Automated Colourisation of Historical Photographs from East Asia with U-Net and Conditional Generative Adversarial Network

Jiarong Li (20230033)
School of Computer Science
National University of Ireland, Galway

Supervisors
Dr. Michael Schukat

In partial fulfillment of the requirements for the degree of
MSc in Computer Science (Software Design and Development)

December 7, 2022

Total 7231 words and 44 pages, which satisfies with the thesis limits of NUI Galway.

DECLARATION I, Jiarong Li, do hereby declare that this thesis entitled "Automated Colourisation of Historical Photographs from East Asia with U-Net and Conditional Generative Adversarial Network" is a bonafide record of research work done by me for the award of MSc in Computer Science (Artificial Intelligence) from National University of Ireland, Galway. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signature: ______________________________

Acknowledgement

I am immersing in Galway's gentle yet beautiful evening glow, her tender breeze touches my hair and generousness harbors my dream. The early vernal twilight is spreading amid the woods and the Galway Bay is shimmering at such moment. In the rise and fall of the waves, they delightfully reach the shores and break tenderly in the glow. At this moment, a million of emotions surges to restless me at a pinnacle. I am feeling sentimental, blissful, delightful and peaceful with tears and smiles in her tranquil, serene and halcyon evening. I just realised I have completed a journey in my life and am about to start a new chapter.

The mist in the Galway Bay is receding and the rouge evening glow is disappearing at this moment. The town is starting to lie in her light dream. I completed my thesis writing and am walking on her serene street yet my memory and thoughts have been dragged back to the point before I went to in Japan. I had been experiencing an episode of dark time -- I was feeling lost and hopeless yet my life was in hardship as well. Seeking excellence and fulfilling the value always motives me to overcome my darkest time without giving up my dreams. I came to Japan to Ireland with big dreams to become a science girl from being not a science girl. Coming to Ireland is the most precious present that the deity of fate game to me. At NUI Galway, I started to pursue my first science Masters program in Software Design and Development and I was also accepted by the renown Masters program in Artificial Intelligence. NUI Galway harbors

my dreams and offers that I can have the opportunity to fulfil my dreams and to succeed in academia. I came to Ireland with nothing, but now I met fabulous lecturers and had my best intimate friend. I found my own wonderful way to live from here.

The support from my parents is indispensable to my pursuit and meaningful to me. Without their great priceless altruistic support, the odds of I arrive at this point will be nearly impossible. They did their best to support their daughter to desperately and boldly pursue her dreams. I am truly grateful to their support, accompanying and encouraging. My future success and glory will not only be mine but also belong to them my parents. They taught me what the meaning of true love and why love is eternal. I am feeling so blissful and lucky we can meet from the vast time and space. Their love helped and will help me to cope with dark time and sooth my soul at any plight. Their love will last forever. They embody the meaning of true love. I now understand that love is the essence of life.

My thesis supervisor Dr. Michael Schukat is a cordial erudite lecturer. In my research and writing, his generous, cordial, patient and priceless help is indispensable to my success of completing this journey. He encourages me boldly pursuing my academic dream and generously gives me priceless support. I am so grateful to be supervised by him.

I also wish to thank the following lecturers: Dr. Attracta Brennan, who taught us Databases; Dr. Seamus Hill, who taught us Computer Networks and Data Communications and Programming I; Dr. Karen Young, who taught us Software Engineering 1; Dr. Patrick Mannion, who taught us Enterprise Java Programming; Dr. Adrian Clear, who taught us Object Oriented Software Design and Development; Dr. Enda Howley, who taught us Research Project and Algorithmics and Logical Methods; Dr. Michael Schukat, who taught us Computing Architecture and Operating Systems; Dr. Enda Barrett, who taught us Internet Programming. They are fabulous lecturers and I can never be at this point without their priceless help.

I deeply thank my best intimate friend and my research collaborator, Jiaolin Luo, who is pursuing her Ph.D in Artificial Intelligence at NUI Galway. She is an excellent female and desperately pursuing her big dreams. Her brave story also inspires my pursuit in academia. Her support and encouragement are priceless to me. She shows me how beautiful science can be and pursuing academic path can be glorious and great. Such science spirit will forever motivate me. I will forever treasure our friendship.

I thank those who gave me help and encouraging in my life. I forever cherish every piece of help and every moment of being with them. I just quote my speech script at WiSTEM event as a part to end my acknowledgements:

> *The instinct to explore is one of the qualities that defines us as human beings. It propelled us across vast oceans and to every corner of every continent. We also sent probers to the place far beyond the frontier of our solar system to seek new civilisations and to quest our origin.*
>
> *The spirit also motivates me to desperately boldly pursue my dreams. Fears are a hurdle in front of my big dreams. Yet, I realised that confronting the fears is a journey of life, and is the key of being happy and peaceful.*
>
> *I'm pursuing Masters and doing AI research at NUI Galway. I was not even a science girl before, but now I am on a road less travelled by others because I am pursuing a hard science subject in Artificial Intelligence. I came from China to Japan to Ireland to pursue my big dreams. Artificial Intelligence is a future technology and key to the sustainable development of human beings. I am so proud of being a female with big dreams, following my hunch and seeking the significance of my life to fulfil my value to me to the world as a global citizen.*

Abstract

Earlier photographs present a glimpse of the historical panorama and exhibit the features of their times in addition to other recordings. Nevertheless, these historical photographs are monochromatic and thus lack further chromatic details. Chromatic details are crucial to the perceptions of humans. Deep learning based automated colourization sheds light on converting monochromatic photographs to their chromatic versions. In this thesis, we conduct an interdisciplinary research of harnessing the potential of deep learning for colourizing the monochromatic photographs by combining state-of-the-art conditional GAN and multi-scale U-Net. We leverage dual loss functions to secure an authentic colorization quality. Furthermore, we devise a novel mechanism of adjusting the weights of the dual loss values across the training process for meeting the various needs of balancing the loss values. We design a hand-crafted colorization training data set for colorizing historical photographs and train our model on it. We finally evaluate our architecture by colorizing photographs from training data set and comparing the colorized versions with their ground truth versions, and, colorizing historical photographs for demonstrating the performance on unseen data.

Keywords: Deep learning, automated colorization, conditional GAN, U-Net, photography

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Summary

In the Section 1.2 of the Chapter 1, we first conduct a brief investigation regarding the history of photography. In Section 1.3, we recapitulate the importance of the chromatic information in photography from the perspective of psychology and the recapitulation can endorse the needs of our research. In Section 1.4, we justify the choice of East Asian area as our research subject by characterizing the traits from the historical photographs in East Asia. In Section 1.5, we summarize the advances and challenges in automated colorization algorithms. In Section 1.6, we briefly present our architecture and research scope. In Section 1.8 and Section 1.9, we clarify our research objectives and claim our contributions respectively.

1.2 A brief history of photography

In research community, it is granted that the history that humans have been paying enormous endeavours to record their quotidian events can antedate as earlier as the fourth millennium BC. Such endeavours leave a profound impact on the progress of human history. The emergence of writing system marks a landmark in such progress. The Sumerians settling in Mesopotamia first created the cuneiform writing system for keeping their history (Crawford, 2013). The recordings provide a glimpse of their splendid civilization and exhibit the tremendous value and potential to humans. The Figure 1.1 shows an example of the Sumerian writings from the fifth tablet of Gilgamesh.

