in partial fulfillment of the require- ment for the degree of B.Sc. in Computer Science on January 17, 2022.

Examining Committee:

Supervisor: (Member)

Md. Ashraful Alam,PhD
Assistant Professor
CSE Department
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD Chairperson
and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

We present a color vision system that utilizes deep neural net- works to normalize pictures using the autoencoder algorithm. Image processing, encoding, and decoding are the three essen-tial processes in the proposed paradigm. An effective image processing approach is utilized to downsize acquired pictures into a finite image resolution equal to the number of input nodes of an autoencoder in the image processing section. En-coding and decoding procedures are included in the Autoen- coder. Second, a deep neural network-based encoding process creates a code for an input picture, and a deep neural network-based decoding process reconstructs the original image from the encoder's code. Convolutional neural networks were used to train the autoencoder with over ten thousand scaled pic- ture datasets. The results of the experiments showed that the suggested model can recreate predetermined normalized pic- tures from original photographs, which may be employed in sophisticated color vision applications.

Keywords: Autoencoder, Image Reconstruction, Deep Neu-ral Networks, Color Vision.

Dedication

We want dedicate our research work firstly to our parents and almighty Allah. We also want to thank our research supervisorMd. Ashraful Alam sir.

Acknowledgement

First and foremost, all glory be to Allah, for whom our thesis was finished without serious hiccups. Second, we would like to thank our supervisor, Md. Ashraful Alam sir, for his kind assistance and advise in our work. He was always willing to assist us when we needed it. Finally, without our parents' ongoing support, it may not be feasible. We are currentlyon the brink of graduating due to their kind assistance and prayers.

Table of Contents

List of Figures

List of abbreviations(alphabetically)

1. ANN - Artificial Neural Network
2. AUC - Area Under The Curve
3. COM - Center of Mass
4. CTC - Connectionist Temporal Classification
5. DDQN - Double Deep Q-Learning
6. DNN - Deep Neural Network
7. DQN - Deep Q-Learning
8. DTW - Dynamic Time Warping
9. E2E - End to end
10. FG - Fine Gained
11. GRU - Gated Recurrent Units
12. HMM - Hidden Markov Model
13. k-NN - K-nearest-neighbor
14. LCS - Largest common substring
15. LDA - Linear Discriminant Analysis
16. LPC - Linear Predictive Coding
17. LSTM - Long Short Term Memory
18. MFCC - Mel Frequency Cepstral Coefficient
19. MIST - Montreal Imaging Stress Task
20. MLP - Multilayer Perceptron
21. MT - Translation modules

22. PWTT - Pulse Wave Transfer Time
23. PWV - Pulse Wave Velocity
24. RNN - Recurrent Neural Network
25. ROC - Receiver Operating Characteristicenvironment.

Chapter 1 Introduction

1.1 Introduction

The study of photographs is one of the most significant fields of deep learning. Images are simple to make and maintain, and they are the ideal form of machine learning data: simple for people to comprehend yet challenging for computers to process. It's no surprise that picture analysis has aided in the creation of deep neural networks for applications such as identifying vehicles, humans or animals, lanes, pedestrians, and so on. If RGB picture files are accessible, they are historically and often utilized in CNN training [1]. These properties may be employed in driverless cars, surveillance, citizen monitoring, and a variety of other applications.

However, light variation is a difficulty when interpreting a pic- ture. When there is enough light on the picture, analysis is easier than when there isn't. Because the sun is absent dur- ing the day, the light is brighter than at night. To address this issue, many approaches for image analysis are now being employed, including both hardware and software solutions. A new test has been developed to provide accurate information on these color contrast perception thresholds [2]. As discussed in, there are several approaches for classification and grading of fruits and agricultural goods utilizing artificial neural net-works and image processing, where parameters such as size, shape, color, texture, and others are employed for grading [3]. Color detection in outside conditions is critical for automatic harvesting [4].

To begin with, owing to its extremely contagious features, coro- navirus (official name Covid-19) is proclaimed a pandemic in 2019-2020. Many developed countries, such as the United States and the United Kingdom, struggled to stop the virus from spreading, whereas South Korea was able to control the pandemic in its country by utilizing technology such as arti- ficial intelligence (AI) to track down contacts via surveillance camera footage and quarantine them immediately [5], [6]. In addition, many industrialized nations utilize artificial intelli- gence to evaluate the speed of cars on the roadway and imme-diately recognize the vehicle number using surveillance cam- eras. It enables them to penalize the accused motorist with near-perfect accuracy while also ensuring road safety. Fur-thermore, utilizing ANN and other image processing methods, early diagnosis of diseases such as Parkinson’s disease may beachieved [7].

As a consequence, low-light image analysis is solved using night vision cameras, grayscale pictures, and other techniques. In other words, picture enhancement increases the amount of vis- ible light. The photographs will be easier to view as a result. Even in the darkest hours, little flecks of light may be seen. All of this light might be infrared light, which is invisible to humans. Using Image Enhancement technology, night vision goggles catch all available light. Instead, they make it more intense so you can see what’s going on at night .

However, numerous approaches such as R-CNN and YOLOV2/V3 are employed to identify an object. When the sun goes down, the difficulty occurs. Detecting things in the dark is challenging for both people and machines. Because a machine cannot be as intelligent as a person, it must be taught or pro- vided with technology to overcome this obstacle. One of the most essential and extensively utilized elements for describing the content of a picture is the color histogram [8]. Shape vari- ance, illumination variance, and object posture variations are all key challenges in object detection [9]. To align the match- ing pictures, a feed forward neural network is used to estimate the transformation, which is described in terms of translation, rotation, and magnification parameters [10].

As a consequence, low light image analysis is accomplished us- ing night vision cameras, gray scale pictures, and other tech- niques. In other words, picture enhancement increases the amount of visible light. The photographs will be easier to view as a result. Displaying the collected picture in a genuine color image is difficult [11]. Even in the darkest hours, little flecks of light may be seen. All of this light might be infrared light, which is invisible to humans. Using Image Enhancement technology, night vision goggles catch all available light. In- stead, they make it more intense so you can see what's going on at night.

Additionally, a thermal imaging option exists. Instead of look- ing for the light that objects reflect, we look for the heat that they emit. In general, live items going in the dark would be hotter than their surroundings; this also applies to auto- mobiles and machinery. Hot things emit infrared radiation, which is similar to light but has a slightly longer wavelength (lower frequency). It's rather simple to build a camera that collects infrared radiation and transforms it to visible light: This works similarly to a digital camera, except that instead of visible light, an image detector chip (i.e., a load-coupled device (CCD) or a CMOS image sensor) reacts to infrared light . It nonetheless creates a clear picture on a screen in the samemanner as a regular digital camera.

On the other hand, photographs taken in low light are identified by the presence of noise. This results in a reduction in the pace of object identification as well as a loss of visual perception of the picture. Denoising a picture is usually associated with filtering. There are a number of well-known approaches for condemning photographs, including the Gaussian smoothing model and the Bilateral filter [12]. Furthermore, in the holo- graphic reconstruction process, noises like speckle and white noise are reduced [13]. Zhou-Wang wavelet total variation, for example, is another approach for noise reduction.

As a result, it is evident that assessing a picture under a variety of lighting circumstances requires a variety of methodologies. Light sparsity tends to shift in a more comprehensive exami- nation. This study, on the other hand, is primarily concerned with the intensity of light. Using an artificial neural network (AE), which is a kind of deep neural network. That is, any picture that has a tendency to When there is little or no light, the model changes the picture to make it seem better.

1.2 Problem Statement

We proposed and demonstrated a color vision approach based on deep neural networks and the autoencoder technique that allows for the conversion of real-world images to normalized images for use in advanced color vision/color machine visionapplications such as robot vision and other industrial applica-tions.

1.3 Research Objective

The objective of this study is to eliminate any attempts made in the name of darkness or low light. The major purpose of this study is to brighten a picture that is dark or lacks brightness. The goals of this project are to:

1. build a collection of photos in different lighting circum-stances

2. reconstruct an image using an autoencoder.

1.4 Scope and Limitations

The purpose of this thesis paper is to design a system capa-ble of completely reconstructing a picture based on its color variation. Image reconstruction may be important in a variety of ways. To overcome the low light picture configuration, sev-eral models and techniques are applied. This approach may be effective for avoiding unnecessary hard labor. Rather of putting in additional effort to recognize an item from a pictureusing multiple models, the method for low lighting allows oneto focus on other duties. However, it is difficult to reverse the process, i.e., this model cannot convert a brilliant picture to a dark one, which is occasionally essential. For example, to create a picture of how it seems to watch in the morning or evening, or to anticipate the gloomy weather and how it looks.Furthermore, forecasting a black picture when the moon shines at night is challenging since the night seems brighter than a typical dark night.

1.5 Thesis Report Outline

The following is the remainder of this research paper's section. The background research and literature review are included in Chapter 2. It discusses previous research on the subjectas well as several kinds of machine learning algorithms. The methodology section of Chapter 3 covers the framework of our whole project. It depicts the whole process of our suggested paradigm using diagrams. The dataset, data preparation, fea- ture extraction, and classification model will all be covered in this section. The findings and analysis of our model are de- scribed in Chapter 4. It explains about the graphical reports in our findings. This section will display the characteristics in our results. Chapter 5 wraps up the research project by summarizing all we've done thus far.

