Pansharpening of Multispectral Images using Dual Attention based Two Stream Network

By

Syeda Roshana Ali Naqvi

(snaqvi.mscs17seecs@seecs.edu.pk)

00000205176

A thesis submitted in conformity with the requirements for the degree of *Master of Science* in Computer Science (MS-CS)

Supervisor

Dr. Muhammad Shahzad

(muhammad.shehzad@seecs.edu.pk)

Assistant Professor

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

August 2020

Approval

It is certified that the contents and form of the thesis entitled "Pansharpening of Multispectral images using dual attention based two stream network" submitted by SYEDA ROSHANA ALI NAQVI have been found satisfactory for the requirement of the

Advisor : Dr. Muhammad Shahzad

Signature: ____________________

Date: 13-Aug-2020

Committee Member 1:Dr. Anis ur Rahman

Signature: ____________________

Date: 14-Aug-2020

Committee Member 2:Prof. Dr. Faisal Shafait

Signature: ____________________

Date: 14-Aug-2020

Committee Member 3:Dr. Muhammad Muneeb Ullah

Signature: ____________________

Date: 13-Aug-2020

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Pansharpening of Multispectral images using dual attention based two stream network" written by SYEDA ROSHANA ALI NAQVI, (Registration No 00000205176), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: M. SHAHZAD

Name of Advisor: Dr. Muhammad Shahzad

Date: 13-Aug-2020

Signature (HOD): ____________________________

Date: ______________________________________

Signature (Dean/Principal): ___________________

Date: ______________________________________

This thesis is dedicated to *my exceptional parents and adored siblings whose remarkable love, cooperation and encouragement led me to this wonderful achievement*

Certificate of Originality

I hereby declare that this submission titled "Pansharpening of Multispectral images using dual attention based two stream network" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:SYEDA ROSHANA ALI NAQVI

Student Signature: _______________

Acknowledgement

First and foremost, I would like to express my deepest praises and gratitude to my creator Allah Almighty (SWT) for his abundant blessings. After that, I am profusely grateful to my research supervisor Dr. Muhammad Shahzad for his invaluable guidance, enthusiasm, immense knowledge and motivation. His supervision and encouragement helped me a lot throughout my research.

I would like to thank my thesis evaluation committee members including Dr. Anis ur Rahman, Dr. Muhammad Muneeb Ullah and Dr. Faisal Shafait for their precious time, insightful comments, and indispensable constructive criticism. Apart from my committee members, I would like to express my sincere gratitude to Dr. Muhammad Murtaza for his tremendous support and guidance on each and every step. This work would have not been possible without his help and supervision. I would also like to pay special thanks to Dr. Sohail Iqbal for giving me his precious time and valuable suggestions that kept me motivated and on track during my research.

I am also grateful to my lab fellows Huma Israr, Qurrat ul Ain and Javeria Hassan, and seniors Sidra Batool Kazmi, Sara Imtiaz and Zakia, who inspired me and gave me their valuable suggestions and continuous support, whenever I got stuck in something and having tough times. I am thankful to my friends Tajwar, Mahreen and Maham for their time and support.

Last but not least, I am extremely obliged to my parents and sisters for their immense love, endless prayers, encouragement, and sacrifices they have made for educating me and making me what I am today. Without their devotion and continuous support, I would not be able to accomplish this milestone.

Table of Contents

Abbreviations

LR	Low Resolution
HR	High resolution
PAN	Panchromatic
MS	Multispectral
HS	Hyperspectral
SR	Super Resolution
CS	Component Substitution
MRA	Multi Resolution Analysis
NN	Convolutional Neural Network
CNN	Convolutional Neural Network
AE	Autoencoder
CAE	Convolutional Autoencoder
CLA	Channel Level Attention
PLA	Pixel Level Attention
TFNet	Two Stream Fusion Network
SAM	Spectral Angle Mapper
ERGAS	Erreur Relative Global Adimensionnelle de Synthese
SCC	Spatial Correlation Coefficient
UIQI	Universal Image Quality Index
SSIM	Structural SIMilarity index

Nomenclature

I	Input image for proposed network
I_M	Input low resolution Musltispectral image
I_P	Input high resolution Panschromatic image
X^i	Target/ Reference image
$\breve{E}(I)$	Feature Extraction network for image I
$\breve{E}(I_M)$	Feature Extraction network for low resolution Multispectral image
$\breve{E}(I_P)$	Feature Extraction network for high resolution Panchromaticl image
$\breve{G}$	Mean pooling
Υ	ReLU activation function
σ	Sigmoid activation function
$\otimes$	Element wise multiplication
$\breve{E}_{CLA}(I_M)$	Channel level attention for feature maps of Multispectral image
$\breve{E}_{PLA}(I_P)$	Pixel Level attention for feature maps of Panchromatic image
$\oplus$	Concacationation
$\breve{E}(I_p, I_M)$	Fused Feature map of Multispectral and Panchromatic image
Ω	Residual function
r_n	Input of residual CNN unit
$\hat{I}$	Identity mapping
ϑ	Network parameters
ℓ_1	Loss function

List of Tables

List of Figures

Abstract

With the advent of technology there are large number of earths observing satellites having sensors, delivering different type of images that can provide a varying useful information about different surfaces of earth and those areas that are not reachable by human. Most popular of these images are panchromatic (PAN) image having the geometric details and multispectral (MS) image that contains reflectance data that represents wide variety of spectral band and wavelengths. Availability of these image encouraged the development and utilization remote sensing application. Regardless of being rich in spectral details, MS images lack in spatial information due to the sensors limitations. Pansharpening approach is devised to enhance the spatial details of MS image using the spatial features of PAN image. This method simply takes the spatial information of PAN and fuse it with MS image and produce a pansharped MS image high in both spectral and spatial information. In this thesis, we explored literature review and following the new trends we proposed a new dual attention based two stream pansharpening approach that extracts the details from both MS and PAN images and fuse them together to produce a pansharped MS image. For analysis we performed different experiments using Pleaides dataset and compared our model with a recently proposed approach with the help of visual results and different quality metrics.

Keywords:
Pansharpening, Multispectral image, Panchromatic image, Convolutional Neural Network, Channel level attention, Pixel level attention, Residual learning, Deep learning, Image fusion, Remote sensing.

Chapter 1

Introduction

Generally a human being perceives his surroundings with the help of five senses. Sometime this requires the direct physical contact with the object and sometime this can be achieved from a distance as well through the sense of sight and hearing. Still there are many places in the world from where a normal human cannot get the required information with ease. Like reaching the large earth surface and uninhabited areas is not that easy for a human. Vision capability of a human is still unbeatable by the most recent cameras but there is a restriction of distance up to where a normal human eye can perceive the details accurately and extent of recorded information also varies with the memorization capability of a human. Covering large earth surface simultaneously is not in human capacity that can provide spatial details needed to learn and discover the relationships between earth surfaces. Providing up to date information all the time is also not feasible for a normal human being. To deal with all such limitations and acquire the information on larger extent numerous tools and techniques are being utilized. One of the major techniques employed for this purpose is remote sensing. Different types of sensors are used in remote sensing process that captures different type of images. Depending upon the type, these images are either high in spatial information or in spectral information and provide relative diverse and rich information about the target area/object. Effectiveness of remote sensing applications is directly proportional to the quality of images it consumes. Due to sensors limitation, a single image cannot be rich in both spatial and spectral information. Numerous mechanisms are devised to enhance the quality of these images so that the resultant image is high in both spatial and spectral information.

1.1 Remote Sensing

Remote sensing refers to the method or activity of obtaining/recording the details regarding the object, earth surface or an event without having any direct contact with it. Information is acquired by perceiving, identifying and recording the energies emitted and reflected by the objects or areas under observation at remote areas. The process of remote sensing focuses on the incident and reflected energies, process and analyze it and convert it into useful information. Following Figure 1.1 illustrates the main building blocks, their interaction and process of a remote sensing system:

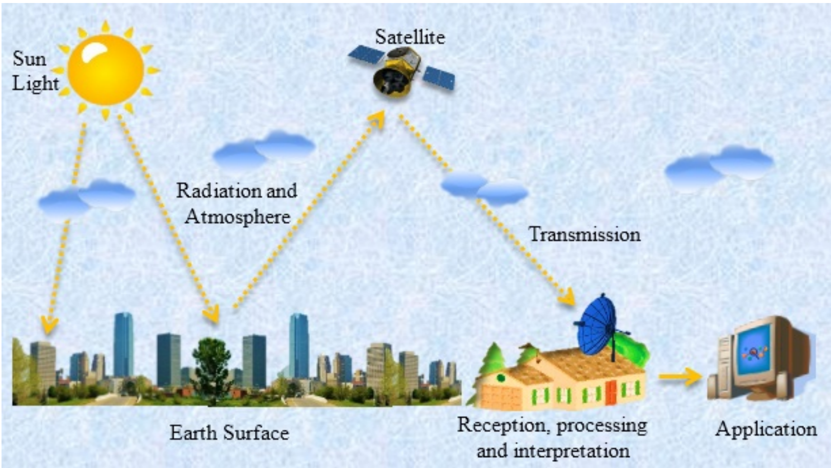

Figure 1.1: Process of Remote Sensing including main componnents of it.

The major steps/components of the remote sensing process and their roles are summarized as follows:

- First and foremost component of remote sensing is the source of illumination that provides the energy to the object of interest.
- When the electromagnetic energy in terms of radiations travels from illumination source to the object and from object to the satellite having sensors, it gets affected by the atmospheric conditions.
- When the energy comes in contact with the object of interest it get emitted or scattered by the object depending upon the properties of object and type of radiation/energy.
- Radiation/energy after hitting the object gets scattered, they are then sensed/recorded by the sensor mounted at some reasonable distance.

- Recorded energy is then transmitted to the receiving station for further processing and analysis in image format.
- Processed image is then further explored and interpreted to get useful information about the object of interest.
- This information is finally used by an application to solve some particular problem or to produce more useful information.

Radiations reflected to the space are affected by the earth surface and atmospheric conditions thats why the sensors capturing these on satellites have information about earth surface as well as atmosphere.