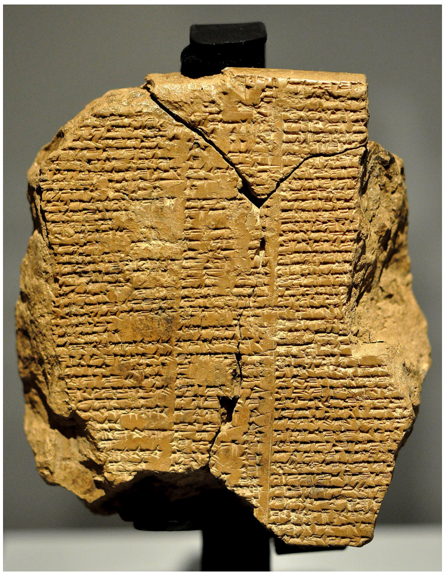


Figure 1.1: The tablet V of the Epic of Gilgamesh. The image is from the Wikipedia (https://en.wikipedia.org/wiki/Epic_of_Gilgamesh).

Notwithstanding, symbol based writing system has its intrinsic weakness, in which the information that the writing system manages to preserve is the indirect information from its writers. As a result, the indirect information conveyed in writing system can hardly reflect the original fashion of its subjects. With the light of the invention of photography, we can therefore retain more information in an authentic way. In 1839, Louis Daguerre invented the Daguerreotype device, which is considered as the earliest photography technique, and can capture monochromatic versions of scenes (Newhall, 1982; Benjamin, 1972). Nevertheless, monochromatic photography can suffer from the lacking of chromatic information until chromatic photography becomes dominant since 1970s.

(a) The Daguerreotype device.

(b) An early photograph by the device in Figure 1.2(a).

Figure 1.2: The earlier photography device and its corresponding photograph. The two images are from Wikipedia (https://en.wikipedia.org/wiki/Daguerreotype).

1.3 Chromatic perception of humans

Chromatic information in photographs is indispensable to the perceptions of humans. First, humans are designed to perceive the world with chromatic vision, which enables humans surviving from evolution and circumventing dangers. The vision system of human eyes consists of trichromatic receptors known as cones and is able to sense various lights with different wavelengths – which are corresponding to various colors (Regan, 2000; Surridge et al., 2003). The human vision system is thus a sophisticated machine designed for chromatic world. The chromatic vision plays an important role for humans outwitting from the long evolution. Second, chromatic information can imply emotions in our perception cited from psychological study, and also purports aesthetic implications. The perception of humans is not only a physical or biological process but also a psychological process (Mollon, 1982). The psychological aspects such as emotions can thus participate into the perception process and affect the understanding of the world.

1.4 East Asian photographs

East Asian photographs have disparate traits from earlier historical photographs in west. Firstly, the surviving historical photographs from East Asia area are scarce due to warfare and the use of photography is late than western world (Thomas et al., 2009). The majority of earlier East Asian photographs was taken by western travelers and evangelising missionaries. Secondly, western cultures and oriental cultures have discrepancy regarding their themes in photography (Wubin, 2016). For instance, portrait photography was less popular since the culture preference in East Asian area. Finally, the well established works from research community for colorizing historical photographs largely concentrate on western photographs. We anticipate that our research of colorizing East Asian historical photographs can have positive inspiration on relevant disciplines and public.

1.5 Automated colorization

Colorizing historical photographs can transform the way that we perceive and sense the past history, and extend insights of the understanding of the past world. Transforming monochromatic photographs into their chromatic versions poses as a huge challenge in both academia and industry. Hand-crafted colorization has been devised and dominant in industry over decades. However, such approach can suffer from the affordability, the time consumption and the availability of colorization artists.

Automated colorization leveraging deep learning sheds a light on the availability of colorizing historical photographs with affordable cost. Such approach can have positive implications and profound impact on ourselves from multifarious aspects: (1) We can sense that past history is as vivid as today; (2) Colorized photographs can also be harnessed in public education purpose and (3) Relevant disciplines – historical research, anthropology and archaeology – can benefit from the colorized photographs.

Deep Colorization is the first attempt of colorizing monochromatic photographs by leveraging the potential of Deep Learning. The Deep Colorization divides the input monochromatic images into patches, extracts the DAISY features and leverages the features to predict the chromatic information (Cheng et al., 2015; Ebcioğlu and Altman, 1997). Nevertheless, the work dose not exploit the potential of state-of-the-art convolutional neural network (CNN) and later works manage to improve the colorization performance by introducing CNN architecture. Colorful Colorization can achieve a better performance over prior works by leveraging the potential of CNN, which consists of multiple layers of CNN (Zhang et al., 2016).

Deep neural networks can suffer from the gradient vanishing problem, which poses as a hurdle of training deep neural networks. He et al. first propose the skip connection structure for solving the gradient vanishing problem and mark a milestone in deep learning architecture (He et al., 2016). Depth Colorization harnesses the state-of-the-art skip connection structure and outperforms the previous state-of-the-art works (Carlucci et al., 2018).

Notwithstanding, automated colorization remains lacking of authenticity until the onset of Generative Adversarial Network (GAN) (Goodfellow et al., 2014). Ci et al. propose an architecture based on conditional GAN for colorizing artifacts in anime line arts and is state-of-the-art (Ci et al., 2018).

1.6 Our research

We conduct an interdisciplinary research regarding the history of photography, the research of East Asian historical photographs and artificial intelligence. We leverage the potential of state-of-the-art deep learning for colorizing historical photographs from East Asia region. Our architecture benefits from state-of-the-art U-Net and conditional GAN for securing the authenticity of the chromatic distribution of historical photographs (Ronneberger et al., 2015; Goodfellow et al., 2014).

Prior works of colorizing photographs exploit dual losses for delivering better colorization performance. However, the aforementioned state-of-the-art works can suffer from the difficulty of determining the hyper parameter regarding the ratio of the dual loss values. We solve this problem by dynamically decaying the ratio factor exponentially to meet the various importance of the two loss values across the training process. We claim this technique as a major contribution in this research. Our evaluation can further verify the efficiency of this technique.

We also present an extensive literature review regarding automated colorization for monochromatic photographs from the perspective of the timeline of the evolution of automated colorization and neural network architectures. We deem the extensive literature review as another contribution.

1.7 Challenges

We identify the following challenges in automated colorization from our investigation on prior state-of-the-art works:

- Securing the authenticity of chromatic distribution;
- Narrowing the chromatic intensity value difference between prediction and ground truth;
- Determining the ratio of two loss values across the training process.

1.8 Research objectives

In this thesis, we seek to answer the following concerns as our research objectives:

- We explore the architecture of automated colorization on historical photographs. In particular, we concentrate on conditional GAN based colorization architecture in tandem with state-of-the-art U-Net;

- We attempt to construct a hand-crafted data set consisting of multiple categories from landscapes, pedestrians and buildings. We train and evaluate our model on the constructed data set;
- We establish the interdisciplinary connection from the perspective of humanities in relation to history and anthropology and the cutting-edge artificial intelligence by extensively conducting the literature research;
- We investigate the impact of the loss balance on the colorization performance. More specifically, we devise a mechanism of decaying the ratio factor of two loss values exponentially and meet with the various needs on the balance of two loss values across the training process.

1.9 Contributions

We pinpoint and claim the followings as our major contributions:

- We conduct an extensive literature review from both perspectives of humanities and artificial intelligence. As a result, we justify the needs of colorizing historical photographs from multiple aspects;
- We implement the proposed automated colorization architecture with state-of-the-art techniques using Tenserflow;
- We devise a mechanism of controlling the loss balance throughout the training process.
- We make available our implementation of automated colorization to open source community. We expect the open source can inspire the research community for further study.