Chapter 2 Literature review

2.1 Literature Review

Computer vision addresses the issues that generic algorithms are incapable of resolving. Computer vision is an artificial in- telligence (AI) discipline that aims to comprehend still picturesand video sequences [14]. Human eyes are extraordinarily sensitive to light, allowing us to distinguish between differenthues. In comparison to humans, the system can obtain superior outcomes using color constancy models and computer graph-ics models, systems of vision. Because of this, there are still many hurdles in the field of computer vision research. RGB pictures only give little information [15]. Lossy compression algorithms are often used to encode images and movies owingto their substantial memory or bandwidth requirements . Re- quirement The retinal system, on the other hand, functions in such a manner that color may shift. In terms of viewpoint, from one individual to the next . We will be able to attain consistent success throughout the domain with the aid of com- puter vision. Our study's goal is to find the right hue. Color illuminance estimation with correct co- efficient variation de- termination. Using CNN architecture and reflectance models CNN that has been educated can extract more information. Characteristics for input photos, as well as the ability to recreate the image for improved feature visualization . The most significant aspect in object detection and recognition is color discrimination. It should be noted that a computer vision test should be performed in advance. [16]. Computers must be able to analyze visual input in a data space defined by the eas- ily observable but less common. Colors, textures, and other distinguishing characteristics [17]. When it comes to object detection, image reconstruction might be a challenge. For this, stereoscopic methods are utilized. Image reconstruction in 3Dpictures is improved . Reconstruction of general pictures is possible. employing neural networks in general. The notion of computer vision system for autonomous cars may include human vision capabilities for color recognition and detection. The chromaticity of light, on the other hand, affects the pic- tures captured by a camera [18]. In today's technology, object identification is critical in machine learning, computer vision, and robotics . As a pre-processing step, color constancy pro- cessing may ensure that the recorded color of the objects in the scene does not vary under varied lighting circumstances [19]. Weather variables like as haze, fog, smoke, and rain might im-pede scene analysis and categorization . Night datasets that make greater use of the RGB matrix and color consistency. The human eye can perceive a red, green, and blue combina- tion [20]. Various physical factors like as lighting, direction, depth, reflectance, and velocity are the criteria that determinehow a scene is represented in an image . The researchers devel- oped a strategy for removing noise from photos using sparse AE in their research study [21]. These devotees have conducted a lot of research to evaluate the efficacy of the techniques in various settings.

These researchers used gray photos and gray copies of datasets as their experimental inputs, and they used a variety of datasets to train their network. Traditionally, the

application of neural networks in image processing has been confined to picture reduction or restoration.
Additionally, they added several sounds based on a variety of criteria in a significant portion of the photos in their datasets.They experimented with visuals having many sounds in a single image in order to fit in diverse circumstances. They honed their model by putting it to the test on datasets with differenttypes of noise, such as Gaussian noise. They also included Salt and Speckle noise in their datasets, as well as Pepper noise for comparison, in this publication. However, when theyapplied it to a testing set with a different kind of noise from the one for which they had not prepared their model, the function began to collapse. They observed, however, that hybrid picture collections may provide more effective results. The peak signal to noise ratio (PSNR) , a frequently used quality metric, is used to assess the performance. The PSNR index of the hybrid noise training sets is always greater than the maximum of othernetworks.
Using mixed datasets to train the sparse denoising AE, these researchers got extremely good performance in a variety of circumstances. These authors introduced a dual AE network model to minimize picture noise and brighten images in low light conditions in this key research [22]. The paired AE net-work model was created by combining the convolutional andstacked AEs, and the study was based on the retinex hypoth- esis. Their proposed network model's main goal is to accom- plish noise reduction and brightness improvement. The stacked AE might be regarded the brightness limitation or smoothness term in their study on the lighting component. To reduce the fake noise, these hobbyists employed convolutional AE on ap-ply a penalty to the reflectance element. To train the neural network,

they employed a collection of areas from both high and low contrast photos.

To eliminate noise from photos, use the stacked denoising au- toencoder. The convolutional autoencoder, on the other hand, may be trained utilizing two-dimensional structure informa- tion. It takes care of the shortage of information. As a result, their suggested dual AE can compute the higher reflectance with the use of convolutionary AE to reduce noise expansion.Because the expected reflectance is inversely proportional to the illumination element, the over-enhanced reflectance com- ponent is extremely likely to be realized because the lowered part of incoming light provides the low lighting portion. In this research, they contrasted the limitations of the variational framework of the retinex approach with their disclosed pairedAE model to prevent overriding of the reflectance components. Using convolutional AE for noise reduction, their suggested ap- proach may reduce enhanced noise in the intensified reflectance component. Furthermore, their suggested network model canprovide output that is equal to input in a self-monitored learn- ing process that may be accomplished by training their model.Because it is possible that obtaining pairings of low and high contrast pictures may be challenging, they generated a data set for experimentation that is backed by pairs of low and high contrast pixels. With the use of a dual AE model based on the idea of retinex without augmentation and saturation, these researchers suggested a design that may provide good outcomes. This model was presented by these researchers, and it can create high-quality images in low-light circumstances for a variety of image processing applications, such as visual monitoring systems (VMS).

These scholars discussed UDAE in this publication [23]. They attempted to preserve accuracy and cost estimates in order to deploy the AE network on their suggested model in real time.

They found it challenging to get data to train the network since clean photos are difficult to come by. They gathered informa- tion from big fish tanks, as well as photographs recorded from software-processed video and underwater images taken at a safe distance from artificially lit structures. These researchers gathered almost 15,000 see-through and distorted photos, re- sulting in 7,055 pairs of images being created and filtered. For visual style transfer, they employed the CycleGAN generative model [24]. On four NVIDIA TITAN X GPU cards, the Cycle- GAN model training took more than a week. After the failed instances were removed, there were 5,194 pairs of photos left. The failed instances were caused by CycleGAN's restrictionsin style transference. Using a single denoising AE, these hob-byists were able to recover the color of the underwater picture. The testing of their UDAE on the NVIDIA Quadro M5000 took roughly a day. They used their suggested network model to direct and train it to restore the color of underwater photos. The UDAE network was trained to test its capacity to simplify real-world data, such as underwater footage taken from social media.

These researchers demonstrated that underwater pictures can be reconstructed using a single AE-based denoising network, and that employing a single AE for real-time implementation is simpler. Furthermore, their network model revealed that UDAE may provide superior color restoration outcomes in pho-tos.

These authors investigated numerous deep learning architec- tures that may be used for compressed sensing in another re- search study [25]. They looked into photon-limited imaging as a possible application of their experiment. To achieve signal reconstruction, the researchers used a stacked AE and a convo- lutional neural network. They created their suggested model using the PAIN architecture to reconstruct compressed signals with Poisson noise. To train their model and support deep neu- ral networks using reconstruction approaches, they employedthe MNIST dataset. They used three distinct deep learning ar- chitectures in their proposed model to recover data from noisy low-dimensional pictures. The backpropagation technique was used to train all three designs, using the mean squared error (MSE) as the loss function. SDA, PICS, and PAIN were the three architectures they recommended. The MNIST dataset contains seventy thousand photographs of handwritten num- bers, of which sixty thousand were used for training and the remaining photos were used for testing. In terms of compres- sion, the PAIN and PICS designs outperformed SDA by a wide margin. They discovered that the PAIN design can recre-ate initial pictures with a greater intensity than previously thought. The MSE frequently produces a lower MSE value for all three designs, according to their testing. The suggested model revealed that the PICS architecture is more beautiful in figures, and that when compression increases, the SDA and PAIN structures resemble the PICS architecture more. They observed that when utilizing MSE as a result measure, PAIN and PICS networks may provide superior results than SDAE when evaluating overall performance.

These enthusiasts suggested a strategy for reconstructing a color picture sequence and decreasing the noise and blurri- ness associated with acquired color photos in this pertinent research work [26]. For testing their network model, they employed near-infrared (NIR) photos taken in extremely low light settings with varying exposure times. They proposed a method in this study that can collect both RGB and NIR pic- tures at the same time. To capture RGB/NIR pictures, they employed an RGB camera and an NIR camera in their imaging system. An RGB color model's long exposure picture may acquire sufficient color information. The short exposure NIR picture, on the other hand, may catch background changes in real time. As a result, their imaging system can capture the necessary picture sequence information for reconstruction.These researchers used both simulated and real photos to traintheir system. In their study work, they also performed com- parisons with guided image filtering and Yan's approach [27],both of which use NIR pictures to reduce noise in color im- ages. This comparison yielded significant findings for them. Their study revealed that their algorithm can eliminate ex- cessive noise in color photos and rebuild images when other approaches have failed. They tested their algorithm on pho- tographs with low light backgrounds and images with extremelow light backgrounds and extended exposure times. Their re- sults demonstrated that they were able to tackle the crucial issue caused by the very dark environment. There are several flaws in their imaging method as well. One of the problems isto entirely eliminate haze in motion and excessive noise. They recommended choosing ideal control settings (k) based on the scene's backdrop. Furthermore, they demonstrated superior performance by manually implementing the settings for areas with significant mobility. Another disadvantage of their net- work approach is that it requires the use of two cameras to properly adjust the pixel spot between the cameras. If the cameras are somewhat different, the system may struggle to deliver proper findings. These authors suggested that over- coming this limitation might be as simple as utilizing a single camera to capture both RGB and NIR pictures.