1.1.1 Types of Remote Sensing

Remote sensing can be generally divided into two types:

- Passive Remote Sensing
- Active Remote Sensing

Brief description of these two is as follows:

1.1.1.1 Passive Remote Sensing

Passive remote sensing is process in which it sense or collect the energy/light naturally reflected or radiated from the object/area of interest. In this type of remote sensing the source of energy is mostly sun.

1.1.1.2 Active Remote Sensing

Active remote sensing is a mechanism that has its own source of energy for illuminating the object/area of interest. Best example of such approaches is Radar.

1.2 Satellites and sensors

These are the platforms that carries different types of sensors. Earth observing satellites are providing invaluable data and information related to earth that can be effectively utilize in numerous remote sensing applications. Sensors on these satellites enable them to capture multiple images at different spectral and spatial resolution. However due to hardware constraints and some other factors, a single sensor is not capable of capturing image high at both spectral and spatial resolution for instance an image captured at higher spatial resolution cannot have high spectral resolution at the same time and vice versa. In other words both spectral and spatial resolutions are complementary to each other. Sensors always have to make an intelligent tradeoff between spectral and spatial resolutions [1]. Brief description of the images captured by these sensors and related to our research work are as follows:

1.2.1 Panchromatic Image

Panchromatic image is captured by the sensors, sensitive to light with large wavelength and have broad bandwidth normally 200nm in VIS-NIR. These images are single band images and have high spatial resolution. PAN images provide single value against each picture element (pixel) and are capable of learning the geometry of landscape like buildings, road etc. with great precision. Meanwhile a single PAN image is not sufficient to identify all the objects in the image. Figure 1.2 shows a sample PAN Image.

Figure 1.2: A sample single band Panchromatic(PAN) Image.

1.2.2 Multispectral Image

Multispectral (MS) image is captured by measuring the reflected light within certain regions of electromagnetic spectrum in discrete format. They usually consist of 3-10 bands, which means each single pixel have 3-15 values against each band. Values recorded by MS sensors are in form of discontinuous bands. The bands in the image of these sensors consist of visible red, visible green, visible blue, near infrared etc. High spectral images can provide more rich information against each band that cannot be captured by visible receptors of human eye. Figure 1.3 shows a sample MS Image.

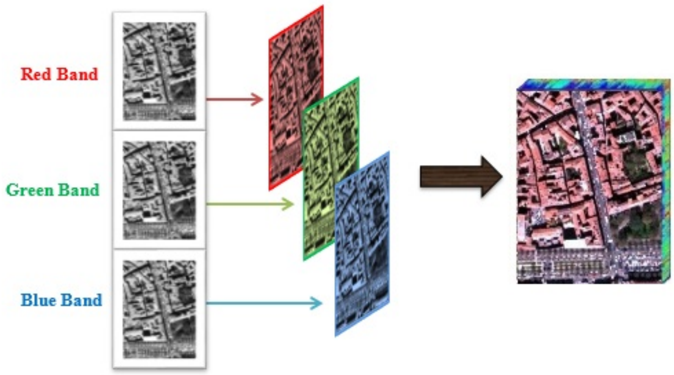

Figure 1.3: A sample three band Multispectral (MS) Image(Red, Green Blue) formed by joining three separate images each image corresponds to a band.

1.2.3 Hyperspectral Image

Hyperspectral (HS) image is captured by the sensor sensitive to light with narrow wave length of 10-20nm. They consist of 100-200 contiguous bands which means for each single point of earth surface a HS image provide many reflectance values corresponding to each wavelength. In simple words we can say that an HS image is collection of images of the same scene taken at different wavelengths and these are stacked together. This make a HS image, a three dimensional cube. Like a general cube two dimensions are width (x), height (y) and a third dimension is depth (λ) that corresponds to the wavelength at which each image is taken. Each pixel in an HS image has its own spectrum which can be used to identify different materials present at that pixel point. Figure 1.4 displays an HS image cube.

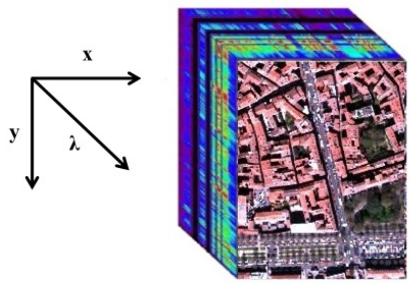

Figure 1.4: A sample Hyperspectral (HS) Image Cube formed by joining λ number of images, each corresponds to a single band.

1.3 Enhancement of MS image

Due to limited capability of the sensors MS images needs to be enhanced prior to be used by any application. As MS image is naturally high in spectral resolution so, improvement is required in terms of spatial resolution. Numerous spatial enhancements techniques are developed until now for multirspectral images but here we categorized them into two main groups on the basis of number of images used for enhancement: super resolution and pansharpening.

1.3.1 Super resolution

It is a single image based enhancement technique. It uses low resolution MS image to enhance the spatial resolution of MS image. Variety of super resolution (SR) methods has been proposed so far [2], [3], [4] and [5]. The major drawbacks of these methods are the absence of target image and training requires HR MS images. Figure 1.3 shows simple architecture of super resolution.

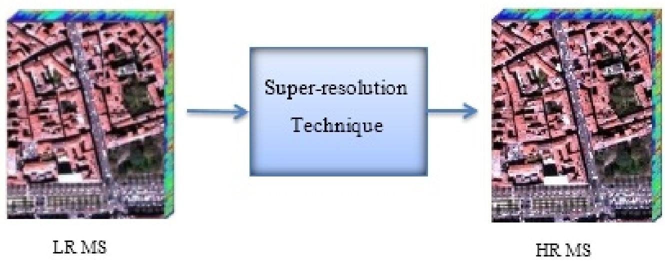

Figure 1.5: A simple architecture of Super-resolution taking LR MS as input and after performing some operations produecs HR MS image.

1.3.2 Pansharpening

It is a two image based enhancement technique. Pansharpening uses Panchromatic or Multispectral image for spatial enhancement of MS image. It fuses high spatial resolution image like PAN with high spectral resolution MS/HS image and the fused image is high in both spectral and spatial resolution by having the spatial resolution of first and spectral resolution of second image. Figure 1.4 depicts the basic architecture of pansharpening.

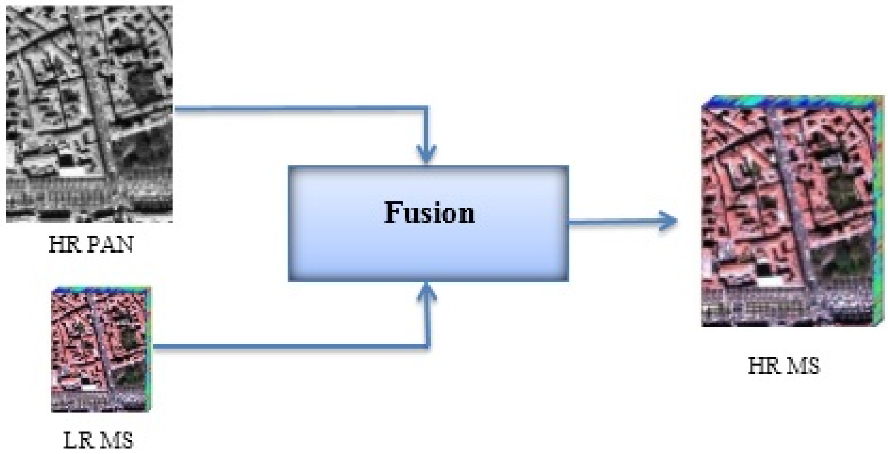

Figure 1.6: A basic architecture of Pansharpening, taking LR MS and HR PAN as input and after fusion produces a HR MS image.

1.4 Motivation

Recently the idea remote sensing is being employed in many domains like mineral exploration and identification, vegetation health monitoring, crop identification, weather forecasting, forest monitoring, surveillance, land cover mapping, natural disasters prediction and target detection etc. They are mostly based on MS images. Apart from these MS images are also being used in many other fields like medicine, document analysis, seed examination, food quality assessment and industries. All these applications are taking advantage of the spectral information present in these images. Along with that numerous enhancement methods have been devised for these to enhance their spatial resolution these methods mostly focus on improving the spatial details of MS image but due to spectral mismatch between the two images (MS and Pan) there is chance of spectral distortion after the fusion. A method is needed that has a parallel check on channel details while enhancing the spatial information and provide an efficient tradeoff between spatial and spectral resolution in the fused image.

1.5 Objectives of our research

Remote sensing is being utilized in many fields of life and these applications require images high in both spatial and spectral resolution as an input to take correct and timely decisions. With the recent advancement in technology makes MS images readily available and encourages their utilization by

different remote sensing applications.Due to sensors limitations MS images lacks in spatial details, so there is a need of an enhancement method, more specifically a pansharpening technique that improves the spatial details while preserving the spectral details.
The key objective of this research is to propose and develop a pansharpening technique that overcomes the sensors limitation in terms of spatial resolution and work well for most of the remote sensing applications. The proposed model utilizes attention mechanism to focuses on significant features and ignore the irrelevant details. This step not only enhances the distinct features but also minimizes the computational cost. The proposed technique also takes into account the details of MS image while enhancement, this will minimize the spectral distortion caused while enhancing the spatial details. The use dedicated loss function guarantees the output produced must be close to the actual result by providing an efficient tradeoff between spatial and spectral resolution. Another important concept of residual learning is also incorporated in this model that keeps the balance of low- and high-level feature causing the proposed model to work well and efficiently.

1.6 Summary of Contributions

In order to accomplish the objectives and incorporate the features mentioned in section 1.5, a new pansharpening model is proposed that enhances the spatial details while maintaining the underlying spectral details of MS image. Following are the overall contributions of this research:

- We propose a novel dual attention based two stream Convolutional Neural Network for pansharpening of Multispectral (MS) image.
- Feature extraction and feature fusion is carried out in feature domain.
- Like Encoder decoder approach, features are extracted first from both input images (PAN and MS), using two different networks.
- To flexibly deal with and enhance the distinct features channel level and pixel level attention mechanism is employed.
- The concept of residual learning is also incorporated, this enables the network to bypass the low frequency details with the help of skip connections and maintain the balance of distinct information.
- To further optimize the proposed approach, we used ℓ_1 as loss function as an alternative of generally used ℓ_2 loss function. This substitution considerably improves the model results.