Chapter 2

Background

2.1 Summary

We conduct an investigation on the related notations regarding color space, Convolutional Neural Network (CNN), U-Net, and conditional Gernerative Adverserial Network (GAN). In Section 2.2, we first summarize the prevailing color space definitions and analyze their applicable scenarios. More specifically, we depict the Lab color space in details and justify the use of Lab color space as the chromatic representation of the proposed automated colorization architecture. In Section 2.3, we recapitulate the development of CNN for later exploitation in which the CNN is a fundamental component in colorization architecture of automated colorization. In Section 2.4, we introduce the U-Net architecture which is crucial for colorizing photographs and can achieve state-of-the-art result. In Section 2.5, we present the final component in our architecture – conditional GAN, which plays an important role for securing state-of-the-art authenticity.

2.2 Color Space

Color space defines a set of the mapping of the nature color and the numerical space (Kuehni, 2001). Various color space definitions have different emphases in scenarios – e.g. CMYK model in press industry and RGB model in computer displays (Baqai et al., 2005; Pascale, 2003). Multiple color space models are therefore proposed for industry over the past decades on various purposes. RGB color space model, HSV color space model and Lab color space model are widely adopted in industry.

In our research, we adopt Lab color space for processing and inference instead of RGB color space, which is widely used in deep learning. Lab color space consists of illuminance component known as L- component, and chromatic components – a- component and b- component. Automated colorization can benefit from the Lab color space due to the predicted information can be decreased to dual components instead of triple components. It agrees on that the training can be hard if the information in prediction is large. Prior works widely use Lab color space to avoid such problem and their works have verified the effectiveness of Lab color space in colorization (Schanda, 2007).

2.2.1 Lab color space

Lab color space was proposed by the International Commission on Illumination (CIE) in 1976 and designed as a perceptually uniform color space. The color changes are similar as the changes in perceived color by humans. The color space consists of three components – L- represents the intensity of perceptual lightness, a- and b- represent the the four unique colors perceived by human eyes – red, green, blue, and yellow (Schanda, 2007).

In our research, the inter-conversion between RGB color space and Lab color space can take place in the colorization pipeline of our architecture. The conversion of RGB color space to Lab color space is implemented by transforming RGB color space to XYZ color space to Lab color space, in which the conversion of RGB to XYZ is computed by:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804275 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.1}$$

and the conversion of XYZ to Lab is given by:

$$L = 116\ f(\frac{Y}{Y_n}) - 16$$

$$a = 500\ (f(\frac{X}{X_n}) - f(\frac{Y}{Y_n}))$$

$$b = 200\ (f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n}))$$

where:

$$f(x) = \begin{cases} \sqrt[3]{t} & \text{if } t > \sigma^3 \\ \frac{t}{3\sigma^2} + \frac{4}{29} & \text{otherwise} \end{cases}$$

in which $\sigma = \frac{6}{29}$ and $t = \frac{X}{X_n}$, $\frac{Y}{Y_n}$ or $\frac{Z}{Z_n}$.

The conversion of Lab color space to RGB color space is computed by transforming the Lab color space to XYZ color space:

$$X = X_n f^{-1}(\frac{L+16}{116} + \frac{a}{500})$$

$$Y = Y_n f^{-1}(\frac{L+16}{116})$$

$$Z = Z_n f^{-1}(\frac{L+16}{116} - \frac{b}{200})$$

where:

$$f^{-1}(t) = \begin{cases} t^3 & \text{if } t > \sigma \\ 3\sigma^2(t - \frac{4}{29}) & \text{otherwise} \end{cases}$$

in which $\sigma = \frac{6}{29}$, and the conversion of XYZ color space to RGB color space is given by:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2404542 & -1.5371385 & -0.4985314 \\ -0.9692660 & 1.8760108 & 0.0415560 \\ 0.0556434 & -0.2040259 & 1.0572252 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.2}$$

.

2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) is pervasively applied in vision tasks from image classification, object detection to semantic segmentation and marks a landmark in deep learning (Goodfellow et al., 2016). Prior to the emergence of CNN, the algorithms of machine learning in vision tasks are hand-crafted in feature extraction, which is domain specific and can be considerable hard to design. For instance, HOG and SIFT features have been popular in image classification and object detection until the emergence of CNN (Dalal and Triggs, 2005; Lowe, 1999). In CNN based networks, the networks can learn the process of feature extraction and outperforms the prior algorithms based on hand-crafted feature extraction.

A convolutional layer is an operation that a convolutional kernel slides across an input image with given stride for outputting a feature map. The number of the parameters of a convolutional kernel is far smaller compared with of fully connected layer. This property can extend huge benefits in training deep neural networks. The convolution computation is denoted as:

$$x^* = x \circledast k$$

where x denotes the input image, k denotes the convolution kernel, x^* denotes the computed feature map from the convolution operation and $\circledast$ denotes convolution operation.

As shown in Figure 2.1, a typical deep learning neural network with CNN as backbone consists of multiple CNN layers which extract features and multiple fully connected layers which serve for further classification.

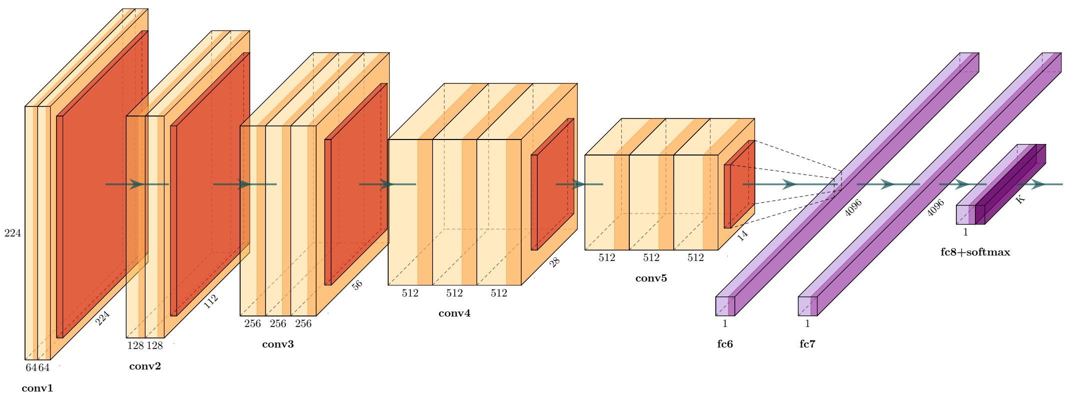


Figure 2.1: The VGG16 architecture, the figure comes from https://github.com/HarisIqbal88/PlotNeuralNet.

Each convolutional layer follows with an activation function for introducing non-linearity. After the activation function, pooling layer and normalization layer can be further attached for better performance. In earlier architectures based on CNN, pooling layer is used to decrease the size of the outputted feature map for reducing the computation for larger feature map. However, pooling layer is eliminated in current architectures due to the availability of more powerful GPU. In current architectures, the hyper parameter stride in convolutional neural network can determine the size of outputted feature map instead of standalone pooling layer.

Batch normalization (Ioffe and Szegedy, 2015) is essential for successfully training deep neural network by constructing new statistical value which transforms to standard normal distribution $\mathcal{N}(0, 1)$ from original feature map. The values of an original feature map can be arbitrarily large which can pose as a problem for software implementation. Training deep neural networks can benefit from batch normalization. A typical computation of batch normalization is given by:

$$\tilde{x} = \frac{x - \bar{x}}{\sigma}$$

where x denotes the values of a feature map, $\bar{x}$ denotes the mean value of the feature map, σ denotes the standard deviation of the feature map and $\tilde{x}$ denotes the normalised feature map by batch normalization algorithm.