Chapter 3

Proposed Methodology

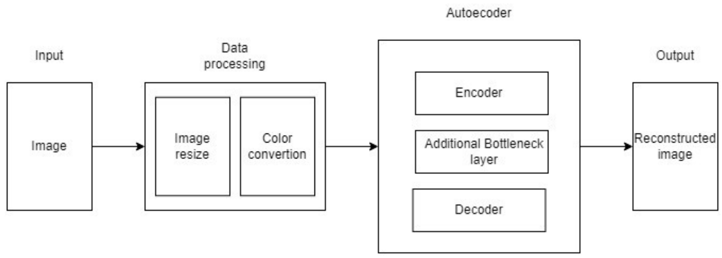

Figure 3.1: Overall System Architecture

We shot photographs in a variety of orientations in order to build an effective reconstruction approach for photos with vary- ing color variations. We concentrated on the color differences of the items in these distinct orientations since there were many different sorts of things. At various times of the day, the hue of the item changes. Five distinct orientations are employed to provide greater color variety. The encoding portion resizes the picture according to its model and refers to the autoencoder's bottleneck by training these color variations in the autoen- coder. The decoding section then receives the enlarged picture from the bottleneck as input and converts it to a normalized image. To verify the picture as output in the relevant layer, each layer employs distinct activation functions, such as ReLu and Sigmoid. The working flow of the autoencoder is explained sequentially in the following parts.

3.1 Image Acquisition

The initial component of the model is picture acquisition, which involves the collection of an image dataset in raw image formatwith a pixel resolution of 1920 x 1080 pixels. A one-of-a-kinddataset has been created for the research. In general, accumu-lating reliable photos for a situation where we can anticipate the same image for varied lighting circumstances is quite chal- lenging. Furthermore, free source datasets have a number of flaws, including a lack of ambient illumination, correct angle variation, time definition, and a clean RGB distribution. In terms of the time dimension, sunshine and a specified angle distribution were critical for our model to be effectively trained and evaluated using our autoencoder structure. As a result, having a distinct dataset allowed us to correctly identify the issue area and operate inside it. Our dataset consists of 10,000photos captured from five distinct perspectives, with 1000 im- ages shot from each viewpoint. Five separate places are used to split and gather the data. To capture the photographs, a gadget called the xiaomi yi lite 4k camera is employed. This camera is capable of producing wide-angle, crisp photos as well as eye-view photographs. In that zone, the maximum recorded temperature was 31 degrees Celsius and the lowest recorded temperature was 23 degrees Celsius on the day the dataset was collected. Weather conditions are linked to light- ing conditions and temperature. As a result, temperature playsa significant role in dataset collecting because the amount of light that can be refracted and reflected from the image to be recorded on the picture is affected by the amount of light that can be refracted and reflected from the image. Furthermore, timing and lighting conditions are linked issues. This data was obtained in Bangladesh between the hours of 9 a.m. and 4 p.m. Our dataset collection is influenced by the temporal di- mension, since it will be difficult to train the first dataset if it is gathered before dawn and after dusk. The photos have a resolution of 1920x1080 pixels (2073600 pixels). In the case ofan image collection, wind conditions might add a lot of noise. The wind conditions remained steady throughout the dataset collection. Unnecessary items with a lot of brilliant color were avoided since they could offer us a color value that causes a spike on the RGB histogram, resulting in dominance of one hue during the training period. And this might lead to im- proved training in one color dimension but poor outcomes in another. Afternoon and noon are the two labels used to cate-gorize the data. Each category has five distinct views that are used to guess on the differences in lighting conditions. Each angle has 1000 photos in our collection. The model employeda ten percent validation split, training 80 percent of the photosand testing the remaining ten percent. Because RGB values sometimes overlap during AE training, this distribution was created to prevent data overfitting.

3.2 Image Processing

Due to technology constraints, our photos have been scaled to two distinct pixel dimensions: 148 x 196 and 380 x 420 pixels, with the findings for 380 x 420 pixels being primarily exhib- ited to support our theoretical method. Following the initial scaling of the picture, the next section comprises the BGR to RGB image color space transition. Because of OpenCv's image importing feature, this method is used in this model. Because of its unique importing capabilities, OpenCV usually imports photos in BGR format.

Our issue area, on the other hand, is primarily concerned with the calculation of RGB characteristics. As a result, we've chosen to convert to RGB color spaceusing an OpenCv architectural feature. It is critical to ensure that our image values do not change throughout the conver- sion process. As the conversion method merely transfers values from one array to another, it was verified for each picture to see whether the conversion was successful.

3.3 Autoencoder

Autoencoders are techniques that seek to reconstruct their out- put as closely as possible to their input. It consists mostly of two components: encoding and decoding. The autoencoder and its processing are shown in detail in Figure 3.2.

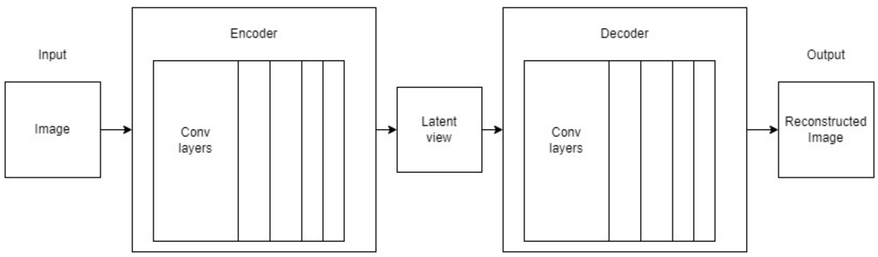

Figure 3.2: Workflow of Autoencoder

3.3.1 Encoder layer

The encoding layer function is accomplished by encoding mul- tiple or single layer encoding layers, which are also referred to as convolution layers. The encoding layer is composed of three convolutional encoding layers. Our approach operates by preceding each encoding layer with a max pooling layer. The third encoding layer routes the third layer's output to the bot- tleneck layer. This section of the model compresses the picture by using the image's main attributes. Convolution layer is em- ployed in our encoding architecture to learn color drops and color value of an input picture. Three encoding layers were employed in our encoding system. We employed 64 filters with a 3x3 weight matrix in our initial encoding convolution layer. In our second encoded layer, we utilize 32 filters with a 3x3 weight matrix. In our third layer, we utilize sixteen filters with a three-dimensional weight matrix. The third tier in the maxpool hierarchy is our latent view, which is the last layer before the decoding layer begins.

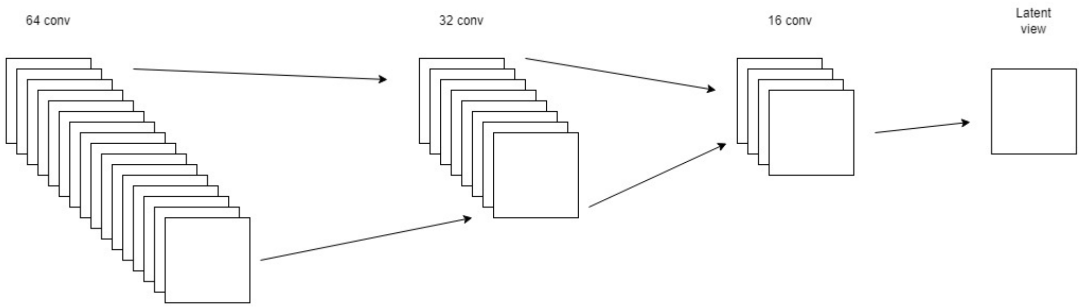

Figure 3.3: Architecture of encoding in proposed model

3.3.2 Bottleneck layer

The bottleneck layer, sometimes known as the latent view layer, represents the final reduced dimension of the input pic- ture that was specified in the image's input section. Three hidden layers are represented in our bottleneck layer. Due to the intricacy of our training, we employed three hidden layers in order to give our final output additional depth. The RGB picture contains its own 256x256x256 pixels, which represents about 16.8 million color choices, making it a complicated do- main. Convolutional layers work with feature maps that have two spatial axes: height, breadth, and depth. Because our photos and issue domain primarily concentrate on red, green, and blue images, we chose 3 as the depth axis dimension for our problem domain. As a result, our output layer is a three-dimensional form tensor.

3.3.3 Decoder layer

The decoded layer operates by completing the convolution pro- cess. Each time the model runs, up sampling is employed to augment the picture size using learned characteristics. Three decoding layers make up the Decoding layer. The output of the third layer defines the final output, which in this case is the normalized picture. We employed three decoding layers in our decoding architecture. We employed 16 filters with a 3x3 weight matrix in the first convolution layer. We employed 32 and 64 filters with a 3x3 weight matrix for the second and third convolutional layers, respectively. We conserved sampling in our decoding architecture by using a 2x2 filter after the convo- lution layer. Stride 2 is utilized to ensure that the maximum pooling ratio is maintained. It aids in the reconstruction of pictures.

