5-1-2022

APPLLICATION OF CONVOLUTIONAL NEURAL NETWORK TOWARDS IMAGE UPSCALING DEMONSTRATED VIA CHEST X-RAY DATASET

Murtaza Aliasgar Khokhar

APPLLICATION OF CONVOLUTIONAL NEURAL NETWORK

TOWARDS IMAGE UPSCALING DEMONSTRATED

VIA CHEST X-RAY

DATASET

by

MURTAZA ALIASGAR KHOKHAR

Presented to the Faculty of the Honors College of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

HONORS BACHELOR OF SCIENCE IN BIOMEDICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2022

ACKNOWLEDGMENTS

I would like to acknowledge and thank my mentors Dr. Lee and Dr. Behbehani for their time and efforts towards helping with this study. I would also like to thank my Bioengineering Senior Design team members: Nicholas Laudermilk, Xi Hau Tan, and Nowshin Faiza. A special thank you to Nicholas for his significant contribution towards the development of these deep learning Neural Network models that were used for this project.

April 29, 2022

ABSTRACT

APPLLICATION OF CONVOLUTIONAL NEURAL NETWORK

TOWARDS IMAGE UPSCALING DEMONSTRATED

VIA CHEST X-RAY

DATASET

Murtaza Aliasgar Khokhar, B.S. Biomedical Engineering

The University of Texas at Arlington, 2022

Faculty Mentor: Juhyun Lee, Khosrow Behbehani

Light Field Microscopy is a 3D imaging technique that sacrifices spatial resolution to capture angular information. A neural network was developed to increase the resolution of images captured. The objective of this study is to show the application of the developed convoluted neural network towards other topics. A dataset containing chest x-ray images will be used to train, test, and analyze the neural network. The neural network will be trained by converting the images to low-resolution and using it as training data. The original high-resolution data will be used as ground truth. PSNR, SSIM and visual tests will be used to test and analyze the data. It is expected that the system will output an image that is upscaled by a factor of 2. Obtaining an upscaled image will show that the developed

system can be used to upscale various images and hence can be applied towards various other fields.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Overview

Machine Learning plays a significant role in powering many parts of modern society. LeCun et al., mentions in a study that the role of machine learning ranges from web searches, to content filtering, to recommendations on e-commerce websites. Machine learning can be used in various systems such as identifying things in images, transcribe speech to text, obtaining relevant results of search, and more. All of these use a technique called deep learning. Representation learning allows a machine to take raw data as input and automatically discover the representations needed for detection or classification. Deep learning systems are representation learning methods with multiple levels of representation (LeCun et al. 436).

One of the most developed and established models of deep learning is Convolutional Neural Networks (CNN). Yamashita et al., argues that CNN is a dominant method used in computer vision and has achieved expert level performances in the medical field. There has been a surge of interest in CNN from radiology researchers, and various studies have been published in areas such as lesion detection, classification, segmentation, image reconstruction, and natural language processing (Yamashita et al. 612).

A Convolutional Neural Network model was developed to increase the resolution of images obtained from a Light Field Microscope (LFM) system. Light Field Microscopy is a 3D imaging technique, where it sacrifices spatial resolution to capture angular information. The objective of this study is to take the CNN model developed for the LFM system and show its applications towards other fields. The aim is to show that upon tweaking the developed model slightly to fit our requirements, the CNN model can be successfully used to upscale any kind of images. A chest x-ray dataset was used for this purpose.

1.2 What is CNN?

CNN is a type of deep learning model that is used for processing data that has a grid pattern. They are designed to process data that come in multiple array formats. Colored images are a good example, as they compose of three 2D arrays that contain the pixel intensities in the color channels (LeCun et al. 439). CNNs are basically a mathematical construct that contains three main building blocks or layers: convolution, pooling, and fully connected layers. Convolutional layers and pooling layers perform feature extraction, whereas a fully connected layer maps the extracted features into the final output. The Convolutional Layer consists of a combination of linear and nonlinear operations such as convolution operation and activation function. In convolution, a small array of numbers called a kernel is applied across the input (i.e., an array of numbers), called tensor. An elementwise multiplication is performed between each element of the kernel and input tensor. The values are summed up to obtain the output values in corresponding position, called a feature map (Yamashita et al. 612).