2.4 U-Net

U-Net (see Figure 2.2) is a state-of-the-art architecture in biomedical image segmentation using less images in training while achieving better performance and becomes a standard architecture in biomedical image tasks (Ronneberger et al., 2015). U-Net has three novel modifications from fully convolutional network: (1) Skip connections (He et al., 2016) exist between multiple layers in which each layer is corresponding to a specific scale, (2) The network has multi-scale for features and (3) The middle layers contain more channels compared with standard CNN.

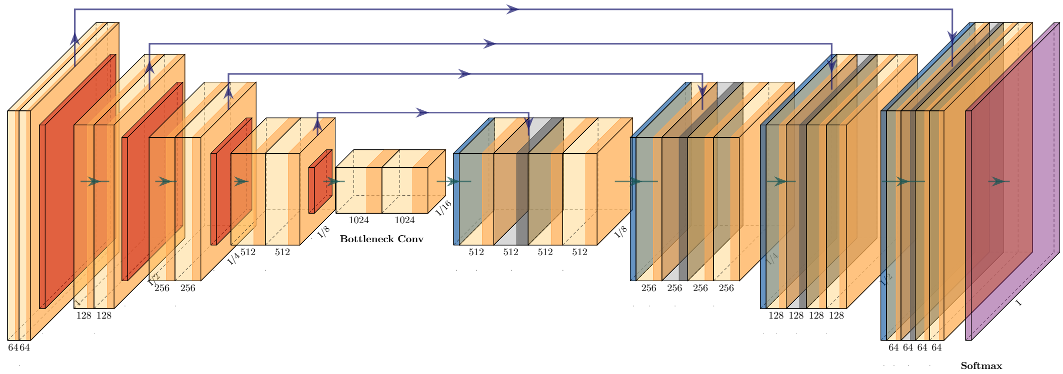


Figure 2.2: A typical U-Net architecture. The figure comes from https://github.com/HarisIqbal88/PlotNeuralNet.

The operations of down-sampling and up-sampling in U-Net can lose information and degrade the performance. A naive U-Net without skip connection can not benefit from the multi-scale design. In the light of skip connection, this problem can be mitigated by passing extra information from former layer with higher resolution to the latter layer suffering from information losing.

Feature extraction can benefit from multi-scale architecture, in which small scale enables the model focusing on global feature while large scale gives more details. Down-sampling creates a set of multi-scale features of the input from large to small respectively. Up-sampling increases the sizes of features to match the desired output dimensions.

The middle layers of multi-scale U-Net contains more channels given same amount of parameters and enable the model being more expressive. Model can benefit from the potential of feature extraction and U-Net can extend such advantages for archiving higher performance with more expressive feature extraction.

2.5 GAN

The emergence of Gernerative Adverserial Network (GAN) (Goodfellow et al., 2014) has profoundly transformed the design of neural networks and has tremendous impact on generative tasks. GAN can secure the authenticity of generative tasks by drawing two distributions close with the idea of zero-sum game. The architecture of GAN consists of generator, discriminator and GAN loss function. The generator is responsible for image generation and the discriminator is responsible for scoring the quality of generated image by generator. Cross entropy loss function is widely used in GAN. Training GAN can be hard and multiple tricks are therefore proposed for improving the training – random label flipping and smoothing label are crucial in successful training. GAN gains huge attention in community and derives multiple variants. Conditional GAN (Mirza and Osindero, 2014) is a such variant and plays an important role in colorization architectures.

Chapter 3

Related works

3.1 Summary

In Section 3.2, we conduct an extensive literature review regarding automated colorization from the perspective of the evolution of the backbone in deep learning. We first summarize the pioneer works prior to the emergence of CNN. We also recapitulate the works using CNN with and without skip connection. At the end of the Section 3.2, we introduce the state-of-the-art architecture from the benefit of U-Net and GAN. We also analyze their limitations in the literature review for extending a deep overview regarding the development of automated colorization.

3.2 Automated colorization

Automated colorization (Larsson et al., 2016) provides an approach of predicting the color distribution from monochromatic input. The research of automated colorization gains tremendous attentions in the deep learning community.

Deep colorization is the first work of fully automated colorization and combining techniques from deep learning (Cheng et al., 2015). This work splits the input image into small patches, extract hand-crafted DAISY features (Tola et al., 2009) and feed the features into multiple layer dense neural network to predict the color distribution. The work also adopts a bilateral filter to improve the output quality. However, this work dose not exploit the potential of the convolutional neural network and relies on hand-crafted features. Colorful image colorization improves the predicted colorization quality by using convolutional neural network (CNN) and trains the model with over a million of color images (Zhang et al., 2016).

Naive CNN has no skip connection which later becomes a standard technique across all architectures and the number of features the CNN can extract is limited. Instead of using trivial CNN without skip connection – which can be hard to achieve realistic colorization distribution from the prior research, Deep Depth Colorization introduces skip connection from residual paradigm to CNN based colorization architecture (Carlucci et al., 2018). The evaluation of this work can verify the effectiveness of the architecture design and the work is state-of-the-art.

In typical CNN architecture, the number of features can be limited given the number of parameters. U-Net creates multi-scale features by down-sampling and up-sampling and can have more features in the middle layers. The architecture of U-Net can benefit certain vision tasks such as colorization. Prior works are state-of-the-art, however, their colorization quality can suffer from lacking of authenticity in color distribution. This problem poses as a challenge in a colorization research community. The emergence of generative adversarial network (GAN) sheds light on this challenge. Ci et al. propose a novel colorization architecture with the benefit from U-Net and GAN and their work is state-of-the-art.

Chapter 4

Data set and pre-processing

4.1 Summary

In the Section 4.2 of Chapter 4, we first characterise the East Asia historical photographs we have chosen for evaluation purpose. In Section 4.3, we summarise the characterization into principles and depict the principles of creating training data set from prior open source data set. In Section 4.4, we create the training data set by applying such principles. In Section 4.5, we state the specifications of our training and evaluating data set.

4.2 Characterizing East Asian historical photographs

The alignment of the historical photographs which are used in our evaluation and the training images can play an important role in the high quality of colorization tasks. The major approach of characterizing East Asia historical photographs is to identify the various categories and the distribution of various categories can be used as reference for choosing our training data set.

We advise the principles from the Berne Convention (Burger, 1988) which is used for the Protection of Literary and Artistic Works (Kalodner and Vance Jr, 1958) and the regulations from European GDPR (Goddard, 2017) for choosing the historical evaluation photographs without the infringement of the copyright and privacy, which is crucial for research community. We choose historical Chinese photographs from the 'Historical Photographs of China' project (Bristol University, 2022). We also use the National Diet Library (Japan National Library, 2022) and Waseda University Library (Waseda University, 2022) as our sources for choosing other East Asian historical photographs.

We limit the scope of the objects in chosen historical photographs into: landscape, building and street from China, Korea and Japan respectively. We choose ten historical photographs for each category with respect to each nation for evaluation purpose. The Figure 4.1 shows the examples we choose for the evaluation purpose, each row is corresponding to a nation and each column is corresponding to a category. The images are chosen from the late nineteenth century to the middle twentieth century. We re-scale the chosen images to 224×224 in dimension with JPEG format.