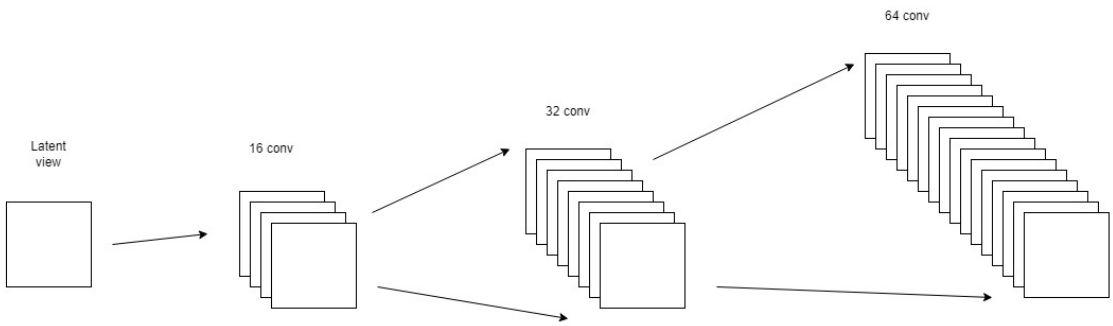

Figure 3.4: Architecture of decoding in proposed model

3.3.4 Output

The recommended model generates an image with a resolu- tion of 380 × 420 pixels as its output. This output is in the RGB color space. The resulting output has a different time- line than the input. For example, if we provide a picture of an afternoon lighting state, our model outputs a noon lighting

condition. The model outputs values for the appropriate angles. RGB values may vary according to the hue of the light source, since color space is a critical factor in learning the result of our model. The lighting condition is critical to the picture producing an effective result. RGB values may vary according to the hue of the light source, since color space is a critical fac-tor in learning the result of our model. As a result, our modelgenerates a normalized picture.

3.3.5 Maxpool and Upsampling

This model is characterized by an encoding and decoding ar- chitecture comprised of many convolution layers. Our model employs an AE structure that was developed after extensive training and testing on our own dataset. Depending on the input, our model may create two distinct types of output. Theafternoon lighting condition is defined by one output, whereas the noon lighting condition is defined by another. Our modelemploys a two-dimensional maxpool to sample the input pic- ture, allowing for easy feature assumptions. We utilized a 2x2 filter on the input layer after each convolution layer stride value 2 is used for smooth picture movement. Max Pooling will aidour model in extracting crisp and smooth features more effi- ciently.

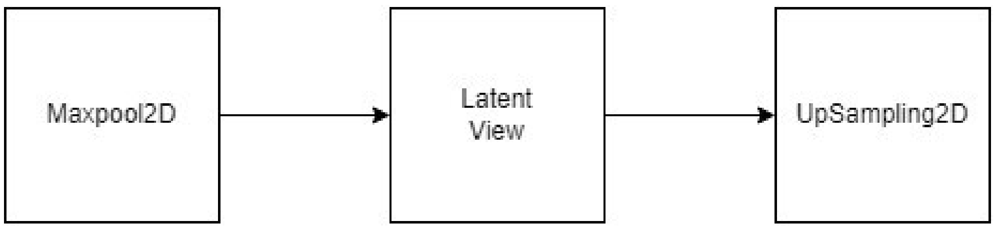

Figure 3.5: Maxpool2D and UpSampling2D architecture in the model

3.4 Activation function

In all levels except the output layer, the proposed model uses the Relu (Rectified Linear Unit) activation function. Relu is utilized to process a portion of our model since it aids in the detection of interaction and non-linear effects. Relu in accor- dance with the bias term increases the efficiency of this model since it is a monotone function. This model becomes quickeras it requires less time to train and run. The ReLu activation function is not affected by the disappearing gradient issue. It is more convergent than other functions. We did not have to de- pend on leaky ReLu since the RGB domain lacks negative val-ues, which would have had a deleterious impact on our model. Due to the absence of a negative value in the RGB domain, the sparse feature will compensate by disregarding the pres- ence of a fake negative value. ReLu is defined mathematically as y=max (0, x) [28]. The sigmoid function is employed in

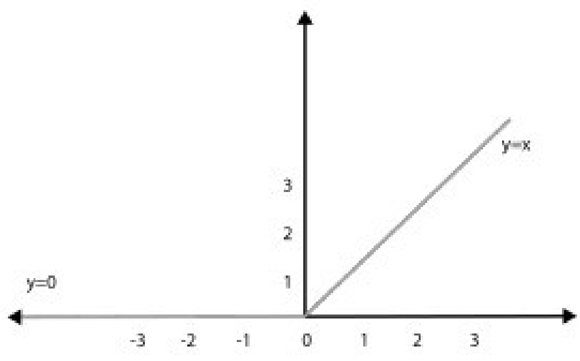

Figure 3.6: : ReLu activation function graph

our architecture's output layer to minimize excessive changes in output due to slight changes in input and to maintain the architecture's integrity. The sigmoid function's formula. By focusing on our model's color characteristics in RGB format and fundamental underlying features, we can have a firm hold on our output thanks to the sigmoid function. If a value de-

viates significantly from the initial value, the sigmoid function enables it to be related to the initial value, hence reducing the variance between the input and output values. In com- parison to our AE model, this feature effectively preserves the characteristics of our source photos throughout training and validation. Following successful training, the model's output is properly distributed when the sigmoid activation function Y
= 1/(1+ex) is used [29].

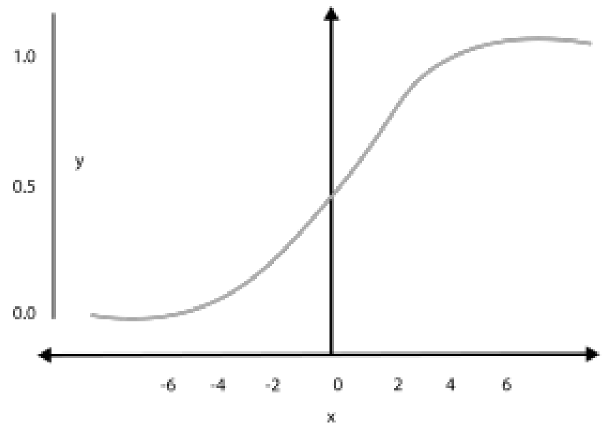

Figure 3.7: : Sigmoid activation function graph

3.5 Adam optimizer and cross validation

Adam (Adaptive Moment Estimation) is utilized as the opti- mizer since it is a solid technique for computer vision image processing. It is employed in our model to repeatedly update network weights using training data. Additionally, Adam opti- mizer took use of both RMSProp and Momentum. In this case, parameter updating is invariant to gradient rescaling. Adam Optimizer does not need a stationary aim. Naturally, the step size is annealed. We employed two distinct epochs to exam- ine if epoch variation affects the training and validation of ourdata. We utilized routine validation of.1 to fine-tune the model hyperparameters appropriately. As a result, our model employs cross validation to prevent overfitting and underfitting. Additionally, we utilized a batch size of twenty. This descent hyperparameter specifies the amount of samples through which our model runs before changing its internal model parameters.

3.6 Dataset

A unique dataset was produced specifically for this investiga- tion. In general, it is quite difficult to gather consistent photos for a situation in which the same image may be speculated for a variety of lighting circumstances. Additionally, open-source datasets have a variety of irregularities, including a lack of ambient illumination, enough angle variation, time definition, and a pure RGB distribution. As with the time dimension, sunshine and a determined angle distribution were critical for adequately training and testing our model based on our au- toencoder structure. As a result, developing a unique dataset enabled us to adequately identify and operate inside our issue area. The dataset consists of 10,000 photos captured from five distinct viewpoints, each having an equal distribution of 1000images.

3.6.1 Data Cleaning

To create a clean dataset consisting of clear photographs with a steady RGB distribution, we excluded images with anomalous noise distortion caused by moving objects such as birds and people. The out-of-focus photos have been deleted to protect the integrity of our training and testing procedures. Addition- ally, photos with almost identical

RGB values were removed from our dataset to prevent training the same image several times. Additionally, over 10% of the photos in our original collection had an overexposed ISO, which was excluded from our dataset. Excessive background zooming was avoided in orderto retain 100% of the source photos and prevent import loss.

3.6.2 Image Resize

Our dataset was prepared in Full HD format and has 2073600pixels (1920x1080). However, owing to hardware restrictions, we were forced to downsize our picture. Additionally, we tested two alternative pixel values, 148x196 (29008) and 380x420 (159600), to demonstrate that greater pixel values might pro- vide superior results. By visualizing two alternative pixel ra- tios, we can discover the varied circumstances and outcomes of our model. We trained our dataset for two epochs for eachpixel ratio in order to ascertain the variance present in our training environment. The following are the advantages of us-ing an autoencoder in the case depicted:

Data specific: An AE may compress data only if it is com- parable to the data on which it was trained. An AE differs from a normal data compression algorithm such as grip in that it learns specialized properties for a given collection of data. Landscape photographs are too large for an AE trained onwriting numbers to compress.

Reconstruction Loss: This is a metric used to determine the efficiency of an AE and the difference between the input and output. The model then seeks to reduce the reconstruc- tion loss via the use of backpropagation.