1.7 Organization of Thesis

The objective of this research is to propose and implement a pansharpening approach that enhances the spatial resolution of MS image while maintaining the underlying spectral details of same image. This thesis is organized into five chapters. Chapter 1 includes brief introduction of remote sensing, satellites/sensors, types of images, image enhancement techniques, lastly the motivation and objectives of this research. In Chapter 2 of this document, literature review of the state-of-the-art pansharpening methods is critically analyzed under the relevant headings. Along with this, detail description of the quality measures used to evaluate and assess the performance of these methods is added. Chapter 3 explains the proposed approach in detail. Chapter 4 elaborates the experimental work carried out along with the detail of dataset used, visual results, quantitative evaluation carried out and discussion of the obtained results. Chapter 5 includes the findings and conclusion of the proposed approach. In the end of this chapter extension and future work of the proposed technique are discussed.

Chapter 2

Literature Review

2.1 State of the art Methods

Numerous pansharpening techniques have been developed so far most are based on pansharpening of Multispectral (MS) images to make use of high spectral information present in these images. The general classification of the pansharpening methods can be found in [6], [7] [8] and [9]. We can divide these methods into six broad categories: Component Substitution (CS), Multi Resolution Analysis (MRA), Hybrid, Matrix factorization, model based, and Deep Learning based methods.

2.1.1 Component Substitution

In component substitution (CS) methods, high spectral resolution image is decomposed into spectral and spatial components by projecting it into another space. The enhancement is performed by substituting the spatial component of HS/MS image with Panchromatic (PAN) image (or its spatial component) and then taking its inverse projection provides enhanced image [9]. The methods that fall under this group are: Intensity-Hue-Saturation (IHS) [10] [11] [12], Principal Component Analysis (PCA) [13] [14] [15], Hyperspherical Color Sharping (HCS) [16], Gram-Schmidt (GS) spectral sharpening [17] [18] and GS Adaptive (GSA) [19].
These methods can produce an image with high spatial resolution, work well in case of spatial mismatch between both images [20] and are easy to implement [12] but the enhanced image often suffers from the spectral distortion due to spectral mismatch between spectral ranges of both images [9]. Use of histogram matching and multivariate regression is proposed to reduce the spectral distortion in these methods [21]. The basic workflow of CS based methods is illustrated in the figure 2.1, given below.

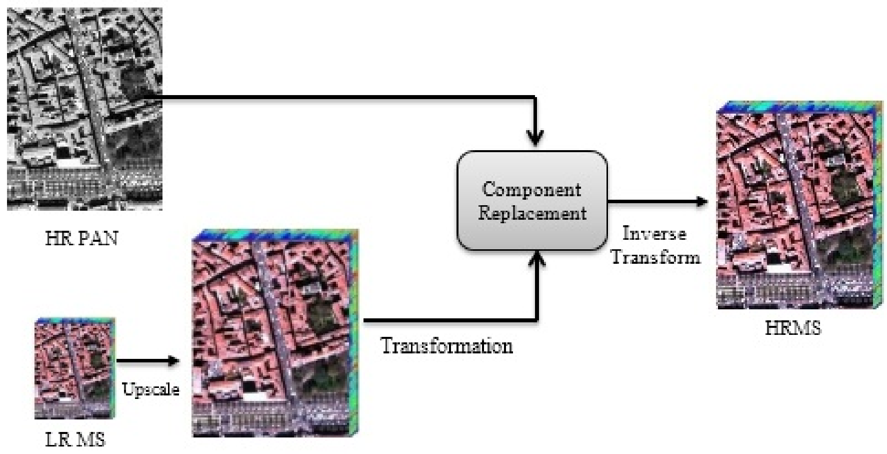

Figure 2.1: A basic workflow of Component Substitution Method that takes LR MS and HR PAN as input and by replacing a componenet of up sample LR MS creates a HR MS image as a result.

2.1.2 Multi Resolution Analysis (MRA)

In Multi Resolution Analysis (MRA) methods, spatial details extracted from PAN image by multi resolution decomposition and injected into interpolated MS image. The figure 2.2 demonstrates the general workflow of MRA method.

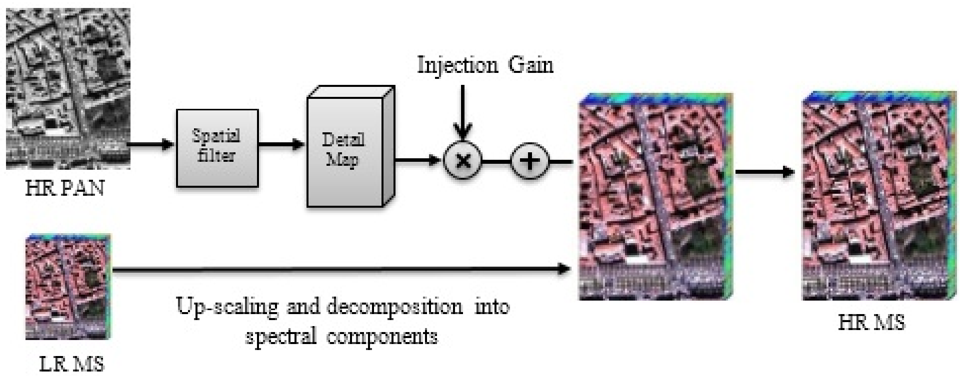

Figure 2.2: A basic workflow of Multi Resolution Analysis Method taking HR PAN and LR MS as input, extracts details form PAN and inject them into up sampled LR MS to generate HR MS image.

Methods that fall under this group are Modulation Transfer Function Generalized Laplacian Pyramid (MTF-GLP) [22], MTF-GLP with High-Pass

Modulation (MTF-GLP-HPM) [23], Hysure [24] [25] and Smoothing Filter-based Intensity Modulation (SFIM) [26]. Some methods proposed to extract the spatial details in MRA based techniques are: Decimated Wavelet Transform (DWT) [27], Undecimated Wavelet Transform (UDWT) [28], "Á-Trous" Wavelet Transform (ATWT) [29], Laplacian pyramid [30], non-separable transforms, either based on wavelet (e.g., contourlet [31]) or not (e.g., curvelets [32]).
These methods are robust in case of spectral consistency and aliasing [20], but implementation of these is computationally hard, as it requires more resources and spatial filters for extracting spatial details has to be designed carefully.

2.1.3 Hybrid Methods

To combine the advantages of both CS and MRA based methods, few hybrid pansharpening techniques are proposed. In this type of methods, substitution of CS based technique is replaced by spatial detail injection of MRA based methods. In this parameters (spatial filter) are carefully set to maintain a balance between the spatial and spectral details. The most prominent example of such methods is Guided Filter PCA (GFPCA) [33]. These methods cannot provide the optimal result, but can maintain the fair trade of between the spatial and spectral information. The figure 2.3 shows the general workflow of Hybrid method.

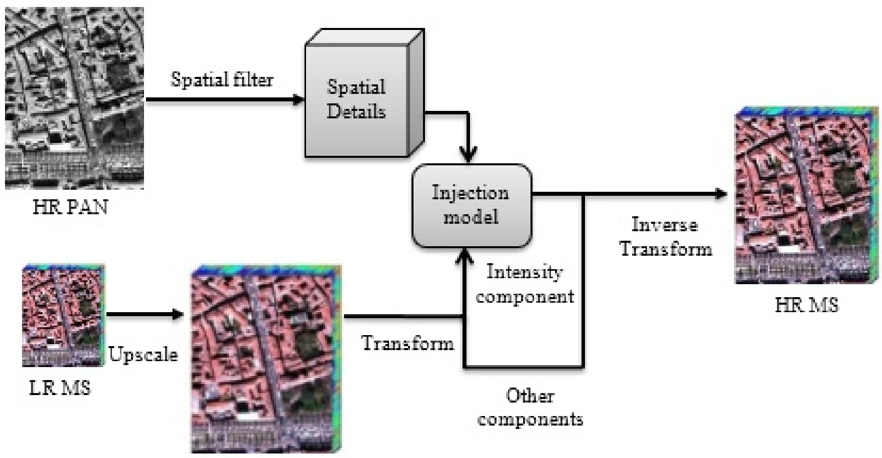

Figure 2.3: A basic workflow of Hybrid Method taking LR MS and HR PAN as input, to produec HR MS extract a detail map type component from PAN and inject it into transformed MS image.

2.1.4 Matrix factorization

Methods of Matrix Factorization group are based on two steps first a dictionary learned from observed HS image, so that the target image is decomposed into factorization that is the product of matrix formed by dictionary and matrix of coefficients of target image. Second using this decomposition, actual coefficient matrix can be recovered and then using it, target image can be obtained.
Most prominent method of this group is coupled nonnegative matrix factorization (CNMF) [34]. Initially this method is proposed for the fusion of HS and MS but in [7] it is used for fusion of HS and PAN. These methods are computationally very expensive.

2.1.5 Model based methods

Most prominent technique that comes under model based methods is Bayesian inference. It is based on the posterior distribution of the reference image given the input images (MS and PAN). This distribution has two main parts: likelihood function (probability density of input images given reference) and prior probability density of the reference image.
Name of noticeable Bayesian based methods are Bayesian naive [35], Bayesian HySure [36] and Bayesian sparse [37]. Many other can also be found in the literature [38] [39] [36] [40]. Sparse [41] and variational models [42] [43] [44] also fall in this group. Model based methods by formulating proper prior knowledge for the region of interest can better confront the ill pose nature of fusion process. On the other hand these methods work well only when the prior knowledge is accurate (representative model is used), require high resolution images for learning and are computationally expensive.

