Colorizing Grayscale CT Images of Human Lung Using Deep Learning

Yuewei Wang

A thesis submitted to the Auckland University of Technology
in partial fulfilment of the requirements for the degree of
Master of Computer and Information Sciences (MCIS)

2021

School of Engineering, Computer & Mathematical Sciences

Abstract

The COVID-19 epidemic has broken out globally and will influence human healthy history perpetually. We notice that Computed Tomography (CT) images are grayscale ones from the point of view of digital signal visualization. Therefore, we contemplate whether the automatically rendering colours for the CT lung images via deep neural networks will contribute to diagnosing any diseases for medics.

At present, our motivation for image colorization is inspired by the advancement of associated techniques, such as machine learning and artificial intelligence (AI), especially video analogies and transfer learning in the domain of deep learning. In this thesis, we experimented with two deep learning networks in completely distinct orientations for implementing the most reliable outcomes for colorizing the CT lung images: VGG-19 and ResNet based on exemplar colorization full-automatic colorization, respectively. For hybrid colorization, we select appropriate reference images so as to combine the style and content of the representations to colorize the target CT lung grayscale images. The colours of meat resemble those of human lungs, so the images of fresh pork, lamb, beef, and even rotten meat (for infected lung) are collected for the hybrid colorization model. Moreover, three sets of training data consisting of painting and meat images are analysed to extract the per-pixel erudition for colorizing the greyscale CT lung images for the fully automatic approach.

Pertaining to the results, we consider numerous methods (human visual analysis, PSNR, and SSIM) to evaluate the proposed deep neural network models. Compared with other techniques of colorizing CT lung images, the results of rendering the CT lung images by using deep learning are significantly genuine. The adoption of deep learning is a striking and adventurous endeavours.

Keywords: Colorization, deep learning, CNN, CT lung images

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: Date: 03 March 2021

Acknowledgment

First of all, I would like to appreciate my supervisor Wei Qi Yan, who provided me with extremely professional guidance. I also have benefited from the theoretical lectures delivered by my supervisor regularly. He offered timely answers and continuous spiritual encouragement through the Internet during lockdown time, which was the greatest motivation to complete this MCIS thesis.

Moreover, I would like to say thanks for the learning facilities provided by the School of Engineering, Computer & Mathematical Sciences, Auckland University of Technology (AUT), New Zealand, such as libraries and laboratories.

Last but not the least, I would like to thank my family for their encouragement in my studies and provide financial support for my research project while living in New Zealand.

Yuewei Wang

Auckland, New Zealand

February 2021

Chapter 1
Introduction

This chapter is essentially split into five sections. The first section primarily clarifies the background and motivation of automatic colorizing of CT lung images from reference images and training datasets by applying deep learning methods. In Sections 1.2 and Section 1.3, we remark on the research problems to be investigated in this thesis and proffer notable contributions to the domain of our research. In Section 1.4, we explicate the significance of this research project and its implementation. Finally, the detailed content of this thesis and the context of each chapter will be sketched in Section 1.5.

1.1 Background and Motivation

Since the end of 2019, the COVID-19 epidemic has thoroughly broken out across the world, which generates everlastingly influences beyond the imagination of our public. For example, our ordinary activities are restricted due to the COVID-19 remarkably contagious nature. Until now, more than 100 million verified cases in the whole world. All clinical trials reflect that human lungs are fundamentally destroyed by the coronavirus.

Computed Tomography (CT) defines a diagnostic imaging examination, which was previously called Computed Axial Tomography (Buzug, 2011). Precisely, 'C' in the CT denotes the abbreviation of computed calculation and 'T' stands for tomography abbreviation for tomography (Kalender, 2006). Although CT also applies X-ray, plain film X-ray is directly scanned and projected images, CT lung images are manifested through computer calculations. A plenty of professional studies have defined that contagious lungs are split into a total of four stages according to the degree of lung damage through inspecting the CT lung images, early-stage, progressive stage, peak stage, and absorption stage, respectively (Pan et al., 2020) (Xia et al., 2020). These stages reveal that COVID-19 has the various scope and extent of impact as the disease of condition worsens.

Under normal and equal circumstances, the texture of CT lung images of infected patients is recognized as more complicated and more confusing than the healthy CT lung images (Fan et al., 2020). Specifically, the multiple patches and strips are visible with slightly higher density shadows in the infected CT lung images. Moreover, the boundaries of COVID-19 CT lung images are unclear, and the density of the lesions are approximately uneven ground glass changes. A considerable amount of CT lung images have emerged on the Internet, we assumed that colorization in the domain of deep learning would be a brave and unique implementation of the grayscale CT lung images.

Deep learning has been developed swiftly in the past five years, a portion of representation in machine learning focused on artificial neural networks. Markle created the method of rendering colours to monochrome movies and television programs in 1970, where the word “colorization” was introduced (Burns, 1997). Practically, the initial assumption of colorizing greyscale images and videos is an extremely ancient notion in early 1902 (Kuehni, 2002). Nevertheless, a troublesome dilemma appears if the new coloured image accurately restores the original ideal colour of the original creators because it is possible to misunderstand the expressions of works without proper restoration technology.

The compelling actions to make colorization come true in the last two decades due to machine learning are utilized in a wide range. Colorizing grayscale images and videos is an extremely intriguing direction to explore with machine learning development nowadays (Camastra & Vinciarelli, 2015). For example, transfer learning is a crucial component of colorization that renders a target image from reference images swiftly and stably. However, rare researchers perceive to colorize the grayscale of CT lung images which still are observed from the monologing of the visualization of digital signals for medical personnel.

CT lung images strikingly are employed to show the human lung situation no matter for a patient or a healthy person. Convolution Neural Networks (CNNs or ConvNets) provide the most superior accuracy in multiple image recognition, including object recognition (Alexandre, 2016) (Cao, Chen, & Khosla, 2015) (Ba, Mnih, & Kavukcuoglu, 2014) (Anwar, Hwang, & Sung, 2015), segmentation (Durand, Mordan, Thome, & Cord, 2017) (Yu et al., 2017) (Cheang, Chong, & Tay, 2017), image superresolution (Dong, Loy, He, & Tang, 2014) (Dong, Loy, He, & Tang, 2015) (Wang, Liu, Yang, Han, & Huang, 2015) and object detection (Lin et al., 2017) (Ren, He, Girshick, Zhang, & Sun, 2016) (Bell, Zitnick, Bala, & Girshick, 2016), and so forth.

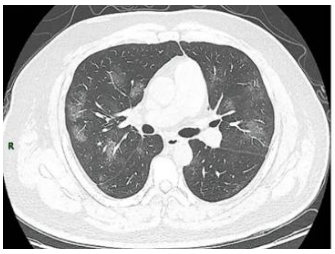
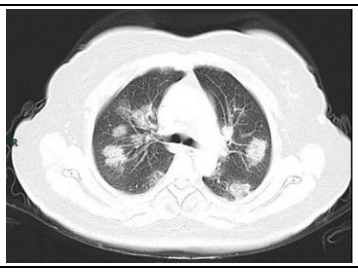
1.2 Research Questions

The foremost research question of this thesis is to colorize the grayscale CT lung images with two distinct approaches in the field of colorization from deep learning. In this research project, we have the following issues that require to be resolved.

In order to accomplish more reliable performance in these ultimately generated panoramic CT lung images, we propose two essentially popularized approaches to generate colourful CT lung images. The initial method is to assemble a diversity of training datasets to fully-automatically colorize the greyscale CT lung images by practicing the ResNet model (Wu, Shen, & Van Den Hengel, 2019). Simultaneously, another system is proposed to utilize the varying reference images to transfer the style and content to the target CT lung images through VGG-19 network (Aimar et al., 2018). VGG-19 and ResNet are both implemented in CNN for deep learning. The ultimate consequences are essentially compared via human eyes and examined with several full-reference quality methods.

Pertaining to the target images, the grayscale CT lung images, we merely select the images of the two stages of COVID-19 infection to analyse in this thesis. Moreover, the CT lung images from healthy people are additionally implemented in our research project to obtain the appropriate colorization method for the image with different status via deep neural networks. Table 1.1 illustrates the CT lung images from healthy people and infected patients, and we are able to observe a plethora of discrepancies from the CT lung images.

Table 1.1 The target CT lung images in this thesis

Healthy CT lung image	COVID-19 infected CT lung image_01	COVID-19 infected CT lung image_02
		

In terms of collecting reference images and datasets, we implemented multiple metrics according to the requirements of each experiment. The reference images are ones we took by using a camera from our supermarket, such as images of pork, lamb and steak. Meanwhile, we stored the meat outside for a long term to gain rotten meat as one of the reference images because the actual condition of the infected lung is not the same as the healthy lung, the colour of the contaminated lung is in the detrimental condition.

Regarding the collected datasets, from the two perspectives (content and quantity) we analyse the datasets if they influence the ultimate experimental results. Firstly, we applied 2,129 painting images as one set of data from the Kaggle website. Additionally, we collect 2,100 meat images individually with a similar quantity to the painting data. In analogous datasets, we implemented 30 meat images, which are vastly divergent from the 2,100 images of the meat dataset for comparisons.

1.3 Contributions

We would like to list the five contributions of this thesis as:

- A novel way to restore the authentic condition of organs in the human body

We colorize greyscale CT lung images that hopefully have the colours as same as human tissues. At present, the professional doctors adjust the width and positions of windows to

enhance the image contrast for CT images, which assist medics visibly distinguish complex tissues in the greyscale image to increase the success rate of diagnosis. However, no medical experts has established that the colourful CT images positively influence doctors' efficiency and accuracy when observing the CT images to diagnose the diseases. As a fact, it is not only for doctors to check the health status of organs but also for patients to comprehend the exact condition of their body without any professional knowledge by simply inspecting the CT lung images. In this situation, the greyscale CT lung images are unquestionably troublesome without any unavoidable medical information to understand and imagine the original condition of organs thoroughly.

In this thesis, we plan to modify this circumstance for patients and doctors and also assist them in inspecting the CT lung images efficiently and expeditiously.

- Framed two colorization models and collected numerous relevant reference images and datasets

The hybrid colorization method and CNN-based colorization method are executed in this research project to compare which method effectively achieves substantial reliable performance in colorizing the grayscale CT lung images. Furthermore, a diversity of reference images and datasets for these two approaches are allocated to select more trustworthy outcomes for obtaining suitable colour CT lung images. Consequently, we configured two sets of classification models, collected numerous associated reference images and datasets.

- The ultimate consequences contribute specific and cloudless opinions for prospective researchers

In the method applied to the reference image, we have introduced multiple valuable suggestions and intend to conduct complementary studies on colorized CT lung images. For example, what kind of meat image is more suitable for colorizing CT lung images；

which technology to practise in this field is more likely to produce high-value results. We also explored detailed information in the experimental design. For example, whether the content and quantity of the data set affect the ultimate colour CT lung image generation or not.

- The state-of-the-art perspectives for the prospective contributions of deep learning in the field of colorization

This thesis also contributes to multiple ways for exploring new solutions in security checks at airport, ferry and railway stations so that security personnel could detect colourized X-ray images. For instance, clothing has an exclusive colour instead of the overall grey screen. We expect this thesis could spur and motivate more researchers to have a productive and efficient work environment.

- A strike to consolidate the connection with medical diagnosis via deep neural networks

Hereinafter, the colourful CT lung images are generated by various researchers in a completely different approach. Deep neural networks are considered as the most forward-looking technique to study colorization in the computational domain. However, these two disciplines are separated, it is the first time they emerged in this thesis project. There is no confirmation if the deep neural networks have the ability to restore the colours of human tissues. This is a bold endeavour to strengthen the combination of CT lung image colorization and deep learning.

1.4 Objectives of This Thesis

The principal objective of this thesis is to colorize the monochrome CT lung images through two separate deep learning networks in our experimental project. It is not the

initial event that researchers dedicate to colorize CT lung images. Homogeneously, deep learning schemes to colorize grayscale images and even videos have been extensively performed by numerous research groups. However, the unification of these two domains is the fearless and intrepid trial in the chronicle.

After reviewed the relevant literature on colorizing black-and-white images and the medical techniques practised to colorize CT lung images, the two standard representative colorization methods are determined to fulfil our work by transferring the reference image style and content automatically, respectively. From the viewpoint of human visual system and mathematical evaluations, full reference image methods are associated with analysing the ultimate consequences of the experiments.

All the processes of implementing models and the outcomes of the analysis are explicitly introduced in this thesis, for example, which kind of emthods are joined in the model; why this method is more authentic than other systems. The limitations of our experiments are interpreted in details, provided specific research directions for future researchers engaged in the field of colorizing CT lung images.

1.5 The Structure of This Thesis

This thesis is organized into seven chapters: Introduction, literature review, CNN-based method, hybrid method, experimental results, resultant analysis and discussions, conclusion and future work.

The first two chapters (i.e., introduction and literature review) reveal the motivations of this research work, introduce the technology developed based on colorizing the CT lung images, and apply all colorization strategies.

This chapter elaborates a fully-automatic model to train the three collections of data for obtaining the excellent colour CT lung images. The architecture of the entire model is

detailly described in this chapter with the most superior technique, which is fitting for our model in this experiment by considering the colorization.

The next chapter concentrates on the hybrid colorization method, which combines the deep neural network and the exemplar-based colorization method. This chapter demonstrates the referenced images are compared with the two distinct deep neural networks, VGG-16 net and VGG-19 net. Moreover, as same as the former chapter, the thoroughgoing structure of this model transfers the style and content of reference images to the CT lung images. Additionally, the loss function is merely effectively practised in this hybrid colorization method to discover which meat image is more suitable for colorizing CT lung images than the other meat images in our research.

Subsequently, our experimental results are demonstrated in a tabular manner and promptly observed the distinctions of diverse reference images and varying datasets. However, the hybrid colorization is laborious to discover the vast variations among the diverse reference images via our human visual system. In order to correlate and compare these two approaches in various reference systems, we introduce the image quality evaluation methods PSNR and SSIM to examine which scheme performs the more authentic from the eventual colour CT lung image results.

The next chapter (Analysis and Discussions) specifically analyses the PSNR and SSIM values that adequately correlate with these two distinctive methods. Furthermore, we discuss the complete findings and the reasonable explanations for the experimental results.

The conclusive chapter (Conclusion and Future Work) points out the research limitations and future work to replicate the listed limitations for anticipated more reliable models and widespread implementations in real-world life.

Chapter 2
Literature Review

With a comprehensive examination of the research questions and thoughtful reviews of the previous studies, in this chapter, we concentrate on the performance of four approaches to explore the problem of image colorization, including scribble-based colorization, exemplar-based colorization, learning-based colorization, and hybrid methods of colorization. Due to the field of colourizing CT lung images being involved in different domains, the detailed information is described initially in this chapter. For the method of deep learning, we implemented our experiments, the CNN-based colorization method and the hybrid colorization method are introduced in the following sections.

.

2.1 Introduction

Before reviewing the literature about monotonous image colorization in the scope of deep learning, we would like to discuss relevant experiments for colourizing the lung CT grayscale images that are entirely inconsistent with the principles of deep learning. Four principal approaches have been proposed to explore the field of colorization in terms of scribble-based colorization, exemplar-based colorization, learning-based colorization, and hybrid methods of colorization. Our methods required reference images and datasets, the deep learning-based method is split into two sections, CNN-based colorization and hybrid colorization. Therefore, this chapter is in total grouped into six categories to introduce the relevant research outcomes in the past two decades.

2.2 Relevant CT Lung Images Colorization

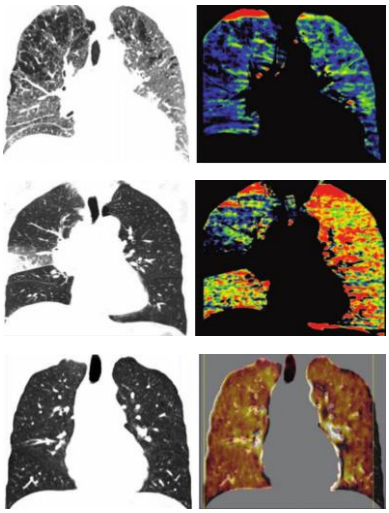
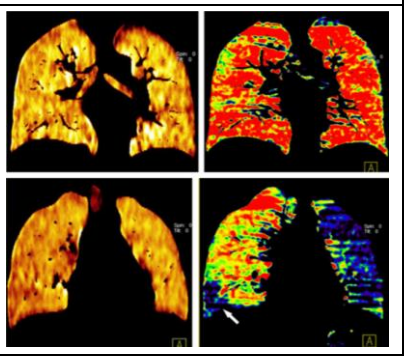
Nowadays, the relevant methods to enhance colours in the CT lung images have been developed by DECT (Dual-energy CT), which unveils the morphological and functional information of the lung in the states of various pulmonary diseases, especially a single contrast-enhanced CT examination (Johnson, 2012). The advantages of this technique are implemented in the clinic applications, such as radiation dose decrease, symptomatic performance enhancement, discovery and characterization of diseases, and image property optimization. However, DECT shortages still exist and demand further investigation (Bongartz et al., 2015).

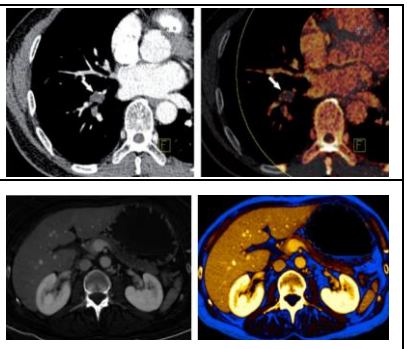
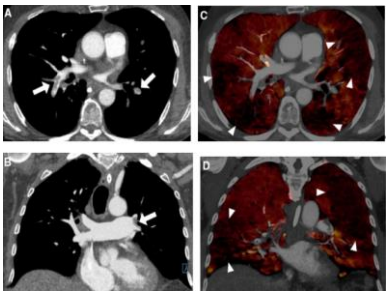
Firstly, DECT is in the infancy, which has enormous room to be explored and developed. Secondly, DECT is restricted in morbidly obese patients because of high image noises regularly intervening with CT lung image analysis. However, with the continuous improvement of living standards and the addition of social pressure, the number of obese patients is also soaring in the whole world (Bleich, Cutler, Murray, & Adams, 2008). Thirdly, the noises of a virtual non-enhanced image are also a limitation

of DECT, which lead to the lack of tiny calcifications. Moreover, a large number of imaging data with DECT are demanded, for instance, 80/100 kVp, 140 kVp, and average weighted virtual 120 kVp and related material- and energy-specific images (Marin, Mileto, & Nelson, 2014). It implies that increasing the storage capabilities of data is required, suitable clinical workflows are essential (Lu, Zhao, Zhang, & Schoepf, 2012) (Zhang et al., 2013) (Zhang et al., 2013) (Marin, Boll, Mileto, & Nelson, 2014).

Table 2.1 illustrates the outcomes in the related work. There are two columns of images we observe from Table 2.1 in each row; the left-hand images are the original images, the right-side ones have eventually generated images after applying the corresponding method.

Table 2.1 The experimental outcomes of the related work

Authors	Time	Methods	Outcomes
Lu et al.	2012	Dual-energy CT	
Zhang et al.	2013	Dual-energy CT	

Marin et al.	2014	Dual-energy CT	
Martin et al.	2018	Dual-Energy CT Pulmonary Angiography	

2.3 Scribble-Based Colorization

In the early stage of colorizing development, the method of scribble-based colorization is universally implemented, which is recognized as a wearisome, time-consuming, and expensive task due to typically executed by manual operation in Photoshop (PS) and other software (Brown, 1990). Consequently, another name is much aptly expressed as human-labeled color scribbles. One grayscale image regularly demands one month to be colorized; for instance, processing face solely requires up to 20 layers of pink, green and blue shades in order to colorize it accurately (Pandey & Sharma, 2019).