Unsupervised: Training an AE isn't all that difficult. Due to the fact that it is an unsupervised learning approach, an AE does not need explicit labels for training. To be trainedon a given set of data, an AE produces its own label.

This is referred to as self-supervised learning.

An AE is first and foremost a completely linked layer; it is via this layer that the bottleneck layer or code is constructed. Thismostly refers to the encoder. Additionally, the decoder com- ponent is an ANN design that generates the output from the bottleneck layer. The result should match the input exactly. While it is true that the encoder and decoder are mirror im- ages of one another, this is not required. The sole mandatory condition for an AE is that the input and output dimensions be comparable. The center section may be adjusted by trial and error.

This section of the AE contains the hyperparameters. Hyper-parameters are variables that describe the network architecture and how the network is trained. In this scenario, the network structure is defined by the number of layers or nodes per layer,and the variables are defined by the loss function, learning rate, and number of epochs, among others. Primarily, four hyper- parameters must be defined before to training an AE.

Code size: The number of nodes in the middle or bottleneck layer. Condensation occurs more readily at a smaller size.

Nodes per layer: The number of nodes per layer decreases with each succeeding encoder layer and increases again in thedecoder. Additionally, the decoder's layer structure is identicalto that of the encoder. As stated before, this is critical, andwe have complete control over the settings. By specifying anumber of hidden units less than the number of inputs, AE isforced to learn a compressed approximation [30].

Loss function: There are just a few loss functions in the neural network. As an example, the mean squared error (mse) and the binary cross entropy. Cross entropy may be utilized if the input values are between 0 and 1, else the mean squared error is employed.

Chapter 4

Results and Discussion

This model's output is not scaled to 1920 x 1080 pixels since it works primarily with shrunk images of 380 x 420 pixels and 148 x 196 pixels as input and runs the autoencoder model to ultimately train and validate in resized image size. Thus, com-pletely scaling the resulting picture to the original 1920 x 1080 resolution will stretch the image and produce erroneous RGB values. RGB values would degrade further when the model does extensive scaling back to the original picture, resulting in a distorted image in output. As a result of the hardware restrictions of training and testing this model at 1920 x 1080 pixels, the approach must function with reduced picture sizes of 380 x 420 and 148 x 196 pixels. In Figure 4.2, the out- put for 380 x 420 pixels is shown since this model performs better with a larger pixel ratio. However, the suggested tech-nique trained and tested this model on a lower pixel size of 148 x 196 to demonstrate and compare how well this model performs on bigger pixel proportions, as seen in Figure 4.3 for smaller pixels. Aspect ratios of 148 x 196 and 380 x 420 pixels were selected for this model's simplicity and to create higher- quality output, since pixel sizes preserve a unique pattern of RGB values. This model validated and specified several char- acteristics using cardinal directions, which include the north, south, east, and west orientations shown in Figure 4.1. The model is constructed to demonstrate the variation in lightingconditions caused by the sun in different directions.

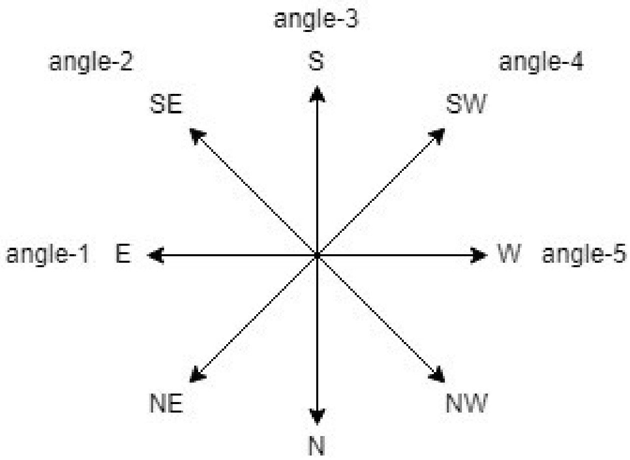

Figure 4.1: : Directions defining our angles

1. Original Image: This section contains the raw data that was gathered as part of our dataset in the 1920 × 1080-pixel ratio. Section (a) in Figures 4.2 and 4.3 is split into five distinct angles denoted by the numbers 1, 2, 3, 4, and
   5. These photographs were taken in the afternoon, which, in the Bangladesh time zone, signifies a decrease in general lighting conditions, finally leading to sunset (dark lightingcondition as dark is also a color spectrum in RGB color space). According to section (a), angle 1 has the brightestcolor setting, whereas angle 5 has the least vibrant color setting.

2. Resized input image: Section (b) of Figures 4.2 and 4.3 shows the scaled picture ratio before to loading it into the AE model. In comparison to the original picture, this section depicts the decrease in pixel size to 380 × 420 and 148 x 196 pixels for Figure 4.1 and Figure 4.2 in the

afternoon. The section has the same five angles as the preceding part (a), and these pictures are put into the architecture's ini- tial encoding layer to start training. We picked afternoon photographs as input to demonstrate our model's perfor- mance since, according to typical color space difficulties, it is simpler to lessen an image's lighting condition. Whereasit is impossible to improve a picture's lighting condition to replicate the original image of a noon situation. As a result, we've opted to display the output of photographs taken at noon with the specified input of afternoon images.

3. Output Image: This section provides output pictures. The output of the pictures in Figure 4.2 and Figure 4.3 is dif- fered by five angles. The outcomes of our model training are shown in the Noon setting here. Noon signifies a more brilliant illumination situation, which is about before 12
   p.m. in Bangladesh. Here, we can distinguish between theafternoon angle 3 picture in section (a) that serves as the input image and the noon angle 3 image in section (b) that serves as the output image. Additionally, lighting condi- tions may be seen from various perspectives in sections (a) and (b) (b). The result is 380x420 pixels in size for Figure
   4.2 and 148x196 pixels in size for Figure 4.3.

4.1 Experimental Results

The model was validated against two distinct picture spaces using five angular images as inputs. The experimental out- come demonstrates that photos with greater resolution pro- vide better-quality reconstructed images. The original input pictures, scaled input photos, and reconstructed output images all have a resolution of 380 x 420 pixels, whereas the collected images have a resolution of 1920 × 1080 pixels. We utilized scaled photos as input to the autoencoder approach for both training and testing. Similarly, we trained and evaluated the suggested model on photos with a resolution of 148 x 196 pix- els. The results of this experiment are shown in Figure 4.2 andFigure 4.3 for pixels 380x420 and 148x196, respectively.

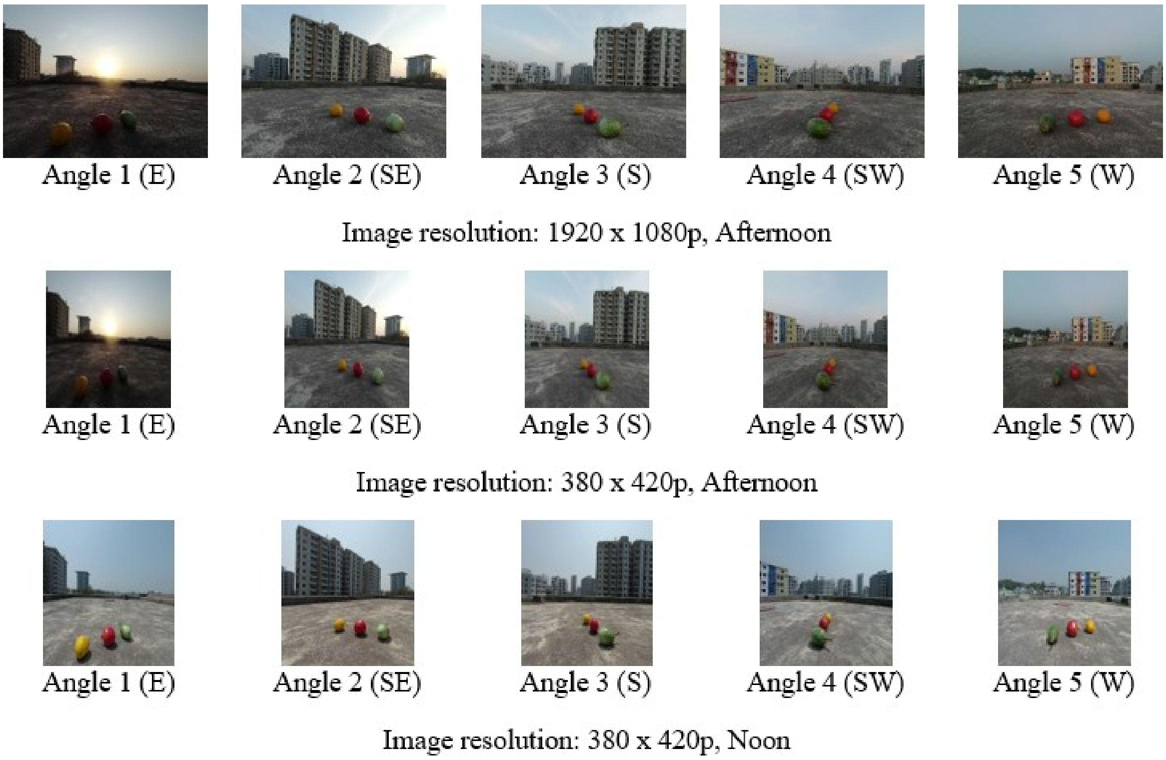

Figure 4.2: : Examples of the proposed model's output pictures for 380 x 420 pixels
(a) original test input photographs from five different perspectives (b) downsized input images from the original images (c) reconstructed output images from after- noon to midday images.