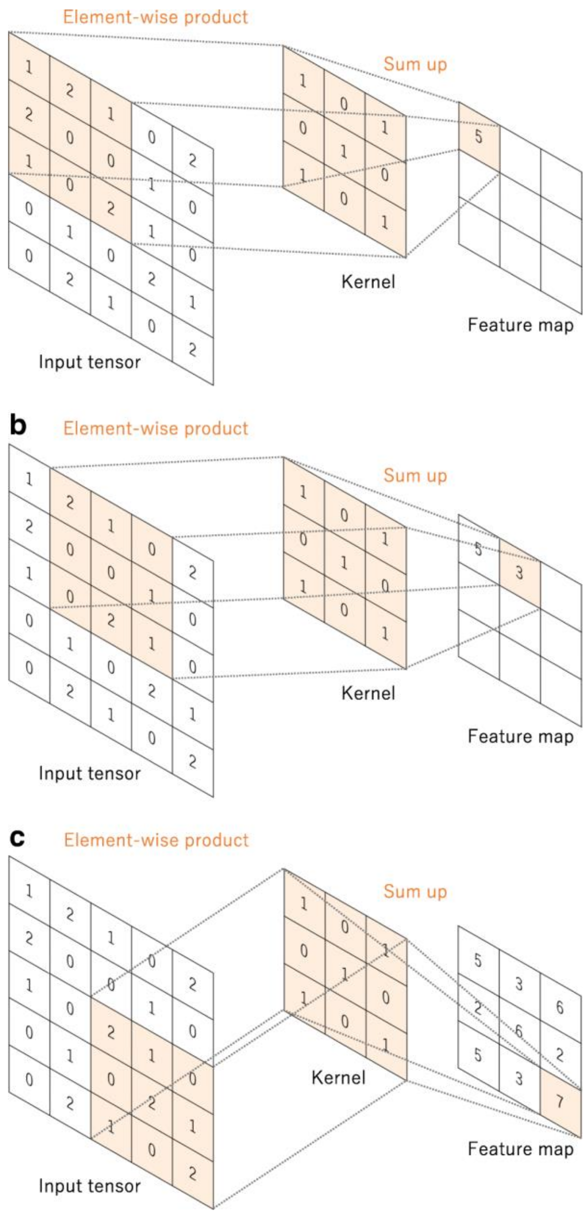

Figure 1.1: Feature map formation from input tensor and kernel (Yamashita et al., fig. 3, p. 614)

The distance between two successive kernel positions is called a stride and is typically 1. Weight sharing is a key feature of CNN's where the kernels are shared across all the image positions. The process of training a CNN model with regards to convolutional layers is there to identify the kernels that work best for a given task and is dependent on a given training dataset. Kernels are automatically learned during the training process in the convolutional layer (Yamashita et al. 614). The outputs of the linear operation are then passed through a non-linear activation function. The most frequently used one is called ReLU, which stands for Rectified Linear Units. This simply computes f(x) = max(0,x) (Krizhevsky et al. 3).

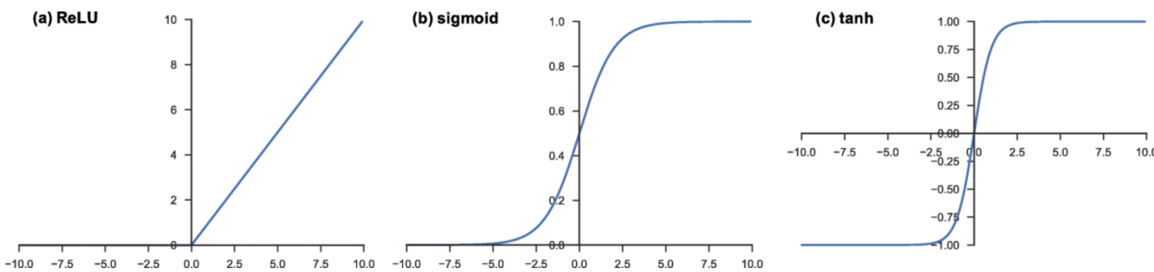

Figure 1.2: Common Activation functions used (Yamashita et al., fig. 5, p. 617)

