DEEP LEARNING BASED COMPUTED SUPER-RESOLUTION OPTICAL MICROSCOPY

Pradeepta Das

DEEP LEARNING BASED COMPUTED SUPER-RESOLUTION OPTICAL MICROSCOPY

Thesis submitted to
Indian Institute of Technology Kharagpur
for the award of the degree

of

Bachelor of Technology
in
Instrumentation Engineering

by

Pradeepta Das

(12IE10021)

Under the supervision of

Dr. Debdoot Sheet

Department of Electrical Engineering
Indian Institute of Technology Kharagpur
May 2016

DECLARATION

I certify that

a. The work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Pradeepta Das

CERTIFICATE

This is to certify that the Dissertation Report entitled, **Deep Learning based Computed Super-Resolution Optical Microscopy** submitted by **Mr. Pradeepta Das** to Indian Institute of Technology Kharagpur, India, is a record of bonafide Project work carried out by him under my (our) supervision and guidance and is worthy of consideration for the award of the degree of Bachelor of Technology in Department of Electrical Engineering of the Institute

Dr. Debdoot Sheet
Department of Electrical Engineering
Indian Institute of Technology Kharagpur
India

Date:

Acknowledgment

The roots of all goodness lie in the soil of appreciation for goodness.

Dalai Lama

I would like to take this opportunity to express my heartiest thanks and immense gratitude to all who supported me and inspired me to come so far and made this journey successful. I would begin by expressing my gratitude to my supervisor, Dr. Debdoot Sheet for his continuing support and valuable feedbacks throughout my thesis. I thank Niladri Garai, Abhijit Guha Roy, Kausik Das, Praveen Kumar, Rachana Satish at Department of Electrical Engineering for their timely guidance and encouraging me to explore and learn. I am grateful to my lab mates and dear friends for helping me stay motivated and complete my work on time. I would also like to mention some of my dear friends and classmates Asif Ahmed, Sebin Mathew, Prabhat Yeluri and Venkat Sainath for their constant support and encouragement. Last but not the least, all this would not have been possible without constant support and love from my mother, father and sister. I am really honored to have been associated with these amazing people and am very grateful for their support throughout this amazing endeavor.

Contents

List of Figures

List of Tables

List of Abbreviations

AE	Auto Encoder
DBN	Deep Belief Network
GMM	Gaussian Mixture Model
HR	High-resolution
LR	Low-resolution
MRF	Markov Random Field
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
RBM	Restricted Boltzmann Machine
RMSE	Root Mean Square Error
SAE	Stacked Auto Encoder
SR	Super Resolution
SSIM	Structural Similarity Measure

List of Symbols

X	HR Image
Y	Decimated and blurred version of original HR image i.e. LR image
D	Decimation matrix
B	Blur matrix
D^h	High-resolution dictionary
D^l	Low-resolution dictionary
$m_H \times n_H$	patch size ($m \times n$) of HR image
$m_L \times n_L$	patch size ($m \times n$) of LR image
R^m	Feature vector of m
$\mathbf{W}$	Weight matrix of an AE
α	learning rate
x^k	k^{th} input to the auto encoder
$h^{(i)(k)}$	i^{th} feature vector corresponding to the x^k input
a_i^l	activation (meaning output value) of unit i in layer l
z_i^l	total weighted sum of inputs to unit i in layer l
W_{ij}^l	the parameter (or weight) associated with the connection between unit j in layer l, and unit i in layer $l+1$
b_i^l	Bias associated with unit i in layer $l+1$

Abstract

HIGH-RESOLUTION (HR) Images are desired and often required in most of the electronic imaging applications. HR means that pixel density within an image is high, so an HR image can offer more details that may be critical in various applications. For example, HR images are helpful to distinguish an object from similar ones in satellite images, the performance of pattern recognition in computer vision can be improved if an HR image is provided. HR medical images are very helpful for a doctor to make a correct diagnosis. Since the 1970s, charge-coupled device (CCD) and Complementary metal-oxide semiconductor (CMOS) image sensors have been widely used to capture digital images. Although these sensors are suitable for most imaging applications, the current resolution level and consumer price will not satisfy the future demand. For example, people want an inexpensive HR digital camera and scientists often need a very HR level which has no visible artifacts when an image is magnified. Thus, finding a way to increase the current resolution level is needed.

Super-resolution (SR) is a class of techniques that enhance the resolution of an imaging system. Many applications in graphics or image processing such as image-based rendering (IBR), texture mapping, enlarging consumer photographs, converting National Television System Committee (NTSC) video content to high-definition television and Medical Image Analysis could benefit from HR Images.

In this project we present a fast computational method for post-acquisition SR of optical microscopy images. We leverage a one-pass method using a k-d tree based nearest neighbour search over a dictionary of HR to LR mapped exemplar pairs. Restroration of high frequency component is better in the proposed method as compared to the traditional methods.

In the second part of the project, we develop a new approach which is based on the Deep Belief Network (DBN) for image super-resolution. The DBN architecture has ability of learning a set of visual patterns, called dictionary elements from a set of training images. The learned dictionary will be then used to synthesize high resolution images.

Keywords: Deep learning, super-resolution, optical microscopy, exemplar based methods.

1 Introduction

If others can see it as I have seen it, then it may be called a vision rather than a dream.

William Morris

1.1 Overview

OPTICAL microscope, often referred to as light microscope, is a type of microscope which uses visible light and a system of lenses to magnify images of small samples. Microscopic image is a tool to examine histopathological images. Histopathological examination of biopsy samples using bright-field optical microscopic investigation is a practice followed by clinicians for diagnosis and staging of tissue abnormality diseases viz. cancers. This requires investigation of the samples at different optical magnifications commonly varying across 10x to 100x at the objective. Although high-magnification investigation provides more diagnostically relevant information, yet it has limited optical exposure due to smaller aperture, limited depth of focus and requires larger number of scanning sweeps for total investigation of the sample slide. Thus it makes the process tedious.

SR algorithms help to counter this limitation by generating similar quality HR images from LR scans by restoring the high frequency content through a signal processing route.

Currently there are multiple ways to achieve super-resolution. The most direct so-

lution to increase spatial resolution is to reduce the pixel size (i.e., increase the number of pixels per unit area) by sensor manufacturing techniques. Another promising approach is to use signal processing techniques to obtain an HR image (or sequence) from observed multiple LR images. Recently, such a resolution enhancement approach has been one of the most active research areas, and it is called SR image reconstruction or simply resolution enhancement in the literature.

SR technology, which restores high-frequency information given a low-resolved image, has attracted much attention recent years. Various super-resolution algorithms were proposed so far: example-based approach, sparse-coding-based, GMM, Back Projection for Lost Pixels(BPLP), and so on. Most of these statistical approaches rely on the training (or just preparing) of the correspondence relationships between low-resolved/high-resolved images. Here in our approach, Deep Belief Nets(DBN) are used to restore high-frequency information from a low-resolved image. The idea is that learning a dictionary from a set of LR and HR image pairs using a deep architecure can store the information about the high frequency components, which are missing in the LR images. Experimental results show the high performance of our proposed method.

1.2 Prior Art

Due to price competition, the need to reduce the cost of the image sensor has become a serious problem, so that the technology to produce high-resolution images using digital image processing is attracting much attention. In general, low-resolution images are enlarged by using interpolation techniques, such as bilinear or bicubic interpolation. The interpolation methods, however, decrease the resolution of the images because their edge information is lost. For appropriate enlarging, it is necessary to restore the high-frequency components of the image. Such techniques are called "super-resolution", one of the most actively-studied topics on computer vision in recent years. Super-resolution techniques restore the original image from the observed image that has lost its high-frequency components for some reason.

Super-resolution techniques are generally divided into two approaches: example-based methods and linear regression methods. Example-based methods (Freeman et al., 2002), (Freeman et al., 2000) simply use pairs of low-resolution and high-resolution patches for the reconstruction. In this approach, a low-resolved input image is decomposed into patches, each of which is compared with the patches in the database and replaced with the corresponding high-resolved patch. Although this approach produces relatively less-deteriorated images, it is not based on any statistical models and lacks versatility.