(a) A landscape in Korea

(b) A building in Korea

(c) A street in Korea

(d) A landscape in Japan

(e) A building in Japan

(f) A street in Japan

(g) A landscape in China

(h) A building in China

(i) A street in China

Figure 4.1: The chosen examples of historical photographs from Korea, Japan and China respectively.

4.3 The principles of creating training data set

We have conducted a succinct characterization for East Asia historical photographs in Section 4.2. We summarise the aforementioned analysis into the following principles:

- **Categories**: We choose three categories in *landscape*, *building* and *street*;
- **Diversity**: We choose tantamount samples from the three categories for balancing the training.

4.4 The process of creating training data set

We use the principles from Section 4.3 to create the training data set. We choose $10,000$ images for each category and combine them as our training data set. We name the created hand-crafted data set as "Triple Category Colorization Dataset" and make it open source available at Kaggle. We hope our contribution can inspire further research in colorization community.

4.5 The specifications of our training and evaluating data set

We depict the specifications of hand-crafted training data – Triple Category Colorization Dataset – and evaluating data set in Table 4.1.

Table 4.1: The specifications of Triple Category Colorization Dataset.

Dataset	Triple Category Colorization Dataset
# of samples	4188
# of channels	3

Chapter 5

Methodology

5.1 Summary

In the Section 5.2 of the Chapter 5, we first introduce the considerations of designing the proposed architecture. In Section 5.3, we depict the overview of the proposed architecture at high level. In Section 5.4, we introduce the design of the data loader which is used for loading images from files and transforming the images into desired format. We also implement image augmentation in the pipeline of the data loader. In Section 5.5, we state the design of the colorization network which is implemented with state-of-the-art U-Net. In Section 5.6, we further state another important component – conditional GAN – which is crucial for the proposed colorization architecture in details. In Section 5.7, we explain the design of the loss functions in the colorization architecture. In Section 5.8, we introduce the mechanism used for monitoring the training process. It has been held for long, knowing the training quality of GAN is hard. We tackle such problem by introducing monitoring network for assessing the training quality. In

Section 5.9, we introduce the design of balancing the two losses in the colorization architecture.

5.2 Considerations

Our proposed architecture consists of a set of state-of-the-art techniques and tricks. Instead of depicting our architecture in a straightforward approach, we first explain the considerations of designing such architecture to justify the choice of candidate techniques and tricks in our research. We summarise our considerations as followings:

- Image augmentation;
- Colorization network;
- Conditional GAN;
- Colorization quality.

Image augmentation: The diversity of images is fundamental for learning the knowledge of the color distribution from the limited number of inputs. Hereby, we devise a dedicated image augmentation (Shorten and Khoshgoftaar, 2019) model to augment the input diversity for further enhancing the training quality. Avoiding trivial solution during the training process is another concern and image augmentation is attested to mitigate such problem (Dosovitskiy et al., 2013). We adopt the following image augmentation methods:

- Random flipping in horizon;

- Random cropping;
- Random rotation.

Colorization network: Colorization network is a key component to automated colorization and the colorization network can largely determine the quality of colorization. As a result, we state our considerations as the followings:

- The candidate architecture of the colorization network is expected to be state-of-the-art;
- The computation should be feasible with limited hardware budget in our research.

Conditional GAN: The colorization quality can suffer from lacking of the authenticity regarding color distribution, we introduce the conditional GAN to improve the colorization quality. The conditional GAN will estimate the colorization quality and predict the authenticity score. The quality score can further be used to optimize the colorization network. However, training conditional GAN is known to be hard. In order to fulfil a successful training, we introduce the following tricks:

- Random label flipping (Kaneko et al., 2019);
- Soft labeling (Gou et al., 2021).

Colorization quality: We identify two factors contribute to the colorization quality in the proposed architecture: (1) The absolute error between the predicted colorization values and the ground truth values and (2) The extent of the predicted color distribution deviates from the ground truth distribution.

5.3 The overview of our architecture

In the light of U-Net and conditional GAN, we combine the two state-of-the-art techniques into our colorization architecture. As shown in Figure 5.1, our architecture consists of the data loader, image augmenter, colorization network and scoring network. We devise a mechanism of dual loss functions for achieving state-of-the-art colorization effects. We also take into account the aforementioned considerations from Section 5.2 to guide our architecture design.

We leverage the data loader framework from TensorFlow to implement our data loader (TensorFlow, 2022). The TensorFlow data loader framework is highly scalable across multiple platforms and such framework can enhance the portability and scalability of our architecture with respect to various data sources. We integrate the image augmentation into our data loader. The image augmentation is indispensable to avoid trivial solutions in the training process. The data loader is also responsible for transforming the original images from RGB space to Lab space and preparing the training pairs of being fed into the pipeline of our architecture.

Colorization network is fundamental to the design of our proposed automated colorization architecture and plays an important role for the ultimate colorization quality. We have conducted an extensive investigation in terms of the colorization architecture. U-Net can deliver state-of-the-art performance in colorization tasks. We leverage the state-of-the-art U-Net as the candidate of the colorization network in our architecture.

U-Net can achieve state-of-the-art colorization quality. However, such architecture can suffer from the lacking of the authenticity of colorization quality. Conditional GAN is a candidate technique of solving this problem. We harness the potential of conditional GAN for further improving the colorization quality to approach the real color distribution.

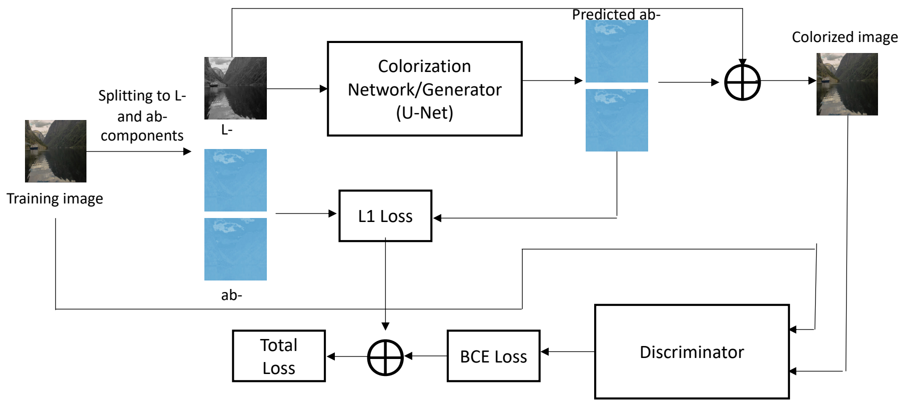


Figure 5.1: The overview of our proposed colorization architecture. The pipeline consists of data loader, image augmenter, colorization network and scoring network. The model uses dual loss functions to deliver state-of-the-art performance.

5.4 The design of data loader

We harness the data loader framework from Keras and inherit the class *keras.utils.Sequence* by reloading the following operations:

- *__init__*: which is responsible for initializing the class;
- *__len__*: which determines the length of the data set;
- *__getitem__*: which is responsible for loading a given image.

We implement three tasks in our data loader: (1) Loading images from files as RGB format, (2) Applying image augmentation transformations on the loaded data from (1) and (3) Converting RGB to Lab and splitting the Lab to L- component and ab- components.

5.5 The colorization network

The design of our colorization network is shown as in Figure 5.2. The colorization network is an U-Net architecture, which consists of five down-sampling blocks, five up-sampling blocks and one bottleneck block. The configuration of U-Net is shown as in Table 5.1. Each down-sampling block contains convolutional layer, Leaky ReLu activation layer, batch-normalization layer and drop-out layer. Each up-sampling block contains convolutional layer, Leaky ReLu activation layer, batch-normalization layer and drop-out layer as well. The bottleneck block contains convolutional layer, Leaky ReLu activation layer and batch-normalization layer.