2.1.6 Deep Learning based methods

As deep learning provided reliable results and efficiency in numerous domains like speech recognition [45] [46] [47] and natural language processing [48] etc, over conventional signal processing techniques it is being used in image processing field as well, like for image classification [49] [50], super resolution [51], pattern recognition [52] etc. Many researchers also proposed the frameworks for pansharpening using deep learning. Most of these are for fusing MS with PAN image. Model using different concepts of deep learning proposed for the pansharpening are categorized as under:

2.1.6.1 Convolutional Neural Network

CNN aimed to process the data presented in the form of multiple arrays like 2D/3D array representing pixel intensity values of a gray/colored image. CNN are capable of approximating the complicated nonlinear relationships. They provide a convenient training mechanism by alleviating the complex task of self-designing of filters. The figure 2.4, depicts the basic architecture of CNN.

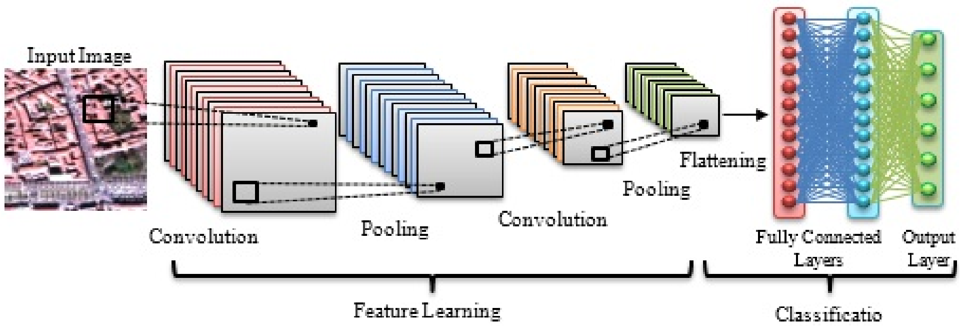

Figure 2.4: A basic architecture of Convolutional Neural Network depicting input, covolution, pooling, fully connected layers and output layer.

Masi et al. [53] proposed the creation of information cube by stacking interpolated MS image and PAN image, then by using CNN mapping is learned between cube and reference image. This mapping is later on used for enhancement. He et al. [54] proposed two detail injection based CNN models, one uses only PAN while other uses both images (MS and PAN) for detail extraction. Guo et al [55] proposed a CNN based model for pansharpening that incorporates the concept of dilated multilevel blocks which means earlier features are combined with local concatenation layer, enabling it to utilize the extracted feature more efficiently. Some machine learning pansharpening models are proposed with special loss functions. Liu et al. [56] proposed a fusion framework with loss function formed by combining the spatial structure similarity and spectral angle mapping that results in achieving optimal result in terms of spatial and spectral details as compared to original than just matching the pixel wise accuracy. Eghbalian et al. [57] presented a DCNN model for pansharpening that uses the DCCN to estimate the detail injection as in MRA techniques and special loss function is proposed that focuses on learning the spectral characteristics.

Yuan et al. [58] proposed a CNN based framework that performs pansharpening at multiscale and multidepth, by using filters of different dimensions at each layer of the network. Huang et al. [59] proposed another multidirection

subband deep learning method that trains with the patches of high frequency subbands of PAN image. In [60] He et al. proposed a multichannel DCNN model for pansharpening that works as channel level enhancement and helps to deal with problem of spectral mismatch between the MS and PAN image. Chen et al. [61] presented an image-to-image (I-to-I) CNN model for feature extraction. I-to-I CNN provides more robust mapping between MS and PAN image and its different hidden layers extract different level and multi-scale features.
Few deep learning based techniques are used to specifically enhance the spatial resolution of HS images with the help of MS image where MS image act as a single band image because it has more spatial details than HS image [62] [63]. Palsson et al [64] presented a deep neural network based fusion framework for HS and MS image. Dimensionality reduction is also performed with the help of PCA before the fusion.

2.1.6.2 Autoencoder

Autoencoder is a type of unsupervised neural network. It consists of two parts; encoder and decoder. Encoder tries to create a compressed representation of the original data known as encoding by focusing on only important details. Then decoder tries to reconstruct the original data from the encoding prepared by encoder. More the reconstructed data is close to the original, more the Autoencoder is good at retaining the important details. Traditional architecture of Autoencoder is illustrated in figure 2.4, given below:

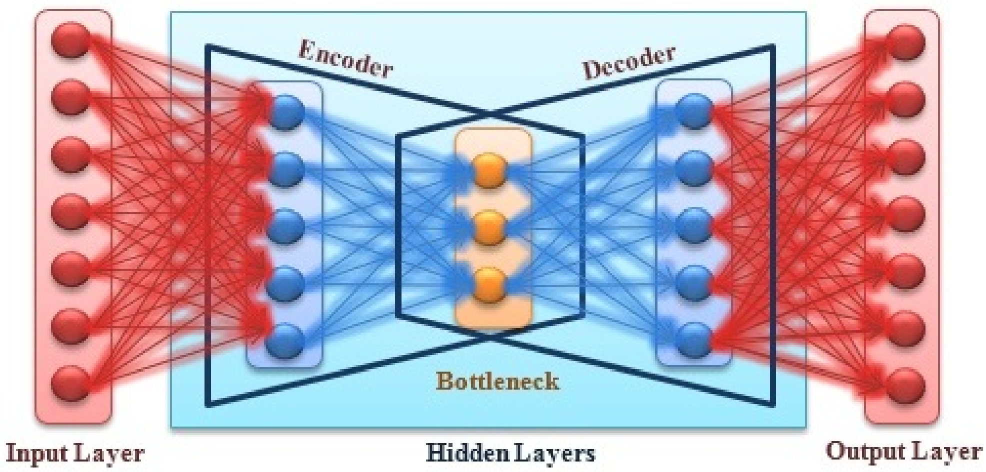

Figure 2.5: A basic architecture of Autoencoders showing input layer, bottle neck and output layer.

AE have been efficiently utilized for image reconstruction, recovering missing data and classification tasks. Promising results in these tasks indicates the capability of AEs in retaining the most representative knowledge. Many researchers have proposed AE based pansharpening models taking advantage of different types of Autoencoders. Cai et al. [65] proposed a deep network that uses the sparse Auto encoders for feature extraction (from input and reference image) and Sparse Denoising Autoencoder for learning the mapping between the extracted features. Xing et al. [66] proposed Stacked Sparse Autoencoders (SSAE) and deep learning based pansharpening technique that uses multiple nonlinear Deep Neural Networks (DNN) to explore the hierarchical features of patches.
Autoencoder can also be used as a dimensionality reduction module because while creating the compressed version it only retains the important details and ignores the noisy and redundant data.

2.1.6.3 Residual learning

Some proposed NN based methods works by incorporating the concept of residual learning. These methods tries to recovery only the missing details and are computationally efficient as they reduce the computational time by not wasting time on recovering the unnecessary details. Wei et al. [67] and Vitale et al. [68] proposed a deep residual learning model for pansharpening of multispectral images. Shen et al. [69] presented a residual learning framework for pansharpening that also incorporates the benefits of model based methods. Scarpa et al [70] proposed another deep learning based target adaptive residual network for pansharpening of MS image.

2.1.6.4 Multiple branch Networks

A number of NN based models that consist of separate branches of NN to extract features from each image provided at the input instead of stacking them together and then passing them to the network. In this way more accurate features can be extracted by applying the image specific filters. Gaetano et al. [71] proposed a simple two branch CNN model for the fusion of MS and PAN image. Liu et al. [72] proposed a two stream CNN based fusion framework, first stream consists of two branches to extract features from both input images, second stream is used to combine the extracted features and finally enhanced image is obtained by fusing the combined features into the interpolated high spectral image. Shao et al. [73] proposed another two branch CNN model for the fusion of MS and PAN image, at both branches use of different level of deeper architecture is proposed to achieve high nonlinearities and obtain high-level features for fusion task. Residual learning based mask is calculated after concatenating the extracted features and final result is obtained by overlaying the mask on the low-resolution MS image. CNN based two branch model is also proposed for the fusion of HS with MS image, first branch takes HS as input and extract details from each pixels spectrum while other branch takes HR-MS image and extracts the spatial information form corresponding neighborhood [74]. This correspondence helps in minimizing the spectral distortion.

2.1.6.5 Attention based Models

Images are composed of pixels and each pixel contains distinct level of information. A mechanism of attention is introduced for images to deal with image pixels according to the level of content they hold, like giving more weightage to the pixels having more information than the one that contains less information. This enables the system to focus only on relevant information and ignore the rest. Like pixels, same case is with the channels in the images. Attention is being recently used in many computer vision tasks such as classification [75] [76] [77], image dehazing [78] [79] [80], recognition [81] [82] [83], super resolution [84] [85] [86] and enhancement [87] [88] [89] etc.
Numerous researchers have used attention mechanism for pansharpening as well. In [90], Jiang et al. proposed a spatial-channel attention pansharpening approach that learns mapping among two image pairs (HR-PAN LR-MS and HR-PAN HR-MS) based on differential information between these. In the input of model, gradient of LR MS is added to enhance the spatial information. Li et al. in [91] presented a residual CNN that incorporates channel attention mechanism for pansharpening which treats each channel differently. In [92] Yang et al. proposed a novel spatial attention-based approach for fusion of HS and MS image. In [93] Zheng et al. presented another HS pansharpening approach that utilizes the deep priors and encompasses the channel-spatial attention module for spatial enhancement. This model also incorporated the idea of residual learning along with attention.

2.1.7 Others

Other than above categories Guided filter and injection gain-based techniques were also proposed for the pansharpening of HS images [21]. In this, guided filters are used to extract the details and injection gain matrix is formulated to estimate the amount of detail to be enhanced.
These methods can maintain an appropriate tradeoff between the spatial and spectral information but require carefully setup of coefficients of guided filters

and injection matrix.

2.2 Dimensionality reduction

Dimensionality reduction is not considered in most of pansharpening techniques. However, while dealing with high dimensional images on large scale, most algorithms have to face the challenge of handling high dimensional feature space [94]. To deal with the curse of dimensionality, remote sensing researchers proposed numerous feature selection methods, most prominent among all is Principal Component Analysis (PCA) [95] [96], Independent Component Analysis [97] Nonnegative Matrix Factorization (NMF) [98] and Autoencoder (AE) [99] [100].
PCA specializes in extraction of linear non correlated components depicting the variance among data. Components have less contribution towards variance are discarded while in some case they are also useful to retain. ICA performs dimensionality reduction by decomposing signals into additive subcomponents while making assumption that all these components statistically independent discarding one does not affect the other while thats not the general case in reality. NMF represents the high dimensional data into two parts: frequent and rarely occurring. Under this technique dimensionality reduction is achieved by retaining mostly occurring parts while discarding the others but we cannot be sure always in advance that rarely occurring parts are not informative. AE is a NN based dimensionality reduction method. It can extract linear as well as nonlinear components.