A YUV color space model was proposed, which provides the undeniable information of colorization (Jack, 2001), where Y indicates the monochromatic luminance channel (simply named as intensity), simultaneously, U and V are defined as the chrominance channels and the color respectively. This is the earliest time during colorizing the greyscale images. Levin et al. (2004) addressed that contiguous pixels with comparable luminance have a similar color to explain a Markov random field (MRF) for generating rare scribble colors. The experimental outcome has a high-value performance. The

specific operations of this approach are to scribble several colors in an image in the user-interactive interface and then colorize the remaining blank space in an optimization method.

Similarly, Yatziv and Sapiro (2006), Sýkora, Buriánek and Žára (2004) applied this method as the essential background knowledge to examine the colorization images. Moreover, the structural characteristics of textures have been considered profoundly in the experiments in the approach of scribble-based colorization (Qu, Wong & Heng, 2006) (Welsh, Ashikhmin, & Mueller, 2002) (Galun, Sharon, Basri, & Brandt, 2003) (Morimoto, Taguchi, & Schölkopf, 2008).

Regarding more mathematical analysis of color reconstruction model, Kang et al (Kang and March, 2007) demonstrated that the penalized versions of variational models by investigating their convergence features and successfully exhibit numerical outcomes, including the extension to texture colorization. A novel chrominance colorization is shown (Huang et al. 2005) with the assistance of an adaptive edge detection method to extract meaningful edge information, which combines the proposed fast colorization scheme based on a scribble colorization system to modify the color transferring way.

The scribble-based colorization is the primary action to explore and exploit the method in the colorization of grayscale images. Accordingly, the foremost drawback of such approaches is that the initialization of the colorization manner is accomplished through inevitable manual work as the prime status. Furthermore, it costs considerable time to colorize a grayscale image with fine-scale compositions by intensive hand-operated work and expert skills for contributing more reliable scribbles. In the meantime, it also spends extra time in computation to obtain high- quality outcomes. Therefore, this method has not been taken into account in this thesis.

2.4 Exemplar-Based Colorization

The exemplar-based colorization has a rigorous prerequisite to select colourfully reference images, which is beneficial for capturing the same scene in black and white target images. Thus, this method is also called patch-based colorization.

The earliest work was proposed (Reinhard, Adhikhmin, Gooch Shirley, 2001) (Levin, Lischinski and Weiss, 2004), which described a more common form of colour revision that imposes the colour characteristics of the purpose images from the reference images in a simple statistical analysis. A comparable approach is also practical (Welsh, 2002), where the specific procedure is to transfer merely chromatic information according to match global colour statistics and maintain the original luminance values of the target image. The performance is heightened by coordinating areas of the two images with rectangular swatches. However, since this approach neglected information of spatial pixel, it produced disappointing outcomes in numerous situations.

Progressively, the accurate colour transfer is accomplished with multiple correspondence techniques are contemplated, including segmented region level (Irony, Cohen, & Lischinski, 2005) (Tai, Jia, & Tang, 2005) (Charpiat, Hofmann, & Schölkopf, 2008) (Chia et al., 2011) (Gupta et al., 2012) (Chia et al., 2011) (Liu et al., 2008) (Bugeau, Ta, & Papadakis, 2013) and semi-automatic retrieval methods (Liu et al., 2008) (Chia et al., 2011). A segmented exemplar image maximizes the probability of entire coloured image of each pixel for colorizing grayscale images at the global level. Compared to the methods of prior automatic colour transfer, this approach exhibits the high performance of spatial consistency. Implementing the tools from machine learning is increasingly discovered to obtain information from the data of coloured illustrations.

A plenty of studies showcase machine learning methods, the reference images are also inevitable to employ in the experimental process. However, Expectation-Maximization (EM) and Gaussian Mixture Model (GMM) are executed in the process of

loss functions. The former objective is to induce natural connectivity among pixels by enhancing both spatial and colour smoothness; the latter distinguishes the colour statistics in each domain (Tai, Jia, & Tang, 2005). By utilizing neural representations to match spatial variation and global coherence, the perceptually similar semantic structure is applied to colour transfer (He, Liao, Chen, Yuan, & Sander, 2017).

This task of colorization provides a more spontaneous approach to minimize the amount of human Labour required by supplying a remarkably similar reference to the input grayscale image. However, the general drawback of this method is the performance of the target image excessively based on the selected reference image. Moreover, a number of procedures applied exemplar-based colorization indicate that the pixels with similar intensities or comparable sections have analogous colours. In Table 2.2, we list the specific information for the work included in this section via the sequence of time.

Table 2.2 The relevant researches are involved in the exemplar-based method

Directions	Authors	Time	Methods	Advantages
Exemplar-based colorization (or patch-based) method	Reinhard et al.	2001	Statistical analysis	More general form of color correction but ignored spatial pixel information
	Welsh et al.	2002	Transfer chromatic information; Retain original luminance values	Reduce extensive user effort by feeding a similar reference, but depends on the selected reference image
	Irony, Cohen, & Lischinski	2005	Segmented region level	More accurate local transfer
	Tai, Jia, & Tang	2005	Segmented region level Expectation-Maximization Gaussian Mixture Model (GMM)	Both spatial and color smoothness to infer natural connectivity among pixels.

	Liu et al.	2008	Pixel level Semi-automatic retrieval methods	Relighting with the illumination component of the target image
	Charpiat, Hofmann, & Schölkopf	2008	Segmented region level L-a-b color space	Using a parameter to enforce the local spatial relationship between pixels
	Chia et al.	2011	Super-pixel level Semi-automatic retrieval methods	A semantic text label and segmentation cues for major foreground objects
	Gupta et al.	2012	Super-pixel level	User only to supply a reference color image which is semantically similar to the target image.
	Bugeau, Ta, & Papadakis	2014	Pixel level	Model the color selection and the spatial constraint problems simultaneously.
	He et al.	2017	Color transfer	Neural representations for matching by similar semantic structures

2.5 Machine Learning-Based Colorization

Machine learning and the state-of-the-art deep learning methods have been extensively implemented to render the black-white images and old videos over the past few years. Specifically, a fully-automatic colorization method relies on locating the numerous patch and pixels of similar images from a massive reference image, transferring colour information from the matched block to the pixel of the final target image. A unique plausible outcome is produced for each input, the intrinsic colorization is an inappropriate problem with multimodal uncertainty outcomes (Charpiat, Hofmann, & Schölkopf, 2008).

Deep neural networks are surprisingly valuable for object classification, target detection, and image segmentation compared to other self-supervised methods. The experimental result successfully misleads humans on 32% of the trials(Larsson, Maire, & Shakhnarovich, 2016). The governing differentiation in the machine learning networks is

shown by setting loss function (Iizuka, Simo, & Ishikawa, 2016) (Zhang et al., 2016) (Isola, Zhu, Zhou, & Efros, 2017) (Salakhutdinov & Hinton, 2009) (Goodfellow et al., 2014).

Each of these systems acquires from large-scale data and unrequired any manual interference. Nevertheless, the volume of data is too enormous to deal with matching noises properly in practice. Table 2.3 exhibits the related work in chronological order.

Table 2.3 *The related work is involved in the machine learning-based colorization*

Directions	Authors	Time	Methods	Advantages
Learning-based colorization (or fully-automatic colorization) method	Deshpande et al.	2015	LEARCH framework Quadratic objective function Chromaticity maps	Defined colorization as a linear system and learned its parameters.
	Cheng et al.	2015	Fully-automatic colorization	Concatenated several pre-defined features and fed them into a three layer fully connected neural network.
	Larsson,Maire, & Shakhnarovich	2016	Classification loss End-to-end learning approaches	CNN to automatically extract features and predict the color result
	Iizuka, Simo, & Ishikawa	2016	Image reconstruction L2 loss; End-to-end learning approaches	Fully-automatic colorization
	Zhang et al.	2016	Classification loss End-to-end learning approaches	Fully-automatic colorization
	Isola, Zhu, Zhou, & Efros	2017	L1 + GAN loss for considering the multimodal colorization; End-to-end learning approaches	Fully-automatic colorization

2.6 CNN-Based Colorization

Deep neural networks in colorization yet have sufficient scope to explore heretofore because deep learning is the most advanced computer vision method. The foundational manner of deep learning fundamentally applies CNNs to extract features full-automatically from the large-scale data and predict the possible colours of the objective image.

Semantic information and a colour histogram instead of merely colorization at each image spot that these two intuitive observations are realized in the design. With the presupposition of an excellent patch matching method, extremely large-scale reference data has been employed for model training, including joint bilateral filtering as a postprocessing step to eliminate artifacts offered by CNN networks (Cheng, Yang, & Sheng, 2015). However, the noises of patch matching are liable to increase with the size of the applied data in inconstant training. As the early investigator, Zhang et al. (Zhang and Efros, 2016) employed CNN to colorize the grayscale images. They leveraged the deep neural network to attest the empirical probability distribution by predicting 313 sets of the gamut, converted into 'A' and 'B' channels of the "LAB" colour model.

In the domain of image colorization by using deep learning, the models such as Inception, ResNet, and VGG net are typically utilized in the experimental models. Zhang et al. (2016) leveraged the Inception ResNet v2 network and retrieved the grayscale image from its ultimate layer rather than training a feature extraction model from the inception layer. By training the deep convolutional architecture model, colour histogram (hue and chroma distributions) of each pixel is employed for developing a fully automatic colorization method (Larsson & Shakhnarovich, 2016).

An intriguing method was proposed (Larsson, Maire and Shakhnarovich, 2016), in which an entirely convolutional version of VGG-16 net without any classification layers was applied in each pixel to create a colour probability distribution. Eight blocks are

accumulated in the approach of linear convolutional stacks. Either two or three convolutional layers followed by a ReLU layer and Batch Normalization (BatchNorm) layer incorporate one block. Similarly, our method also utilized linear activations, but ten blocks are combined by a series of Conv-Leaky-ReLU layers and BatchNorm layer to configure the VGG-19 model. However, striding is applied to decrease the size of the image instead of pooling. A nonlinear activation function (Leaky-ReLU activation) has been applied to image colorization (Maas, Hannun, & Ng, 2013) (Xie, Girshick, Dollár, Tu, & He, 2017).

2.7 Hybrid Methods for Image Colorization

Multiple researchers have utilized the advantages of various methods for image colorization. The scribble-based colorization methods are more robust than machine learning-based colorization methods. Zhang et al. (2017) proposed that developing a comparable colour model is an intuitive approach that automatically applies image retrieval to automatically accomplish the reference collection (Sangkloy, Lu, Fang, Yu, & Hays, 2016). The hybrid of colorizing model via CNNs and Inception-ResNet v2 is employed by importing the inputs to the model (Baldassarre, Morín, & Rodés-Guirao, 2017). The focus of these models is on neural networks by using 20 images to enable colour images and attain 30% satisfactory results after the process of image colorization.

Chapter 3

CNN-Based Methodology

In this chapter, fully automatic colorization for grayscale CT lung images is introduced whilst practicing a deep residual network (ResNet) from the CNN-based colorization manner. There are three training data collections in this chapter, two sets of meat images with enormously varying quantities and one painting dataset. A collection of reference images are from our supermarket, where we acquired the images of fresh pork, lamb and steak. The rest of the painting images we collected online manifests the number of data is enormous. For configuring the model, LeakyReLU is performed as an activation function by combining with batch normalization to render the CT lung greyscale images. Moreover, the loss functions are calculated for each dataset, one dataset is much fitting for colourizing the CT lung images in this model.

3.1 Training Data

Quality and quantity are the two significant factors to measure a dataset. We collect two sets of image data with completely distinct content to discover the influence of the dataset's content for rendering CT lung images. One of the datasets is not relevant to the human organs, including the painting images with 2,129 images from the Kaggle. Another dataset seems much suitable for colorizing CT lung images by practicing 2,100 meat images, which is possible to obtain genuine outcomes after colorizing grayscale CT lung images.

Additionally, we are curious whether the amount of the dataset affects the ultimate outcomes. Therefore, we randomly assemble 30 images of meat images compared to the meat images to witness the assumption.

3.2 Image Preprocessing

The trichromatic system remains the well-known and most straightforward implemented in virtually. However, the computationally intensive RGB pixel is not the ideal selection for colour image processing. When a series of problems appear in complex forms, such as the essential association between channels, perceptual nonlinearity with visual perception, device dependence, and chrominance and luminance data mixing. Consequently, we demand to seek CIE-Lab and CIE-LUV colour spaces that are more decent for digital image processing and more uniform perceptually. Two necessary procedures deal with original images (training data) for rendering grayscale mode into coloured images.

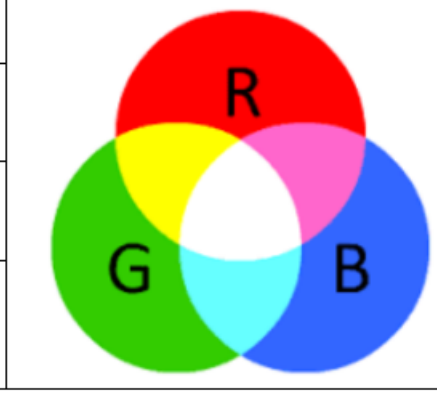
Initially, we transform an RGB image into an LAB image and separate the values ‘L’ and ‘a’, ‘b’ from the image to train the model for predicting the values of ‘a’, ‘b’. Secondly, we convert a RGB image into a LUV image, likewise, separate the values ‘L’ and ‘U’, ‘V’ from the training images and then configure the relevant model to obtain the values of ‘U’ and ‘V’. The three sets of colour models will be introduced specifically.

3.2.1 RGB Colour Model

In the middle of 19th century, the theory of three scalar channels evolved the foundation of the RGB colour model proposed (Bueno et al., 2013). The principle of the RGB model is associated with our human visual system. The RGB model defines three variables components namely R (red), G(green), and B(blue), respectively. These three colours components are comparable to the three visual vertebral cells in the human eyes.

This is an additive colour model in which three colours of light are combined to produce numerous colours. According to the usage, the share of colour is notified in percentage (0 to 1 or 0% to 100%), absolute value (0 to 255), or hexadecimal (00 to FF). The hexadecimal notation in the form of #RRGGBB is comprehended essentially through HTML and CSS (Duvvuri, Laggoni, & Karinga, 2018) (Siddiqui, 2019). For instance, if all the shares are zero (#000000), the colour is black; if all the shares are the highest (#FFFFFF), the colour is white. Between these two values, there are various colours in this method. The fundamental and standard colour superposition rules as shown in Table 3.1.

Table 3.1 The colour superposition of RGB colour model

Red + Green = Yellow;	
Red + Blue = Purple;	
Blue + Green = Cyan;	
Red + Blue + Green = White	

RGB colour model is one of the most important colour spaces, which is extensively applied in electronic devices, such as displays of computers and televisions. RGB is an

additive colour system model composed of these three primary colours that can superimpose all possible colours.

3.2.2 Lab Colour Model

Lab is a device-independent colour system in accordance with human physiological properties, which utilizes digital systems to describe human visual understanding (Harman, 2002). In this 3D, the letter 'L' represents the brightness of the pixel, which controls the colour is lighter or darker, which is practised to represent the applicable range of value is [0,100], which indicates from pure black to pure white. For instance, the larger 'L' refers to the higher brightness of colour, 'a' stands for the value of colours of red and green which the value of colour range from red to green is [127, -128]; Likewise, 'b' refers to the value of colours of blue and yellow which the value range from yellow to blue also is [127, -128] (Burger & Burge, 2016).

As a matter of fact, the Lab colour model has been developed in the 1950s; Hunter established a tri-stimulus model, which is evaluated to achieve near-constant spacing of distinguished colour variations (Wrolstad & Smith, 2017). While this version of the Lab was adopted as the actual model for plotting colour coordinates and variations between colours, it was never formally accepted as an international standard. Designing a colour model that approximates human vision, Lab colour theory is constructed upon the Munsell colour system, Hunter colour space, and CIE colour space. Lab is not device-dependent which is utterly dissimilar from the RGB and the CMYK. The software applications currently implement the CIE-Lab or the CIE-Lab D50 (Baldevbhai & Anand, 2012) (Fairchild & Berns, 1993).

The grayscale CT lung image we strive to colour is considered as 'L' channel of the image in the Lab colour space, which is to adjust the lightness contrast applying 'L'

component. Researchers eager to achieve precise colour balance corrections by altering the output curves of ‘a’ and ‘b’ components.

3.2.3 LUV Colour Model

In colorimetry, the LUV colour space is named by CIE 1976 or CIE-LUV, which was approved in 1976 (Schanda, 2007) by the International Commission. The simply calculational conversion of the 1931 CIE-XYZ colour model endeavoured perceptual uniformity with the visual combination. The CIE-LUV signifies an Adams chromatic valence colour space, an updated version of the CIE 1964 (‘U’, ‘V’, ‘W’) colour space (CIE-UVW). The variations constitute a slightly adjusted luminance scale in each coordinate and a redesigned uniform chromaticity scale.

CIE-LUV and CIE-LAB are both selected simultaneously by the CIE colour space while no explicit agreement could be constructed following singly one or the other two-colour channels. It is extensively employed for applications to deal with coloured lights, such as computer graphics. However, additive compounds of copious coloured lights descend on a line in the uniform chromaticity description of CIE-LUV. Contrary to the opinions, unless the mixtures are constant in brightness, such additive compounds will not drop along a line in the CIE-LUV colour model. For prevailing images, the range of ‘U’ and ‘V’ is within [-100,100] and the range of ‘L’ is [0,100].

3.2.4 Calculations of Converting Colour Systems

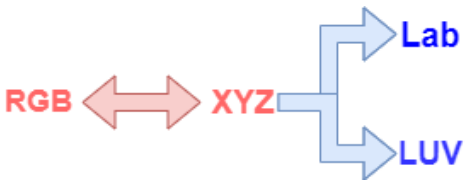


Figure 3.1 The procedure of converting the colour systems

The RGB colour space is unable to be directly transformed to the Lab or LUV colour space as demonstrated in Figure 3.1. The XYZ system is an unavoidable mechanism to function in preprocessing at the root of whole colorimetry (Ford & Roberts, 1998). Full visible colours refer to employing merely positive values and 'Y' is lightness. Firstly, we follow the detail functions to convert RGB to the XYZ colour space. The conversion matrix is given by (Asmare, Asirvadam, & Iznita, 2009)

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.53715 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} x \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.1}$$

We comprehend that the value range of RGB is [0,255]. If the coefficient sum is comparable to one, the necessity value scale of X is between 0 to 255 (Li & Li, 2019). Therefore, XYZ and RGB are mapped in the equivalent scope, modifying the coefficients in proportion to construct their sum equal to 1.

Aftermath, by converting XYZ colour space to Lab colour system, the forward transformation between XYZ and Lab (Asmare, Asirvadam, & Iznita, 2009):

$$L = 116\ f\left(\frac{Y}{Y_n}\right) - 16 \tag{3.2}$$

$$A = 500\ \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \tag{3.3}$$

$$B = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \tag{3.4}$$

where the tristimulus X_n, Y_n, Z_n are the standard values for the colour stimulus of white objects.