Linear regression techniques solve the linear problem of $Y = L * X$, where Y, X and L are a low-resolution image, a high-resolution image, and a degrading filter, respectively, and * denotes convolution. Generally, it is impossible to obtain the exact high-resolved image because the filter is not known. To approximate this filter, various approaches have been proposed so far: a sparse-coding method (Elad et al., 2010), (Yang et al., 2012), total-variation regularization (Babacan et al., 2008), an Markov Random Field(MRF) -based approach (Rajan and Chaudhuri, n.d.), a Gaussian Mixture Model(GMM) -based approach (Ogawa et al., 2012), a Neural Networks(NN) - based approach (Huang and Long, 2007), and so on. Some of these statistical approaches rely on the training (or just preparing) of the correspondence relationships between low-resolved/high-resolved images. Therefore, if one wants to enlarge an image with the desired scale, the relationships between the low and high resolution with that scale need to be trained beforehand.

Meanwhile, Hinton et el. introduced an effective training algorithm of Deep Belief Nets (DBNs) in 2006 (Hinton and Teh, 2006), and the use of DBNs rapidly spread in the field of signal processing with great success. DBNs and related models have been, for example, used for hand-written character recognition (Hinton and Teh, 2006), 3-D object recognition (Deselaers et al., 2009), machine transliteration (Deselaers et al., 2009), and speech recognition tasks (r. Mohamed et al., 2012). DBNs are probabilistic generative models that are composed of multiple layers of stochastic latent variables, and have a greedy layer-wise unsupervised learning algorithm. DBNs are not only used

for classification tasks, but also for the completion of an image or for collaborative filtering. Eslami et el. adopted a type of DBNs (called SapeBM) to complete the missing region in an image (Eslami et al., 2012). Salakhutdinov et el. used 2-layer DBNs (i.e., Restricted Boltzmann Machines; RBMs) for collaborative filtering (Salakhutdinov et al., 2007), which has the benefit of the DBNs dealing with missing data.

1.3 Aims and Objectives

In view of growing usage of deep learning and the need for a robust super-resolution algorithm for the medical image analysis constraints due to (i) bulky infrastructure, (ii) costly equipment, or (iii) challenges with integrating into conventional clinical workflow, we set the aim of this thesis to develop an algorithm for image super-resolution for the LR bright field optical microscopy images using deep learning. The objective is to achieve super resolution using DBN, Implementation of Super Resolution using Interpolation, MRF, Example Based Learning and Deep Learning both in Spatial and Frequency Domain. The following *objectives* have been set to achieve this aim:

1. Obj.one : Different method of generating the training set from training images and different structure of training patches.
2. Obj.two :Using the k-d tree to search the best matching high-resolution patches.
3. Obj.three : Using a method that is using the dictionary encoding to attempt to improve the result.
4. Obj.four : Using a deep learning approach to improve the time complexity and performance of the dictionary based approach implemented in the previous step.

1.4 Roadmap to the Thesis

In Chapter one, Introduction, the research background and importance of this project, arrangement of the tasks and the research scoping are introduced, followed by the introduction of the structure of the whole thesis.

The main contributions of this thesis are organized into 2 coherent chapters providing proof-of-concept with supporting theoretical and experimental validation to acheieve the image super-resolution. The chapters are organized as follows:

Chapter 2: Super-resolution Bright Field Optical Microscopy Imaging Using a k-d Tree Based One-pass Exemplar Search Algorithm introduces the study on the example-based super-resolution method. In this chapter, the principle and process of this method is specifically introduced and explained, so as to highlight the difference between the method in this thesis and the original method.

Chapter 3: Super-resolution using for Bright Field Optical Microscopy Using Dictionary and Deep Learning Based Method represents the attempt of using the compressive sensing to improve the experimental result. This chapter probes into the possibility of improving the results by using the dictionary encoding technique and then unfolds the discussion on the deep learning method to achieve it.

Chapter 4: Conclusion and Future Work presents summary of contributions in this thesis and scope of future research.

2

Super-resolution Bright Field Optical Microscopy Imaging Using a k-d Tree Based One-pass Exemplar Search Algorithm

> We can't solve problems by using the same kind of thinking we used when we created them.
>
> Albert Einstein

LEARNING based algorithms are being explored for the past several years for enlarging images. In a training set, the algorithm learns the fine details that correspond to different image regions seen at a low resolution and then uses those learned relationships to predict fine details in other images. (Baker and Kanade, 2000) (Liu et al., 2001)

Repeatability of patterns in images are exploited in the learning based algorithms. We use small pieces of training images, modified for generalization by appropriate preprocessing, to create plausible image information in other images. Without restriction to a specific class of training images, it is unreasonable to expect gneration of correct high-resolution information. We aim for the more attainable goal of synthesizing visually plausible image details, such as sharp edges, and plausible looking texture.

2.1 Exposition to the Solution

In this chpter, first we model the spatial relationships between patches using a Markov network, which has many well-known uses in image processing. In the later part, a one-pass algorithm that gives results that are nearly as good as the iterative solution to the Markov network, is used to compute high-resolution patch compatibilities for neighboring high-resolution patches that are already selected, typically the patches above and to the left, in raster-scan order processing.

Due to the high dimension of the search space, finding the absolute best match would be computationally prohibitive. Instead, we use a kd-tree-based, approximate nearest neighbor search.

2.2 MRF Based Super Resolution

We model the spatial relationships between patches of an image using a Markov network. MRF models have been widely used to solve vision problems because of their ability to model context dependency, since interpretation of visual information necessitates an efficient description of contextual constraints. The utility of MRF models arises from the Hammersley-Clifford theorem which describes the equivalence of the local property that characterizes an MRF and the global property which characterizes a Gibbs Random Field (GRF).

2.3 Example Based Learning

We collect pairs of LR and HR image patches from a set of images as training. An input LR image is decomposed to overlapping patches on a grid, and the inference problem is to find the HR patches from the training database for each LR patch. We use the kd-tree algorithm, which has been used for real-time texture synthesis, to retrieve a set of high-resolution, k-nearest neighbors for each LR patch. Lastly, we run a max-product belief propagation (BP) algorithm to minimize an objective function that balances both

local compatibility and spatial smoothness (Freeman et al., 2002).

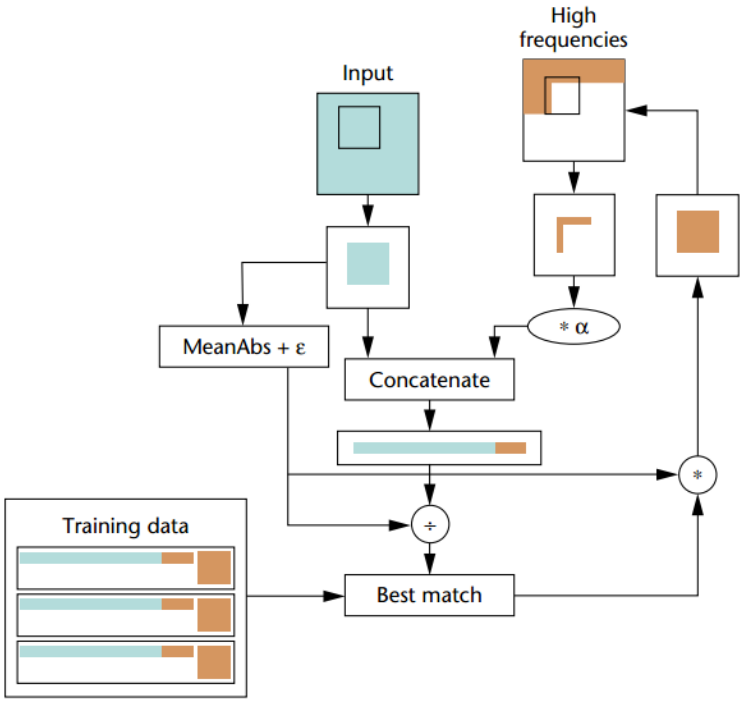


Figure 2.1: Graphical representation of Algorithm Used in Example Based Learning (Freeman et al., 2002)

2.4 Observation Model

In order to comprehensively analyze the super-resolution image reconstruction problem, we need to build a model that relates the original high-resolution image to the observed low-resolution to simulate the process of obtaining the low-resolution image from the high-resolution image. This model is called observation model. It's critical to build a precise system model for the image super-resolution reconstruction technology. The common observation model for images is shown in Figure 2.3.