2.3 Qualitative and Quantitative Assessment

Pansharpening process fuses two images of different resolutions to obtain a single high resolution image but due to difference in dimensions it may introduce spectral and spatial distortion in the fuse image. The robustness of the pansharpening method is evaluated by analyzing the fused image created by it but such assessment is not an easy task due to unavailability of reference image. Simple method that can be followed for the assessment is to compare each band of the final image with PAN image on the basis of some suitable criteria for spatial quality assessment and for spectral assessment final image is downgraded to the resolution of original high spectral image and then these two are compared [7].

2.3.1 Walds Protocol

To deal with the problem of no reference, Walds protocol is followed, that proposed the use of original image as a reference and its downgraded version is used as an input [101]. This protocol is formed on the principal that each pansharpened (fused) image must possess two major characteristics: reversibility and synthesis. Reversibility means that the original image can be acquired by downgrading the fused image. Synthesis means that the fused image should be similar to the image that would be acquired by the same sensor of the same location but at a high resolution. The general Walds Protocol followed for quality assessment is depicted in the figure given below:

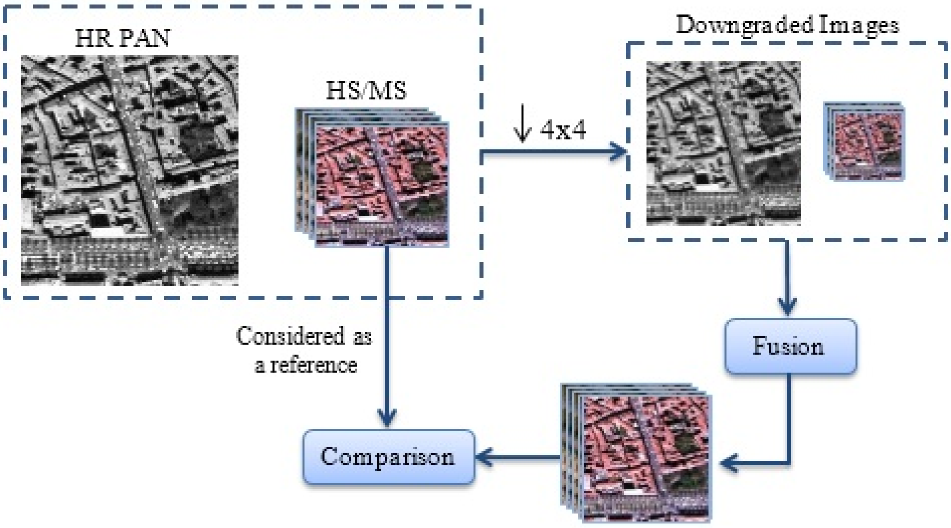

Figure 2.6: A general architecture Walds Protocol that takes HR PAN and LR MS, downsample it, perform operation on it and use the original data as a reference.

2.3.2 Assessment Convention

Generally two types of assessment is performed on fused image; qualitative and quantitative assessment. Qualitative assessment is also known as visual analysis. In this type of assessment strategy, a group of people compare the input with the fused image and then assign quality scores on the basis of different parameters like geometric outlines, amount of spatial details, color distribution, size of objects and local contrast etc. This kind of assessment is simple and honest but it is based on the experience of the viewer. Evaluation could be more precise if the observer has knowledge of ground truth. Its main focus is on finding the visual artifacts like blur etc. the viewer usually grade the image based on their quality. These grades and mesure against each grade is summarized in the table below: Quantitative measures compute some numerical values and on the basis of these it assess the quality of fused image. Many quantitative measures have been proposed to assess the quality of the fused image or performance of the pansharpening process. These are described in detail under the heading of quality metrics.

2.3.3 Criteria of Analysis

There are some criteria of analysis to compare the results of different pansharpening methods. Three type of analysis is performed on the pansharpened image to evaluate the capability of pansharpening process: Visual, spectral and numerical.

2.3.3.1 Visual

Visual inspection is actually a process of finding artifacts present in the resultant images. Most common visual artifacts are blur and spectral distortion.

- Blur generally appears close to small objects and transition parts in the image. It occur when inadequate spatial details are added and due to spectral mixing.

- Spectral distortion occurs due to spectrum modification caused by pansharpening method.

2.3.3.2 Spectrum

This analysis carefully examines the spectral performance of numerous pansharpening methods at different areas in the images like transition and homogeneous parts.

2.3.3.3 Numerical

A numerical criterion is actually application of different quality metrics on the image to perform the quantitative assessment of the pansharpening result. This criterion is further sub divided into four classes:

- Spatial analysis concentrates on spatial details that are preserved or enhanced.

- Spectral analysis concentrates on the spectral details preservation.
- Global analysis is considers both spatial and spectral details at the same time.
- No reference based analysis is performed when there is no target image available for comparison or assessment.

2.3.4 Quality Metrics

For the quantitative assessment of pansharpening technique different quality metrics are employed. They provide a single value against each metric representing the quality of result produce by pansharpening method. Quality metrics can be categorized into two groups on the basis of use of reference image [52].

2.3.4.1 With reference

With reference metrics are also known as reduced resolution metrics because the original image is downscaled and then used as input. Metrics under this group compares the fused image with reference image by computing some numerical measures. These metrics require the reference image for comparison. Quality metric that fall in this group are stated below with short description working on fused $\hat{H}$ image and H reference image, both belong to $\mathbb{R}^{M \times N}$, where i and j corresponds to row and column index of the image. μ and σ are the mean and standard deviation of the respected images.

2.3.4.1.1 Root Mean Square Error (RMSE)

This measure simply computes the difference between pixel values of fused image and reference image. RMSE falls under the spectral analysis criterion that tells how well spectral information is preserved. The ideal value of this measure is 0 and lower is better [102]. Following equation states the formula used to compute RMSE:

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{ij} - \hat{H}_{ij})^2}$$

2.3.4.1.2 Universal Quality Index (UQI)

This measure computes the transformation of known data from the original image into the pansharpened image [103]. The range of values for this metric is -1 to 1. Ideal value is 1, which means that the both images are similar. Formula for this metric is given below:

$$UQI = \frac{4\sigma_{\hat{H}H}(\mu_H + \mu_{\hat{H}})}{(\mu_H^2 + \mu_{\hat{H}}^2)(\mu_H^2 + \mu_{\hat{H}}^2)}$$

2.3.4.1.3 Spectral Angular Mapper (SAM)

It calculates the spectral angle between the pixels of referenced image and fused image. This metric tells how well the pansharpening method has preserved the spectral details. It is a pixel level process final output is obtained but taking mean of all the values and output is either in radians or degree [104]. Ideal value is 0.

$$SAM = \arccos\left(\frac{H_j, \hat{H}_j)}{\parallel H_j \parallel, \parallel \hat{H}_j \parallel}\right)$$

2.3.4.1.4 Mutual Information (MI)

It calculates similarity between the intensity values of the reference and fused image. Higher value of MI indicates that both image intensities are alike [105].

2.3.4.1.5 Signal to Noise Ratio (SNR)

This metric computes the ratio between the relevant details and noise in the fused image. Higher value of this measure indicates that fused image is similar to reference image. Formula for this measure is stated below [106]:

$$SNR = 10\log_{10}\left[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{ij})^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{ij} - \hat{H_{ij}})^2}\right]$$

2.3.4.1.6 Relative Dimensionless Global Error

It calculates the fused image quality by computing the normalized mean error of every band of fused image. Ideal value is 0. Higher value indicates the presence of distortions in fused image. The equation for computing Relative Dimensionless Global Error (ERGAS) is as follows [107]:

$$ERGAS = 100\frac{dk}{dl}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{RMSE_i}{mean_i}\right)^2}$$

2.3.4.1.7 Mean Bias (MB)

It calculates the difference among the mean of fused and reference image. Ideal vale for this measure is 0 indicating no difference [108]. Formula for it is as follows:

$$MB = \frac{\mu_H - \mu_{\hat{H}}}{\mu_H}$$

2.3.4.1.8 Peak Signal to Noise Ratio (PSNR)

This measure tells about the fusion quality. Number of gray levels is used as a numerator in the formula of this measure. Higher value indicates the better quality of fusion. Formula for the measure is as follows [109]:

$$PSNR = 20 \log_{10} \left[\frac{L^2}{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (H_{ij} - \hat{H}_{ij})^2} \right]$$

2.3.4.1.9 Correlation Coefficient (CC)

This measure falls under spatial criteria of analysis. It focuses on the spectral features similarity and geometric distortions between the fused and referenced image. Ideal value is 1, indicating the accurate match. Formula for CC measure is stated below [110]:

$$CC = \frac{1}{M} \sum_{i=1}^{M} (\hat{H}^i, H^i)$$

2.3.4.1.10 Structural Similarity Index Measure

It is also a similarity criteria between fused and reference image in terms of structure, contrast and luminance. Range of value is -1 to 1 and ideal is 1 representing both images are similar. Equation below states the formula for Structural Similarity Index Measure(SSIM) [111]:

$$SSIM = \frac{(2\mu_H \mu_{\hat{H}} + C_1)(2\sigma_{H\hat{H}} + C_2)}{({\mu_H}^2 + {\mu_{\hat{H}}}^2 + C_1)({\mu_H}^2 + {\mu_{\hat{H}}}^2 + C_2)}$$

2.3.4.1.11 Percentage Fit Error (PFE)

It calculates the norm of the variation among the pixels of fused and reference image. Ideal value is 0 indicating that both images are same. Formula for PEE is as under [112]:

$$PEE = \frac{norm(H - \hat{H})}{norm(H)} \times 100$$

2.3.4.1.12 Spatial Correlation Coeficient (SCC)

The Spatial Correlation Coefficient (SCC) is a quality metric designed to assess the spatial quality of pansharped image. In this metric, the spatial details of pansharped and reference image are extracted and then compared by computing a correlation coefficient between the extracted details. For detail extraction Laplacian filter is employed and correlation computed band by band. Ideal value for this metric is 1 and high value indicates, more spatial correlation between the two images [113].