The calculation of converting XYZ colour space to LUV colour space as follows (Asmare, Asirvadam, & Iznita, 2009):

$$u' = \frac{4X}{X + 15Y + 3Z} \tag{3.5}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \tag{3.6}$$

$$L = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, & \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \\ \left(\frac{29}{3}\right)^3 \left(\frac{Y}{Y_n}\right), & \frac{Y}{Y_n} \le \left(\frac{6}{29}\right)^3 \end{cases} \tag{3.7}$$

$$U = 13L\left(u' - u'_n\right) \tag{3.8}$$

$$V = 13L\left(v' - v'_n\right) \tag{3.9}$$

where the numbers u_n' and v_n' are the (u', v') chromaticity coordinates of the white position. Under the D65 illuminate in the reflection style, it is generally practised as the exemplary reflection diffuser (u', v'). For 2° observer and standard light source, u_n' = 0.2009, v_n' = 0.4610 (Asmare, Asirvadam, & Iznita, 2009).

3.3 Deep Residual Network (ResNet)

From a theoretical opinion, more layers of the network, the more dependable the model will perform. Since the quantity of network layers needs to improve, the network extracts more complicated feature patterns. However, the experimental truth is that as the depth of the model rises, the accuracy of the network would be saturated, the accuracy of the model would be diminished. The analysis for this situation is due to the vanishing and exploding gradients, which deliver the training remarkably challenging to converge. This problem is currently alleviated by using normalized initialization and intermediate normalization layers.

Additionally, if the quantity of layers is extended repeatedly for deep neural network, the accuracy of this model will decline swiftly (both training error and test error). However, CIFAR-10 and ImageNet are examples not to be created by the model over fittings. Specifically, one deep neural network has a total of *L* layers, the attenuation of Plain Net gradient correlation without residual is $1/2^L$, while the attenuation of ResNet is merely $1/\sqrt{L}$.

Relying on gradient descent, the loss function is transferred down to obtain the minimized weight. It should be notified that the quantity of layers is as appropriate as possible unless repeated multiplication will assure the gradient smaller until it eventually "disappears", resulting in performance degradation with each additional layer oversaturated. The solution to this problem for ResNet is to instantly establish an association within the input layer and the output layer. Initially, ResNet fails to perform any operation layer when it stacks the identity mapping. It merely demands learning new features predicated on the original input layer. After that, the previous activation layer is reused to compress the network into several layers to absorb the residual faster. When the network is trained again, all layers will be expanded, and the "residual" part of the network will explore more sources from image feature spaces.

The ResNet network is modified based on VGG-19 network and attached to the residual unit via the short path device. It principally reflects the straightforward application of stride (e.g., 2) convolution for sampling, and the global average pool layer displaces the sufficiently relevant layer. The essence intention of ResNet is to assure the complexity of the neural network layer. When the scope of feature map is diminished by half, the quantity of feature maps is multiplied correspondingly. Deep residual learning for image identification takes use of a residual learning structure, which exceedingly shrinks the training time of this model, allowing the model to gain more profound training in a fixed time. This method accomplishes the most desirable outcomes on ILSVRC2015.

The ResNet network employs two kinds of residual units, external network and deep network, respectively. When the input dimension and output dimension are the equivalents for short-circuit connection, the input dimension is straightly appended to the output dimension. Vice versa, if the dimensions are inconsistent (corresponding to double the dimensions), this is impossible to be consolidated directly. The primary distinction between ResNet v2 and ResNet v1 are the batch normalization of ResNet v2 before per weight layer (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017).

3.4 Architecture

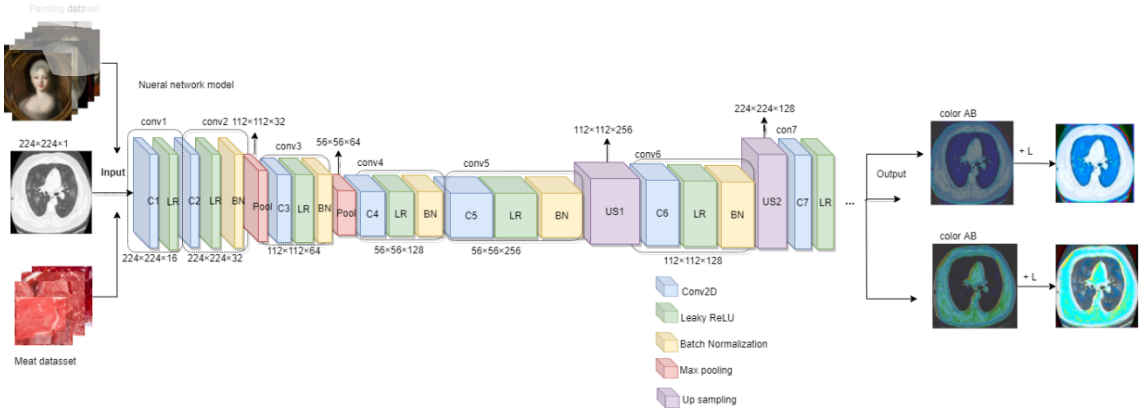


Figure 3.2 The architecture of CNN-based method for colorizing the CT lung images

Generally, a neural network (filter or model) constitutes a link between inputs and the output. In our research project, the inputs refer to the CT lung grayscale images, the output indicates the numbers of filters and two-colour channels (‘a’ and ’b’ in Lab colour model). Considered the principal as the blue and red filters in 3D glasses, each filter directly determines what variety of colour is formed in the image. They prominent or eliminate information to transfer the significant information to the objective images. The ultimate CT lung image is generated by combining these two filters. The original input CT lung images are adjusted the size to 224×224. The deep neural network inputs a grayscale image and predicts “ab” pairs of the gamut, reveals the empirical probability distribution, which is transformed to ‘a’ and ‘b’ channels of the Lab colour space. Eventually, it is

unsampled to the original image size. Meanwhile, appending the lightness channel ‘L’ will generate the CT lung images.

The structure of our CNN-base colorization model is presented in Figure 3.2. The black and white layer signifies our input information; meanwhile, these two colours belong to the output layers. With the correlation between the grayscale images and the colourful training images, we decrease the size of the images to extract the features better.

For example, when we apply a 3×3 filter to per image in the dataset and combine the innovative pixels with the simplistic filters, the complicated patterns are able to be detected, such as a semicircle, a small point, a curve line and even a tiny straight line. Frequently, the model extracts the identical pixel from the images and then generates 128 unique purified images at this time.

The neural network model starts operating a random foresight for each pixel in a heuristic manner. The deviation for each pixel operates rearward through the interface to promote the strength of feature extraction whether to colour and how to position the different objects initially adjust for the most error-producing situation. The loop firstly counts the entire image names in the file, which iterates through the image index. The converted images in an array of pixels generate a giant vector.

The objective of this method is to conduct the standard elementary auto-encoder that would perform rendering CT lung images. This full autoencoder model is implemented with high-level or complicated methods to solve image colorization by employing the TensorFlow (Hodnett & Wiley, 2018). The network linearly stacks convolutional layers forming eight blocks and each block comprises either two or three convolutional layers followed by using a Leaky ReLU layer and terminating in a Batch Normalization layer. Moreover, we apply upsampling to enlarge the number of feature maps instead of max pooling during the latter operations.

3.4.1 Activation Functions - LeakyReLU

First and foremost, two activation functions in our experiment are determined to manage all network structures unaltered with activation functions, ReLU(Nair & Hinton, 2010) and Leaky ReLU (Maas, Hannun & Ng, 2013). By using the non-saturated activation function brings two aspects of advantages: The primary goal is to solve the problem of "exploding" (or "vanishing gradient"), ReLU is a piecewise linear function that prunes the negative part to zero and retains the positive part. It is commonly considered that sparsity is the superior performance of ReLU. (Glorot, Bordes & Bengio, 2011). Another one is to speed up the convergence of the network (Scardapane, Van, Totaro & Uncini, 2017). The trend of LeakyReLU activation function is illustrated in Figure 3.3.

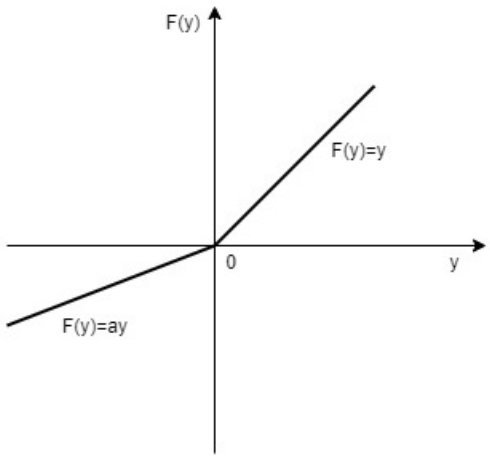


Figure 3.3 The LeakyReLU activation function

Leaky rectified linear (Leaky ReLU) activation is referred as (Xu et al., 2015):

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \frac{x_i}{a_i} & \text{if } x_i < 0, \end{cases} \tag{3.10}$$

where a_i is a fixed parameter in interval (1, +∞). Xu et al. (2015) set a value of a_i like 100. Additionally, they also experiment smaller value a_i. The eventual results eye that the performance of leaky ReLU is comparable to that of ReLU, while leaky ReLU with larger a is much better than the standard ReLU. The Leaky ReLU function is defined as follows

(Xu, Huang, & Li, 2016),

$$f(x) = \max(0.01x, x). \tag{3.11}$$

Contrary to the negative part of ReLU, the leaked ReLU is assigned a zero-noon slope. Thus, leaky ReLU effectively solves the dead ReLU phenomenon. A small value like 0.01 is applied to initialize the neuron so that ReLU is inclined to activate in the negative area. After comparing these two activation functions, we ultimately determined to utilize the leaky ReLU as our essential partial of network.

3.4.2 Batch Normalization

Owing to the parameters of the last layers, the distribution of each layer's inputs also alters during training deep neural networks (Ioffe & Szegedy, 2015). This complicated process virtually slows down the velocity of model training. The batch normalization technique effectively alleviates this problem and unmistakably accelerates training deep neural networks that standardize each input variable in a layer to assist in coordinating the update of multiple layers in the training model. Specifically, the activation of a node from the previous layer to halve, scale the epochs or the output of the layers to provide regularization and reduce generalization error (Bengio, Goodfellow, & Courville, 2017).

In this CNN-based colorization method, for a layer with two-dimensional entry $\mathbf{x} = (x_1, x_2)^T$, we standardize each dimension as (Ioffe & Szegedy, 2015),

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}} \tag{3.12}$$

where the batch normalization is calculated in reference to minibatch SGD. The original activation x corresponding to a particular neuron is converted by subtracting the mean value $E(x)$ acquired by m activation x obtained from m instances in the minibatch and dividing by the obtained variance Var(x). In our study, the function of Batch

normalization is to build the network by forwarding calculation. Once the network has been equipped, we implement the normalization in the following (Ioffe & Szegedy, 2015)

$$\widehat{x} = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} \tag{3.13}$$

3.4.3 Upsampling

By configuring unspecified colorization networks, all of the image size and ratio are employed in the model, which ought to guarantee the equivalent throughout the deep neural network. On the contrary, in other types of deep neural networks, the lower the realism of the image is performed as closer to the terminal layer. For instance, the researchers insist on shrinking the image dimension and quality since they merely recognize the definitive categories in the classification system. The max-pooling layers enhance the information quantity while deteriorating the quality of the images implemented in the classification systems. It merely estimates the information and content of the image, but it makes no attempt to consider the layout of the original image. In the practical experiment, we alternately conduct a stride of two to halve the width and height.

Consequently, appropriating the up sampling in the convolution network enlarges the feature maps, the model maintains the original size of the input image. The fundamental principle of image magnification frequently employs interpolation methods that are attributed to the original image pixels. This system is not involved in introducing trainable parameters, while a simple interpolation and solely a vital parameter (size) is practised (Hui, Loy, & Tang, 2016). The algorithms of interpolation are performed to insert new elements between the pixels. Interpolation algorithms are stumblingly classified into three representations: Traditional interpolation, edge-based interpolation and area-based interpolation severally.

3.4.4 Optimizations

According to the loss values (introduced in the next section) in the first epoch, we obtain the following specific matrix in the different optimizations. In Table 3.2, Adaptive Moment Estimation (Adam) as the optimizer in our model is significantly high-value performance. Adam is applied to our method as an optimizer, which is a random objective function optimization algorithm based on a step degree according to the adaptive estimation of low-order moments (Kingma & Ba, 2014). The empirical results demonstrate that Adam algorithm has outstanding accomplishments and prominent advantages compared to other classifications of random optimization algorithms in practice (Bae, Ryu, & Shin, 2019). Thus, it is a particularly prevalent algorithm in the field of deep learning.

Table 3.2 The losses in the first epoch step with diverse optimizers

	Painting dataset	**Meat dataset (2,100 pictures)**
SGD	0.0761	0.0751
Adam	0.0652	0.0720
Adadelta	0.0856	0.0836
Adagrad	0.0657	0.0742
Adamax	0.0755	0.0796
Nadam	0.0752	0.0869
Ftrl	0.0659	0.0879

3.4.5 MSE Loss

Mean Squared Error (MSE) is analysed in the model as loss values for evaluating the models in different optimization and ultimate results of various datasets, defined as (Rougier, 2016)

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[R(i,j) - G(i,j)]^2 \tag{3.14}$$

where *i, j* represents the position (width and height) corresponding to each pixel severally. Parameter *R*(•) indicates the initial image and *G*(•) refers to the outcome image. The loss

value is equal to the distance between these two images, $R(\bullet)$ denotes the noise approximation of $G(\bullet)$.

Analysing the final results in different datasets, the losses in individual epochs are displayed in Figure 3.4, Figure 3.5 and Figure 3.6. It confirms that a considerable amount of data is essential to obtain more trustworthy outcomes than 30 meat images in Figure 3.6. The painting data and the meat data (2,100 images) have the similarity epochs times because the number of painting images is up to 2,129, and meat data is 2,100 images. Moreover, compared with Figure 3.4 and Figure 3.5, both datasets perform high performance when epochs are approximately three. However, the meat data is more stable than the painting data from comparing the tendency of these two charts. The painting images have a higher value of loss of approximately nine and range from 15 to 18. Consequently, the conclusion from the loss functions that the generated CT lung images were trained by the painting dataset have more reliable performance than trained by the meat data.

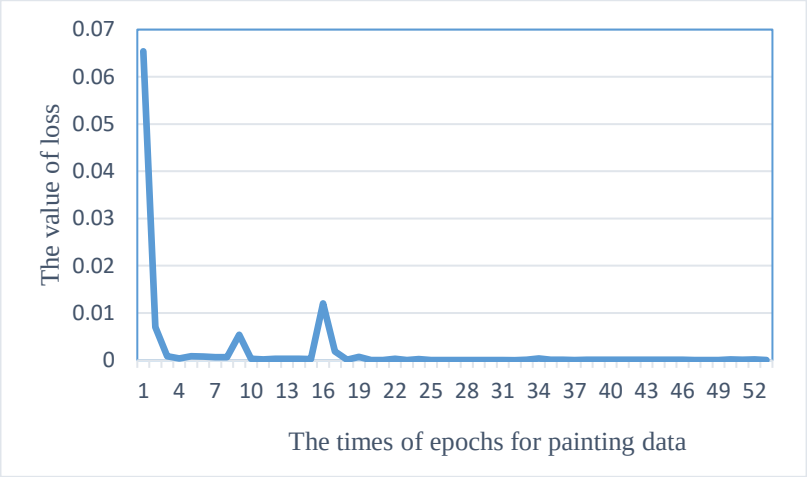


Figure 3.4 The losses in the rendered image data

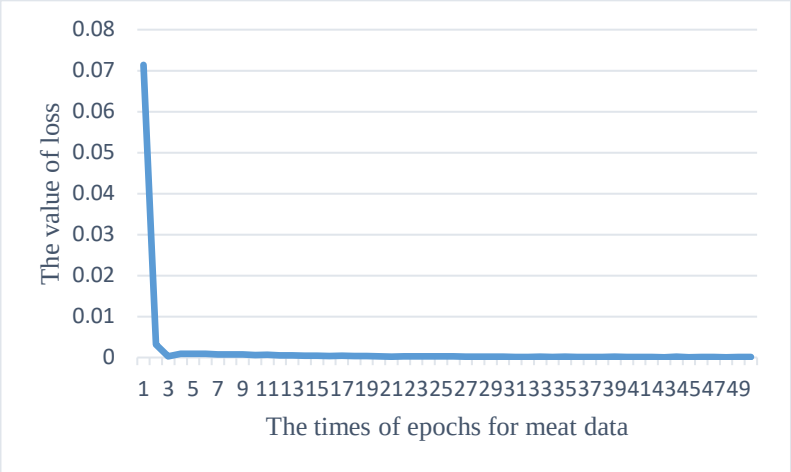


Figure 3.5 The losses in the meat data (2,100 pictures)

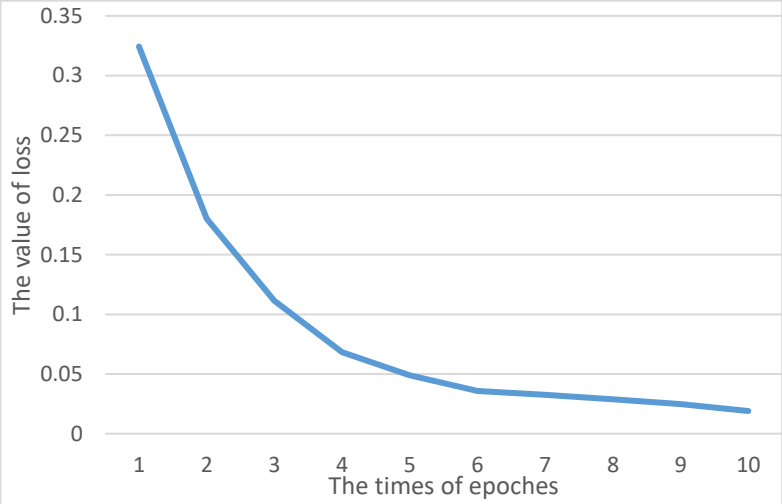


Figure 3.6 The losses in the meat data (30 pictures)

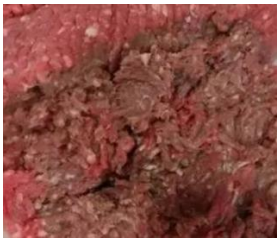
Chapter 4
Hybrid Methodology

This chapter chiefly showcases the hybrid method that transfers the style and content of reference images into grayscale CT lung images by leveraging VGG-19 and CNNs (or ConvNets) in deep learning. The reference images (meat images) are acquired and collected from the supermarkets. The images of rotten meat are obtained and stored without any protection measures. For constructing the model, the neural network VGG-19 is executed in this method, displaying the exemplar-based method of colorization to transfer colours from reference images to target grayscale CT lung images. Moreover, the loss function is merely fit for this model to estimate which one reference image is the most suitable selection to render the grayscale CT images.

4.1 Referenced Data

In the method of hybrid colorization, the reference images are all collected by ourselves, we acquired images of fresh pork, lamb, steak at supermarket. Due to COVID-19 CT lung images in our research project, considering the infected lungs probably is not in a healthy situation. That means, we bought and took home without any measures to store for obtaining the rotten meat. Table 4.1 illustrates the meat images we eventually acquired as the reference images for the hybrid colorization method.

Table 4.1 The reference images in this hybrid image colorization

Steak	Lamb	Pork	Rotten meat
			

4.2 VGG Net

The VGG net is foremost introduced by the Visual Geometry Group from Oxford University (Dhamari, Sudirman, & Mahmood, 2020). The 'V' structure of the VGG net is exceptionally uncomplicated, the entire network includes the same size of the convolution kernel (3×3) and max pooling (2×2). Employing piled tiny convolution kernels is more dependable than the applications of widespread convolution kernels considering complicated nonlinear layers enhance the net depth to guarantee that more intricate patterns are knowledgeable. Another advantage is that the cost is relatively economical, especially with fewer parameters. For example, three combinations of 3×3

convolution kernels as opposed to 7×7 convolution kernels, two collections of 3×3 convolution kernels are appropriated instead of 5×5 convolution kernels in the VGG net (Wang et al., 2018).