The above observation model can be expressed with the following formula (Park et al., 2003):

$$y_k = DB_k x + n_k, 1 \leq k \leq p \tag{2.1}$$

Where p denotes the number of the low-resolution images, x , y_k and n_k denote the high-resolution image, the kth observed low-resolution image and the noise vector,

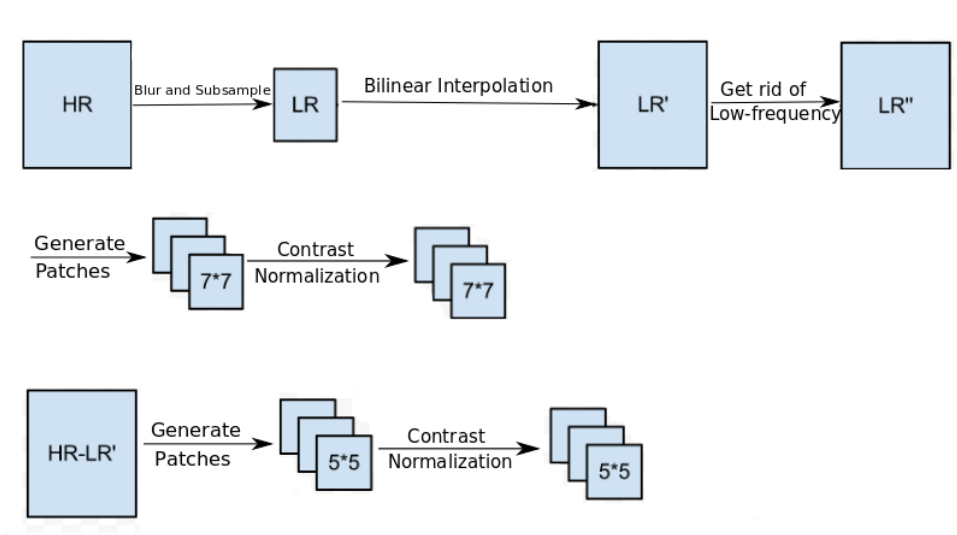


Figure 2.2: Training Set Generation (Freeman et al., 2002)

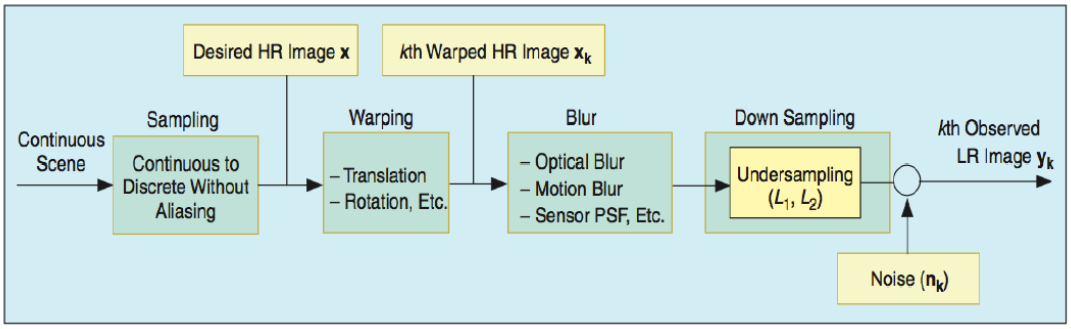


Figure 2.3: Observation model relating LR images to HR images

respectively. The matrixes of D and B_k denote the subsampling matrix and blur matrix. We let

$$H = DB_k$$

, the Equation 2.1 can be simplified as

$$y_k = Hx + n_k, 1 \leq k \leq p \tag{2.2}$$

SR technology in the early research only

2.5 Experiment

In a SR framework, an upscaled image is reconstructed from observed LR image (Y) which is assumed to be decimated and blurred version of original HR image (X) such that we have

$$Y = DBX \tag{2.3}$$

where D is Decimation matrix and B is Blur matrix.

We consider the task as being that of reconstructing a HR image X from a LR image Y such that a patch of size $m_H \times n_H$ in HR corresponds to a patch of size $m_L \times n_L$in LR and a pair of such representative patches are used to create a dictionary of exemplars. In our SR approach, given a LR bright-field optical microscopy image Y test , for a given pixel location, we represent it using a patch of size $m_L \times n_L$. Using a k-d tree based nearest neighbour search method, we find the HR patch of size $m_H \times n_H$ which is closest to the patch from qll the training patches. The closest matching HR patch is subsequently used to generate the SR version of the test image represented as X test. The flowchart is described graphically in Figure 2.4.

In our experiments we have used LR images acquired at 10× and HR images acquired at 40× to create a dictionary of exemplars which is optimized for fast k-d tree implementation with N = 625,000, and we have used H&E stained biopsy samples of the oral mucosa for our experiments.

So all the above steps can be summarized as :

2.5.1 Training Phase

Generation of training set $\{y^i, x^i\}_{i=1:N}^N$which contains N number of patch-pairs from high pass filtered LR and high pass filtered HR images, where $y_i \in R^{m_L X n_L}$ and $x_i \in R^{m_H X n_H}$.

2.5.2 Deployment Phase

Deployment phase consists of 4 steps.

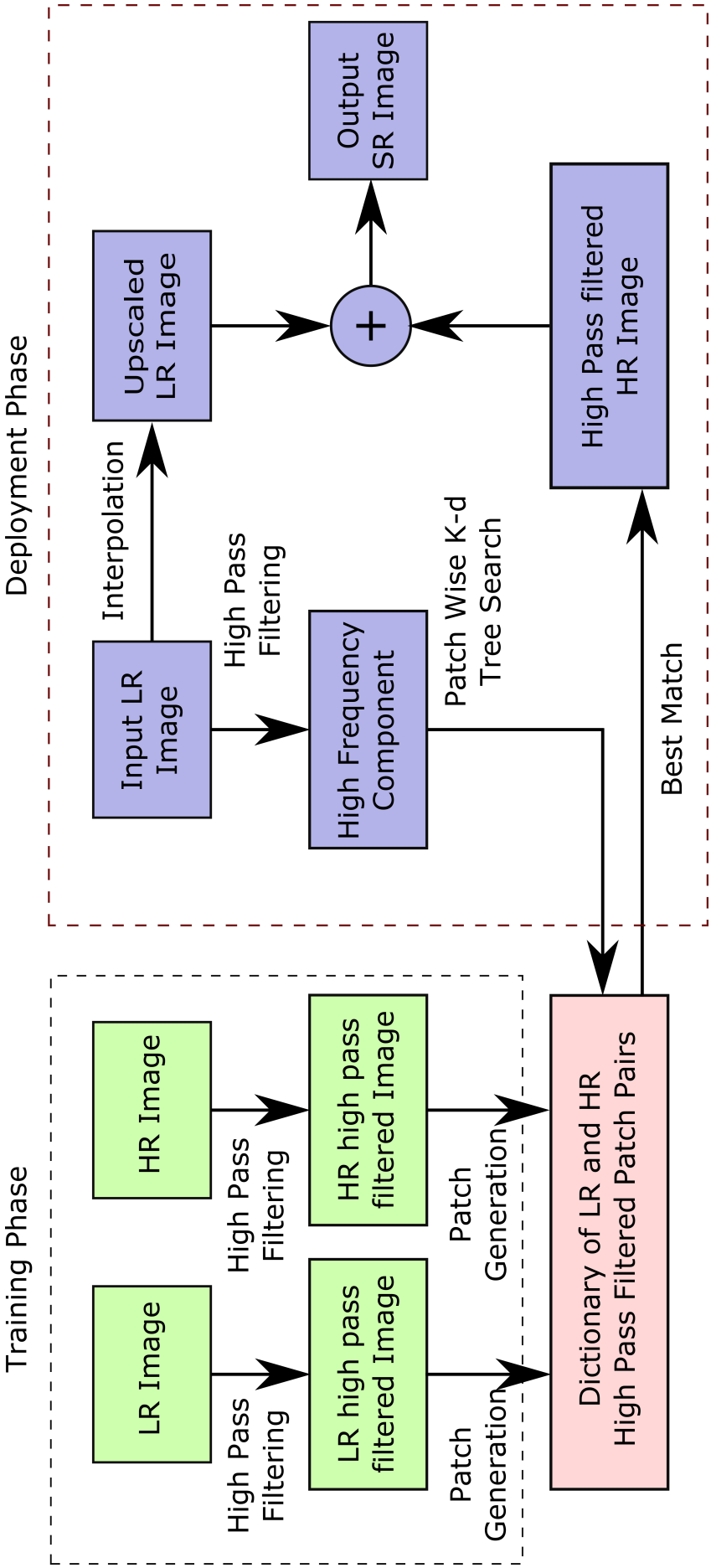


Figure 2.4: Process overview for generation of Training dataset, Training a kd-tree and reconstructing SR image using our approach.