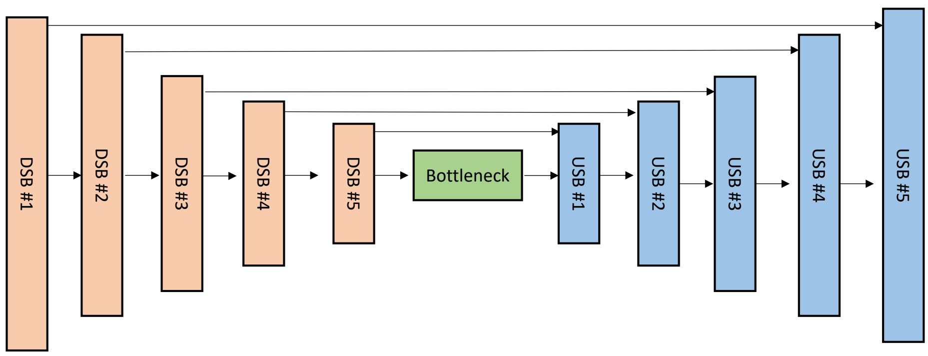


Figure 5.2: The architecture shows the design of our colorization network. In the diagram, the DSB stands for Down-sampling Block and the USB stands for Up-sampling Block. Each DSB or USB can consist of convolutional layer, activation function layer, drop-out layer and batch-normalization layer.

Table 5.1: The configuration of the colorization network U-Net in our architecture.

Block	Layer ID	Layer Type	Parameters
DSB #1	1	Conv2D	filters=32,strides=2,padding = 'same',kernel_size=5
	2	LeakyReLU	alpha=0.2
	3	BN	
DSB #2	4	Conv2D	filters=64,strides=2,padding = 'same',kernel_size=5
	5	LeakyReLU	alpha=0.2
	6	BN	
DSB #3	7	Conv2D	filters=128,strides=2,padding = 'same',kernel_size=5
	8	LeakyReLU	alpha=0.2
	9	BN	
DSB #4	10	Conv2D	filters=256,strides=2,padding = 'same',kernel_size=5
	11	LeakyReLU	alpha=0.2
	12	BN	
DSB #5	13	Conv2D	filters=512,strides=2,padding = 'same',kernel_size=5
	14	LeakyReLU	alpha=0.2
	15	BN	
Bottleneck	16	Conv2D	filters=1024,strides=1,padding = 'same',kernel_size=5
	17	LeakyReLU	alpha=0.2
	18	BN	
USB #1	19	Conv2DTranspose	filters=512,strides=2,padding = 'same',kernel_size=5
	20	Conv2DTranspose	filters=512,strides=1,padding = 'same',kernel_size=5
	21	LeakyReLU	alpha=0.2
	22	BN	
USB #2	23	Conv2DTranspose	filters=256,strides=2,padding = 'same',kernel_size=5
	24	LeakyReLU	alpha=0.2
	25	BN	
USB #3	26	Conv2DTranspose	filters=128,strides=2,padding = 'same',kernel_size=5
	27	LeakyReLU	alpha=0.2
	28	BN	
USB #4	29	Conv2DTranspose	filters=64,strides=2,padding = 'same',kernel_size=5
	30	LeakyReLU	alpha=0.2
	31	BN	
USB #5	32	Conv2DTranspose	filters=32,strides=2,padding = 'same',kernel_size=5
	33	LeakyReLU	alpha=0.2
	34	BN	
Final	35	Conv2DTranspose	filters=2,strides=1,padding = 'same',kernel_size=3,activation = 'tanh'

5.6 The conditional GAN

The design of the conditional GAN is shown as in Figure 5.3 and the Figure 5.1. We deem the U-Net as the generator of the GAN and combine a patch based discriminator into the ultimate GAN architecture. The configuration of discriminator is shown as in Table 5.2. The discriminator consists of four blocks, in which the first three blocks consist of convolutional layer, Leaky ReLU layer, drop-out layer and batch-normalization layer respectively and the final block merely consists of convolutional layer.

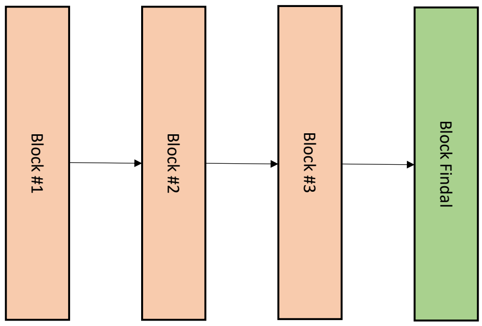


Figure 5.3: The architecture shows the design of our conditional GAN. Block #1 consists of convolutional layer, activation function layer and drop-out layer. Block #2 consists of convolutional layer, activation function layer, drop-out layer and batch-normalization layer. Block #3 consists of convolutional layer, activation function layer, drop-out layer and batch-normalization layer. The final block merely consists of convolutional layer.

Table 5.2: The configuration of the discriminator in our architecture.

Block	**Layer ID**	**Layer Type**	**Parameters**
Block #1	1	Conv2D	filters=32,strides=2,padding = 'same',kernel_size=5
	2	LeakyReLU	alpha=0.2
	3	Dropout	p=0.3
Block #2	4	Conv2D	filters=64,strides=2,padding = 'same',kernel_size=5
	5	LeakyReLU	alpha=0.2
	6	Dropout	p=0.3
	7	BN	momentum = 0.1,epsilon = 1e-05
Block #3	8	Conv2D	filters=128,strides=2,padding = 'same',kernel_size=5
	9	LeakyReLU	alpha=0.2
	10	Dropout	p=0.3
	11	BN	momentum = 0.1,epsilon = 1e-05
Final	12	Conv2D	filters=1,strides=1,padding = 'same',kernel_size=5
	13	Flatten	

5.7 The loss design

The design of loss functions is a gist to secure colorization quality. We identify two factors contributing to the ultimate colorization quality: (1) The distance norm between ground truth and the prediction and (2) The color distribution.

$\mathbb{L}1$ measure in colorization is widely used and achieves state-of-the-art performance. In our research, we also adopt $\mathbb{L}1$ measure in our colorization network to measure the distance of the prediction deviates from the ground truth. We formulate the distance as the following:

$$d(x_{a,b}, \hat{x}_{a,b}) = ||x_a - \hat{x}_a||_1 + ||x_b - \hat{x}_b||_1$$

where $d(\cdot,\cdot)$ denotes the distance sum, $x_{a,b}$ denotes the ground truth of the ab-components of input images, $\hat{x}_{a,b}$ denotes the prediction of ab- components of given input images, x_a denotes the ground truth of a- component in $x_{a,b}$, $\hat{x}_a$ denotes the ground truth of a- component in $\hat{x}_{a,b}$, x_b denotes the ground truth of b- component in $x_{a,b}$ and $\hat{x}_b$ denotes the ground truth of b- component in $\hat{x}_{a,b}$.

The second major factor affecting the colorization quality is the distance of the color distribution of the prediction of given input images and the samples from training data set. However, it is infeasible to estimate the distance by using analytical approach in mathematics. We harness the potential of GAN to estimate the distribution distance and the training process can optimize the discriminator, which can learn how to estimate the distribution distance.