The pooling layer does the downsampling operation. This helps introduce a translation invariance to small shifts and distortions, and hence decreases the number of learnable parameters. The most common pooling layer used is max pooling. It extracts patches from input feature maps and outputs the maximum value in each patch. The general filter size used in this is 2 x 2 (Yamashita et al. 616).

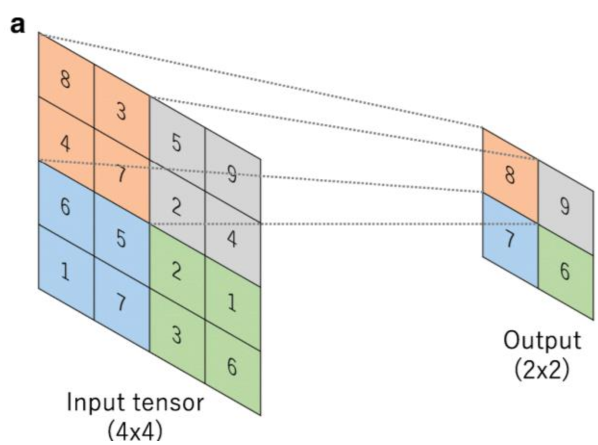

Figure 1.3: Max pooling (Yamashita et al., fig. 6, p. 618)

After the features are extracted by the convolutional layer and downsampled by the pooling layers, they are mapped by a subset of fully connected layers to the final outputs of the network (Yamashita et al. 616). The fully connected layer is a feed forward neural network where the input is the output that is obtained from the final pooling or convolutional layer. That output is flattened and fed into the fully connected layer (Arc).

CHAPTER 2

METHODOLOGY

2.1 Neural Network model

A 10-layer Convolutional Neural Network was created where each Convolutional layer has ReLU activation. Jupyter Notebook in Anaconda Environment was used to code, run, and test the software.