On the other hand, the VGG net consumes more extra computing resources and applies more parameters, resulting in tremendous memory usage. The vast majority of the parameters are from the first fully connected layer, but the VGG net typically has three fully connected layers. Many pretrained models instantly work on the VGG model (mainly VGG-16 and VGG-19). Compared with other methods, the VGG net has a broad parameter space, the ultimate model has more than 500 MB (Li, Zhang, & Liu, 2016). AlexNet is utterly 200 MB, and GoogLeNet is even more insufficient (Liu et al., 2017). Therefore, it regularly demands a longer time to train a VGG Net model.

In order to enhance the neural network under the same perceptual field and betterment the effectiveness of the neural network. Although VGG net has more parameters and a deeper level than AlexNet, VGG only demands a few iterations to start converging. There are two principal reasons: Depth and miniature filter size are conducive to hermit regularization and pre-initialisations of layers, indicating the Gaussian distribution weight (Simonyan & Zisserman, 2014).

The VGG net is configured in TensorFlow from one of the libraries in Python. TensorFlow is the second-generation Artificial Intelligence (AI) system developed by Google (Abadi et al., 2016). It is broadly implemented in multiple machine deep learning domains such as speech recognition and image recognition. Its name was originated from the operating principle, which combines two components: Tensor and Flow together. “Tensor” indicates an *n*-dimensional array, “Flow” involves a calculation based on the data flow (Abadi et al., 2016). TensorFlow describes the computation process of TensorFlow from one side of the image to the other, transmitting complicated information constructions to the artificial neural network for processing (Widrow, Rumelhart, & Lehr, 1994). Although the VGG structure is simplistic, it includes a considerable number of

weights and parameters (weights of the convolution kernel and fully connected layer) (Liu, Wang, Li, & Han, 2017). For the leading layer of convolution, the quantity of channels in the input image is equal to three means that the network acquires a size of 3×3 convolution kernel and three channels. A total of 64 convolution kernels and the relevant parameters are (3×3×3) ×64 = 1728 (Wang, Wu, Zhang, & Yao, 2020).

4.2.1 VGG-16 Network

VGG-16 network and VGG-19 network are both derived from the VGG net structures. There is no distinction in essence, the foremost differentiation between these two models is the number of convolutional layers. One improvement of VGG16 is to replace the kernels by using successive 3×3 convolution kernel (11×11, 7×7, 5×5) in AlexNet (Salakhutdinov & Hinton, 2009).

The outstanding feature of VGG-16 network is simplicity. First of all, in the convolutional layers, the same convolution kernel parameters are completely managed. The convolutional layers (kernel size) are ultimately expressed as “conv3-XXX”, where “conv3” symbolizes that the width and height of the convolution kernel are both three (three fully connected layers), “XXX” signifies the quantity of flows of the convolutional layer. Moreover, other parameters (stride, padding) are stable and unchanged, 3×3 signifies a tiny convolution kernel dimension so that each convolutional layer maintains the same width and height as the preceding layer. According to the specific investigation of VGG-16, it is affirmed that VGG-16 comprises 13 convolutional layers. The total number is equal to 16 (13+3=16), which is the root of 16 in VGG-16 (Han, Mao, & Dally, 2015). Since the pooling layer ultimately is uninvolved weights, it is not reckoned to the weighting layer.

Another feature is that the pooling layers entirely utilize the same pooling core parameters. For example, the pooling layer parameters are entirely 2×2, the stride is equal to two so that the size of per pooling layer is 12 of the former layers (Yu et al., 2016).

The model is composed of several convolutional layers and pooling layers stacked (stack). It was straightforward to form a more extensive neural network construction in 2014, and 16 layers are profound.

4.2.2 VGG-19 Network

VGG-19 network has the function and principle as same as the VGG-16 network. Nevertheless, the VGG-19 net demands more space to deal with the data of images due to more convolutional layers and pooling layers. The activation function efficiently outlines the input neuron to the output layer (Agatonovic, Kustrin, & Beresford, 2000). Consequently, we eventually implemented the VGG-19 network to configure our colorization model in this thesis. The VGG19 acts from convolutional networks to decompose reference images further and extract associated styles and content for colorizing the target black and white CT lung images.

The first two layers are convolutional layers with 3×3 filters, 64 filters are implemented. The beginning numerals of the height and width of layers are 400 and 585, respectively. Thus, it ends up with a volume 400×585×64. Moreover, the content of conv64×2 indicates that 2 convolutional layers with 64 filters. The filters are implemented continuously with the same convolutions 3×3 with the stride of 1. A pooling layer assists the height and width of a volume to be decreased. The information of convolutions signifies that the height and width of layers are evenly spilled. For example, 400×585×64 would come down to 200×292×64 with the guidance of the pooling layer.

Similarly, after coupling more convolutional layers with 128 filters and a new pooling layer, the dimension 200×292×128 is more powerful than 100×146×256 due to the same distance of convolutions. A pooling layer is added and then the new dimension will be 50×73×256 (Sun, Song, Jiang, Pan, & Pang, 2017). The following procedure also followed this, 4 convolutional layers with 256 filters and the pooling layer. In the end, the total of 12×18×512 into the fully connected layer (FC) with 4096 units.

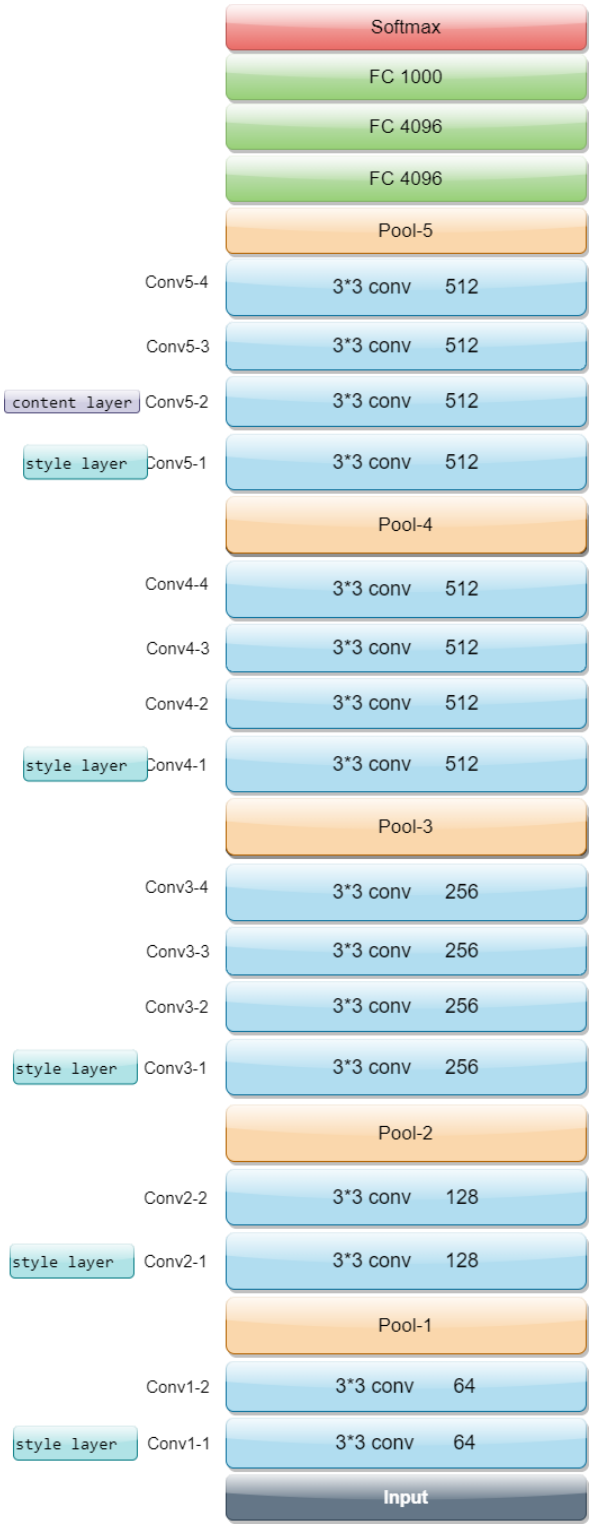


Figure 4.1 The architecture of VGG-19 network in this thesis

4.3 Architecture

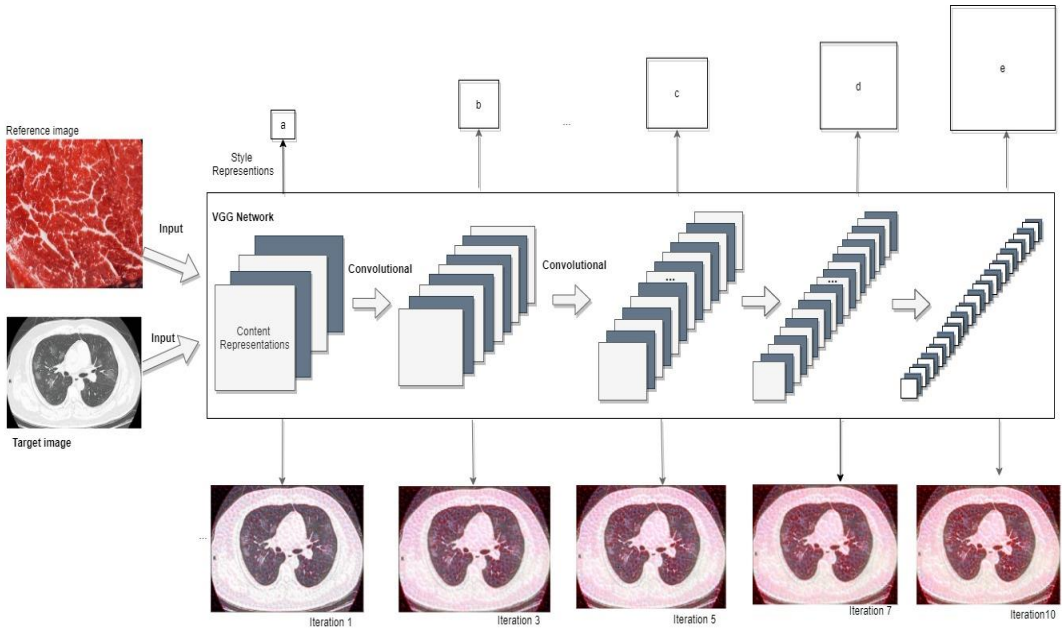


Figure 4.2 The specific procedure of colour transfer from the reference images

With the processing of the neural network hierarchy, the input image is converted into the sequence that is interpreted by the computer compared with the target CT lung image. When the model is trained on object recognition, the target CT lung images are reconstructed to transfer the object information along the processing hierarchy gradually. The image is reconstructed by using feature mapping so that the input image can be visualized in each layer of the neural network. In Figure 4.2, a defined input image is accepted to represent a set of filtered images for each processing stage of the neural network.

Content reconstruction is to reconstruct the input image by simply comprehending the response of each layer of the network so as to visualize the information of multiple processing stages in the model. The high-level and excellent content in the network is captured according to the object and its arrangement in the input image without limited accuracy of the reconstructed image. In contrast, content reconstruction from lower layers will merely imitate the exact pixel values of the original image. Accordingly, we refer to

the characteristic response generated by the high-level network as the ultimate content representation.

In order to obtain a representation of the input image style, we apply a feature space initially designed to capture texture information, which is produced on the filter response of each layer of the neural network. It consists of the correlation between various filter responses within the spatial range of the feature map. The image merely captures its texture information rather than the global arrangement of information processing. In Figure 4.2, we gradually refer to larger squares to represent style representation (a, b, c, d, e).

In total, five basic blocks of convolutions are adopted in the model. In the first step, the input image is trained to apply VGG19 so as to achieve the information of reference image for rendering the target image. The input image attributes to the texture information of the image and takes advantage of functional space based on the filter response of each layer network. The functional space is composed of the correlation between various filters in the feature map space. Minimizing the mean-squared distance between the Gram matrices of the input image and the output image to obtain the ideally eventual consequences.

Specifically, the two sets of convolutional layers with 3×3 filters and 64 filters are performed in the model. The height and width of layers are 400 and 585, respectively and it terminates up with the size 400×585×64. Moreover, the convolution 64×2 indicates that the second convolutional layer with 64 filters. The filters are operated continuously with the same convolutions 3×3 with a stride of 1. A pooling layer that determines a volume will be diminished, the height and width of layers are spilled equally. Gram matrix is implemented to measure the inconsistency of the target image and the reference image. Eventually, the style is transmitted and the ultimate images are generated by selecting the smallest variation in the background of the content of the input image is guaranteed.

Algorithm 4.1: The iteration is realised in this method
Require: Preprocessed the target CT lung images
Ensure: The mixed images with time and loss value
For i in ***iterations do***:
record the number of ***iterations: then***
compute the loss value
image *is reshaped*
if image is not in folder: ***then***
image *is renamed*
image *is saved*
End if
End for
return *images, time* and *loss value*

The number of iterations for transferring the colours of reference images into the grayscale CT images of infected lungs is set as 10. If we would like to have more accurate outcomes, we could arrange more iterations in our further study. Moreover, the associated loss values and iteration times are illustrated in each iteration. Algorithm 4.1 outlines the measures of image colorization presented in this thesis.

4.3.1 Activation Function - ReLU

The activation function plays key role in deep learning models, especially for complicate nonlinear functions. The strengthen of the nonlinear activation function is the capability of the neural learning network by increasing the nonlinear factors. Moreover, the activation function supports representing arbitrarily complex nonlinear function mapping between inputs and outputs (Chen & Pock, 2016). In a neural network, the inputs are weighted and summed, which necessitate being imported into an activation function. Since per output layer is a function of the former input layer, the activation function is

considered an essential component of the model, no matter the quantity of layers of the filters that are superimposed in the deep neural network, the output layer is a linear combination which is the most primitive perception.

The activation function contributes a nonlinear determinant to the neuron so that the neural network approximates any nonlinear function arbitrarily. Ultimately, the neural network is employed for nonlinear models. Activation functions involve the functions like sigmoid, hyperbolic tanh, ReLU, Leaky ReLU, PReLU, ELU, maxout, etc. The rectified linear unit (ReLU) is the most generally accepted one in deep learning, especially in CNN networks. Most feedforward neural networks utilize the activation function by default. The ReLU function is specifically described as shown in Figure 4.3 and positive number is output (unsaturated), the negative number is directly set to zero (hard saturation).

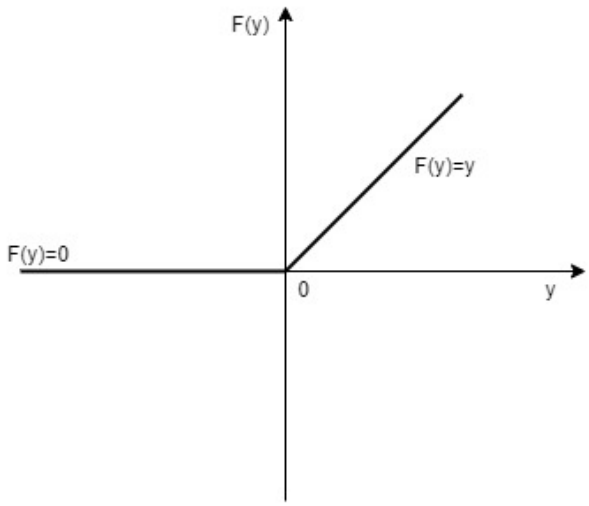


Figure 4.3 The ReLU activation function

Standard rectified linear (ReLU) activation is defined as (Xu, Wang, Chen & Li, 2015)

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ 0 & \text{if } x_i < 0. \end{cases} \tag{4.1}$$

The advantages of ReLU activation function are:

1) The convergence speed of the SGD algorithm employing ReLU is faster than most of the activation functions, such as sigmoid and tanh (Wang, Giannakis, & Chen, 2019).
2) In the field of positive numbers, there is no obstacle with gradient saturation and gradient disappearance (Qiumei, Dan, & Fenghua, 2019).
3) The computational complexity is typically lower than the other activations. The activation function is accomplished by providing a threshold, and exponential calculations are not required to the computing (Weng et al., 2018).

The disadvantages of ReLU activation function are:

1) The output of ReLU activation function is not a zero mean.
2) "Dying ReLU" problem (neuron necrosis) suggests us that an extended gradient flow through a ReLU neuron triggers a weight update that probably provokes the neuron to be inactive for other data points throughout the future training. Ultimately, numerous neurons will be dead without updating their weights, and the network will remain to provide the same output during the iteration. When ReLU is negative, the value of the gradient is equal to zero. The gradient of this neuron and subsequent neurons is frequently naught and no longer responds to any data, causing the corresponding parameters never to be updated.

    Parametric initialization problems and exceedingly high learning rates lead to too large parameter updates during the training process resulting in ReLU being "fragile" during the model training, in this case, utilizing the initialization method of Xavier, setting the learning rate lower than common situations and practicing AdaGrad or other algorithms that automatically adjust the learning rates to solve this problem. For example, when the gradient is enormous, the ReLU problem will recur frequently, but the disappearance of the gradient generated by the sigmoidal function assists in preventing the ReLU unit from dying during the backpropagation process (Chen, Sathe, Aggarwal, & Turaga, 2017).

4.3.2 Gram Matrix

Gram matrix is treated as the eccentric covariance matrix from various features (Almeida, Asada, & Garcia, 2008), a convolution of a modified filter where the per number appears at a specific location in the feature map. An output consequence is initialized by using white noises within the given image.

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} \, F_{jk}^{l} \tag{4.2}$$

where *i, j, k* corresponds to the height, width, and channel respectively; *F* (*i, j*) represents the *k-th* pixel of the *i-th* feature map; G_{ij}^{l} is the output at the position *i, j, k.*

$$G=A^{\tau}A \tag{4.3}$$

where $A = (a_1, a_2, \cdots, a_n)^{\tau}$, a_i is a vector, $i = 1, 2, \cdots, n$. After the inner product of Gram matrices, diagonal elements provide principle components of distinctive feature maps (a_1, a_2, $\cdots$, a_n). Moreover, it reveals that each feature appears in the image. Gram matrix is applied to calculate the correlation between the two features, which accommodates to grasp the general style of the undivided image.

4.3.3 Optimization

Optimizations contribute to resolving the optimal problem of the objective function under a constraint (Karaboga & Basturk, 2007). Most of the problems in machine learning and artificial intelligence are sprightly associated with optimizations. SciPy is an open-source python library practically implemented for mathematics, science, and engineering calculations released under the BSD license (Clewley et al., 2014). The SciPy library relies on NumPy, which provides linear algebra, Fourier transform and random number generation. The essential data structure employed is the multidimensional array produced by the NumPy module, which is available and speedy for *n*-dimensional array operations

(McKinney, 2012). Its optimized module provides various affirmed numerical optimization algorithms. Classic optimization algorithms are included in linear regression, solving function extremes and roots, and determining the coordinates of the intersection of two functions.

The chief methods recommended in the Scipy library, such as Newton-CG, BFGS, L-BFGS-B. Quasi-Newton methods are implemented to estimate the approximate of Hessian (curvature) method based on the gradient, such as BFGS and L-BFGS (Sohl-Dickstein, Poole, & Ganguli, 2014). This gradient is also assumed to attain a much reliable estimate regarding the curvature of the point, which is more affordable in time. Therefore, Newton-CG converges for quadratic functions; for non-quadratic functions, the functions of quasi-Newton converge more satisfying.