5.8 The mechanism of monitoring the training process

Training GAN based models is well known to be rather difficult because there are no indicators of tracking the training process. We design a novel mechanism of monitor the training process as an auxiliary facility to improve the efficiency of model debug and development. The auxiliary facility consists of two layers: (1) A dense layer with ReLU as activation function, and, (2) A linear dense layer combining to one score the patch scores from the discriminator of the conditional GAN. The auxiliary model is responsible for scoring the training process and displays the scores as an important indicator to the undertaking training. The model can learn to score the training quality in the training process.

5.9 The dynamic loss balancing

We have conducted experiments regarding the loss balancing and have a finding that at various training stages the needs of the ratios of the loss balancing can vary from stage to stage. Furthermore, regarding the successful training in GAN, we hope the generator can be stronger than the discriminator. In addition, we also hope at the initial stage of training process, our architecture can focus on the learning of the colorization by only using $\mathbb{L}1$ measure. We take into accounts the aforementioned concerns regarding successful training and devise a dynamic mechanism for balancing two loss values across the training process. The loss value weights are computed as the following:

$$\alpha_{1,k} = \alpha_0^k$$

$$\alpha_{2,k} = 1 - \alpha_0^k$$

$$\mathbb{L} = \alpha_{1,k}\mathbb{L}_1 + \alpha_{2,k}\mathbb{G}$$

where α_0 denotes the initial value of the weight of the $\mathbb{L}_1$ loss and we initialize this value to 0.9999, $\alpha_{1,k}$ denotes the value of α_1 at the k-th epoch, $\alpha_{2,k}$ denotes the value of α_2 at the k-th epoch, $\mathbb{L}_1$ denotes the $\mathbb{L}_1$ distance of the prediction and the ground truth and $\mathbb{G}$ denotes the GAN loss. The two alpha values will be updated at each epoch during the training process by using the framework from TensorFlow.

Chapter 6

Experimental settings

6.1 Summary

In the Section 6.2 of Chapter 6, we first depict our experimental specifications of software and hardware. In the Section 6.3, we introduce the settings of the colorization experiments on training data set. In the Section 6.4, we introduce the settings of the colorization experiments on testing data.

6.2 Test bed

Our experimental settings are shown as in Table 6.1. We install Anaconda version 3 on a Mac Book Pro 2018 which equips with 2.3 GHz Dual-Core Intel Core i5, 8 GB 2133 MHz LPDDR3 and runs macOS Monterey with version 12.1. We develop our models by using the Jupyter version 3.0.14 as our development environment. The Jupyter runs with Python version 3.8.8 compiled with Clang 10.0.0. The deep learning frameworks inside the Jupyter are TensorFlow version 2.6.0 and Keras version 2.6.0.

We follow the principle of the experimental reproducibility in scientific research. We initialize the random seeds to 2022 for scientific reproducibility.

Table 6.1: Experiment environment

Operating System	macOS Monterey with version 12.1
Processor	2.3 GHz Dual-Core Intel Core i5
Memory	8 GB 2133 MHz LPDDR3
Storage	120GB Apple NvME SSD
Python version	3.8.8
Tensorflow version	2.6.0
Keras version	2.6.0
Random seed	2022

6.3 The colorization experimental settings on training data set

We design this experiment to assess the training quality by comparing the prediction of given images from the trained model with their ground truth.

Firstly, we randomly – we fix the random seeds – choose 30 samples from the training data set for evaluation purpose. We scale the chosen samples to 256×256 with three channels in Lab space. We then split the chosen samples into L- component as evaluation input and ab- components as control. We normalize all components to the range of $[0, 1]$.

Secondly, we feed the L- component of the chosen samples to the trained model and the model outputs the predicted ab- components for the given input. We de-normalize the outputs to the values in Lab space. We combine the de-normalized ab- components with the L- component to form the predicted full color images.

Lastly, we use human-guided approach to assess the training quality by comparing the predicted full color images and their corresponding ground truth versions.

6.4 The colorization experimental settings on historical photographs

We design this experiment to evaluate the potential of our architecture on historical photographs. If the training is successful, we hope the colorization can exhibit a good performance for historical photographs.

Firstly, we randomly – we fix the random seeds – choose 40 samples from the training data set for evaluation purpose. We scale the chosen samples to 256×256 with three channels in Lab space. We then split the chosen samples into L- component as evaluation input and ab- components as control. We normalize all components to the range of $[0, 1]$.

Lastly, we feed the L- component of the chosen samples to the trained model and the model outputs the predicted ab- components for the given input. We de-normalize the outputs to the values in Lab space. We combine the de-normalized ab- components with the L- component to form the predicted full color images.