Step 1

Given a test LR image, we generate an upscaled version corresponding to it using any interpolation technique.

Step 2

We also generate a high pass filtered version of the input LR image and patches are subsequently extracted from the high pass filtered image.

Step 3

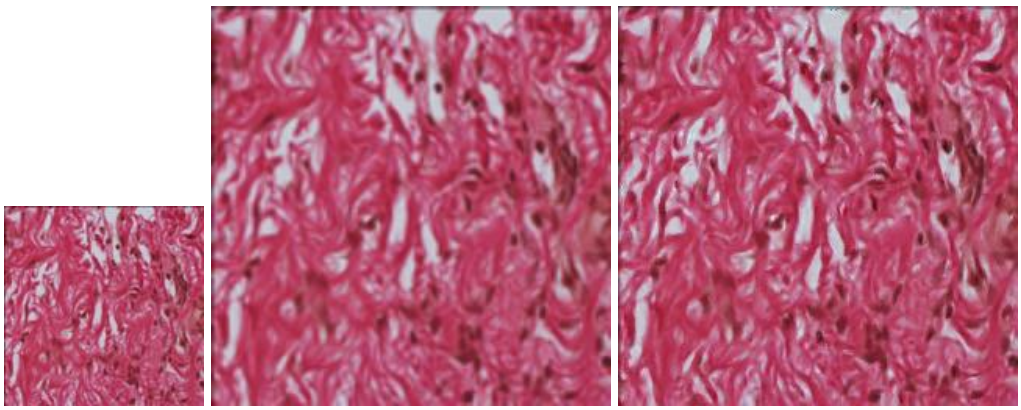
The nearest element is found in the dictionary of LR high pass filtered image space using k-d Tree based search algorithm. The corresponding HR high pass filtered patch is used for reconstruction of high pass filtered HR image.

Step 4

High pass filtered HR image generated in Step 3 is added to the upscaled image generated in Step 1.

2.6 Result

Experimental evaluation as presented in Figure 2.5. Structural similarity measure (SSIM) using simple bilinear interpolation is 0.7287 while using our approach it is 0.7912. The corresponding pSNR are 74.12 dB and 75.78 dB respectively thus substantiating the high-frequency restoration in SR.



(a) Input (b) Interpolation (c) Super-resolution

Figure 2.5: Comparison of SR reconstruction vs. bilinear interp. from 10× to 40×

2.7 Discussion

Table 2.1: Quantiative comparison of bicubic and one pass based SR.

	SSIM	PSNR(in dB)
Bilinear Interpolation	0.7287	74.12
One Pass Based SR	0.7912	75.78

In this experiment we have super resolved image data of luminance channel instead of RGB image, as searching time is less without significant loss in performance. Experimental result is shown in Figure 2.5 where we find edge preservation is better in the proposed method compared to bicubic interpolation. High scores of structural similarity (SSIM) index and peak signal-to-noise ratio (PSNR) substantiates the high-frequency restoration in our SR approach Table 2.1.

3

Super-resolution for Bright Field Optical Microscopy Using Dictionary and Deep Learning Based Method

> The only person who is educated is the one who has learned how to learn and change.
>
> Carl Rogers

EXAMPLE based learning approaches aim to learn the co-occurence prior between the HR and LR image local structures from an external database (Zhao et al., 2016). A learning based scheme which is based on sparse representation was proposed in (Yang et al., 2010). In this chapter, we implemented the SR via sparse representation by solving an l_1 optimization. Then, a new approach based on deep learning is developed for encoding the dictionary. (J. Gao and Yin, 2013)

The Restricted Boltzmann Machines (RBM) architecture has ability of learning a set of visual patterns, called dictionary elements from a set of training images. The learned dictionary will be then used to synthesize high resolution images. DBNs can be viewed as a composition of simple, unsupervised networks such as RBMs, where hidden layer of each sub-network serves as the visible layer for the next. This also leads to a fast, layer-by-layer unsupervised training procedure.

We test the proposed algorithm in DBN as well as Stacked Auto-encoder (SAE). The visual quality of the results has also been assessed by both human evaluation and

quantitative measurement.

3.1 Dictionary Based Learning

The SR problem based on a single-image is to recover a HR image X for given a LR image Y of the same scene. Recovering is considered as an inverse process of generating LR image from the HR image, i.e., $Y = F(X)$. For example, F may represent a blurring filter, followed by downsampling. Because of the information loss in process, in order to recover the original HR image X, some prior knowledge about X is preffered.

A patch x $\in R^m$ (or the feature of each patch) of the HR image X is assumed to be a sparse linear combination of elements in a given dictionary D_h where the dictionary is usually pre-learned from sample HR patches of training images.

$$x \approx D_h \alpha \tag{3.1}$$

for some $\alpha \in R^K$ with $||\alpha||_0 << K$. Although the dictionary size K is quite large, for a particular patch x, the linear combination in the above is very sparse.

The core idea used here is, the sparsity coefficient defined by Equation 3.1 can be conveyed to its LR image patches determined by the assumed model. In other words, the corresponding LR image patch $y \in R^n$ on the LR image Y will have a similar sparse representation over an appropriate LR dictionary D_l, i.e.,

$$y \approx D_l \alpha \tag{3.2}$$

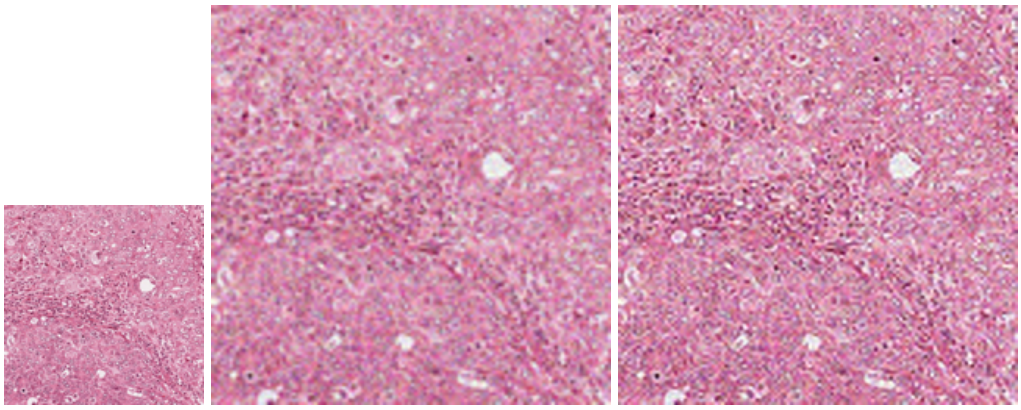
Thus an image super-resolution algorithm can be formed based on both Equation 3.1 and Equation 3.2. Suppose we have already known the HR dictionary D_h and its relevant LR dictionary D_l respectively. For any patch y of an input LR image Y, we can recover its sparse representation Equation 3.1 by solving for a sparse vector α which can be

done by the state-of-the-art l_l minimization algorithms such as

$$min_\alpha = \| y - D_l\alpha \|^2 + \lambda \| \alpha \|_1 \tag{3.3}$$

Then with this sparse representation coefficients α, the corresponding HR patch x can be recovered by Equation 3.2.