2.3.4.2 Without reference

These metrics are also known as full resolution metrics because the original image is used as input for enhancement. This group consists of metrics that are used when reference image is not available. Without reference measures only require fused image (final output) for computation. Quality metric that fall in this group are:

2.3.4.2.1 Entropy (He)

Entropy metric calculates the amount of information contents in the pansharpened image. High value of this measure ensures that the fused image is rich in information [114].

2.3.4.2.2 Standard Deviation

The contrast of the fused image is computed in terms of standard deviation. High contrast id depicted by the high value of standard deviation [114].

2.3.4.2.3 Spatial Frequency (SF)

It calculates the column and row frequency of the fused image. Higher value is better indicating more resemblance between reference and fused image [115].

2.3.4.2.4 Cross Entropy (CE)

This measure is used to compute the content similarity between both images. Lower value of this measure shows that both images have similar information [116].

2.3.4.2.5 Distortion (D)

As name indicates it computes the amount of distortion occurred in fused image. Lower value is better representing less distortion [110].

$$D = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \mid H_{ij} - \hat{H_{ij}} \mid$$

2.3.4.2.6 Fusion Quality Index (FQI)

It measures the quality of the fused image. Range of value for FQI is from 0 to 1. 1 is the best values showing that the final image has all the information of the input image [117].

2.3.4.2.7 Fusion Similarity Metric (FSM)

This measure falls under the category of spatial analysis criteria. It computes the spatial resemblance between the reference and fused image. Range of value for FSM is from 0 to 1. 1 is the best values showing that the fused image has all the information of the input image [105].

2.3.4.2.8 Fusion Mutual Information (FMI)

This metric computes the amount of reliance between the input and fused image. Higher value of this measure indicates the high quality of the resultant image [118].

2.3.4.2.9 Quality Not Requiring a Reference

It is the most common measure used in situations where reference image is not available [103]. It is composed of two terms D_λ and D_s. D_λ computes the amount of spectral variation present between the up-sampled input and fused image by using the formula of UIQI metric. The ideal value is 0 indicating minimum spectral difference between the both images. D_s measure the amount of spatial variation present between the PAN and fused image by using the formula of UIQI metric. Optimal value is 0 indicating no difference.

$$QNR = (1 - D_\lambda)^\alpha (1 - D_s)^\beta$$

α and β are the positive integers that can be modified to have optimal balance between the spectral and spatial information. Ideal value for this metric is 1 when both D_λ and D_s are 0, indicating no distortion.

Apart from above all quality metrics, there are many others proposed by several researchers but they are somehow interlinked with these. Most commonly employed metrics include ERGAS, CC, SAM, RMSE and UIQI. We will also compute only these for our model and its results assessment.

Chapter 3

Proposed Methodology

This chapter of manuscript elaborates the motivation and proposed model in detail. Each module of the proposed approach is explained with help of diagram.

3.1 Motivation

It is generally recognized that both PAN and MS/HS images hold distinct information and details. PAN is rich in spatial details that encompass the geometric information whereas MS is a carrier of spectral details that provides the capability to distinguish two visual similar objects. Pansharpening is an image enhancement process, often considered as an essential preprocessing step in most of the remote sensing applications, that takes spatial information from PAN image and spectral information from MS/HS image to create a MS/HS that is rich in both spatial and spectral details. Still, it is not easy to explicitly state what these details are and how they can be presented self-sufficiently. Most of the pansharpening techniques proposed employ same kind of detail extraction mechanism for both images while the underlaying information is not same. So, the result suffers from spectral irregularities and other artifacts. Both images have distinct information and must be dealt according. Even each feature of the respective images, must be given different weightage in order to achieve optimal fused image.

3.2 Proposed Model

Inspired by use of CNN in many other image processing domains and its powerful capability of representing an image into its hierarchal feature and then reconstructing image back from those features flawlessly, we propose a pansharpening method that deals separately with both PAN and MS image for feature extraction with help of attention mechanism for further enhancement of extracted features, fuse them in feature domain and then from fused features reconstruct a high resolution MS image as a result while incorporating the benefits of residual learning for fast convergence of network.General architechure of the propose model is depicted in the conceptual diagram given below:

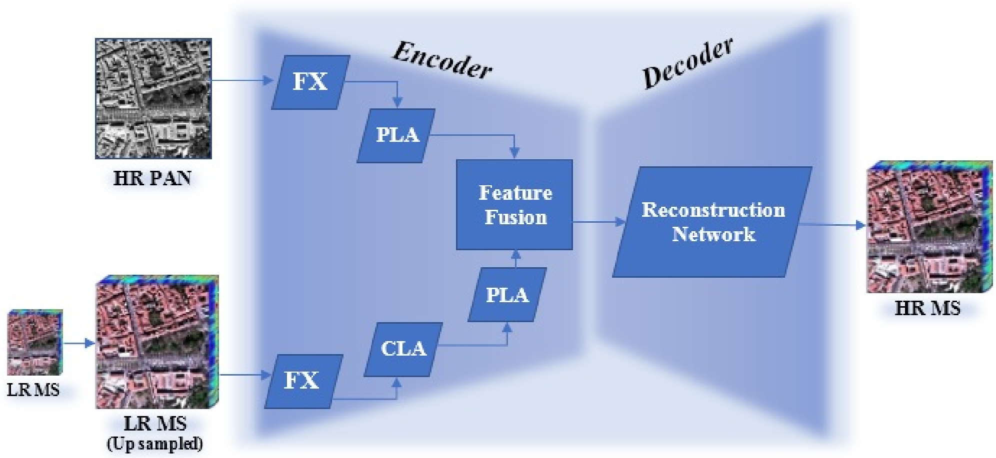

Figure 3.1: Conceptual Diagram of proposed approach representing all important network components, inputs and output.

3.2.1 Problem Formulation

Let $\breve{E}$ be a feature extraction network that transform the image I into feature space, such that image I can be reconstructed without loss of any information from $\breve{E}(I)$. For proposed approach we employ two separate CNN based feature extractors to transform PAN and MS images into feature domain. The two images PAN and MS are denoted by I_P and I_M respectively. The two CNN based feature extractors for PAN and MS images are denoted as $\breve{E}(I_P)$ and $\breve{E}(I_M)$, where l letter as subscript specify the corresponding layer of features extractor.

3.2.2 Feature Extraction

For pansharpening first step is to extract details also known as features. There are numerous procedures are proposed to extract these features from corresponding images. First are hand crafted features for instance SIFT, LBP and HOG, but they do not fulfill the requirement of reconstructing the original image from the obtained features. In other words, these techniques can be categorized as lossy transformations. Another most used feature crafting technique is sparse encoding, which utilizes a dictionary created during training to represent an image. There are several sparse encodings based pansharpening approaches proposed but the fusion is still carried out in image domain. Because fusing features in spare domain is still uncertain. Also, this technique requires a high-resolution image during training for dictionary creation, which is not available in most of the cases. So, this technique will not work well where high resolution image is not available for training. CNN has the power to extract features from the image and then can also combine them in feature domain.

3.2.3 Attention

Many studies have shown the promising result by using attention mechanism in construction of neural networks [119] [120] [121]. Recently it is being employed for a leading image restoration task i.e. super resolution [84] [85] [86] . Many researchers have used it for pansharpening as well [90] [91] [92]. Inspired by their work we have also incorporated attention in our model but separately for each image to deal the distinct information contained in them. After basic feature maps are created by feature extraction network, to further enhance these feature maps attention mechanism is employed. PAN and MS image contain different information thus attention mechanism must also be designed accordingly. Two type attention mechanisms are designed to flexibly deal with the information of each image; pixel and channel level attention. Pixel level attention is applied on the feature maps generated from PAN image and channel level attention is applied on features maps constructed from MS image. In this way we deal with pixel level and channel level features differently by giving distinct weights to features of different levels before passing it to the feature fusion network.

3.2.3.1 Channel level Attention

From many studies it is globally accepted that each channel of an image encompasses distinct information and to deal with this diverse information optimal weightage must be given to corresponding features. In pansharpening we need to recover the lost spatial information so in channel level attention module, we first consider the channel level spatial information by employing the global mean pooling function. The Figure 3.2 illustrates the architecture of Channel Level Attention(CLA) module.

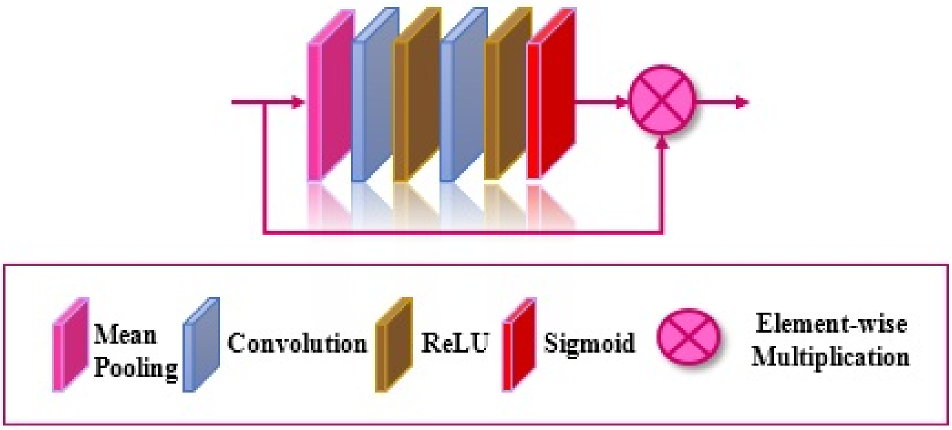

Figure 3.2: Architecture of Channel Level Attention Moduel of the proposed approach.