Chapter 7

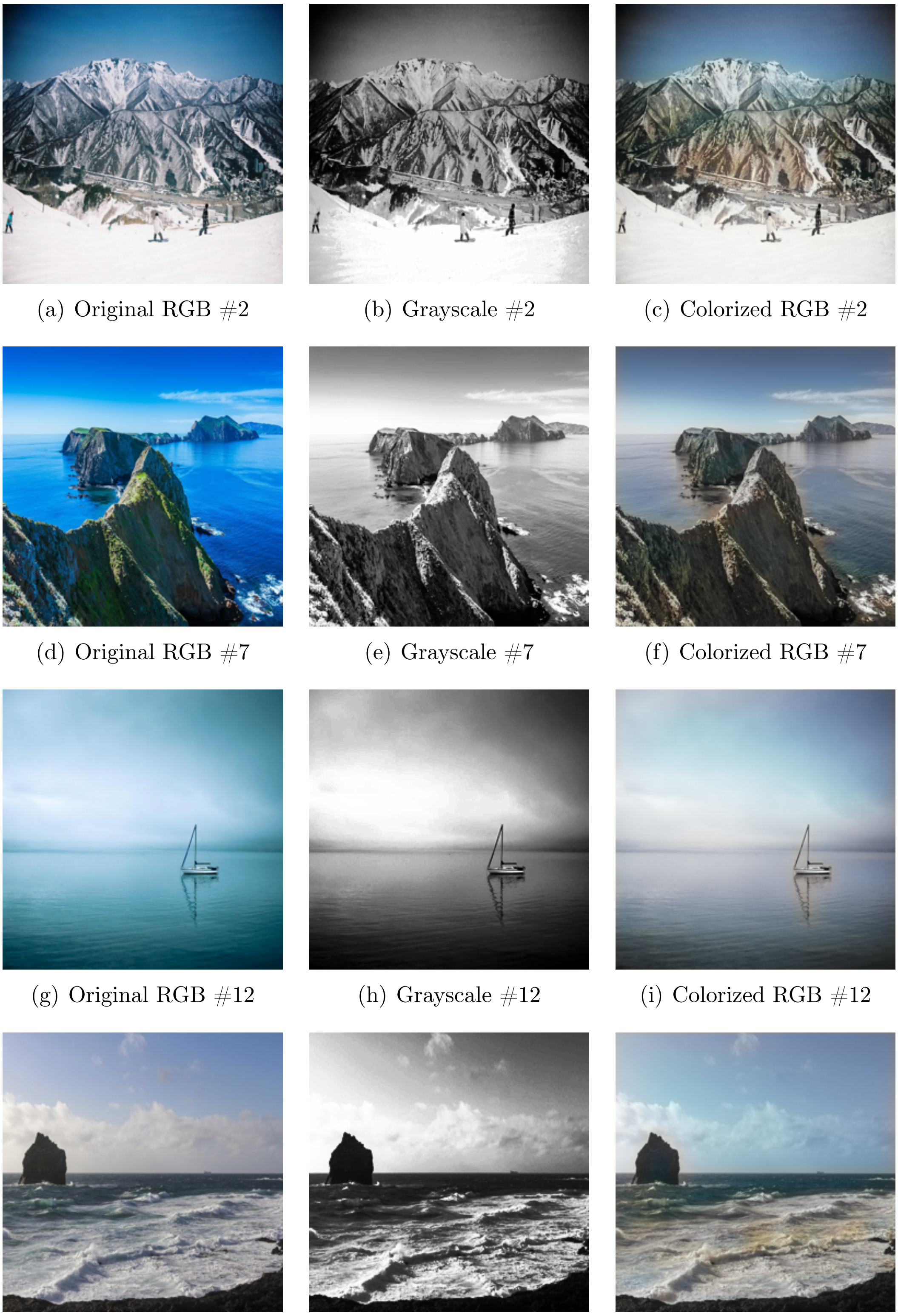
Results and analysis

7.1 Summary

In the Section 7.2 of Chapter 7, we first demonstrate the results of the colorized images on training data set and conduct a brief analysis regarding the results. In the Section 7.3, we show the ultimate results of our architecture on historical photographs and also conduct a brief analysis regarding the results.

7.2 The results on training data set

The Figure 7.1 shows the results of the colorized images chosen from the training data set. Each row shows a triple tuple of the ground truth version, the L-component version and the colorized version.



(a) Original RGB #2 (b) Grayscale #2 (c) Colorized RGB #2

(d) Original RGB #7 (e) Grayscale #7 (f) Colorized RGB #7

(g) Original RGB #12 (h) Grayscale #12 (i) Colorized RGB #12

(j) Original RGB #27 (k) Grayscale #27 (l) Colorized RGB #27

Figure 7.1: Each row is a triple tuple consisting of full color ground truth version, gray scale L- component version and colorized version.

7.3 The results on historical photographs

The Figure 7.2 shows the results of the colorized images chosen from the historical photographs. Each row shows two pairs and each pair consists of the colorized image and its gray scale version.

(a) Historical photograph #5

(b) Colorized photograph #5

(c) Historical photograph #7

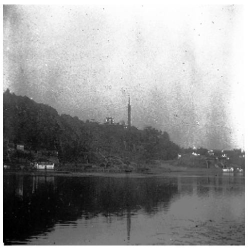
(d) Colorized photograph #7

(e) Historical photograph #16

(f) Colorized photograph #16

(g) Historical photograph #26

(h) Colorized photograph #26

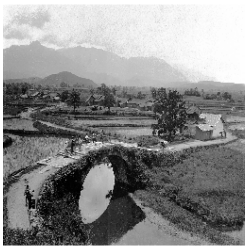


(i) Historical photograph #32

(j) Colorized photograph #32

(k) Historical photograph #35

(l) Colorized photograph #35



(m) Historical photograph #36

(n) Colorized photograph #36

(o) Historical photograph #37

(p) Colorized photograph #37

Figure 7.2: The colorized examples of choosing historical photographs.

Chapter 8

Conclusion

8.1 Summary

Automated colorization technique sheds light on multiple interdisciplinary subjects from history research to anthropology research to social science research. Such technique can also inspire the public to be aware of our past vivid world.

We first highlight the impact of the invention of the photography on human society and the history can justify our research. We recapitulate the relevant topics and challenges regarding automated colorization. We also conduct an extensive literature review regarding our research topic. We create a hand-crafted data set for the training purpose and evaluation purpose. We implement our architecture and evaluate on our own data set.

Chapter 9

Further work

Our research can be further improved by using novel cutting-edge techniques in colorization network and harnessing the pre-trained weights on ImageNet data set which has more than 1.2 million images, and using state-of-the-art GAN such as spectral normalization GAN.

References

Farhan A Baqai, J-H Lee, A Ufuk Agar, and Jan P Allebach. Digital color halftoning. *IEEE Signal Processing Magazine*, 22(1):87–96, 2005. 10

Walter Benjamin. A short history of photography. *Screen*, 13(1):5–26, 1972. 3

Bristol University. Historical Photographs of China. `https://www.hpcbristol.net`, 2022. Online; accessed 22 February 2022. 21

Peter Burger. The berne convention: its history and its key role in the future. *JL & Tech.*, 3:1, 1988. 21

Fabio Maria Carlucci, Paolo Russo, and Barbara Caputo. Co: Deep depth colorization. *IEEE Robotics and Automation Letters*, 3(3):2386–2393, 2018. 6, 18

Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015. 5, 18

Yuanzheng Ci, Xinzhu Ma, Zhihui Wang, Haojie Li, and Zhongxuan Luo. User-guided deep anime line art colorization with conditional adversarial networks. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1536–1544, 2018. 6, 18

Harriet EW Crawford. *The Sumerian World*. Routledge London, 2013. 2

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005. 12

Alexey Dosovitskiy, Jost Tobias Springenberg, and Thomas Brox. Unsupervised feature learning by augmenting single images. *arXiv preprint arXiv:1312.5242*, 2013. 26

Kemal Ebcioğlu and Erik R Altman. Daisy: Dynamic compilation for 100% architectural compatibility. In *Proceedings of the 24th annual international symposium on Computer architecture*, pages 26–37, 1997. 5

Michelle Goddard. The eu general data protection regulation (gdpr): European regulation that has a global impact. *International Journal of Market Research*, 59(6):703–705, 2017. 21

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 6, 16

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. 12

Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021. 27

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6, 14

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 14

Japan National Library. National Diet Library. `https://www.ndl.go.jp/en/`, 2022. Online; accessed 22 February 2022. 21

Howard I Kalodner and Verne W Vance Jr. The relation between federal and state protection of literary and artistic property. *Harv. L. Rev.*, 72:1079, 1958. 21

Takuhiro Kaneko, Yoshitaka Ushiku, and Tatsuya Harada. Label-noise robust generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2467–2476, 2019. 27

Rolf G Kuehni. Color space and its divisions. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 26(3):209–222, 2001. 10

Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *European conference on computer vision*, pages 577–593. Springer, 2016. 17

David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999. 12

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 16

John D Mollon. Color vision. *Annual review of psychology*, 33(1):41–85, 1982. 4

Beaumont Newhall. *The history of photography*. Museum of modern art New York, 1982. 3

Danny Pascale. A review of rgb color spaces... from xyy to r'g'b'. *Babel Color*, 18:136–152, 2003. 10

David Regan. *Human perception of objects*. Sunderland, MA: Sinauer, 2000. 4

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 14

János Schanda. *Colorimetry: understanding the CIE system*. John Wiley & Sons, 2007. 10, 11

Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019. 26

Alison K Surridge, Daniel Osorio, and Nicholas I Mundy. Evolution and selection of trichromatic vision in primates. *Trends in Ecology & Evolution*, 18(4):198–205, 2003. 4

TensorFlow. TensorFlow Data Generator. `https://www.tensorflow.org/api_docs/python/tf/keras/utils/Sequence`, 2022. Online; accessed 22 February 2022. 28

Nicholas Thomas, Patricia Spyer, James L Hevia, and James T Siegel. *Photographies East: The camera and its histories in East and Southeast Asia*. Duke University Press, 2009. 4

Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):815–830, 2009. 18

Waseda University. National Diet Library. `https://www.waseda.jp/library/en/`, 2022. Online; accessed 22 February 2022. 21

Zhuang Wubin. *Photography in Southeast Asia: a survey*. Nus Press, 2016. 4

Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016. 5, 18