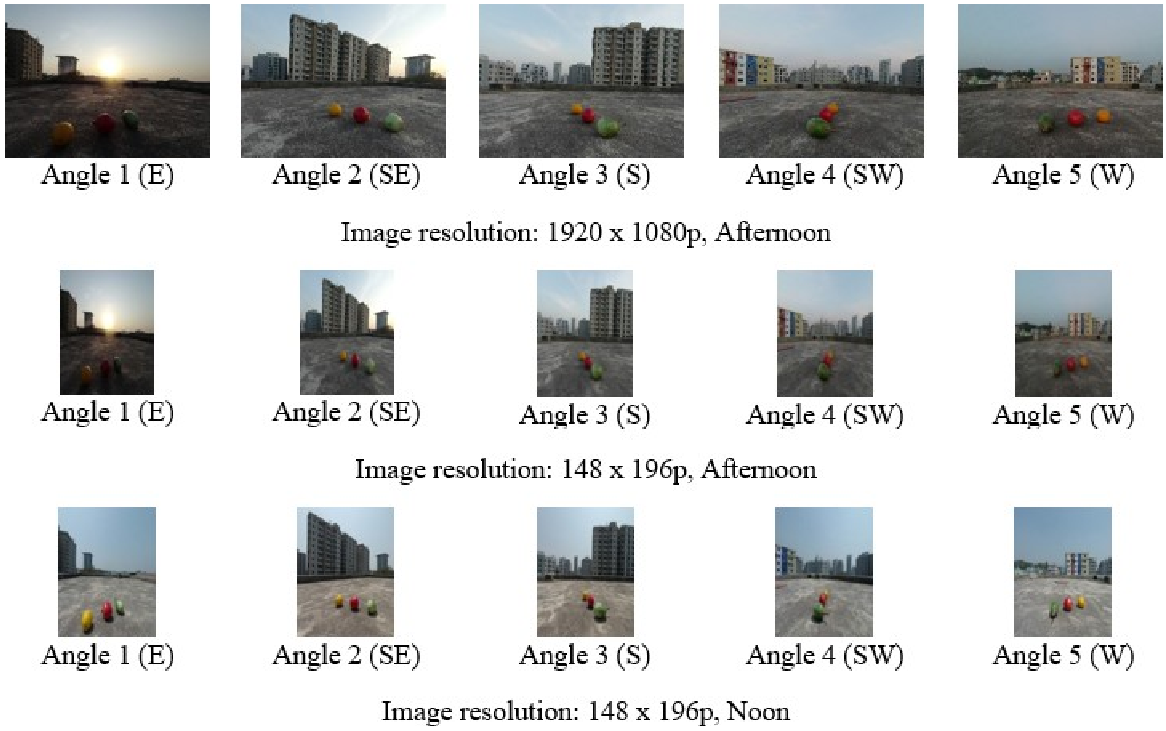

Figure 4.3: : Examples of the proposed model's output for 148 x196 pixel pictures: (a) the original test input photos from five different angles (b) the scaled input images from the original images (c) output pictures were rebuilt from afternoon to midday photographs.

As seen in Figures 4.2 and 4.3, the better the resolution of the input dataset, the higher the quality of the reconstructed output pictures. Additionally, the Result Analyses portion includes a histogram and additional analysis.

4.2 Result Analysis

Our model's analysis and visualization are shown in the images and figures. Our analysis of our findings establishes the validity of the suggested.

4.2.1 MSE Representation for Lower Epoch

MSE denotes the mean square error experienced throughoutthe training and validation phases. It displays the square root of the difference between the actual and estimated values. The

model has shown the MSE representation for the lowest epoch, which is 20. The next figures (Figures 4.4 and 4.5) illustrate the mean square error for angles 1 and 5. The angles 1 and 5 denote the east and west directions, respectively.

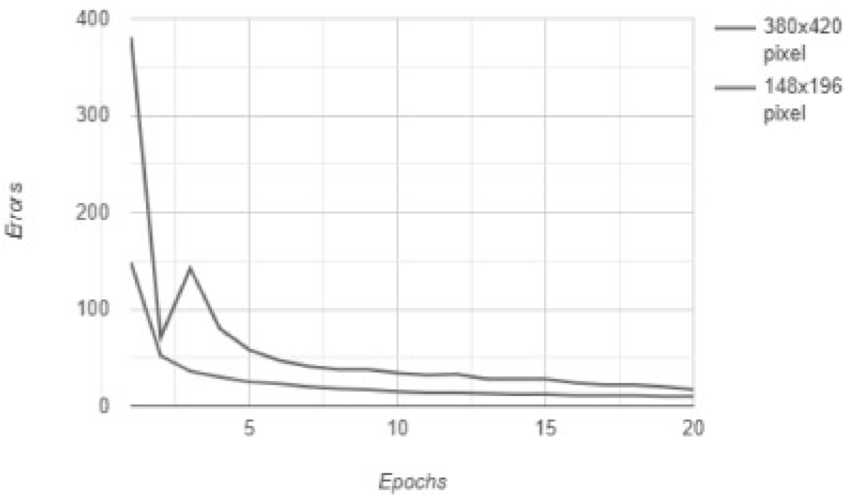

Figure 4.4: : Angle1(E) MSE versus Epoch graph afternoon to noon

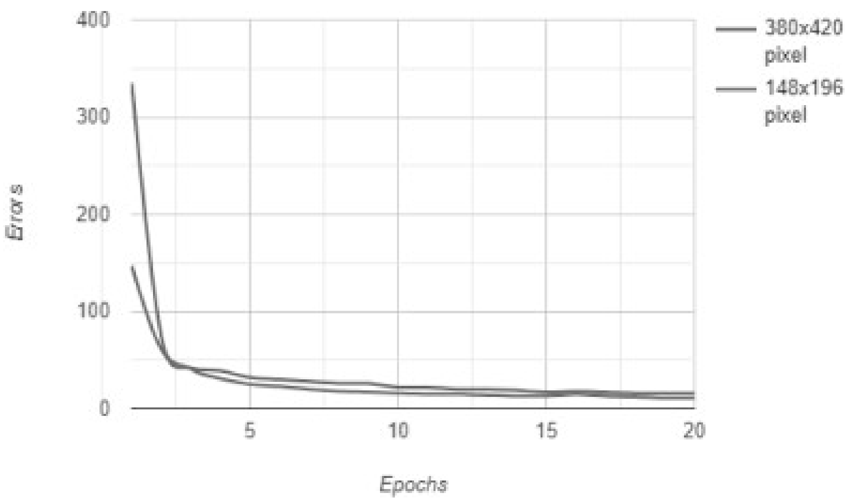

Figure 4.5: : Angle1(E) MSE versus Epoch graph noon to afternoon

The training difference between 380 x 420 and 148 x 196 pixelsizes is determined by this epoch graph. In each of the above pictures, the representation for 20 epochs is displayed together with the associated mean square error. As can be observed, our model's error decreases as the period size increases. In terms of Figure 4.5, the curve virtually flattens out and stays con- stant after the fifth epoch throughout the midday to afternoon training period, while MSE increases and decreases abruptly in Figure 4.4 for pixel value 380 x 420.

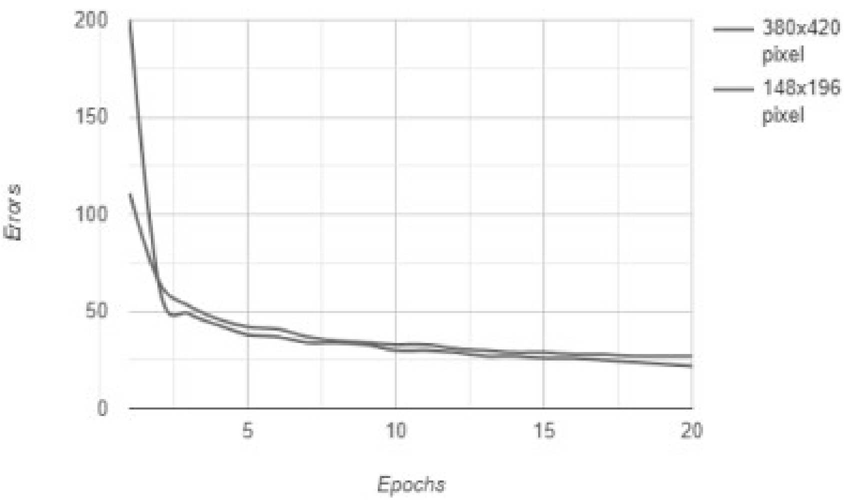

Figure 4.6: : Angle5(W) MSE versus Epoch graph afternoon to noon

MSE (mean square error) vs. epochs graphs for the training domains Afternoon to Noon (angle 5) and Noon to Afternoon (angle 5) are displayed in Figure 4.6 and Figure 4.7, respec- tively. Angle 5(W) denotes the west direction in this case. This epoch graph computes the training difference between 380 x 420 and 148 x 196 pixel sizes. In each of the above pictures, we have shown a representation for twenty epochs and its asso- ciated mean square error. It can be shown that as the period