- Newton-CG

The Newton method leads a local binary approximation to calculate the direction of the jump. It mainly relies on the first two derivatives of the function gradient and Hessian matrix. The binary approximation is accurate, Newton methods are operated at high speed. In the library of Scipy, the Newton-CG optimizer demands less function and more gradient due to appropriate approximation. The disadvantage of the Newton method is that Hessian flipping is expensive and unpredictable (large scale> 250) in ultra-high dimensions. We would like to remark that the Newton optimization algorithm should not be confused with the Newton root discovery method in reliance on the same principle. This inverse Hessian would spend more time than other algorithms.

- BFGS

The BFGS algorithm was developed by Broyden, Fletcher, Goldfarb and Shanno, which was defined as the first letters of their names. The Broyden-Fletcher-Goldfarb-Shanno algorithm improves the approximation of the Hessian matrix at each step. Although BFGS is not as speedy as the Newton method, it is still more agile than other optimization

methods in a proper binary function. One practical approach to calculate the minima of this function is to operate the BFGS algorithm with a starting point. The algorithm calculates the gradient descent of the function from the starting point presented by the parameters and then outputs the minimum value with a zero gradient plus a positive second derivative. The disadvantage of the BFGS algorithm is that the algorithm would discover these local minimums instead of global minimums due to various starting points when the function receives local minimums.

- L-BFGS

L-BFGS denotes a low-memory version of BFGS, which maintains low-rank (Gemulla, Nijkamp, Haas, & Sismanis, 2011). Because the memory stored in each action is enormously less than the entire $N \times N$ matrix, its performance is more electric than BFGS. The memory-limited of L-BFGS is between BFGS and conjugate gradient. The expense of computing and flipping Hessian matrices in remarkably high dimensions (> 250) is pretty huge. In the setting of this optimization, *maxfun* (•) determines that the maximum number of functions is equal to 20.

According to the loss values of the first iteration in these three optimizers in Figure 4.4, we discovered the L-BFGS has a better performance than the other two optimizers in our hybrid colorization model.

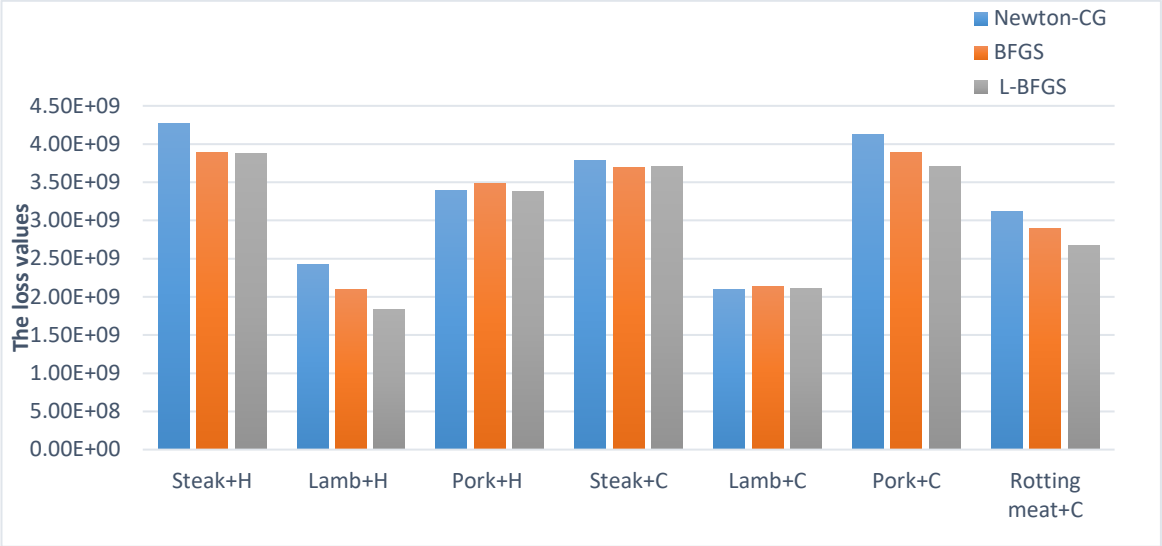


Figure 4.4 The losses in the first iteration with a diversity of optimizers

4.3.4 Loss Functions

The loss function is also applied to the evaluation of the models in a diversity of optimizations as same as the CNN-based colorization method. In this section, we utilized the loss function to obtain a result about which reference image is more suitable for the CT images in these two cases (healthy lung and infected lung). The loss function in the hybrid colorization method is interpreted as

$$L = d_{reference} + d_{generated} \tag{4.4}$$

where $d_{reference}$ refers to the original image, $d_{generated}$ stands for the generated image. L is the value of the distance between the original image and generated image. Ideally, minimizing this loss function means that the rendering influence of colorization outcomes has the most outstanding performance. After simple integration, equation (4.4) simply is displayed as follows:

$$L_{loss} = V_{loss} + C_{loss} \tag{4.5}$$

$$style(generated) \approx style\ (reference) \tag{4.6}$$

and

$$content(generated) \approx content\ (original) \tag{4.7}$$

Table 4.2 The losses of ten times iterations for various data in the healthy CT lung images

Values	Mixed pics	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Time	Steak+Healthy	597	591	574	567	578	577	583	573	571	574
Loss		3.88E+09	1.82E+09	1.29E+09	1.07E+09	8.83E+08	7.66E+08	6.85E+08	6.19E+08	5.64E+08	5.13E+08
Time	Lamb+Healthy	574	574	576	587	579	575	615	584	571	570
Loss		1.84E+09	9.22E+08	6.67E+08	5.36E+08	4.59E+08	4.05E+08	3.68E+08	3.39E+08	3.10E+08	2.86E+08
Time	Pork+Healthy	595	583	578	578	380	308	317	345	328	341
Loss		3.38E+09	1.50E+09	1.11E+09	9.39E+08	8.09E+08	7.44E+08	6.91E+08	6.38E+08	6.01E+08	5.70E+08

Table 4.3 The losses of ten times iterations for various data in the COVID-19 CT lung images

Values	Mixed pics	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Time	Steak+COVID19	354	365	366	361	356	375	371	355	368	376
Loss		3.70E+09	1.77E+09	1.27E+09	1.05E+09	9.04E+08	7.97E+08	7.01E+08	6.36E+08	5.75E+08	5.23E+08
Time	Lamb+COVID19	698	663	663	668	679	666	659	701	667	671
Loss		2.11E+09	8.87E+08	6.24E+08	5.11E+08	4.37E+08	3.90E+08	3.44E+08	3.14E+08	2.90E+08	2.68E+08
Time	Pork+COVID19	356	371	355	427	407	413	389	422	404	423
Loss		3.71E+09	1.55E+09	1.10E+09	8.96E+08	7.91E+08	7.00E+08	6.39E+08	5.92E+08	5.57E+08	5.19E+08
Time	Rotting meat+	383	386	387	383	354	397	368	379	377	376
Loss	COVID19	2.67E+09	1.15E+09	7.83E+08	5.76E+08	4.82E+08	4.27E+08	3.85E+08	3.49E+08	3.22E+08	2.99E+08

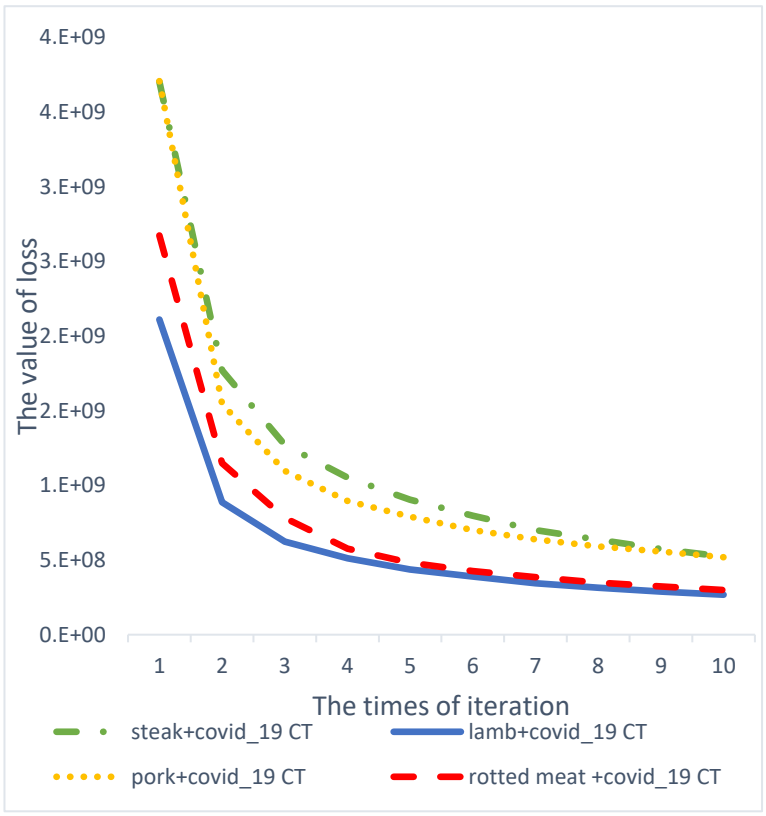


Figure 4.5 The iterations in COVID-19 CT lung images rendering

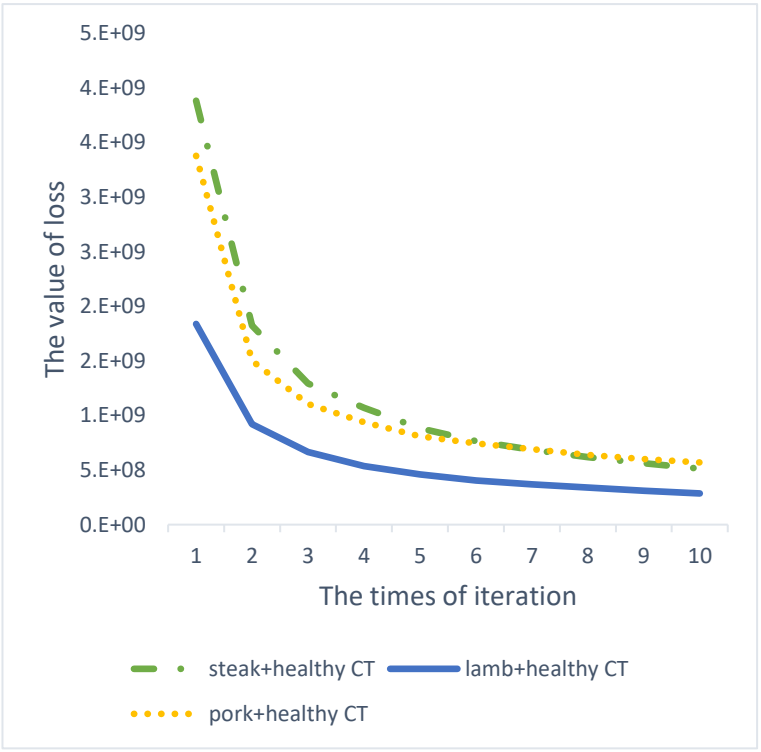


Figure 4.6 The iterations in healthy CT lung image rendering

Table 4.2 and Table 4.3 exhibit the specific data of loss values and time for various reference images in each iteration procedure. For convenient and straightforward observation, we created two line graphs, as shown in Figure 4.5 and Figure 4.6, which show the loss value of lung images infected by COVID-19 and healthy lung images in the varying reference images, respectively. Lamb image as a reference image has the lowest loss value than steak, pork and rotten meat images either the COVID-19 CT lung images or healthy lung images, which indicates the image of lamb attains the most trustworthy and high-quality achievement in these reference images for colorizing the CT lung images. Contrarily, the steak image as the reference image has the most unsatisfactory performance in these four collections of experiments no matter the status of lung CT images. Considered the theory of loss functions, the reasonable deduction is the fat (white) part of the lamb image has more extensive space than the rest of the meat images. It has more overlapped components than the rest of the reference images. Therefore, the value of loss function in lamb image as the reference image is more satisfying than the other generated colour CT lung images.

Moreover, an exciting result appeared in the COVID-19 CT lung images in terms of rotting meat images. It practically has the same losses by employing the lamb meat images as the reference. The reasonable explanation is the texture of rotting meat images is similar to the target images. The outcome has revealed that the texture of CT lung images of the human lungs mimics rotten meat images. Accordingly, the rotting meat is the second trustworthy achievement in these exemplar images for colorizing the infected CT lung images. The steak image and pork image as the reference images have a similar trend in these two figures, but they both are not more reliable than the lamb image. Consequently, the steak image and pork image are not recommended for future experiments in both conditions of CT lung images.

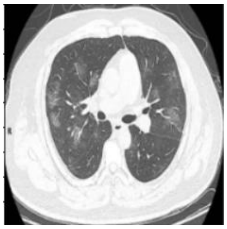
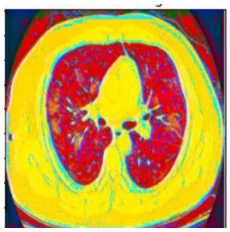
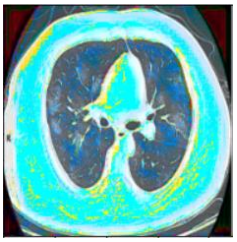
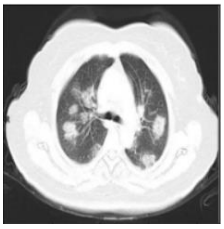
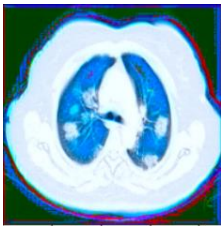
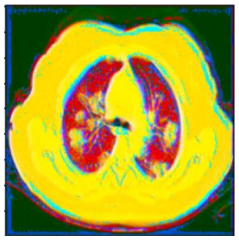
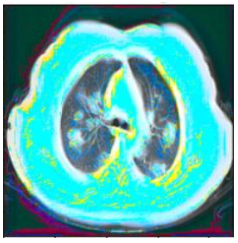
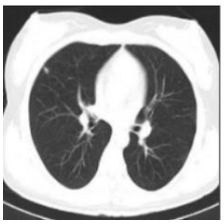
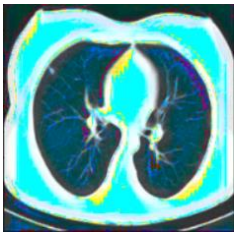
Chapter 5

Results

The content of this chapter is to show the results of the selected two distinct models and list the performance by using tables for each reference image and dataset reflecting the research outcomes. In terms of evaluations, the colorization results are essentially compared based on PSNR and SSIM, which are employed for objective assessments, subjective evaluations, image metrics instead of human visual system.

5.1 Colorizing CT Lung Images in the CNN-Based Method

Table 5.1 The final results of the CNN method in terms of painting and meat datasets

	CT-lung images (original)	Painting dataset (2,196 pics)	Meat dataset (2,100 pics)	Meat dataset (30 pics)
COVID-19 lung_01 image				
COVID-19 lung_02 image				
Healthy lung image				
Model	-	epochs=53,verbose=1, steps_per_epoch=38	epochs=50,verbose=1, steps_per_epoch=42	epochs=10,verbose=1, steps_per_epoch=3

The quantity information in the data is correspondingly grouped into more epoch steps to guarantee that data can be trained sufficiently. The number of epochs determines the number of images to train in each epoch step. For example, 1,000 training images necessitate 50 images in once epoch and a total of 20 steps per epoch.

5.1.1 Painting Works as Training Data

In comparison to the row of original CT lung images and colorizing by painting dataset in Table 5.1, it is explicit that the outcomes of CT lung images are primarily rendered to blue and white colour. Under the examination of visual analysis, it fails to colorize the

colour of a genuine lung in real life. However, the texture after colorizing is more precise than the original standard CT lung images, which have been modified enormously, especially for the healthy lungs, are more precise and sharper. From the colour point of view, the previous CT lung image has multiple black and grey colour levels, and the coloured CT lung images have altered blue levels, though.

5.1.2 2,100 Meat Pictures as Training Data

Utilizing a significant amount of meat images to colorize the greyscale CT lung images, the colours of the generated CT lung images have a strong contrast that essentially combines red with yellow (meat dataset with 2,100 pictures in Table 5.1). Compared with the COVID-19 lung images and the healthy lung image, the fundamental colours distinctions of these two conditions are in red and blue, individually. It is remarkably gratifying that this is an undeniable method to distinguish healthy lung images from infected lung images. If there are additional experimental target images following, the conclusion will be much accurate and reliable. In general, the effect of lung colour blooming in meat data with 2,100 images is relatively satisfactory compared to the painting data. Both the bright colours and the details of the lungs are displayed at an extraordinary performance. Although a larger number of blue colours are attached to the healthy CT lung images, the texture of the lung images is judged in this trained model.

5.1.3 30 Meat Images as Training Data

With 30 images of various meats, the finally-generated images are extensively in light blue. On the details of the processing, we notify that the colour CT image set of meats is not as good as that of others. For instance, the outer layer of the infected lung is dotted with yellow rather than full-colour filling. The texture of the colour healthy CT lung image is not as distinct as the previous two training data collections. Taken into consideration that the consequences of this set of experiments are merely generated by

training on 30 meat images, the outcome has better performance than other machine learning methods.

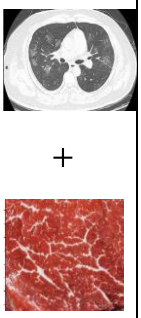
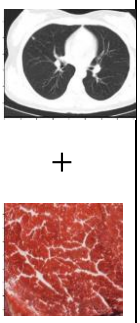
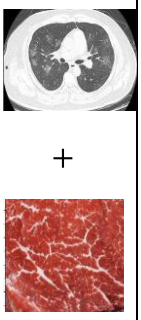
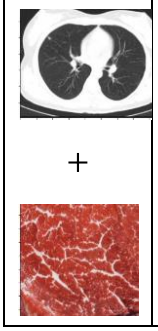
From a simple visual analysis, we believe that a training set with enormous data performs more accurately than a small amount of data. In addition, compared the two collections of data with paintings and meat images, the consequences of these two sorts of data exhibit their corresponding advantages. In other meat data, we discovered that multiple colours individually indicate healthy lungs and infected lungs. However, these conclusions demand to be evidenced in further operations and numerous experiments.

5.2 Colorizing CT Lung Images in the Hybrid Method

5.2.1 Steak images as the Reference Ones

The results in each iteration of CT lung images (COVID-19 lung and healthy lung) with steak image as the reference image is illustrated in Table 5.2. Apparently, the colours of generated images are gradually intensifying as the number of iterations increases. It resembles that CT lung images are progressively converting authentic due to the more in-depth neural network. From the visual evaluation of this ultimate result, the colour deviation to the most standard red colour is the closest to the colour of the human organ. The original white part in the CT grayscale images, and presently the colour red has replaced it. Moreover, the red is frequently superimposed colour in the black part of initial CT lung images. On the whole, the effect of the CT lung image after colorizing is extraordinarily beneficial, and even surpasses the original intention.

Table 5.2 The hybrid method by using steak images as the reference ones

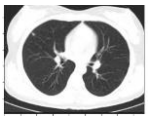
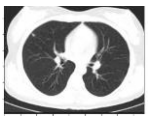
Times	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
					
					
Times	Iteration 6	Iteration 7	Iteration 8	Iteration 9	Iteration 10
					
					

5.2.2 Lamb as Reference Images

According to the reference image colour, unmistakably, the lamb image colours are more moderate than the colour of steak images. Therefore, the coloured CT lung images generated by applying lamb as a reference image are more balanced than the steak image.

Visually, the trustworthiness that the fat part of lamb is more extended (white part) that the generated target image is more accurate than the rest of the reference images.