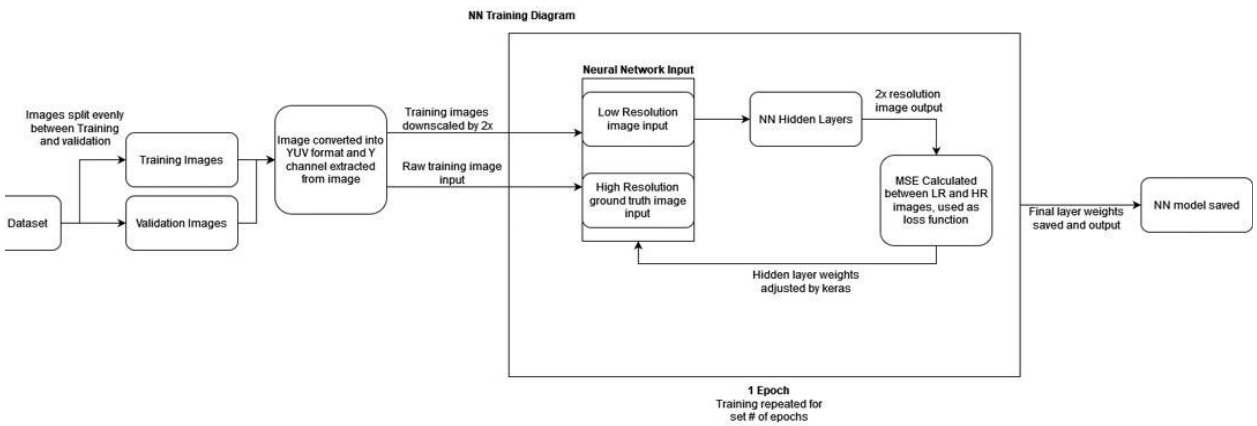

Figure 2.1: Neural Network training model

The way the neural network was trained is mentioned in figure 2.1. The chest x-ray dataset is split into two: a set of training images and a set of validation images. The training images are downscaled by two and passed through the neural network where the 10 CNN layers along with ReLU activation act on the images. The mean square error between the low-resolution training image output from the NN and that from the validation images is calculated and the layer weights are adjusted by keras. This entire cycle is 1 epoch. 30, such epochs were done for this study. The final layer weights and output are then saved. This is the trained NN model.

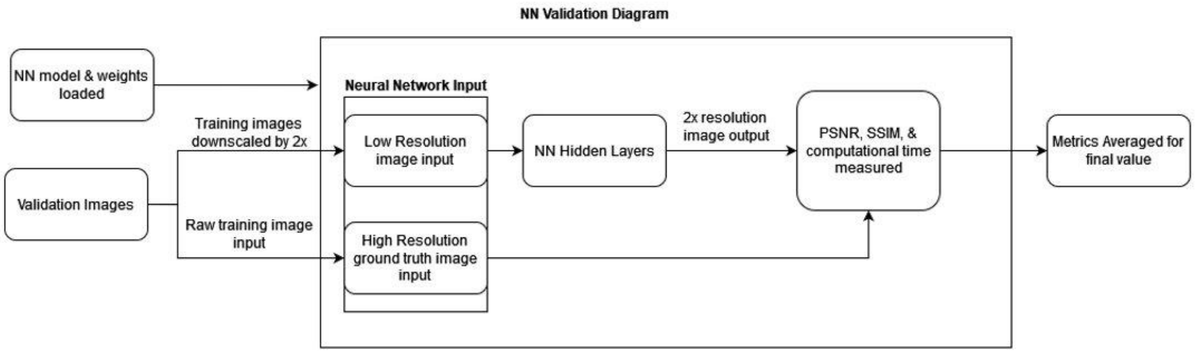

Figure 2.2: Neural Network validation model

The NN is tested and validated as shown in figure 2.2. The trained NN model is loaded. Low resolution images are input and passed through the NN hidden layers. Peak Signal-to-Noise Ratio and Structural Similarity Index were calculated using the output of the NN and high-resolution ground truth images. The values are averaged and saved as the final metric values.

2.2 Additional Features

To make the developed NN model more robust, k-fold cross validation and Image Data Augmentation were implemented.

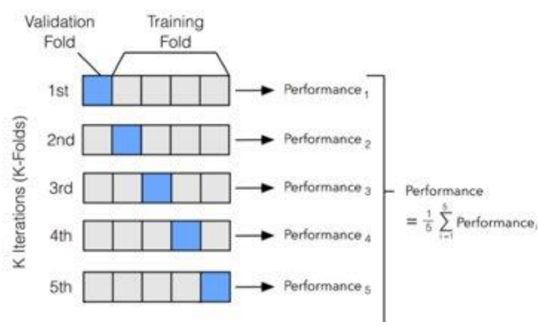

Figure 2.3: K-fold Cross Validation (Brownlee)

K-fold Cross Validation is a resampling procedure that is used to evaluate machine learning models. It is primarily applied in machine learning to estimate the skill of a model on unseen data. The general procedure is that the dataset is shuffled randomly and split into k groups (Brownlee). For each unique set, it is taken as a validation set while the rest are taken as a training dataset. A model is fit on the training dataset and evaluated on the validation set. The evaluation score is retained (Brownlee). A 10-fold cross validation was performed in training this model.