The used mean pooling function is formulated as:

$$\breve{G}_c = F_c(\breve{E}_c) = \frac{1}{h \times w} \sum_{j=1}^{h} \sum_{j=1}^{w} I_c(i,j)$$

Where$I_c(i,j)$ represents the pixel at (i,j) place in c th channel of I_c, F_c is mean pooling function. This transforms the feature map from $h \times w \times c$ to $1 \times 1 \times c$. This transformed feature map is fed to a dual layer convolutional model that has ReLU activation after each convoultional layer and sigmoid at the end.This model computes weights for distinct channels and is formulated as:

$$W_c = \sigma(conv(\Upsilon(conv(\breve{G}_c))))$$

Where Υ is ReLu activation function and σ is sigmoid activation function. Last step of this module is element wise multiplication of $\breve{G}(I_M)$ and W_c as depicted in equation below:

$$\breve{E}_{CLA}(I_M) = \breve{E}(I_M) \otimes W_C$$

3.2.3.2 Pixel Level attention

Like channels information content also varies among different image pixels. Pixel level attention module is designed to explicitly give model ability to focus on the pixels that have more information. The feature maps obtained from PAN image feature extractor are fed to a dual layer convolutional model that has ReLU activation after each convoultional layer and sigmoid at the end. This step is also employed after Channel Level Attention of MS image so further enhance those features. The Figure 3.3 depicts the architecture of Pixel Level Attention(PLA) module.

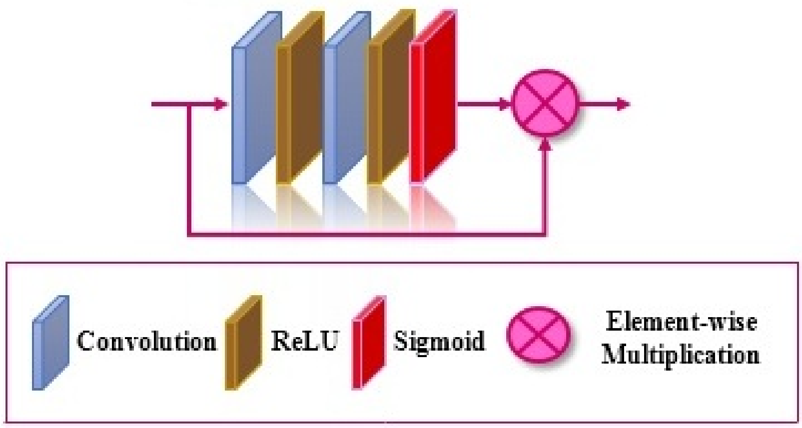

Figure 3.3: Architecture of Pixel Level Attention Moduel of the proposed approach.

This model computes weights for distinct pixels as:

$$W_p = \sigma(conv(\Upsilon(conv(\breve{E}(I_P)))))$$

Last step of this module is also element wise multiplication of $\breve{E}(I_P)$ and W_P, as depicted in equation below:

$$\breve{E}_{PLA}(I_P) = \breve{E}(I_P) \otimes W_P$$

3.2.4 Fusion Module

After enhancing the feature maps with the help of attention network, the next step is fusion of these features. The general protocol followed in this case is to first apply some pooling technique such as mean or max pooling on these feature maps, but this approach leads to the loss of information, while in our task that is pansharpening we need to retain every bit information. So, to avoid the information loss caused by pooling, here we utilize effective alternative to pooling that is concatenation. This procedure is formulated as:

$$\breve{E}(I_p, I_M) = \breve{E}_{CLA}(I_M) \oplus \breve{E}_{PLA}(I_P)$$

In above equation the symbol $\oplus$ denotes the concatenation and $\breve{E}(I_p, I_M)$ is the required fused feature map. This fused feature map act as an input to the fusion network of our model. This network creates a more precise encoded representation of the concatenated features with the help for three convolutional layers and output of this network is a tensor of size $[w/4, h/4, 256]$ which a compact encoding of spectral and spatial details of the MS and PAN images respectively. This tensor now will act as an input for final reconstruction network of proposed model that will reconstruct the required pansharped image.

3.2.5 Reconstruction Network

The feature maps contain only one fourth of input in terms of height and width. Interpolation technique is desirable here to match the spatial resolution requirement. A simple linear interpolation technique can be used but the learnable approach will perform much better [122]. This network is composed of transposed convolutional layers that performs the up sampling in terms of spatial resolution to obtain the required resolution of reconstructed image. The Figure 3.4, illustrates the complete architecture of the proposed pansharpening approach:

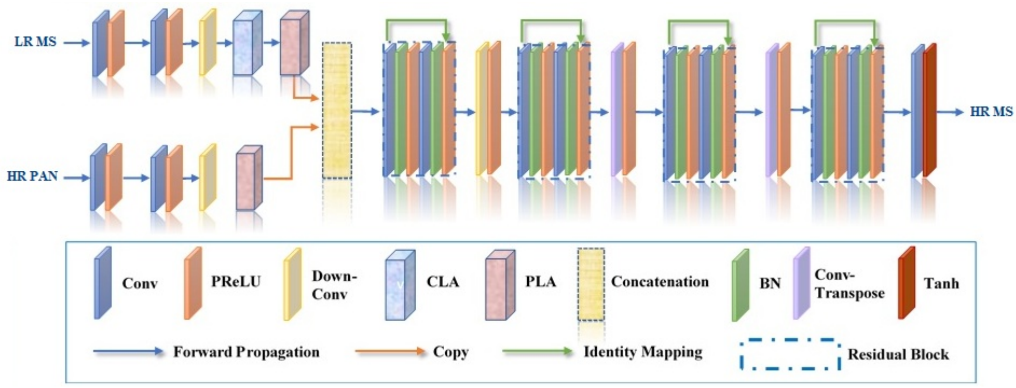

Figure 3.4: Detailed architecture of proposed Dual Attention based Two Stream Pansharpenig Approach, showing all the layers and tyeps of links.

3.2.6 Skip connection

Generally high-level features are used for the recovery of information while they often fail to provide the texture details because they only contain abstract details of the input [123]. To obtain a rich pansharped image we need to include information from all level of representations, but this will increase the computational complexity of our model. To deal with this problem a skip connection is incorporated that copies the information (feature maps) from encoder directly to decoder after each up-sampling phase and concatenate it with equivalent feature map to infuse the detail lost during down sampling.

3.2.7 Residual Learning

Recently residual learning approach [124] [125] is being effectively utilized in many computer vision tasks [126] [127]. A Residual network (ResNet) is a specialized form of highway networks; whose skip connections have no gates. This aids the flow of information from low level layers to high level layers. In our network we also employed residual CNN units instead of simple CNN units. A residual unit can be expressed as:

$$x_n = \hat{I}(r_n) + \Omega(r_n, \omega_n)$$

$$r_n + 1 = \phi(x_n)$$

Where r_n and $r_n + 1$, are the input and output of nth residual CNN unit, Ω is residual function, $\hat{I}(r_n)$ is the identity mapping, $\phi(x_n)$ is the desired activation function. All the weights in our model are adaptive learned that works much better than the one specified directly.

3.2.8 Loss Function

The integral part of training a neural network is loss function that computes the model error by evaluating a set parameter (ϑ). The value of these weight parameters is optimized by minimizing the model error. In case of pansharpening the loss function also significantly affects the quality of pansharped image. Mostly ℓ_2 norm is employed as loss function in image restoration tasks [51] [128] [53] [126], however this causes the blurriness in the restored image. Many researchers have used and recommended the use of ℓ_1 loss function for better training of image restoration models [127] [129]. Inspired from their work we also employ ℓ_1 for optimal training of our network.
The loss function ℓ_1 generally formulated as:

$$\ell_1(\vartheta) = \frac{1}{n} \sum_{i=1}^{n} \mid \breve{E}(I_P^i, I_M^i; \vartheta) - X^i \mid_1$$

Where I_P is PAN, I_M is low resolution multispectral image (LR MS) and X is high resolution multispectral image, n is batch size of sample images for training.

Chapter 4

Experimention and Results

This chapter of the manuscript summarizes the experimental work done for the performance evaluation of the proposed approach. Different quality measures are computed to assess the performance of the proposed model. Visual and quantitative results are compared with a recently proposed approach.

4.1 Dataset Descrpition

Experiments are carried out on the Pleaides dataset. Images of two different locations are used for training and testing. The table below reports the details of the dataset used.

Table 4.1: Attributes of the Pleiades dataset used for training and testing.

Attributes of the Sensor	
Year	2011
Mission type	Earth observation
Resolution	Panchromatic 0.8m Multispectral 3.2m
Bands(N)	P: 470 - 830 nm Red: 590 - 710 nm Green: 500 - 620 nm Blue: 430 - 550 nm Near-infrared: 740 - 940 nm
Dimension	1024 x1024
Swath width	20km

4.2 Implementation Details

There are 50800 image patches (LR MS, HR PAN, up sampled LR MS and HR MS) of size 128×128 are created for training. The larger sizes like 256/512 of an image can also be supported by our model, but it requires more memory so to avoid complexities optimal size is selected. Likewise, is the case with the batch size so its minimum value is set to 16. The proposed architecture is implemented using Pytorch, a machine learning package of python and trained on NVIDIA GeForce GTX 980 GPU. Adam optimizer is used for minimizing the loss with the 0.5 momentum and 0.0001 learning rate. Model training will take about 12 hours.
Test dataset consist of 260 image patches. As we have original HR MS image so used with full reference metrics for quantitative evaluation such as SAM, ERGAS, SCC, UIQI and SIMM. Abberivation of these metric are given in Table 4.2 and remaining details can be found in Chapter 02 under Qualitative and Quantitative Assessment section.

Table 4.2: List of metrics along with their apbreviation, computed for quantitative evaluation of experimental results.

Full Reference Metrics	
SAM	Spectral Angle Mapper
ERGAS	Erreur relative globale adimensionnelle de synthse
SCC	Spatial Correlation Coefficient
UIQI	Universal Image Quality Index
SIMM	Structural Similarity Index Measure

4.3 Experimental results and analysis

In this section, we will compare our model with a recently proposed pan-sharpening algorithm TFNet [72] and report the quality metrics computed against each experiment. The Figure 4.1 is the sample image used for this case, from left first is HR PAN, next is LR MS, then up sampled LR MS and finally at last HR MS image.