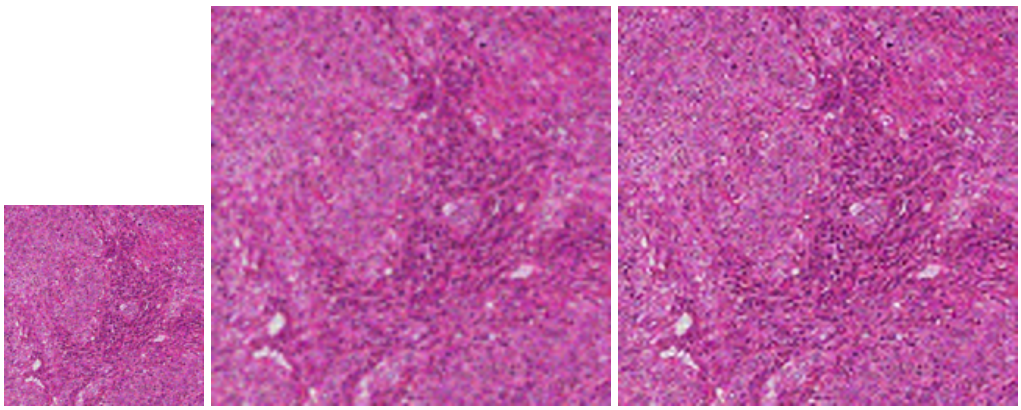
Dictionary was created consisting of 1024 elements (J. Yang and Ma, 2008). Here from the image we have randomly sampled 10x10 low resolution patches from the low resolution image for magnification factor of 2, and learned the HR/LR dictionaries each of size K = 1024 using the algoritm provided by (J. Gao and Yin, 2013). From that by using state of the art l_1 norm minimization problem α was found out. The l_1 minization problem is defined as $min_\alpha = \| y - D_l\alpha \|^2 + \lambda \| \alpha \|_1$. The optimization problem is computationally expensive and has to be performed for each and every patch. The results for this problem are shown in Figure 3.1 and Figure 3.2 for two images collected from scanner A and scanner H Section 3.4.



(a) Input (b) Bilinear interpolation (c) SR Reconstruction

Figure 3.1: Dictionary Based Learning - Scanner A image

Now this information is tried to be encoded in DBN. Before going into its details, some idea about DBN and RBM is provided below.



(a) Input (b) Bilinear interpolation (c) SR Reconstruction

Figure 3.2: Dictionary Based Learning - Scanner H image

3.2 Deep Belief Network

In machine learning, a deep belief network (DBN) is a generative graphical model, or alternatively a type of deep neural network, composed of multiple layers of latent variables ("hidden units"), with connections between the layers but not between units within each layer.

When trained on a set of examples in an unsupervised way, a DBN can learn to probabilistically reconstruct its inputs. The layers then act as feature detectors on inputs. After this learning step, a DBN can be further trained in a supervised way to perform classification.

DBNs can be viewed as a composition of simple, unsupervised networks such as RBMs or autoencoders, where each sub-network's hidden layer serves as the visible layer for the next. This also leads to a fast, layer-by-layer unsupervised training procedure.

3.2.1 Restricted Boltzmann Machine

The RBM is a probabilistic model that uses a layer of hidden binary variables to model the distribution of a visible layer of variables. Boltzmann Machines (BMs) are a particular form of log-linear Markov Random Field (MRF), i.e., for which the energy function is linear in its free parameters. RBMs further restrict BMs to those without

visible-visible and hidden-hidden connections i.e. no connections are present between the nodes of the same layer. A graphical depiction of an RBM is shown below.

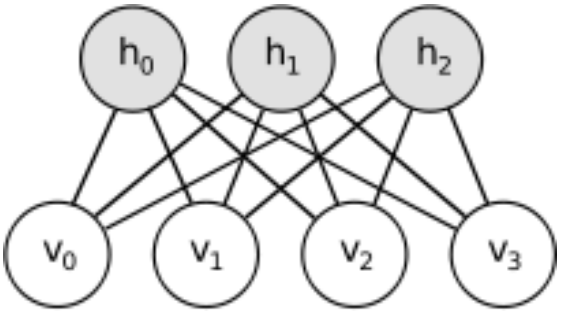


Figure 3.3: general RBM structure

$$\mathcal{F}(x) = -\log \sum_{h} e^{-E(x,h)} \tag{3.4}$$

$$P(x) = \frac{e^{-\mathcal{F}(x)}}{Z} \text{ with } Z = \sum_{x} e^{-\mathcal{F}(x)} \tag{3.5}$$

So, for RBM it will be

$$\mathcal{F}(v) = -b'v - \sum_{i} \log \sum_{h_i} e^{h_i(c_i + W_i v)} \tag{3.6}$$

First we note that the paired sparse representations (1) and (2) share the exactly same sparse coefficients. This fact inspires us to construct an RBM as follows: The hidden layer corresponds to the unknown sparse coefficients and both the HR patch x and LR patch y together define the observed layer. Thus the RBM consists of K hidden units, called the feature detectors, and weights D, stored “primary" elements which can be used as patch dictionaries. The RBM networks that we are interested in is shown in Figure 5 below:

A joint configuration (x, y, α) of the observed and hidden units has an energy given by, [13],

$$E(\mathbf{x}, \mathbf{y}, \alpha) = -\sum_{i} a_i x_i - \sum_{i} b_j y_j - \sum_{k} c_k \alpha_k - \sum_{i,k} x_i \alpha_k d^h_{ik} - \sum_{j,k} y_j \alpha_k d^l_{jk} \tag{3.7}$$

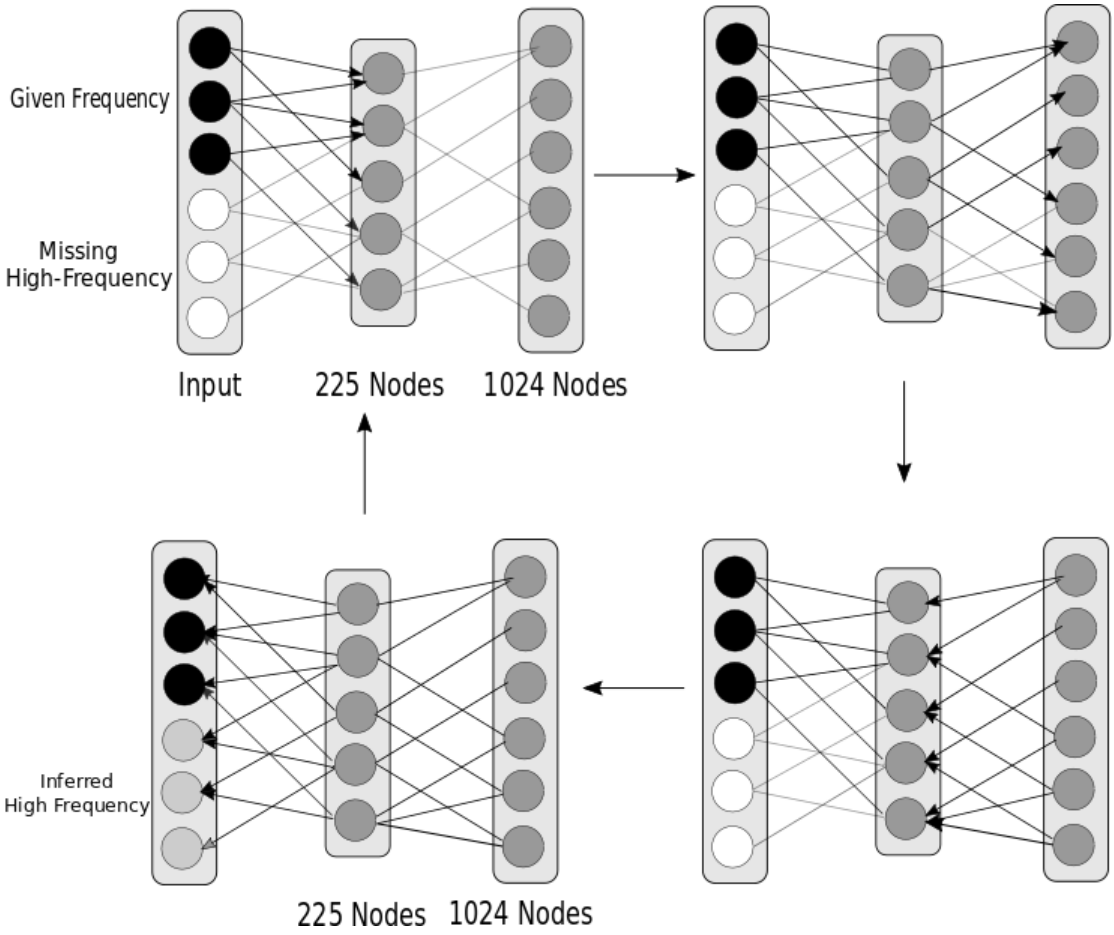


Figure 3.4: Reconstruction of high frquency components using a DBN

where x_i is the i^{th} component of the HR patch feature x, y_j the jth component of the LR patch feature y, α_k the kth coefficient of the linear representation, d_{ik}^h the connecting weights between x_i and α_k and d_{jk}^l the connecting weights between x_i and α_k are their biases between the connections. The network assigns a probability to every triple of observed LR and HR patch pair and their hidden coding via this energy function $p(x,y,\alpha) = \frac{1}{Z}e^{-E(x,y,\alpha)}$ where partition function Z is given by summing all over possible triple of observed and hidden vectors, i.e.,