Table 5.3 The hybrid method by using lamb images as the reference ones

Times	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
 + 					
 + 					
Times	Iteration 6	Iteration 7	Iteration 8	Iteration 9	Iteration 10
 + 					
 + 					

5.2.3 Pork as Reference Images

Table 5.4 The hybrid method by using images of pork as reference one

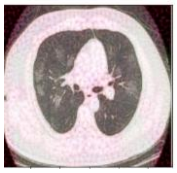
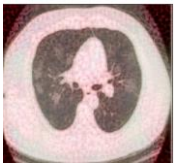
Times	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
 + 					
 + 					
Times	Iteration 6	Iteration 7	Iteration 8	Iteration 9	Iteration 10
 + 					
 + 					

The reference colours of pork images are essentially pinkish, the ultimately generated images are also approximately pink. From the results of CT lung images colorizing, we

recognize that pink is an evenly bloomed colour in the reference image. Since there is no gradual colour, the details of the CT lung images are possible, not as robust as other reference images.

5.2.4 Rotten Meat as Reference Images

Table 5.5 The hybrid method by using images of rotting meat as reference ones

Times	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
 + 					
Times	Iteration 6	Iteration 7	Iteration 8	Iteration 9	Iteration 10
 + 					

For decaying meat images as a reference one to colorize the infected lung CT images, we observe that the colours of the entire lung images have attained the superior result. The colours of rotten flesh are composed of brown and red. Consequently, the COVID-19 CT lung images are evenly colourized by these two colours. It confirms that our model adequately transfers the colour components of the reference images to the target images in a particular situation.

It is challenging to differentiate from visual observations which one of these reference images presents a more dependable outcome. We discerned that the fat part of lamb images forms a particular advantage compared to others as the reference images.

We notice the infected lung images that the rotten meat images also produced a remarkable outcome. The model is fully employed to render the entire CT lung images iteratively. Meanwhile, it also demonstrates that the supposition of this model is remarkably effective and beneficial.

5.3 Evaluation Methods

Evaluating the quality of colorizing CT lung images in the perspective of the human visual is possible subjective and irrational way due to the variety from person to person. Accordingly, it is essential to establish a quantitative and empirical approach to examine the influence of the ultimately generated images on the quality of the original CT lung images. Measuring the quality of CT images depends on four fundamental factors: Image contrast, spatial resolution, image noise and artifacts (Verdun et al., 2015). PSNR is customarily employed to estimate and compare the quality between the reconstructed image and the original image. The evaluation of structural SIMilarity (SSIM) metric consolidates the original image contrast, luminance, and structure into a unique local property score.

Moreover, PSNR and SSIM have a high frequency to measure image quality in the deep neural network (Horé & Ziou, 2013). If the values of PSNR and SSIM demonstrate a reference image or a collection of visual data perform the CT lung images with colour closer to the original images, we accurately resolve that they are more accountable algorithms and opportunity for future extensive research.

5.3.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR is typically manifested in the logarithmic decibel scale manner because numerous signals have a surprisingly extended dynamic range, for instance, the proportion between the highest and least possible values of a varying amount (Najafipour, Babaee, & Shahrtash, 2013). In order to measure the PSNR, the block firstly computes the Mean

Square Error (MSE) (Tanchenko, 2014), which is generally defined by the following equation (5.1)

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [R(i,j) - G(i,j)]^2 \tag{5.1}$$

where *i* and *j* represent the position corresponding to each pixel. *R* and *G* indicate two sets of monochrome images. If *R* is the noisy approximation of *G*, the *MSE* between these two images is defined equation (5.2).

$$PSNR = 10 \cdot log_{10}(\frac{MAX_R^2}{MSE}) = 20 \cdot log_{20}(\frac{MAX_R}{\sqrt{MSE}}) \tag{5.2}$$

where the smaller *MSE*, the higher *PSNR*, the smaller distortions after compression or reconstruction. This trend of the value indicates that the higher PSNR, the more excellent quality of the reconstructed image (Sara, Akter, & Uddin, 2019).

5.3.2 SIMilarity (SSIM)

Contrasted with the PSNR utilizing the ratio to measure the similarity between the original images and reconstructed images, SSIM takes advantages of the matrix to mimic the human visual system and measure these two given images. For human visual perception, the system is profoundly capable of defining structural information from an image. Thus, it performs to distinguish the varying information extracted from the example image and the formed image. Moreover, SSIM is frequently regarded as a loss function in deep neural networks since it was emerged in 2004. However, in this thesis, it is performed to quantify the variation in each of the corresponding pixels between the original CT lung images and the images after colorization. The basic principle of SSIM is that raw images are deeply structured, which are adjacent pixels with a robust correlation. This correlation proves the structural information of images in the scene (Zujovic, Pappas, & Neuhoff, 2013).

Two images are defined *x* and *y*, *x* is our reference image and *y* is the image obtained after the rendering. Equation (5.3) shows three comparative components within samples *x* and *y*, luminance *l*, contrast *c*, and structure *s*. These three principal features are extracted from the supplied two images. We calculate the score SSIM as

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha}[c(\mathbf{x}, \mathbf{y})]^{\beta}[s(\mathbf{x}, \mathbf{y})]^{\gamma} \tag{5.3}$$

where $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ signify the relevant importance in each of the metrics. If we believe $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, we obtain,

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} . \tag{5.4}$$

During the calculation of SSIM between two supplied images, the range of value is from -1 to +1. An amount of labels indicate that the two performed images are incredibly similar and even the equivalent. These SSIMs are adjusted within the interval [0, 1]. The higher amount in SSIM signifies that the smaller image distortion and the more attractive image quality.

Chapter 6

Analysis and Discussions

In this chapter, our experimental results will be clearly analysed and discussed. Comparing to the colorized images from the human visual perspective and standard evaluation methods, we applied RGB colour model to explain the results of the ultimate generated image scientifically and effectively. Additionally, the standard evaluation methods are PSNR and SSIM to investigate the gap between the original data sets or reference images and the colorized CT lung images. Based on the analysis related to the experimental results, we eventually consider to obtain the most suitable methods for rendering the greyscale CT lung images.

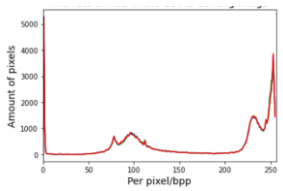
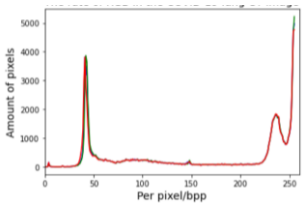
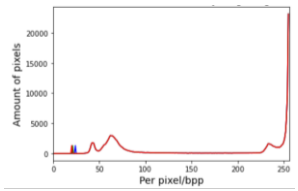
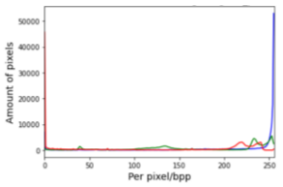
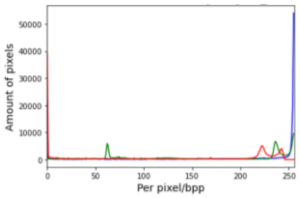
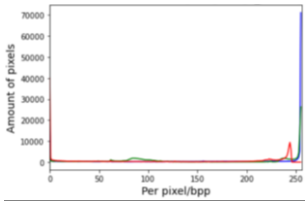
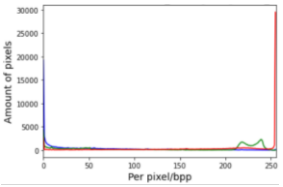
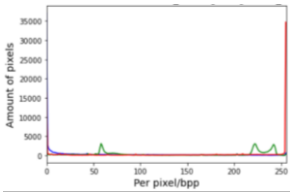
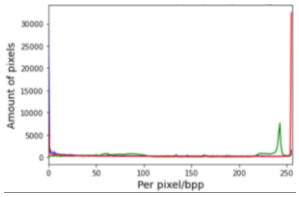
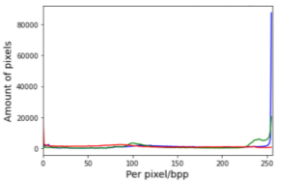
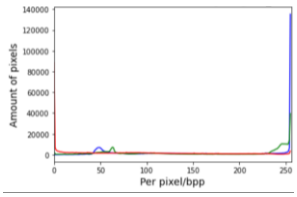
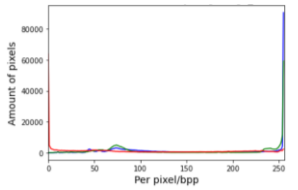
6.1 Analysis CNN-Based Method

6.1.1 Human Visual Analysis in Terms of CNN-Based Method

RGB colour image is decomposed into three channels: Red, green and blue. The colour proportions of red, green and blue in the CT lung images are exhibited in Table 6.1, which presents the original state of the original CT lung images and after colorizing the CT lung images. The abscissa represents the interval from 0 to 255, which is applied by the BGR colour model; the ordinate denotes the proportion of these three colours in an image. In the original uncoloured CT lung images, the three channels are overlapped, that means, these three colours are implemented evenly on the image, resulting in the overall colour being black, white, and greyscale in various lightness. In the rendered images and the meat data with 30 images, the colour blue is presented with high intensity in the blue channel but low in the green channel and red channel. It reveals the ultimate results in these two datasets that the colorizing CT lung images are close to the colour blue. While separating the three colours, we realize that meat images with additional reference images convert much redder, occupied the entire generated colour images.

Overall, according to the RGB ratio in each final image, the colorization is distinct from the reference images. It shows that in the ultimate CT lung images, merely the colour of the meat data collected from 2,100 images in red, while the colour of the meat data collected from 30 images and painting images is biased towards blue. We expect to generate colours directly towards authentic human lung colours. Thus, the influences generated by an abundance of meat data are more supreme and more fitting from the perspective of human vision in the CNN-based colorization.

Table 6.1 The RGB values in original images and generated images

Images	COVID-19_01 CT lung	COVID-19_02 CT lung	Healthy CT lung
Original image			
Painting data			
Meat data 2,100pics			
Meat data 30 pics			

6.1.2 PSNR and SSIM in Term of CNN_Based Method

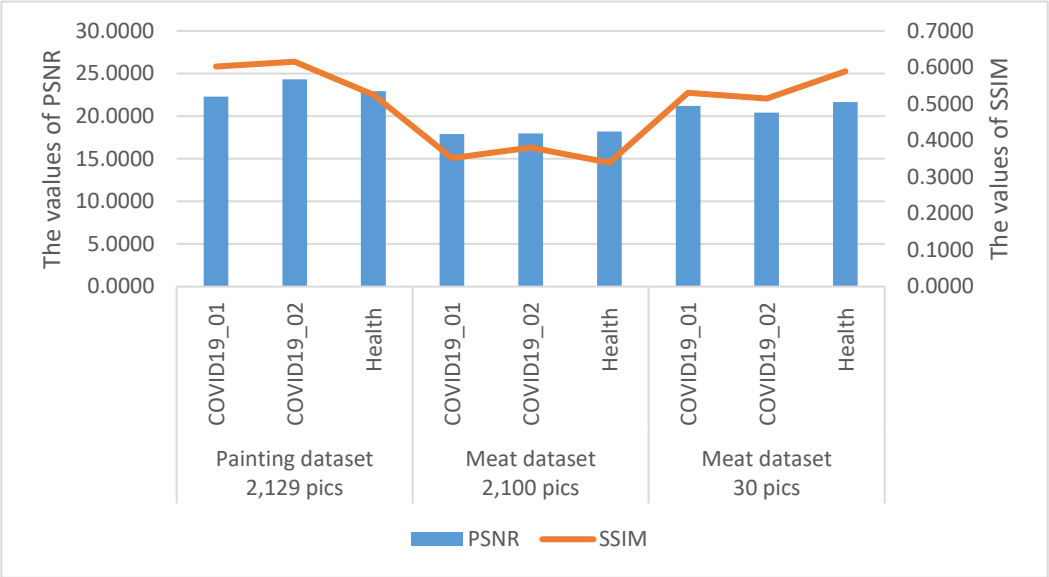


Figure 6.1 The PSNR and SSIM in the CT lung images based on three datasets

The PSNR and SSIM are applied to evaluate the CT lung images (COVID-19 and healthy) after colorizing based on three datasets as shown in Figure 6.1. Initially, compared to PSNR and SSIM in resembling amounts of the data sets, regardless of the conditions of the CT lung images, PSNR and SSIM showcase both higher values in the painting dataset than corresponding values in meat data. This situation confirms that the painting data is practical to train the CNN-based method model and transfer the similar content and style to the original CT lung images.

Compared to the ultimate values of PSNR and SSIM gained by a generous amount of data (2,100 images) and a small quantity of data (30 images), we observed the lines and bars that the CT lung images, colorized from 30 meat images, adequately restore the original image in terms of details and texture equalling to the data of collecting 2,100 meat images.

Overall, these three data collections are applied to the CNN-based methods. This collection of rendered images show the best in COVID-19 CT lung images. It retains

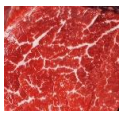
much textures and features of black and white CT lung images, while 30 images in the meat data possible are deemed much suitable for training the healthy CT lung images.

6.2 Analysis Hybrid Methodology

6.2.1 Human Visual Analysis in Term of Hybrid Method

The COVID-19 CT lung images and the healthy CT lung images have higher percentages between 0 and 250 of the RGB colour values. Regarding the principal rules in the RGB colour space, if the whole of portions denotes (0, 0, 0), which is also indicated as 000000, the corresponding colour is black; if all the shares are (255, 255, 255), which is also referred to #FFFFFF and the corresponding colour is white (Leon, Mery, Pedreschi, & Leon, 2006). Between these two extreme values, there are various colours demonstrated in this approach. The percentages of colours are expressed 0 to 1 or 0% to 100%, the values are between 0 and 255, or between hexadecimal 00 and FF.

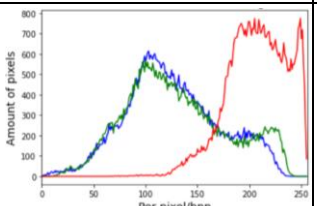
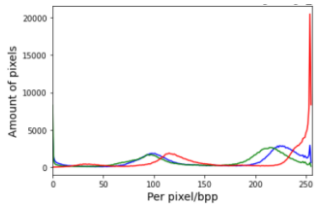
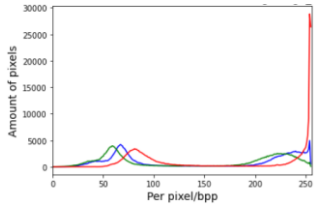
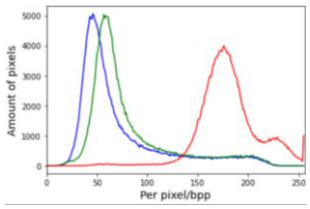
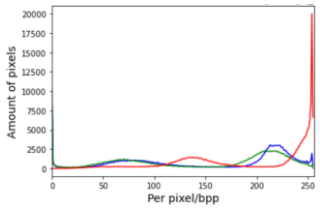
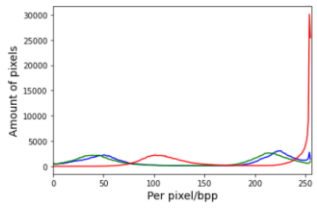
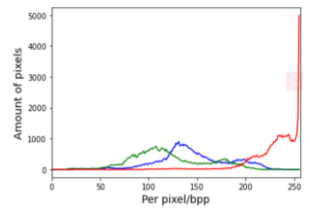
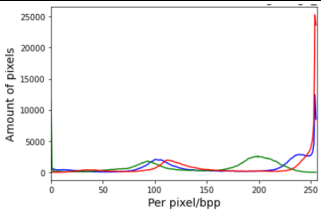
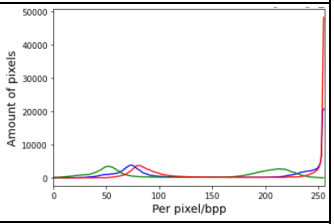
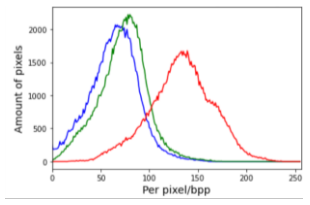
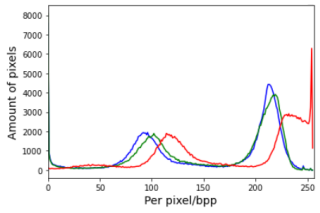
Table 6.2 The synthesized RGB values corresponding to the colours in reference images

Sample images	RGB colours	Hexadecimal code	Colours	Images
Lamb	(225, 100, 100)	#E16464		
Steak	(180, 50, 50)	#B43232		
Pork	(255, 125, 125)	#FF7D7D		
Rotting meat	(150, 75, 75)	#964B4B		

The hexadecimal notation in #RRGGBB is comprehended essentially through HTML and CSS (Goodman, 2002) (Fernandez, Ruiz, Gil, & Perez, 2015). For example,

at the rate of RGB in the lamb image in Table 6.2, the colour of red in the higher value (from 150 to 255) occupies the more tremendous rate than the other two colours, while the rate of green and blue is similar. Thus, the average value of RGB is (225, 100, 100), which is equal to #E16464 (the colour is close to red). Compared with the original CT lung image, the average colour resembles a mixed colour of the lamb image.

Table 6.3 The ratios of RGB of reference images and generated images

	Reference images	COVID-19 CT lung images	Healthy CT lung image
Lamb			
Steak			
Pork			
Rotting meat			Nah

In order to further confirm the outcomes, according to the maximum value maintained by practicing each colour in the BGR distribution of the reference image in Table 6.2, the results are illustrated in the last column, which clarifies the foremost colour of the image remains substantially comparable to the colours perceived by the overall vision. From Table 6.3, we see that though the proportions of the three colours of red, green, and blue are entirely dissimilar in the reference images, the variation in the ultimate-coloured

image is not particularly remarkable. In addition to the COVID-19 CT lung images resembling rotten meat, a vast number of green and blue colours, approximately 220 colours are also implemented in the colour model. The hexadecimal colour code (225, 210, 210) displays that the colour corresponds to light pink by checking the colour model of RGB. However, red colour attains as the average colour tone in whole ten iterations.

Table 6.4 Comparisons of final results in various referenced images

CT-lung images	Steak images	Lamb images	Pork images	Rotted meat images
Covid-19 lung				
Time(s)	376	671	423	376
Loss	522658080.0	267758700.0	518910200.0	298991550.0
Healthy lung				Nothing
Time(s)	574	570	341	
Loss	512695070.0	285693800.0	569641800.0	

Based on human visual analysis, it is challenging for us to determine which reference image captures a more realistic and reliable result based on the proportion of each colour in the RGB value. In Table 6.4, it is produced to examine the generated CT lung images by employing various reference images to compare and observe the influence of the concluding-colorized image in a more intuitive manner. The coloured images are represented in the time and loss values corresponding to the last iteration so as to generate the relevant images. The steak images are the example in this model, the lines and texture of its CT lung images appear much distinct and the colours are quite realistic.