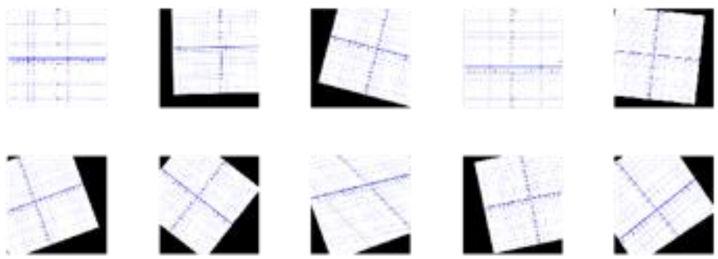

Figure 2.4: Keras ImageDataGenerator (Chollet)

Keras ImageDataGenerator augments data via several random transformations to produce new images. This helps in the model, never seeing the exact same image twice and prevents overfitting. This class helps configure random transformations and normalization operations to be done on the image data. This class provides the option to rotate, shift, rescale, zoom, and flip the image and more (Chollet).

2.3 Analysis Tests

Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) were used to analyze the image quality. PSNR is often used when measuring reconstruction quality, where a higher value is considered better but is not perfect in evaluating the image's quality. MSE is taken by squaring the average value of the difference between values (Sara et al. 8).

For a reference image f and test image g, the PSNR and MSE can be calculated as:

$$PSNR(f,g) = 10\log_{10}(255^2/MSE(f,g))$$

where

$$MSE(f,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_{ij} - g_{ij})^2$$

Here, M x N represents the size of the images. The PSNR value approaches infinity as the MSE approaches zero. This shows that a higher PSNR value corresponds to a higher quality image.

Structural Similarity Index (SSIM) was also calculated to obtain a better perception-based metric than PSNR and MSE (Sara et al. 11). In this perception-based model, image degradation is considered as the change of perception in structural information. The positive values of SSIM range from 0 to +1. A value of 0 means that there is no correlation between the images and a value of +1 means that the images are very similar or the same (Hore et al.). SSIM can be expressed as:

$$SSIM(f,g) = l(f,g)c(f,g)s(f,g)$$

where, luminance, contrast, and structure of an image are expressed as:

$$
\begin{cases}
l(f,g) = \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1} \\
c(f,g) = \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2} \\
s(f,g) = \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3}
\end{cases}
$$

Here, μf and μg are the local means, σf and σg are the standard deviations, and σfg is the cross-covariance for images x and y sequentially (Sara et al. 12). The first term represents the luminance comparison function and measures the closeness of the two images' mean luminance. The second term represents the contrast comparison function and measures the closeness of contrast of the two images. The third term represents the structure comparison function, which measures the correlation coefficient between the two images (Hore et al.).

A visual quality test was also performed where 2 sets of images, each set containing Low-Resolution image, Bicubic interpolation image, and Output image from developed Neural Network, were sent to 20 people. These people were tasked with ranking the images in each of these sets from 1st to last in terms of visual quality. Since 2 sets of images were sent to 20 people, our number of trials was n = 40. The images were not labeled and were ranked based solely on visual quality.

CHAPTER 3

DISCUSSION

The chest X-ray dataset was input into the developed NN. The output image was compared to that obtained from bicubic interpolation. PSNR and SSIM values were calculated, and visual quality test was conducted. The figures below show the set of images that were sent for visual inspection.