$$Z = \sum_{x,y,\alpha} exp(-E(x,y,\alpha)) \tag{3.8}$$

and the likelihood of observed data (x,y) is given by

$$p(x,y) = \frac{1}{Z}\sum_{\alpha} exp(-E(x,y,\alpha)) \tag{3.9}$$

The bottle neck of our algorithm is the computation of the sparse code from the l_1-norm minimization for each LR input image patch. In the SR via sparse representation, for each LR patch input y, one has to solve an l_1 optimization problem defined as

$$min_\alpha = \| y - D_l \alpha \|^2 + \lambda \| \alpha \|_1$$

where $\| \alpha \|$ is the so-called l1 norm. So, we can exploit the RBM's elegant network structure (undirect connections between the observed and hidden units) to speed up this process. Given an input y, our algorithm is to use any simple super-resolution algorithm such as interpolation based methods to get an estimate x^0 of the desired super-resolution patch x. Then take as the input both $y^0 = y$ and x^0 to the observed layer in RBM and run the RBM training algorithm to transfer message from the observed layer to the hidden layer to get the variable. Following that, the message can be transferred back from the hidden layer to the observed layer and the SR results can be taken from those x units. If necessary, this process can be run several cycles to reach the equilibrium. Figure 3.4

3.3 Stacked Auto Encoder

A SAE is a neural network consisting of multiple layers of sparse autoencoders in which the outputs of each layer is wired to the inputs of the successive layer. Formally, consider a stacked autoencoder with n layers.Let $W^{(k,1)}, W^{(k,2)}, b^{(k,1)}, b^{(k,2)}$ denote the parameters $W^{(1)}, W^{(2)}$, $b^{(1)}$, $b^{(2)}$ for kth autoencoder. Then the encoding step for the SAE is given by running the encoding step of each layer in forward order:

$$a^{(l)} = f(z^{(l)}) \tag{3.10}$$

$$z^{(l+1)} = W^{(l,1)} a^{(l)} + b^{(l,1)} \tag{3.11}$$

The decoding step is given by running the decoding stack of each autoencoder in reverse order:

$$a^{(n+l)} = f(z^{(n+l)}) \tag{3.12}$$

$$z^{(n+l+1)} = W^{(n-l,1)} a^{(n+l)} + b^{(n-l,2)} \tag{3.13}$$

The information of interest is contained within $a^{(n)}$, which is the activation of the deepest layer of hidden units. This vector gives us a representation of the input in terms of higher-order features.The features from the SAE can be used for generating the high frequency patch information for the HR image by feeding $a^{(n)}$ with the high fruqency information of the LR Image.

Training:

A good way to obtain good parameters for a SAE is to use greedy layer-wise training. To do this, first train the first layer on raw input to obtain parameters $W^{(1,1)}, W^{(1,2)}, b^{(1,1)}, b^{(1,2)}$. Use the first layer to transform the raw input into a vector consisting of activation of the hidden units, A. Train the second layer on this vector to obtain parameters $W^{(2,1)}, W^{(2,2)}, b^{(2,1)}, b^{(2,2)}$. Repeat for subsequent layers, using the output of each layer as input for the subsequent layer.

The following are the steps to train a SAE with 2 hidden layers for learning the relation between the HR nad the LR patches.

- First, train is done on a sparse AE on the raw inputs $x^{(k)}$ to learn primary features $h^{(1)(k)}$ on the raw input
- Next, the raw input is fed into this trained sparse AE, obtaining the primary feature activations $h^{(1)(k)}$for each of the inputs $x^{(k)}$

  These primary features are then used as the "raw input" to another sparse AE to learn secondary features $h^{(2)(k)}$ on these primary features
- Following this, primary features are fed into the second sparse AE to obtain the secondary feature activations $h^{(2)(k)}$ for each of the primary features $h^{(1)(k)}$ (which correspond to the primary features of the corresponding inputs $x^{(k)}$

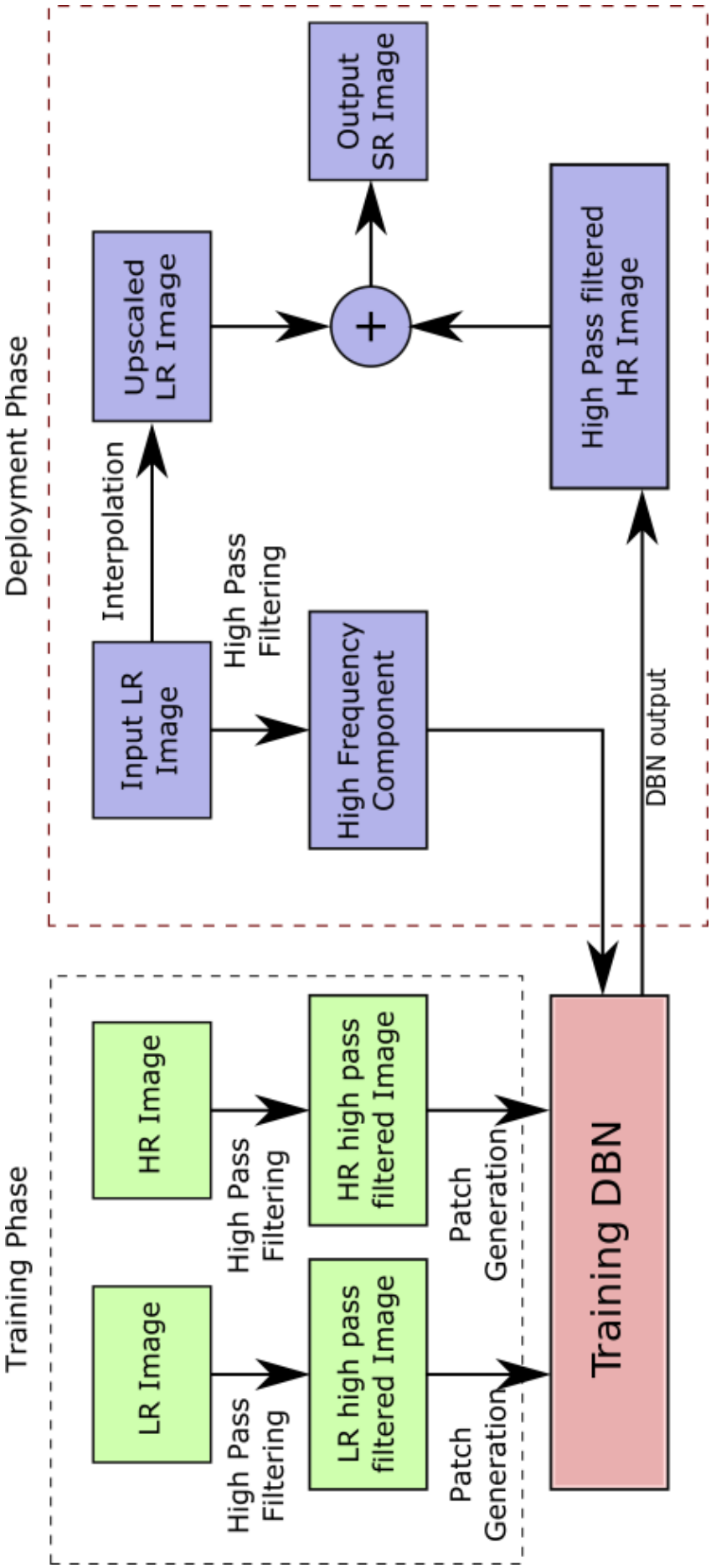


Figure 3.5: Process overview for generation of Training dataset, Training a DBN and reconstructing SR image using our approach.