6.2.2 PSNR and SSIM in Terms of Hybrid Method

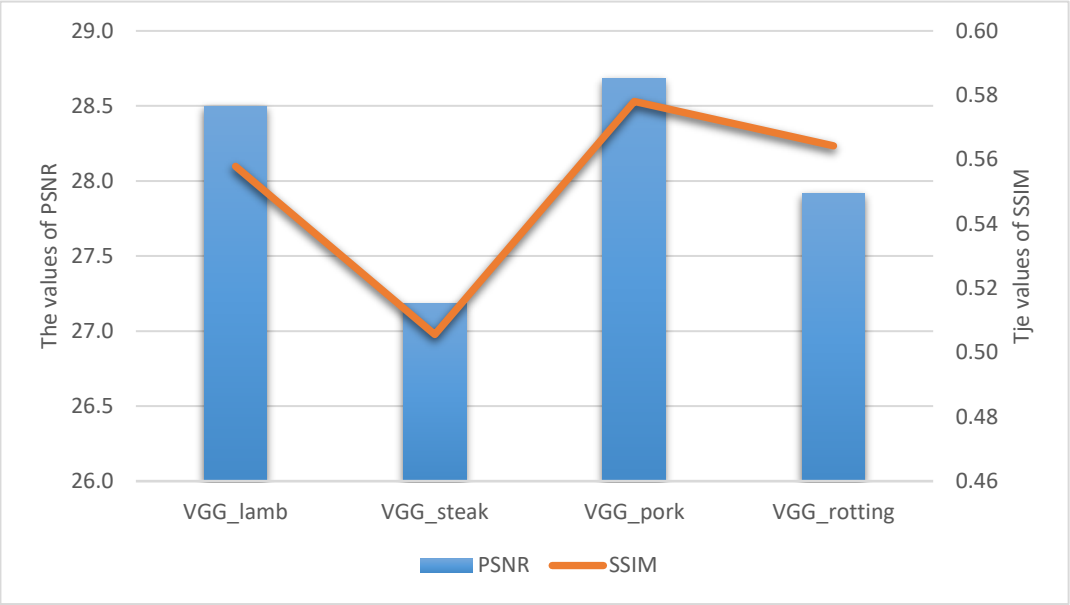


Figure 6.2 The PSNR and SSIM for the COVID-19 CT lung images

The pork image and lamb image both have more genuine outcomes in rendering the grayscale CT images than the rest of the meat images as the reference based on a series of calculations of PSNR and SSIM as shown in Figure 6.2. Despite the similarity of PSNR values (around 28.5) in these two reference images, the pork image is more dependable than the other reference images considering the value of SSIM (0.58 is the highest value). Furthermore, in the value of SSIM, the rotting meat image has a much reliable performance (over 0.56) but the PSNR of the rotting meat image exposes merely more trustworthy than the steak image (nearly 27.9). As for the steak image, it is significantly considered as the worst referenced image from both values, while it is observed bright and striking from the perspective of human vision.

In respect of healthy CT lung images, Figure 6.3 significantly explicates that the lamb image is considered more reliable than the other two classifications of meat images as the reference images. The PSNR in the lamb image (more than 28.2) is remarkably higher than other two reference images. Nevertheless, these reference images all have

outstanding accomplishments in the values of SSIM (around 0.6). The steak image examines the least recommended as the references from the consequences.

From the analysis, the lamb image is deemed as the most suitable selection for either infected CT lung image or healthy CT lung image in the CNN-based colorization method. Notwithstanding the pork image as the source image did not reach the ideal status in the value of PSNR for the healthy lung CT images. It is an undeniable truth that the pork image ranks second to form the dependable performance than the rest of meat images. Moreover, the steak image is a surprisingly unreliable source to apply because it seems thoroughly authentic from the human visual system.

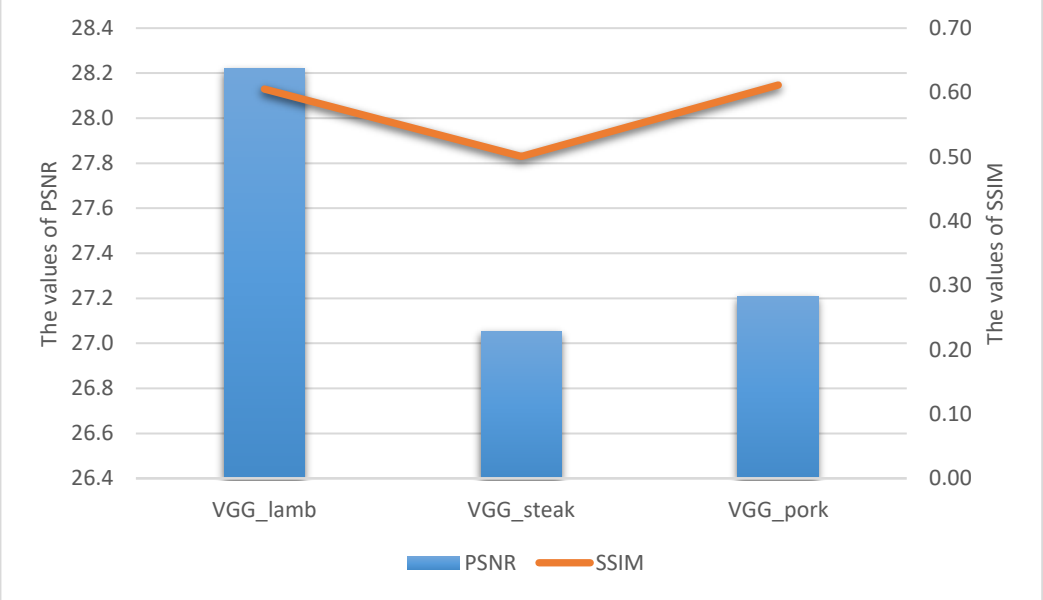


Figure 6.3. The PSNR and SSIM for the healthy CT lung images

6.3 Analysis of the PSNR and SSIM of the Infected and Healthy Lung Images

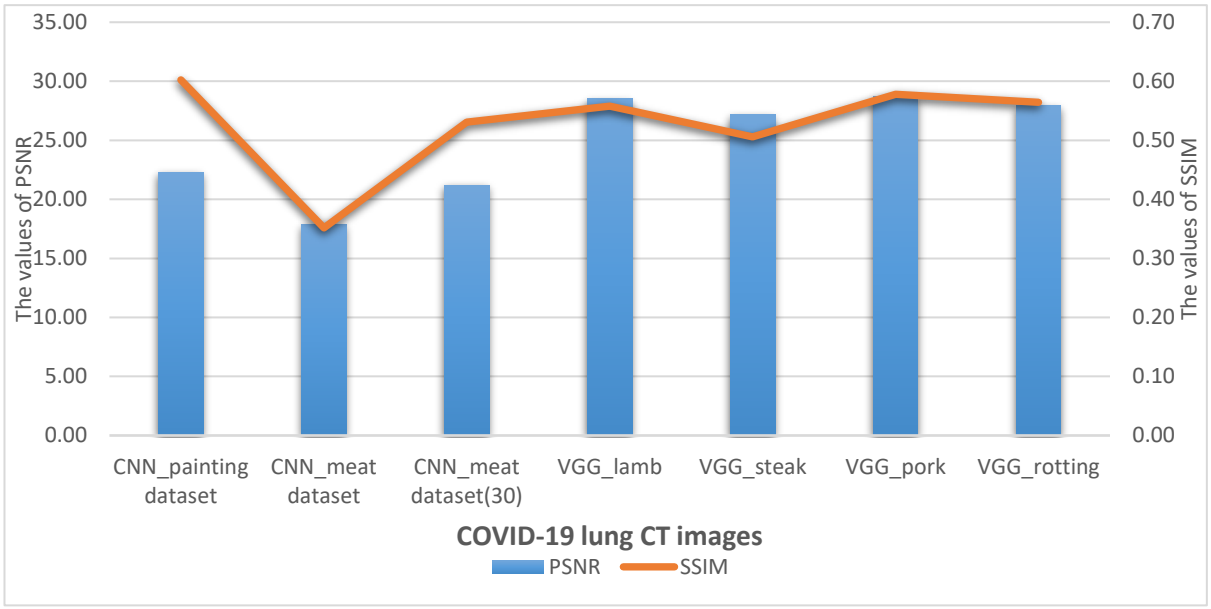


Figure 6.4 The comparisons of PSNR and SSIM in the colorized COVID-19 CT lung images

In order to explore the most suitable method to colorize the COVID-19 infected CT lung image and healthy CT lung images, we deploy the colorizing consequences after applying the reference images and training data sets and original target images (COVID-19 lung and healthy lung respectively), bars and lines are corresponding the values of PSNR and SSIM for comparison in the same figure.

The CT lung images are infected by COVID-19 as the objective image to produce the colorizing images in Figure 6.4. Employing the reference images is significantly closer to the original greyscale CT lung image. The average of the hybrid methods is approximately 28.0, while the CNN-based method is approximately 21.0 in the values of PSNR. Moreover, the optimal values of SSIM are generated in the fully automatic method by training painting data, while the SSIM values are stable for the method of applying reference images.

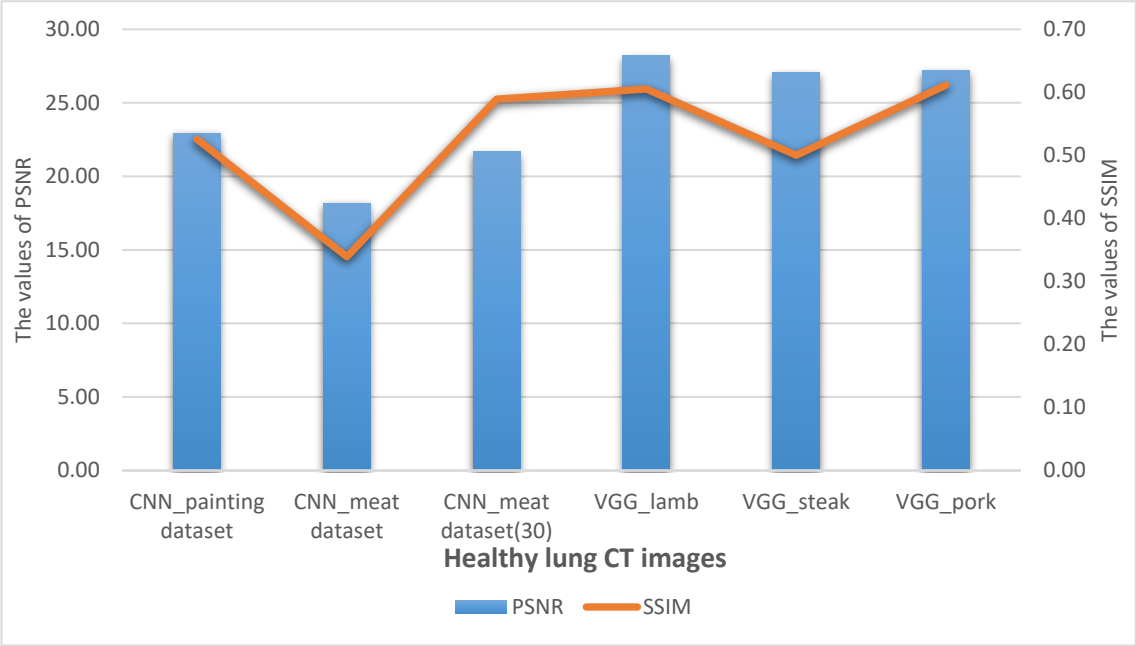


Figure 6.5 The comparisons of PSNR and SSIM in the healthy CT lung images

Similarly, the healthy CT lung images are employed as the target image for colorization. The colour CT lung images are formed by using the hybrid colorization approach that have more high-grade quality than the automatic colorization model (CNN-based method). The average value of SSIM in the method of utilizing reference images closes to 0.55. The CNN-based model is designed to apply the meat data composed of 30 images in the CNN-based methods, the SSIM values proved that meat data (30 images) has more considerable than other datasets. The PSNR values reveal the hybrid method via the reference images to generate more trustworthy images than the CNN-based method that utilizes the training datasets to configure the model.

Compared to the CNN-based colorization methods with the hybrid methods, the generated images are much stable and more eminent in the hybrid method by using reference images. Moreover, from simply analysing the human visual observation, the similar results we achieved from these two methods.

6.4 Discussion

In this thesis, our primary objective is to colorize CT lung images of the lungs through deep learning techniques. Based on the trend of colour research in up-to-date times, we have selected two kinds of methods for the purpose of comparisons between the neural networks, single reference images, and massive data sets.

One method is to configure a fully automated CNN-based typical model to train three sets of data for gaining the colour CT lung images. Specifically, the first particular collected data after colorizing that has no assistance to the colours of the human lungs, which witnesses whether the content of the data affects the colorizing results. The second set of data is derived from supermarket meat images where we took the relative images. The number of meat images is comparable to the number of rendered images in the first collection. Therefore, the first set of comparisons is formed by varying content of data. For the third collection of data, due to curiosity about the influence of the quantity of data, we collected barely 30 meat images for training, which confirms whether the number of images in each data collection will affect the ultimate images, the image generated by using this method could be confronted with a significant number of meat data.

After utilized a fully automatic method to process two COVID-19 lung and healthy lung CT lung images, the colourized CT lung images reveal a blue colour in the rendering dataset and the 30 images dataset. Even though in these two datasets, the first set has a large number of images, the second one is a small number of datasets dominated by meat images from the supermarket. From a visual point of view, a substantial amount of meat images is the most reliable one as the colourizing effectively is much approaching to the real reality. Meanwhile, compared to the hue, red and yellow-based generated images are more apparent than the blue-based created images in the other two data collections. However, from the scientific methods in terms of PSNR and SSIM, we noticed that the most realistic results are created from the rendering work and the gap with the original

image is the smallest. That means, the gap between the generated images and the original image is the largest one of these two methods. In addition, it is challenging to assume that the content and quantity of data have a numerous influence on the ultimate images. In the process of automatically adding colours, the specific content and amount of the datasets have not significantly influences based on the tone of the final colours of the CT lung images. Another approach is to colorize the target images considering the different reference images. We discovered that the generated CT lung images have plenty of links with the corresponding reference images. The principal colour tone is extracted from the reference images and blended into the greyscale CT lung images. For reference images, we applied to take pictures of pork, lamb, and beef images from our supermarket. Because this research work also has CT lung images of coronavirus lung, a rotten meat image as a reference image is applied to secure the coloured image of the lung closer to the authentic environment.

In the secondary method, we noticed a similar predicament. The beef images seem the most suitable as a reference image in the term of human perception. Nevertheless, the scientific analysis (PSNR and SSIM) proved that beef images as the reference image modify the original image a lot. In contrast, there may be white colour regions on the image due to the broader area of fat, such as lamb. Furthermore, this probably results in producing a smaller variance associated with the original image.

Overall, the hybrid method employs the reference images as a more actual outcome in these two approaches because the colours of the colorized CT lung images resemble the original human lung colour. Furthermore, from the consequences of PSNR and SSIM, there are not enormous variations of contrast to the original images. We deduced that the best-looking results frequently exchange the original image through the analysis from direct visual observation and scientific approaches (PSNR and SSIM). Therefore, we determined that whether the consequences are reliable or inadequate based on the mathematical system, human visual system is simply applied as an additional reference.

Chapter 7
Conclusion and Future Work

In this project, two types of deep learning models for colorizing the CT lung images are assessed, the research outcomes and methods are delineated in detail. In this chapter, we also integrate the conclusion into context and explain the limitations of this research. Meanwhile, we emphasize on our future work at the end of this thesis. Moreover, reasonable suggestions to the researchers who have the ambition to study CT lung image colorization further are recommended.

7.1 Conclusion

Two separated deep networks are applied to our experiments. One is essentially to configure out a fully automated colorization model based on ResNet network in CNNs and train three sets of distinctive data to obtain the corresponding three kinds of outcomes. In order to discover the correlation between the images in the training datasets and the objective CT lung images, we implemented the reduction in the size of the images to detect the complicated combination of each pixel. The operation is also named the feature extractions and then colorized the corresponding elements of pixels in the target CT lung images and the datasets. Eventually, it is discovered that the consequence of the images generated by the training image data is the most trustworthy. It also demonstrates that the content and quantity of the data have not a vast variety of influence on the finally generated images. Moreover, it witnesses that even fewer data also attains the outcomes after the entire training procedure. Deep neural networks are literately more magnificent than other machine learning methods.

In the second approach, we utilized VGG-19 to form a deep neural network model separately to extract the content of the reference images and then transfer both features to the target grayscale CT lung images. This fundamental principle of this method is initially proposed (Gatys, Ecker and Bethge (2015). In order to gain the input reference images, we designed a feature space in accordance with the filter response of each layer of the VGG-19 network to capture texture information. It consists of the correspondence among various filter responses within the spatial range of the feature map.

The image reconstruction is to reconstruct the input image only by using the network in a specific layer so as to visualize the information of varying processing stages. According to the objective and arrangement of the input image to capture the high-level content in the network, this operation does not limit the precise pixel value of the reconstruction. Eventually, our model assists in mixing the content and style

representation of reference meat images transfer to the target CT lung images. From the findings, we found that lamb image as the reference image has more beneficial than the rest of meat images because of the fewer modifications in the original image. Furthermore, the reference image is directly correlated to the colorizing of the generated CT lung images. Eventually, through comparing these two methods, we discovered that the hybrid method employing the reference images gained more favourable outcomes than the fully automated method employing the training data either analysed from a human visual viewpoint or calculated by scientific and mathematical methods (PSNR and SSIM).

Nowadays, it is the high frequency to implement deep learning approaches to colorize grayscale images or videos. Numerous methods have been undertaken to render colours to digital images. However, applying deep neural network model to colour grayscale CT lung images is a novel and new research work. In this thesis, we took two categories of lung images into consideration. The images from a healthy lung and others are the most thematic lung image currently infected by COVID-19 at one of the stages.

7.2 Limitations

The primary objective of this research project is to colorize CT lung images through deep neural networks. In this process, the CT lung images are rendered through training a model by using the reference images or datasets. In reference to the previous findings, we also discovered several limitations for our entire study.

For the methods with datasets, merely three data collections are selected for training the model. Although a collection of data performs well in terms of PSNR and SSIM, they do not match the actual colours of human lungs in terms of human visual perception.

For reference images, merely one of the various meat images is referred to as a reference image for the experiment, so it is impossible to presume that all images of such meat images are the most suitable for colorizing the CT lung images. In addition, due to

the limited environment of computer facilities, we only conducted ten times iterations in the experiment. We do not know whether more iterations result in more reliable colorizing outcomes or not.

Generally speaking, ResNet and VGG-19 net are both the neural networks for configuring the experimental models to colorize the CT lung images. However, these two deep neural networks are extensively applied to colorize the grayscale images so far, there possibly a more immeasurable method by configuring a new neural network model for colorizing CT lung images.

7.3 Future Work

In future experiments, regarding the configuration of the model, we would propose multiple deep neural networks as a training model to implement the most suitable approach for colorizing grayscale CT lung images. In the process of applying data, we will increase the training samples of various data in terms of quantity and quantity.

For reference images, we would like to take a considerable number of varying types of meat images to conduct experiments further and then discover which meat images are much suitable for colorizing CT lung images, such as chicken or fish images, etc. Meanwhile, we will enhance the number of training images of each meat image because we merely applied one image to each class of meat images. Moreover, we will collect the most fitting reference images for colorizing grayscale CT lung images in reliance on this further research work.

Additionally, the quantity of model iterations demands to increase in the future study. At present, we will endeavour to the number of iterations in 15, 20, 25, and 30 times instead of simply iterated ten times. For the number of iterations, which is not necessarily as the situation of overfitting is avoided being occurred in the training model. In the evaluations, we compared the values of PSNR and SSIM to recognize the distinctions

between the generated CT lung images and the original images even after implementing more times of iterations.

After this series of efforts, we will expose our most outstanding results to professional doctors and expert representatives to check whether our output colourful CT lung images are remarkably fit for medical applications. After that, we will examine other X-ray-related fields in depth. For example, after passed the security check, the security inspector perceives the colourful items instead of a single black, white, or grey colour image.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *USENIX Symposium on Operating Systems Design and Implementation*, 16, 265-283.