Figure 3.1: Set 1 - low resolution image

Figure 3.2: Set 1 - Bicubic Interpolation image

Figure 3.3: Set 1 - 10 Layer CNN image

Figure 3.4: Set 2 - low resolution image

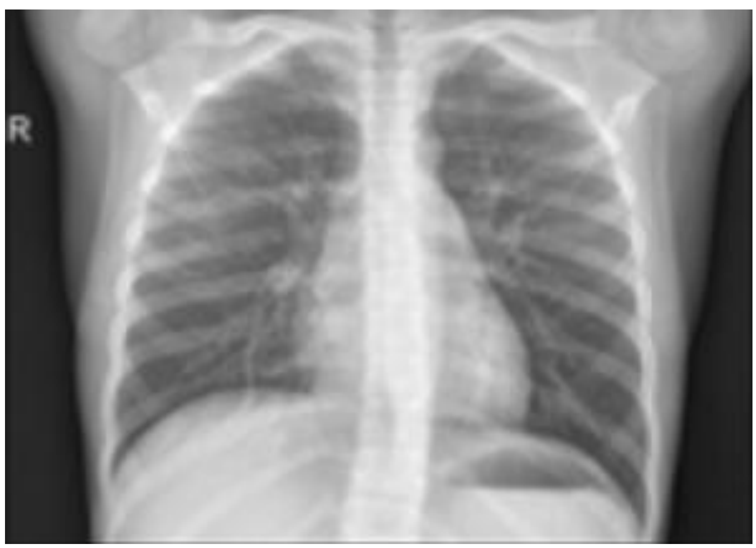

Figure 3.5: Set 2 - Bicubic Interpolation image

Figure 3.6: Set 2 - 10 Layer CNN image

The PSNR and SSIM values obtained from validating the 10-layer CNN developed were compared with that from bicubic interpolation. The values are shown in the table below.

Table 3.1: PSNR and SSIM values for 10 Layer CNN and Bicubic Interpolation model

MODEL	PSNR	SSIM
10 Layer CNN	35.3	0.947
BICUBIC	36.3	0.955

The results obtained from the visual quality test are shown in table below:

Table 3.2: Results from visual tests

VISUAL TEST	10 LAYER CNN	BICUBIC
No. of People	38	2

Our model was compared against a trained bicubic interpolation model, which is also used to upscale images. Upon training and testing our model, the PSNR and SSIM values for the 10-layer CNN were slightly lower than that from the bicubic interpolation model. However, the results from the visual quality test suggested otherwise. Out of the 40 trials conducted, 95% of people rated the images obtained from our 10-layer CNN model as the highest quality image out of the set. While training, it was observed that increasing the dataset size and the number of epochs trained, increased the PSNR and SSIM values of the images from 10-layer CNN model. However, due to limitations with computer RAM and running into Out of Memory Limitations, the training was limited to 30 epochs per fold.

CHAPTER 4

CONCLUSION

The higher the PSNR value and closer the SSIM value to 1 the better. The results show that our model achieved a PSNR and SSIM value slightly lower than the bicubic model. While these metrics are important in determining image quality, visual tests performed by human inspection also play an important part. 95% of people rated the images produced by the 10 Layer CNN model as the image with the better resolution. This shows that with minor changes to the developed model, and upon proper training, our developed model can be used to accurately upscale most images of any type. This shows the application of our model towards various other fields. Further development of the model can be done by adding a detection part to the developed model. This additional development might be useful in detecting whether the x-ray belongs to an individual with pneumonia or not.

REFERENCES

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.

Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging* 9.4 (2018): 611-629.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

Arc. “Convolutional Neural Network.” *Medium*, Towards Data Science, 26 Dec. 2018, towardsdatascience.com/convolutional-neural-network-17fb77e76c05.

Brownlee, Jason. “A Gentle Introduction to k-Fold Cross-Validation.” Machine Learning Mastery, 2 Aug. 2020, machinelearningmastery.com/k-fold-cross-validation/.

Chollet, Francois. "Building powerful image classification models using very little data." Keras Blog 5 (2016).

Sara, Umme, Morium Akter, and Mohammad Shorif Uddin. "Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study." Journal of Computer and Communications 7.3 (2019): 8-18.

Hore, Alain, and Djemel Ziou. "Image quality metrics: PSNR vs. SSIM." 2010 20th international conference on pattern recognition. IEEE, 2010.

BIOGRAPHICAL INFORMATION

Murtaza Khokhar is a senior pursuing an Honors Bachelor of Science degree in Biomedical Engineering. Along with that, he is also working on a minor in Computer Science. His senior design research included working on building a Light Filed Microscope system and on doing research on developing a Convolutional Neural Network to upscale the resolution of images. He has also worked on various fabrications and molding for development of digits, wrist and elbow actuators, and helmet liners.