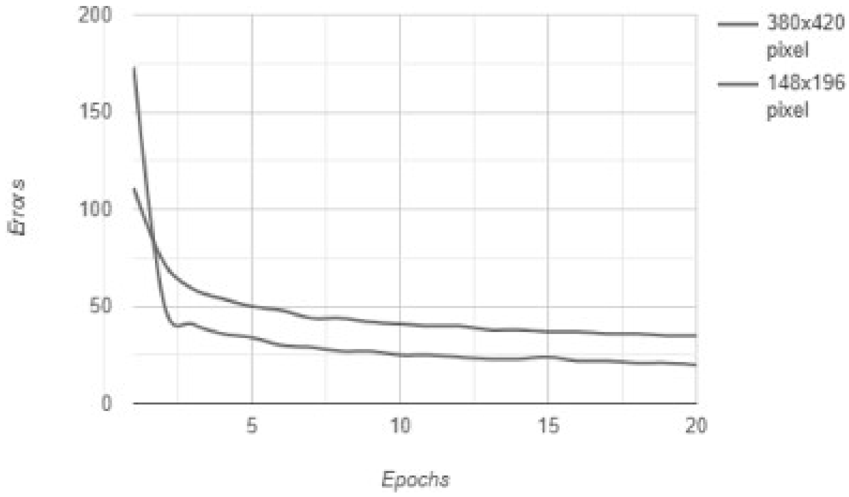

Figure 4.7: : Angle5(W) MSE versus Epoch graph noon to afternoon

length increases, the inaccuracy of this model decreases. After the seventh epoch, the curve almost flattens out and stays con- stant. MSE is greater for 380 x 420 pixels than for 148 x 196 pixels in Figure 4.7 owing to the increased quantity of RGB processing per pixel domain.

4.2.2 MSE Representation for Higher Epoch

MSE denotes the mean square error experienced throughout the training and validation phases. It has been presented herethe MSE representation for the higher epoch, which is 40. The following graphs illustrate the mean square error associated with angle 1(E) and angle 5. (W).

MSE (mean square error) vs. epochs graphs for the training domains Afternoon to Noon (angle 1) and Noon to Afternoon (angle 1) are displayed in Figure 4.8 and Figure 4.9, respec- tively. This epoch graph computes the training difference be-

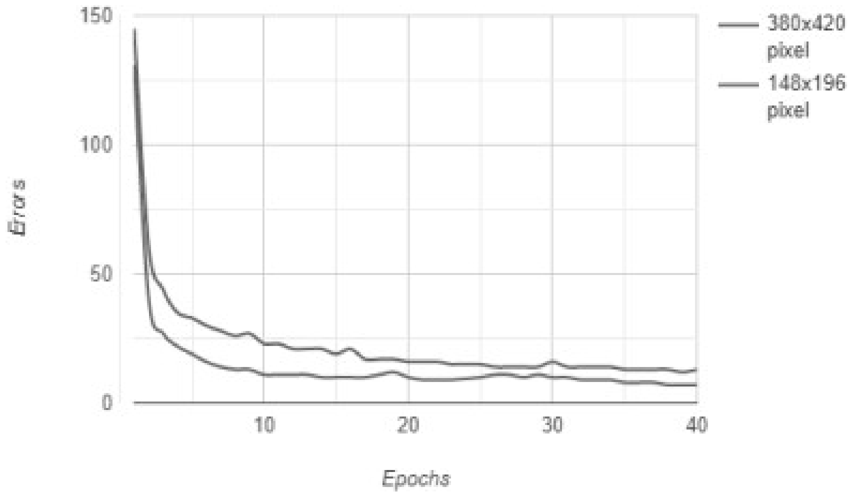

Figure 4.8: : Angle1(E) MSE versus Epoch graph afternoon to noon

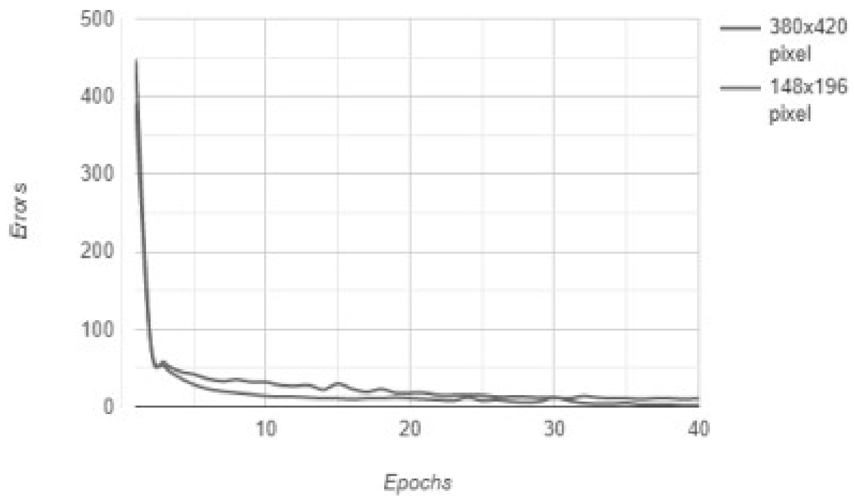

Figure 4.9: : Angle1(E) MSE versus Epoch graph noon to afternoon

tween 380 x 420 and 148 x 196 pixel sizes. Both of the above graphics show a depiction for 40 epochs and the associated mean square error. This demonstrates that when epoch is raised, our model's error decreases across the epoch iterationsand almost flattens for Figure 4.9. The MSE of the 380 × 420 pixel in Figure 4.5 is greater than that of the 380 x 420 pixel in

Figure 4.9 owing to the higher picture values required for pro- cessing. Additionally, it establishes that additional processing is required owing to the difficulties of generating lighting cir- cumstances comparable to those of the afternoon, where the output picture is of the noon time frame with a greater light intensity Figure 4.8.

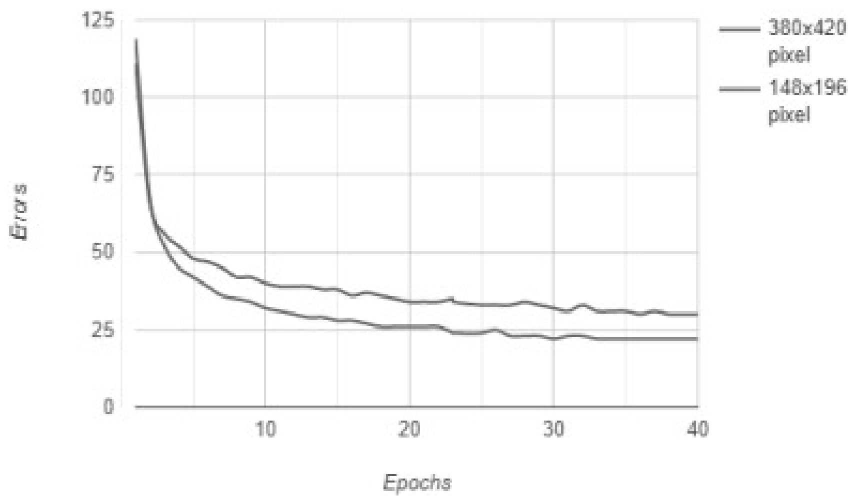

Figure 4.10: : Angle5(W) MSE versus Epoch graph afternoon to noon

On the other hand, in Figures 4.10 and 4.11, the MSE (mean square error) vs. epochs graphs for the training domains After- noon to Noon (angle 1) and Noon to Afternoon (angle 1) are displayed. This epoch graph computes the training differencebetween 380 x 420 and 148 x 196 pixel sizes. Both of the above graphics provide a representation for 40 epochs and its asso-

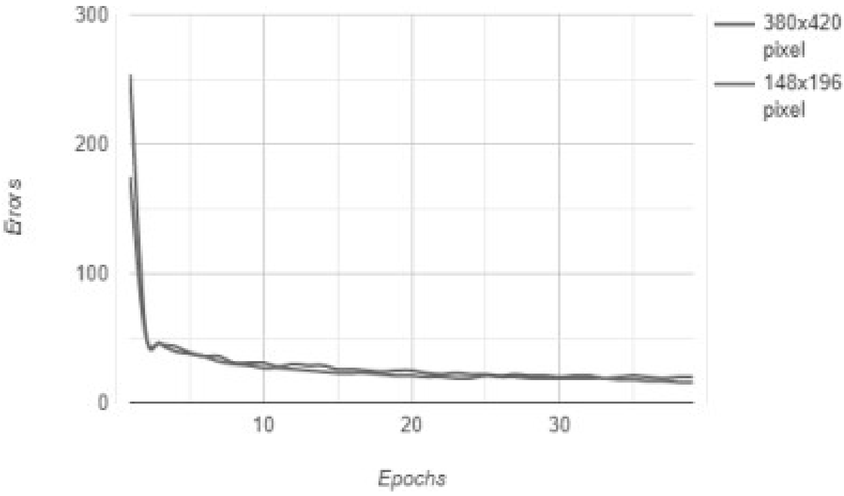

Figure 4.11: : Angle5(W) MSE versus Epoch graph noon to afternoon

ciated mean square error. As epoch is raised, it can be seen that our model's error decreases during the epoch iterationsand almost flattens for Figure 4.11. The MSE of 380 x 420 pixels in Figure 4.10 is greater than that of Figure 4.11 owing to the larger picture values required for processing. The lat-ter figure is more linear because more information is required to analyze a picture when it is converted to higher intensity lighting condition values. As a result, the greater pixel valuealso has a larger MSE value, which corresponds to the epoch variation seen in Figure 4.10.