Aimar, A., Mostafa, H., Calabrese, E., Rios-Navarro, A., Tapiador-Morales, R., Lungu, I. A., ... & Delbruck, T. (2018). Nullhop: A flexible convolutional neural network accelerator based on sparse representations of feature maps. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(3), 644-656.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modelling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717-727.

Alexandre, L. A. (2016). 3D object recognition using convolutional neural networks with transfer learning between input channels. In *Intelligent Autonomous Systems 13* (889-898). Springer, Cham.

Almeida, M. C., Asada, E. N., & Garcia, A. V. (2008). On the use of Gram matrix in observability analysis. *IEEE Transactions on Power Systems*, *23*(1), 249-251.

Anwar, S., Hwang, K., & Sung, W. (2015). Fixed point optimization of deep convolutional neural networks for object recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (1131-1135).

Asmare, M. H., Asirvadam, V. S., & Iznita, L. (2009). Color space selection for color image enhancement applications. In *International Conference on Signal Acquisition and Processing* (pp. 208-212).

Ba, J., Mnih, V., & Kavukcuoglu, K. (2014). Multiple object recognition with visual attention. *arXiv preprint arXiv:1412.7755.*

Bae, K., Ryu, H., & Shin, H. (2019). Does Adam optimizer keep close to the optimal point?. *arXiv preprint arXiv:1911.00289*.

Baldassarre, F., Morín, D. G., & Rodés-Guirao, L. (2017). Deep koalarization: Image colorization using CNNs and inception-ResNet-v2. *arXiv preprint arXiv:1712.03400.*

Baldevbhai, P. J., & Anand, R. S. (2012). Color image segmentation for medical images using L* a* b* color space. *Journal of Electronics and Communication Engineering*, *1*(2), 24-45.

Bell, S., Zitnick, C. L., Bala, K., & Girshick, R. (2016). Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition* (2874-2883).

Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep Learning, MIT Press*.

Bleich, S. N., Cutler, D., Murray, C., & Adams, A. (2008). Why is the developed world obese? *Annu. Rev. Public Health*, *29*, 273-295.

Bongartz, T., Glazebrook, K. N., Kavros, S. J., Murthy, N. S., Merry, S. P., Franz, W. B., ... & McCollough, C. H. (2015). Dual-energy CT for the diagnosis of GOUT: An accuracy and diagnostic yield study. *Annals of the Rheumatic Diseases*, *74*(6), 1072-1077.

Brown, A. M. (1990). Development of visual sensitivity to light and color vision in human infants: A critical review. *Vision Research*, *30*(8), 1159-1188.

Bueno, G., Déniz, O., Salido, J., Fernández, M. M., Vállez, N., & García-Rojo, M. (2013). Colour model analysis for histopathology image processing. In *Color Medical Image Analysis* (165-180). Springer.

Bugeau, A., Ta, V. T., & Papadakis, N. (2013). Variational exemplar-based image colorization. *IEEE Transactions on Image Processing*, *23*(1), 298-307.

Burger, W., & Burge, M. J. (2016). Colorimetric colour spaces. In *Digital Image Processing* (341-365). Springer.

Burns, G. C. (1997). Museum of broadcast communications: Encyclopedia of television. *World Wide Web*.

Buzug, T. M. (2011). Computed tomography. In *Springer handbook of medical technology* (311-342). Springer.

Camastra, F., & Vinciarelli, A. (2015). *Machine learning for Audio, Image and Video Analysis: Theory and Applications.* Springer.

Cao, Y., Chen, Y., & Khosla, D. (2015). Spiking deep convolutional neural networks for energy-efficient object recognition. *International Journal of Computer Vision*, *113*(1), 54-66.

Charpiat, G., Hofmann, M., & Schölkopf, B. (2008). Automatic image colorization via multimodal predictions. In *European Conference on Computer Vision*, 126-139.

Cheang, T. K., Chong, Y. S., & Tay, Y. H. (2017). Segmentation-free vehicle license plate recognition using ConvNet-RNN. *arXiv preprint arXiv:1701.06439.*

Chen, J., Sathe, S., Aggarwal, C., & Turaga, D. (2017). Outlier detection with autoencoder ensembles. In *International Conference on Data Mining*, pp.90-98.

Chen, Y., & Pock, T. (2016). Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1256-1272.

Cheng, Z., Yang, Q., & Sheng, B. (2015). Deep colorization. In *IEEE International Conference on Computer Vision* (415-423).

Chia, A. Y. S., Zhuo, S., Gupta, R. K., Tai, Y. W., Cho, S. Y., Tan, P., & Lin, S. (2011). Semantic colorization with internet images. *ACM Transactions on Graphics (TOG), 30*(6), 1-8.

Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., ... & Moghaddam, M. (2014). A python-based open-source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables. *Remote Sensing*, *6*(7), 6111-6135.

Dhamari, A., Sudirman, R., & Mahmood, N. H. (2020). Transfer deep learning a long with binary support vector machine for abnormal behaviour detection. *IEEE Access*, *8*(61085-61095).

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision* (184-199). Springer.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295-307.

Durand, T., Mordan, T., Thome, N., & Cord, M. (2017). Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (642-651).

Duvvuri, V., Laggoni, S., & Karinga, S. (2018). Mathematical formulation for programmers to select background and foreground colours in designing websites. *Int. J. Web Appl.*, *10*(2), 47-50.

Fan, D. P., Zhou, T., Ji, G. P., Zhou, Y., Chen, G., Fu, H., ... & Shao, L. (2020). Inf-Net: Automatic COVID-19 lung infection segmentation from CT images. *IEEE Transactions on Medical Imaging*, *39*(8), 2626-2637.

Fairchild, M. D., & Berns, R. S. (1993). Image colour-appearance specification through extension of CIELAB. *Colour Research & Application, 18*(3), 178-190.

Fernandez, G. C., Ruiz, E. S., Gil, M. C., & Perez, F. M. (2015). From RGB led laboratory to servomotor control with web sockets and IoT as educational tool. In *International Conference on Remote Engineering and Virtual Instrumentation* (32-36).

Ford, A., & Roberts, A. (1998). Colour Space Conversions. *Westminster University, London, 1998*, 1-31.

Fornasier, M. (2006). Nonlinear projection recovery in digital inpainting for color image restoration. *Journal of Mathematical Imaging and Vision, 24*(3), 359-373

Garhwal, A., Yan, W. (2015) Evaluations of image degradation from multiple scan-print. *Int. J. Digit. Crime Forensics* 7(4): 55-65

Garhwal, A., Yan, W. (2015) Evaluations of image degradation from multiple scan-print. *International Journal of Digital Crime and Forensics,* 7 (4), 55-65

Galun, M., Sharon, E., Basri, R., & Brandt, A. (2003). Texture segmentation by multiscale aggregation of filter responses and shape elements. In *IEEE International Conference on Computer Vision*, 716.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576.*

Gemulla, R., Nijkamp, E., Haas, P. J., & Sismanis, Y. (2011, August). Large-scale matrix factorization with distributed stochastic gradient descent. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (69-77).

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *International Conference on Artificial Intelligence and Statistics* (315-323).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems,* 2672-2680.

Goodman, D. (2002). *Dynamic HTML: The definitive reference: A comprehensive resource for HTML, CSS, DOM & JavaScript*. O'Reilly Media, Inc.

Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., & Zhiyong, H. (2012). Image colorization using similar images. In *ACM International Conference on Multimedia*, 369-378.

Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *arXiv preprint arXiv:1510.00149.*

Harman, P. V. (2002). *U.S. Patent No. 6,496,598*. Washington, DC: U.S. Patent and Trademark Office.

He, M., Liao, J., Chen, D., Yuan, L., & Sander, P. V. (2019). Progressive color transfer with dense semantic correspondences. *ACM Transactions on Graphics (TOG), 38*(2), 1-18.

Hodnett, M., & Wiley, J. F. (2018). *R Deep Learning Essentials: A Step-by-step Guide to Building Deep Learning Models Using TensorFlow, Keras, and MXNet*. Packt Publishing Ltd.

Horé, A., & Ziou, D. (2013). Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? *IET Image Processing*, *7*(1), 12-24.

Huang, Y. C., Tung, Y. S., Chen, J. C., Wang, S. W., & Wu, J. L. (2005). An adaptive edge detection-based colorization algorithm and its applications. In *ACM International Conference on Multimedia*, 351-354.

Hui, T. W., Loy, C. C., & Tang, X. (2016). Depth map super-resolution by deep multi-scale guidance. In *European Conference on Computer Vision* (353-369).

Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (ToG), 35*(4), 1-11.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Ironi, R., Cohen-Or, D., & Lischinski, D. (2005). Colorization by example. In *Rendering Techniques*, 201-210.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1125-1134.

Johnson, T. R. (2012). Dual-energy CT: General principles. *American Journal of Roentgenology, 199*(5), S3-S8.

Kang, S. H., & March, R. (2007). Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Transactions on Image Processing, 16*(9), 2251-2261.

Kalender, W. A. (2006). X-ray computed tomography. *Physics in Medicine & Biology, 51*(13), R29.

Karaboga, D., & Basturk, B. (2007). Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. In *International Fuzzy Systems Association World Congress* (789-798). Springer.

Kingma, D. P., & Ba, J. (2014). ADAM: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kuehni, R. G. (2002). The early development of the Munsell system. 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, 27(1), 20-27.

Larsson, G., Maire, M., & Shakhnarovich, G. (2016). Learning representations for automatic colorization. In *European Conference on Computer Vision* (577-593). Springer, Cham.

Le, H., Nguyen, M., Yan, W. (2020) Machine learning with synthetic data - a new way to learn and classify the pictorial augmented reality markers in realtime. In *IEEE IVCNZ* 2020.

Leon, K., Mery, D., Pedreschi, F., & Leon, J. (2006). Color measurement in Lab units from RGB digital images. *Food Research International*, *39*(10), 1084-1091.

Levin, A., Lischinski, D., & Weiss, Y. (2004). Colorization using optimization. In *ACM SIGGRAPH*, 689-694.

Li, F., Zhang, B., & Liu, B. (2016). Ternary weight networks. *arXiv preprint arXiv:*1605.04711.

Li, X., & Li, A. (2019). An improved image enhancement method based on lab color space retinex algorithm. In *International Conference on Graphics and Image Processing (ICGIP 2018)* (Vol. 11069, p. 110692N).

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition* (2117-2125).

Liu, X., Kan, M., Wu, W., Shan, S., & Chen, X. (2017). VIPLFaceNet: an open-source deep face recognition SDK. *Frontiers of Computer Science, 11*(2), 208-218.

Liu, X., Wan, L., Qu, Y., Wong, T. T., Lin, S., Leung, C. S., & Heng, P. A. (2008). Intrinsic colorization. In *ACM SIGGRAPH Asia*, 1-9.

Liu, Z., Wang, L., Li, C., & Han, Z. (2017). A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway. *IEEE Transactions on Industrial Informatics*, *14*(3), 1067-1077.

Long Jiang Zhang, Gui Fen Yang, Sheng Yong Wu, Jian Xu, Guang Ming Lu, & Schoepf, U. J. (2013). Dual-energy CT imaging of thoracic malignancies. *Cancer Imaging, 13*(1), 81.

Lu, G. M., Zhao, Y. E., Zhang, L. J., & Schoepf, U. J. (2012). Dual-energy CT of the lung. *American Journal of Roentgenology, 199*(5), S40-S53.

Luan, Q., Wen, F., Cohen-Or, D., Liang, L., Xu, Y. Q., & Shum, H. Y. (2007). Natural image colorization. In *Eurographics Conference on Rendering Techniques*, 309-320.

Marin, D., Boll, D. T., Mileto, A., & Nelson, R. C. (2014). State of the art: Dual-energy CT of the abdomen. *Radiology, 271*(2), 327-342.

Martin, S. S., van Assen, M., Griffith, L. P., De Cecco, C. N., Varga-Szemes, A., Bauer, M. J., ... & Schoepf, U. J. (2018). Dual-energy CT pulmonary angiography: quantification of disease burden and impact on management. *Current Radiology Reports, 6*(10), 36.

McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc..

Morimoto, Y., Taguchi, Y., & Naemura, T. (2009). Automatic colorization of grayscale images using multiple images on the web. In *SIGGRAPH*.

Najafipour, A., Babaee, A., & Shahrtash, S. M. (2013). Comparing the trustworthiness of signal-to-noise ratio and peak signal-to-noise ratio in processing noisy partial discharge signals. *IET Science, Measurement & Technology*, *7*(2), 112-118.

Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., ... & Zheng, C. (2020). Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*.

Pandey, A. T. S., & Sharma, P. D. N. (2019). Image colorization using deep learning. *International Journal for Scientific Research and Engineering Trends*.

Qiumei, Z., Dan, T., & Fenghua, W. (2019). Improved convolutional neural network based on fast exponentially linear unit activation function. *IEEE Access*, *7*(151359-151367).

Qu, Y., Wong, T. T., & Heng, P. A. (2006). Manga colorization. *ACM Transactions on Graphics*, *25*(3), 1214-1220.

Radhakrishna, A., Yan, W., Kankanhalli, M. (2006) Modelling intent for home video repurposing. *IEEE MultiMedia*, 13 (1), 46-55

Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, *21*(5), 34-41.

Ren, S., He, K., Girshick, R., Zhang, X., & Sun, J. (2016). Object detection networks on convolutional feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(7), 1476-1481.

Rougier, J. (2016). Ensemble averaging and mean squared error. *Journal of Climate*, *29*(24), 8865-8870.

Salakhutdinov, R., & Hinton, G. (2009). Deep Boltzmann machines. In *Artificial Intelligence and Statistics*, 448-455.

Sangkloy, P., Lu, J., Fang, C., Yu, F., & Hays, J. (2017). Scribbler: Controlling deep image synthesis with sketch and color. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5400-5409.

Sara, U., Akter, M., & Uddin, M. S. (2019). Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *Journal of Computer and Communications*, *7*(3), 8-18.

Scardapane, S., Van, V, S., Totaro, S., & Uncini, A. (2017). KafNets: Kernel-based non-parametric activation functions for neural networks. *arXiv preprint arXiv:1707.04035*.

Schanda, J. (2007). *Colorimetry: Understanding the CIE System*. John Wiley & Sons.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sohl-Dickstein, J., Poole, B., & Ganguli, S. (2014). Fast large-scale optimization by unifying stochastic gradient and quasi-Newton methods. In *International Conference on Machine Learning* (604-612).

Sohl-Dickstein, J., Poole, B., & Ganguli, S. (2014). Fast large-scale optimization by unifying stochastic gradient and Quasi-Newton methods. In *International Conference on Machine Learning* (604-612).

Sun, M., Song, Z., Jiang, X., Pan, J., & Pang, Y. (2017). Learning pooling for convolutional neural network. *Neurocomputing, 224*(96-104).

Sýkora, D., Buriánek, J., & Žára, J. (2004). Unsupervised colorization of black-and-white cartoons. In *International Symposium on Non-photorealistic Animation and Rendering*, 121-127.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *AAAI Conference on Artificial Intelligence 31* (1).

Tai, Y. W., Jia, J., & Tang, C. K. (2005). Local colour transfer via probabilistic segmentation by expectation-maximization. In *IEEE Conference on Computer Vision and Pattern Recognition 5* (1),747-754.

Tanchenko, A. (2014). Visual-PSNR measure of image quality. *Journal of Visual Communication and Image Representation, 25*(5), 874-878.

Verdun, F. R., Racine, D., Ott, J. G., Tapiovaara, M. J., Toroi, P., Bochud, F. O., ... & Edyvean, S. (2015). Image quality in CT: From physical measurements to model observers. *Physica Medica, 31*(8), 823-843.

Wang, G., Giannakis, G. B., & Chen, J. (2019). Learning ReLU networks on linearly separable data: Algorithm, optimality, and generalization. In *IEEE Transactions on Signal Processing, 67*(9), 2357-2370.

Wang, J., Wu, H., Zhang, X., & Yao, Y. (2020). Watermarking in deep neural networks via error back-propagation. *Electronic Imaging, 2020*(4), 22-1.

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018). Understanding convolution for semantic segmentation. In *IEEE Winter Conference on Applications of Computer Vision,* 1451-1460.

Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. In *IEEE International Conference on Computer Vision* (370-378).

Welsh, T., Ashikhmin, M., & Mueller, K. (2002). Transferring colour to greyscale images. In *Annual Conference on Computer Graphics and Interactive Techniques,* 277-280.

Weng, L., Zhang, H., Chen, H., Song, Z., Hsieh, C. J., Daniel, L., ... & Dhillon, I. (2018). Towards fast computation of certified robustness for ReLU networks. In *International Conference on Machine Learning*, 5276-5285. PMLR.

Widrow, B., Rumelhart, D. E., & Lehr, M. A. (1994). Neural networks: Applications in industry, business and science. *Communications of the ACM, 37*(3), 93-106.

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Centre for Disease Control and Prevention. *JAMA, 323*(13), 1239-1242.

Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the ResNet model for visual recognition. *Pattern Recognition, 90*(119-133).

Wrolstad, R. E., & Smith, D. E. (2017). Color analysis. In *Food Analysis* (545-555). Springer.

Xia, W., Shao, J., Guo, Y., Peng, X., Li, Z., & Hu, D. (2020). Clinical and CT features in pediatric patients with COVID-19 infection: Different points from adults. *Pediatric Pulmonology, 55*(5), 1169-1174.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition* (1492-1500).

Xu, B., Huang, R., & Li, M. (2016). Revise saturated activation functions. *arXiv preprint arXiv:1602.05980.*

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Yan, W., Kankanhalli, M. (2002) Detection and removal of lighting & shaking artifacts in home videos. In *ACM Multimedia*, 107-116

Yan, W., Kankanhalli, M. (2003) Colorizing infrared home videos. In *IEEE ICME*, pp. 97-100

Yan, W., Kankanhalli, M., Wang, J. (2005) Analogies based video editing. *Multim. Syst.* 11(1): 3-18

Yan, W. (2019) *Introduction to Intelligent Surveillance - Surveillance Data Capture, Transmission, and Analytics*, Springer International Publishing.

Yan, W. (2021) *Computational Methods for Deep Learning Theoretic, Practice and Applications*, Springer International Publishing.

Yatziv, L., & Sapiro, G. (2006). Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing, 15*(5), 1120-1129.

Yu, L., Yang, X., Chen, H., Qin, J., & Heng, P. A. (2017). Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. In *AAAI Conference on Artificial Intelligence 1*(31).

Yu, W., Yang, K., Bai, Y., Xiao, T., Yao, H., & Rui, Y. (2016). Visualizing and comparing AlexNet and VGG using deconvolutional layers. In *International Conference on Machine Learning.*

Zhang, J., & Cho, K. (2016). Query-efficient imitation learning for end-to-end autonomous driving. *arXiv preprint arXiv:1605.06450*.

Zhang, L., Zhou, C., Schoepf, U., Sheng, H., Wu, S., Krazinski, A., Silverman, J., Meinel, F., Zhao, Y., Zhang, Z., & Lu, G. (2013). Dual-energy CT lung ventilation/perfusion imaging for diagnosing pulmonary embolism. *European Radiology*, *23*(10), 2666.

Zhang, R., Isola, P., & Efros, A. A. (2016). Colorful image colorization. In *European Conference on Computer Vision* (649-666). Springer.

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

Zujovic, J., Pappas, T. N., & Neuhoff, D. L. (2013). Structural texture similarity metrics for image analysis and retrieval. *IEEE Transactions on Image Processing*, *22*(7), 2545-2558.