3.4 Experiment

The dataset Mitos-Atypia-14 is used for the experiment. A set of breast cancer biopsy slides are used to build the dataset. The slides are stained with standard hematoxylin and eosin (H&E) dyes and they have been scanned by two slide scanners: Aperio Scanscope XT (Scanner A)and Hamamatsu Nanozoomer 2.0-HT (Scanner H). In each slide, the pathologists selected several frames at X20 magnification located inside tumours. These X20 frames are used for scoring nuclear atypia. The X20 frames have been subdivided into four frames at X40 magnification. The X40 frames are used to annotate mitosis and to give a score to six criteria related to nuclear atypia.

At the beginning, from the high resolution iamges given in the dataset, we create the low resolution counter part by the generative process as described in Section 2.4. In addition to that, image preprocessing was also done. First the images were converted to YCbCr from RGB. Color space was changed to work on illuminance only so as to reduce the number of parameters and easier training of the network.

After the generation of the LR images, we both possess the HR nad LR images. For the training set generation, first all the images were highpass filtered. Then, 10×10 patches from the low resoltuion images and the corresponding 5×5 patches from high resoltuion images are stored simultaneously. The low resoltuions images are fed as input to the generative DBN network. The high resolution patches served as target labels during training of the network.

The images were high pass filtered first so as to make the deep learning architecture learn only the relation between the high frequency component of HR and LR patches.

Greedy layer-wise unsupervised training is applied to train the DBN. The first layer is trained as an RBM that models the raw input patches as its visible layer. That first layer is used to obtain a representation of the input that will be used as data for the second layer. The second layer is trained as an RBM, taking the transformed data (samples or mean activations) as training examples (for the visible layer of that RBM).

DBNs are graphical models which learn to extract a deep hierarchical representation of the training data. Here the DBN network is constructed using two RBM layers with

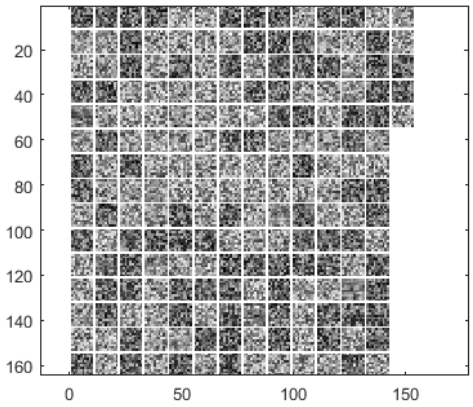
100-225 and 225-1024 nodes. During the training, the number of epochs is set to 500 for each of the layer and sparsity target is set to 0.5 so that it will ensure sparse encoding of the input data. The reconstruction of a image patch in the DBN is described in the Figure 3.4.

To validate the results and for comparision, the experiment was also performed using the SAE. The data preparation and training set generation emthod is the same as explained above. Instead of using an DBN to learn the relation between LR nad HR patches, here we use SAE for that purpose.

In the SAE, one hidden layer of 200 nodes is used with a spasity factor of 0.5. During the training, the number of epochs is set to 500. After training, the weights values when visualized in the SAE is shown in Figure 3.6.

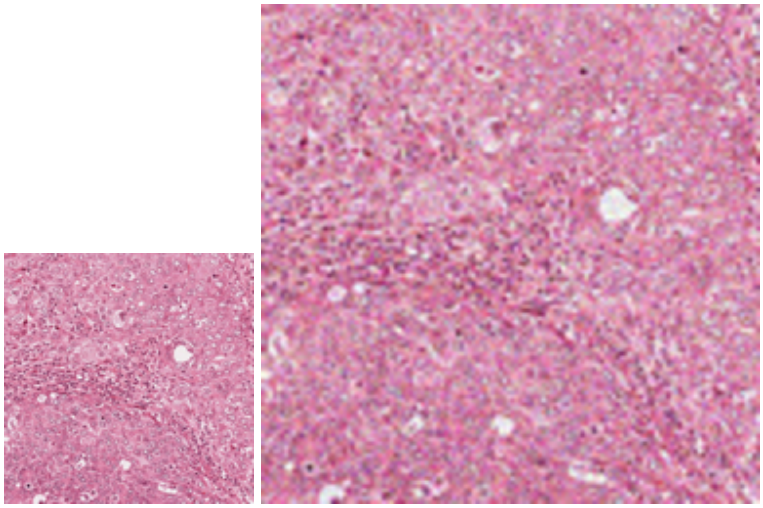
The results obtained for DBN encoding and SAE based SR when applied for Scanner - A image, are shown in Figure 3.7. For another image from scanner H, the results are shown in Figure 3.8.