Chapter 5Conclusion

5.1 Conclusion

This study is primarily concerned with reconstructing picturesin order to normalize them to a preset format under a vari- ety of color vision or color machine vision situations. When evaluating a picture in low light, the key problem is to use new procedures when the result isn't up to par. Because the image's light reflection is so little. As a result, object detec- tion from a picture requires greater affordance. The goal of this study is to find a practical solution to the problem. In this study, a novel technique is presented. As the picture is shot in daylight, any photograph with a lack of light may be transformed into a brighter one using this method. That is, the picture does not need the use of additional procedures oralgorithms to aid in the detection of an item. In this method, a picture is used as the input to SAE, which then feeds the encoding section. After certain characteristics extraction, the picture passes through multiple hidden layers of encoding sec- tions and reaches its bottleneck. The bottleneck then shows as the decoding component of SAE's input. Similarly, it passes through numerous layers of decoding components before being ready to display the results. The important finding is that af- ter using this procedure, the output is almost identical to the input. Several ways, however, are employed to demonstrate that the output is near enough to its input. This study uses a graph to demonstrate the amount of loss for two distinct pixel pictures, one of which is 148x196 pixels and the other of which is 340x420 pixels. The graph shows that after a few epochs, the amount of input to output loss is reducing. It's es-timated for two distinct numbers of epochs (20 and 40). Also, since the model works well, the histogram we use to demon- strate the difference between input and output is practically identical. A super sampled heatmap, RGB color variation, 3D scatter plot, and RGB color code are also included to demon- strate the output's correctness in comparison to its input and original picture. In summary, the research provides a novel technique to picture reconstruction that shows promise.

Bibliography

[1] K. Jhang and J. Cho, "Cnn training for face photo based gender and age group prediction with camera," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 548–551. doi: 10.1109/ICAIIC.2019.8669039.

[2] I. Benkhaled, I. Marc, and D. Lafon, "Colour contrast detection in chromatic- ity diagram: A new computerized colour vision test," in *2017 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*, 2017, pp. 1– 4. doi: 10.1109/WNYIPW.2017.8356262.

[3] T. Nishi, S. Kurogi, and K. Matsuo, "Grading fruits and vegetables using rgb- d images and convolutional neural network," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–6. doi: 10.1109/SSCI.2017.8285278.

[4] B. Öztürk, M. Kirci, and E. O. Güneş, "Detection of green and orange color fruits in outdoor conditions for robotic applications," in *2016 Fifth Interna- tional Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 2016, pp. 1– 5. doi: 10.1109/Agro-Geoinformatics.2016.7577641.

[5] L. Bicker, *Coronavirus in south korea: How 'trace, test and treat' may be saving lives*, Mar. 2020. [Online]. Available: https://www.bbc.com/news/world-asia-51836898.

[6] M. Fisher and C. Sang-hun, *How south korea flattened the curve*, Mar. 2020. [Online]. Available: https://www.nytimes.com/2020/03/23/world/asia/ coronavirus-south-korea-flatten-curve.html.

[7] M. Rumman, A. N. Tasneem, S. Farzana, M. I. Pavel, and M. A. Alam, "Early detection of parkinson's disease using image processing and artificial neural network," in *2018 Joint 7th International Conference on Informatics, Elec-tronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 2018, pp. 256–261. doi: 10.1109/ICIEV.2018.8641081.

[8] D. D. Pukale, S. Bhirud, and V. Katkar, "Content-based image retrieval using deep convolution neural network," in *2017 International Conference on Com- puting, Communication, Control and Automation (ICCUBEA)*, 2017, pp. 1–5. doi: 10.1109/ICCUBEA.2017.8463691.

[9] D. Barik and M. Mondal, "Object identification for computer vision using im- age segmentation," in *2010 2nd International Conference on Education Tech- nology and Computer*, vol. 2, 2010, pp. V2-170-V2–172. doi: 10.1109/ICETC.2010.5529412.

[10] A. B. Abche, F. Yaacoub, A. Maalouf, and E. Karam, "Image registration based on neural network and fourier transform," in *2006 International Confer- ence of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 4803– 4806. doi: 10.1109/IEMBS.2006.260342.

[11] F. Tanriverdi, D. Schuldt, and J. Thiem, "Hyperspectral imaging: Color re-construction based on medical data," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2018, pp. 194–199. doi: 10. 1109/IECBES.2018.8626614.

[12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*,1998, pp. 839–846. doi: 10.1109/ICCV.1998.710815.

[13] L. T. Bang, W. Li, M.-L. Piao, M. A. Alam, and N. Kim, "Noise reduc-tion in digital hologram using wavelet transforms and smooth filter for three- dimensional display," *IEEE Photonics Journal*, vol. 5, no. 3, pp. 6 800 414– 6 800 414, 2013. doi: 10.1109/JPHOT.2013.2265979.

[14] D. Gerónimo, J. Serrat, A. M. López, and R. Baldrich, "Traffic sign recognition for computer vision project-based learning," *IEEE Transactions on Education*, vol. 56, no. 3, pp. 364–371, 2013. doi: 10.1109/TE.2013.2239997.

[15] Z. Cai and L. Shao, "Rgb-d data fusion in complex space," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 1965–1969. doi: 10.1109/ICIP.2017.8296625.

[16] I. Benkhaled, I. Marc, and D. Lafon, "Colour contrast detection in chromatic- ity diagram: A new computerized colour vision test," in *2017 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*, 2017, pp. 1– 4. doi: 10.1109/WNYIPW.2017.8356262.

[17] B. Zhang, "Computer vision vs. human vision," in *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, 2010, pp. 3–3. doi: 10.1109/ COGINF.2010.5599750.

[18] J.-H. Yoo, Y.-H. Ha, and S. K. Subhashdas, "Color constancy method based on local chromaticity distribution and illuminant influence for hue angle," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, 2017, pp. 1–5. doi: 10.1109/GCCE.2017.8229488.

[19] S. Bianco, C. Cusano, and R. Schettini, "Single and multiple illuminant estima- tion using convolutional neural networks," *IEEE Transactions on Image Pro-cessing*, vol. 26, no. 9, pp. 4347–4362, 2017. doi: 10.1109/TIP.2017.2713044.

[20] P. K. Nigam and M. Bhattacharya, "Colour vision deficiency correction in image processing," in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, 2013, pp. 79–79. doi: 10.1109/BIBM.2013.6732581.

[21] X. Ye, L. Wang, H. Xing, and L. Huang, "Denoising hybrid noises in image with stacked autoencoder," in *2015 IEEE International Conference on Information and Automation*, 2015, pp. 2720–2724. doi: 10.1109/ICInfA.2015.7279746.

[22] S. Park, S. Yu, M. Kim, K. Park, and J. Paik, "Dual autoencoder network for retinex-based low-light image enhancement," *IEEE Access*, vol. 6, pp. 22 084–22 093, 2018. doi: 10.1109/ACCESS.2018.2812809.

[23] Y. Hashisho, M. Albadawi, T. Krause, and U. F. von Lukas, “Underwater color restoration using u-net denoising autoencoder,” in *2019 11th Interna- tional Symposium on Image and Signal Processing and Analysis (ISPA)*, 2019,pp. 117–122. doi: 10.1109/ISPA.2019.8868679.

[24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image trans- lation using cycle-consistent adversarial networks,” in *2017 IEEE Interna- tional Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.

[25] O. DeGuchy, F. Santiago, M. Banuelos, and R. F. Marcia, “Deep neural net-works for low-resolution photon-limited imaging,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3247–3251. doi: 10.1109/ICASSP.2019.8682767.

[26] D. Sugimura, T. Mikami, H. Yamashita, and T. Hamamoto, “Enhancing color images of extremely low light scenes based on rgb/nir images acquisition with different exposure times,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3586–3597, 2015. doi: 10.1109/TIP.2015.2448356.

[27] Q. Yan, X. Shen, L. Xu, *et al.*, “Cross-field joint image restoration via scale map,” in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1537–1544. doi: 10.1109/ICCV.2013.194.

[28] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *ICML 2010*, 2010, pp. 807–814.

[29] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning.,” in *IWANN*, J. Mira and F. S. Hernández, Eds., ser. Lecture Notes in Computer Science, vol. 930, Springer, 1995, pp. 195–201, isbn: 3-540-59497-3. [Online]. Available: http://dblp.uni-trier.de/db/conf/iwann/iwann1995.html#HanM95.

[30] L. Gondara, “Medical image denoising using convolutional denoising autoen- coders,” in *2016 IEEE 16th International Conference on Data Mining Work- shops (ICDMW)*, 2016, pp. 241–246. doi: 10.1109/ICDMW.2016.0041.