3.5 Result

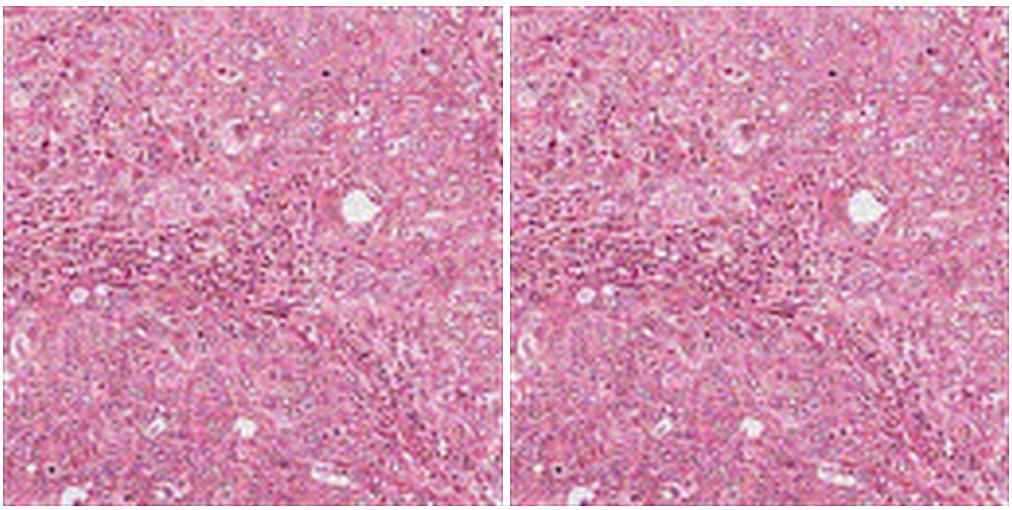


(a) Layer 1

Figure 3.6: Visualization of the weight matrices of first layer in the SAE

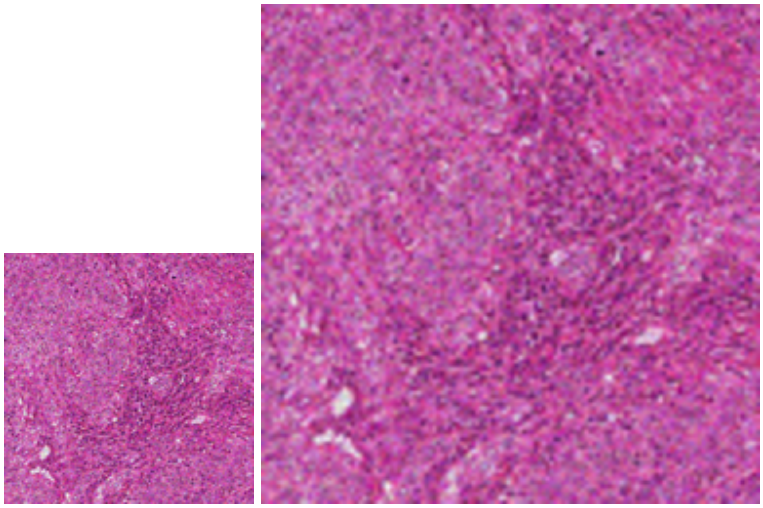


(a) Input (b) Bilinear interpolation

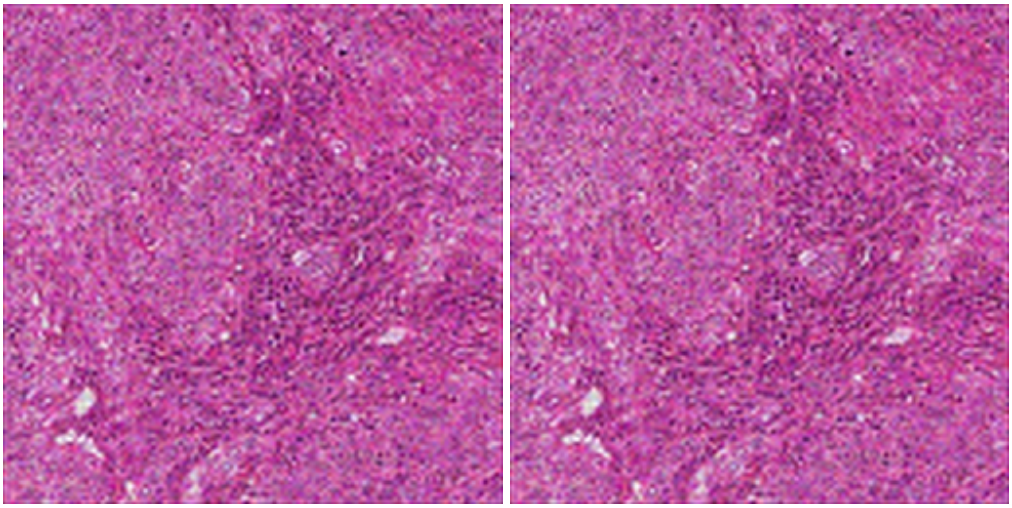


(c) SAE SR Reconstruct (d) DBN SR Reconstruct

Figure 3.7: Output for Deep Learning based SR for Scanner A image



(a) Input (b) Bicubic interpolation



(c) SAE SR Reconstruct (d) DBN SR Reconstruct

Figure 3.8: Output for Deep Learning based SR for Scanner H image

Table 3.1: Comparison of Deep learning based SR using various Image Quality Metrices for Scanner A image

	Bilinear Interpolation	Dictionary Based SR	SAE SR	DBN SR
PSNR (in dB)	23.0965	23.4543	23.3212	23.3436
RMSE	17.8533	17.1327	17.4521	17.3524
SSIM	0.8156	0.8403	0.8206	0.8368

Table 3.2: Comparison of Deep learning based SR using various Image Quality Metrices for Scanner H image

	Bilinear Interpolation	Dictionary Based SR	SAE SR	DBN SR
PSNR (in dB)	25.5483	25.6697	25.6586	25.6684
RMSE	13.4624	13.2755	13.2948	13.2831
SSIM	0.6225	0.6798	0.6503	0.6605

3.6 Discussion

In this experiment also, we have used image data of luminance channel instead of RGB image, as searching time is less without significant loss in performance and the training is also easier as there are less number of parameters and nodes in the neural network. Experimental results for sparse coding, SAE based SR and DBN based SR are shown in Figure 3.7 and Figure 3.8. where in both the cases we find edge preservation is better in the proposed method compared to bicubic interpolation. However, the testing and training time in DBN network is significantly lower. High scores of structural similarity (SSIM) index and peak signal-to-noise ratio (PSNR) substantiates the high-frequency restoration in our SR approach. Similarly, the deep learning approach was also tested using SAE, where we found out that the DBN results outperform the SAE. The comparitive results are shown in Table 3.1 and Table 3.2.

4 Conclusion and Future Work

It's not what you look at that matters, it's what you see.

Henry David Thoreau

4.1 Summary of Contributions

THE thesis presents the work on developing a super resolution framework for generating HR images from

The contribution of this project involves the three following areas:

1. Different method of generating the training set from training images and different structure of training patches.

2. Using the k-d tree to search the best matching high-resolution patches.

3. Using a method that is using the compressive sensing theory to attempt to improve the result.

4. Developing deep learning based approach to achieve the super-resoluton using DBN and SAE

4.2 Future Scope

MRF based SR, Example based SR, Sparse encoding based SR, Deep learning based SR has been successfully applied. DBN network is giving high resolution output as expected but more improvement has to be tried out. The tuning of the DBN and SAE

has to be carried out to determine the optimal structure of the architecture. In addition to that, Parallel training of DBN has to be carried out in order to reduce the computation time.

Bibliography

Babacan, S. D., Molina, R. and Katsaggelos, A. K. (2008). Total variation super resolution using a variational approach, *2008 15th IEEE International Conference on Image Processing*, pp. 641–644.

Baker, S. and Kanade, T. (2000). Limits on super-resolution and how to break them, *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, Vol. 2, pp. 372–379 vol.2.

Deselaers, T., Hasan, S., Bender, O. and Ney, H. (2009). A deep learning approach to machine transliteration, *Proceedings of the Fourth Workshop on Statistical Machine Translation*, StatMT '09, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 233–241.

Elad, M., Figueiredo, M. A. T. and Ma, Y. (2010). On the role of sparse and redundant representations in image processing, *Proceedings of the IEEE* **98**(6): 972–982.

Eslami, S. M. A., Heess, N. and Winn, J. (2012). The shape boltzmann machine: A strong model of object shape, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 406–413.

Freeman, W. T., Jones, T. R. and Pasztor, E. C. (2002). Example-based super-resolution, *IEEE Computer Graphics and Applications* **22**(2): 56–65.

Freeman, W. T., Pasztor, E. C. and Carmichael, O. T. (2000). Learning low-level vision, *International Journal of Computer Vision* **40**(1): 25–47.

Hinton, G. E., O. S. and Teh, Y. (2006). A fast learning algorithm for deep belief nets, *Neural Computation* **18**: 1527–1554.

Huang, Y. and Long, Y. (2007). Super-resolution using neural networks based on the optimal recovery theory, *Journal of Computational Electronics* **5**(4): 275–281.

J. Gao, Y. G. and Yin, M. (2013). Restricted boltzmann machine approach to couple dictionary training for image superresolution, *IEEE International Conference on Image Processing* **pp.**: 499–503.

J. Yang, J. Wright, T. H. and Ma, Y. (2008). Image super-resolution as sparse representation of raw image patches, *IEEE Conference on Computer Vision and Pattern Recognition* **pp.**

Liu, C., Shum, H.-Y. and Zhang, C.-S. (2001). A two-step approach to hallucinating faces: global parametric model and local nonparametric model, *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1, pp. I–192–I–198 vol.1.

Ogawa, Y., Ariki, Y. and Takiguchi, T. (2012). Super-resolution by gmm based conversion using self-reduction image, *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1285–1288.

Park, S. C., Park, M. K. and Kang, M. G. (2003). Super-resolution image reconstruction: a technical overview, *IEEE Signal Processing Magazine* **20**(3): 21–36.

r. Mohamed, A., Dahl, G. E. and Hinton, G. (2012). Acoustic modeling using deep belief networks, *IEEE Transactions on Audio, Speech, and Language Processing* **20**(1): 14–22.

Rajan, D. and Chaudhuri, S. (n.d.). An mrf-based approach to generation of super-resolution images from blurred observations, *Journal of Mathematical Imaging and Vision* **16**(1): 5–15.

Salakhutdinov, R., Mnih, A. and Hinton, G. (2007). Restricted boltzmann machines for collaborative filtering, *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, ACM, New York, NY, USA, pp. 791–798.

Yang, J., Wang, Z., Lin, Z., Shu, X. and Huang, T. (2012). Bilevel sparse coding for coupled feature spaces, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2360–2367.

Yang, J., Wright, J., Huang, T. S. and Ma, Y. (2010). Image super-resolution via sparse representation, *IEEE Transactions on Image Processing* **19**(11): 2861–2873.

Zhao, X., Wu, Y., Tian, J. and Zhang, H. (2016). Single image super-resolution via blind blurring estimation and anchored space mapping, *Computational Visual Media* **2**(1): 71–85.

Awards and Publications

Conference Abstracts

1. Das, Pradeepta ; Gupta, Avi ; Garai, Niladri ; Karri, Sri Phani Krishna ; Chatterjee, Jyotirmoy ; Sheet, Debdoot (2015) Fourier Domain Markov Random Field for Super Resolution Optical Coherence Tomography, *Proc. Int. Symposium of Biomedical Imaging*

2. Das, Pradeepta ; Garai, Niladri ; Ghosh, Biswajoy ; Chatterjee, Jyotirmoy ; Sheet, Debdoot (2016) Super-Resolution Bright Field Optical Microscopy Imaging Using a K-D Tree Based One-Pass Exemplar Search Algorithm, *Proc. Int. Symposium of Biomedical Imaging*