Figure 4.1: Sample image used for Case 01 (From left to right HR PAN, LR MS, upsampled LR MS and HR MS image).

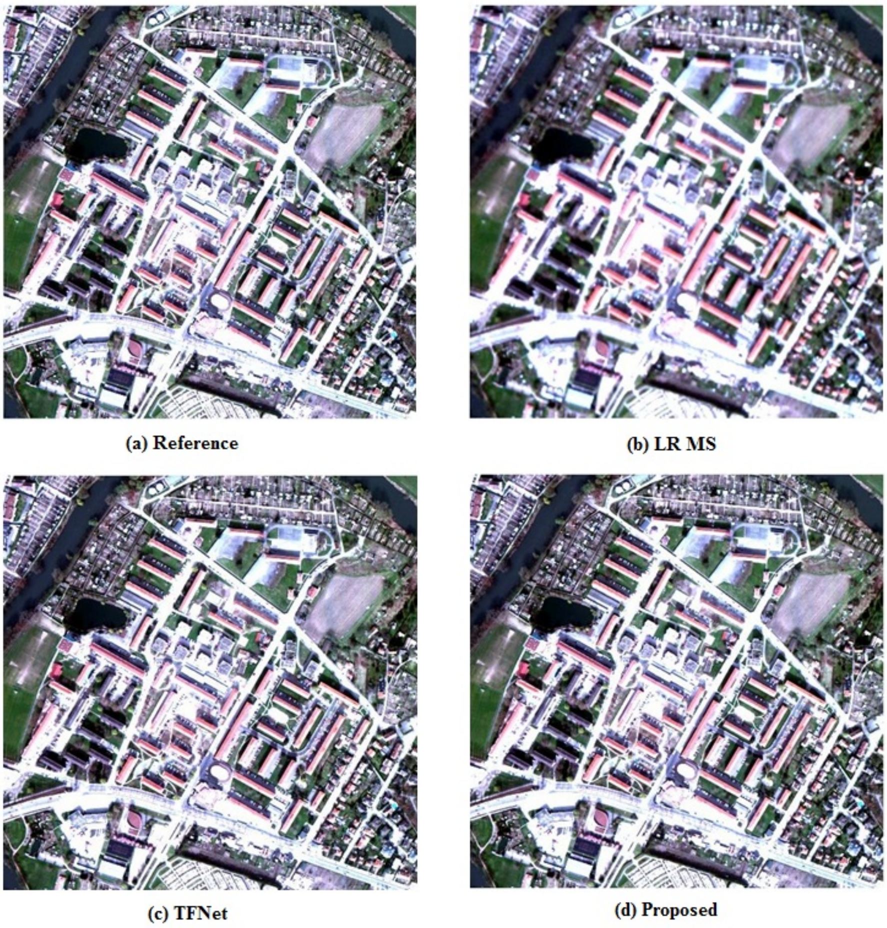

Figure 4.2: Visual result of Case 01 (left to right, First row: reference image and upsampled LR MS image, second row output of TFNet and Proposed).

In our first experiment, we consider the most favorable case, where training and test images are of same location taken by same sensor.Visual examination of the results aids in undrestanding the behaviour of model.

The Figure 4.2 is the visual results obatined by proposed approach and TFNet along with the Reference and LR MS used as input, in false color synthesis (3 bands as 3, 2, 1). Table 4.3 includes the quality measures computed for quantiative evaluation. Bold values in the table are the best values achieved against each metric. From visual analysis it can be seen that both networks performed well, produced visually acceptable pansharped MS image and recovered most of details while maintain the underlying structure of image. Both images are are free from blurrness and all edges are reconsturcted very well.

Table 4.3: List of quality Metrics computed for quantitative evaluation of Case 01 along with their ideal values.

Metrics	**Ideal Value**	**TFNet**	**Proposed**
SAM	**0**	**0.0230**	**0.0230**
ERGAS	**0**	0.7082	**0.7012**
SCC	**1**	**0.6161**	0.6144
UIQI	**1**	**0.9991**	**0.9991**
SSIM	**1**	0.9837	**0.9840**

From Table 4.3, it can be seen that value of quality metrics indicate that overall proposed approach performed well than TFNet with a minor difference. The value of SAM indicate the preservation of spectral details is same in both models, while distortion is less in proposed model as depicted by ERGAS metric. High value of UIQI and SSIM indicate that both models learned parameters very well for this particular location of the corresponding sensor and pansharped is very similar to reference image.

For further analysis, we consider a typical case of evaluation, where training and test images are taken by same sensor but are of different locations. The Figure 4.3 is the sample image used for testing in second case, from left first is HR PAN, next is LR MS, then up sampled LR MS and finally at last HR MS image.The Figure 4.4 is the visual results obatined by proposed approach and TFNet along with the Reference and LR MS used as input, in false color synthesis (3 bands as 3, 2, 1).

Figure 4.3: Sample image used for testing in Case 02 (From left to right HR PAN, LR MS, upsampled LR MS and HR MS image).

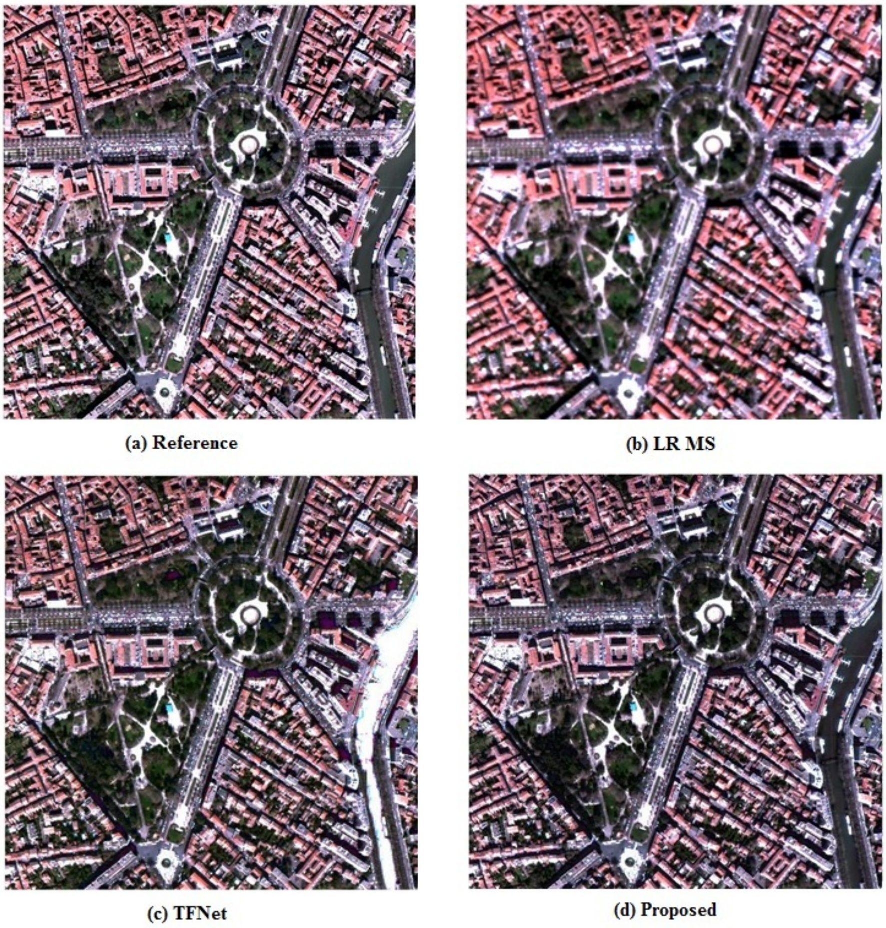

Figure 4.4: Visual result of Case 02 (left to right, First row: reference image and upsampled LR MS image, second row output of TFNet and Proposed)

Even training and test images are not same, still both models performed well and effectively recovered most of the details while the maintain underlying details to some extent. Visually it can be seen that the resultant image of TFNet has some distortion in right side of image. In this case TFNet model lacks in preserving the underlying details of image while pansharpening, in other words TFNet lacks in learning the characteristics of sensor as compared to proposed approach. Proposed approach maintained the underlying structure of image and performed maximum enhancement. Table 4.4 reports the quality measures computed for quantitative evaluation. Bold values in the table are the best values achieved against each metric.

Table 4.4: List of quality metrics computed for quantitative evaluation of Case 02 along with their ideal values.

Metrics	Ideal Value	TFNet	Proposed
SAM	**0**	0.1913	**0.1462**
ERGAS	**0**	5.4783	**4.8305**
SCC	**1**	0.4187	**0.4290**
UIQI	**1**	0.9555	**0.9614**
SSIM	**1**	0.6746	**0.6926**

Now from the values of quality metric reported in above table, it is confirmed that our propose model achieved much better results than TFNet. High value of ERGAS, in case of TFNet indicates the presence of distortion in pansharped image produced by this model. SAM value indicate that spectral distortion is less in the resultant image produced by proposed approach. Value of SCC shows that result of proposed approach spatially corelated to the reference image. High values of UIQI and SIMM also confirms that the proposed approach produces image more close to target.

4.4 Summary

We have presented the visual and quantitative results for the two cases, to assess the performance of the pansharpening approach proposed in this manuscript and compared it with the most recent model taken from the same category of pansharpening approaches. Overall, the visual and quantitative results confirm the effectiveness and proficiency of the proposed approach. It can maintain the underlying spectral details while enhancing the spatial details. Proposed model also efficient in learning the sensor characteristics. So, once trained for one sensor, it can pansharp images of different location taken by same sensor.

Chapter 5

Conclusion and Future Perspective

This chapter encapsulates the overall work done along with the future directions for the extension of the proposed approach.

5.1 Conclusion

The main aim of this research is to propose and implement a new pansharpening approach that fuses a HR PAN and LR MS image and delivers a pansharped MS image high in both spatial and spectral information. To begin with, state of the art is analyzed thoroughly in order to have a clear idea about the current trends and limitations of proposed approaches. Based on techniques employed by numerous researchers, the literature review is divided into six broad families of pansharpening.

Lately deep learning is being effectively utilized for working with images and other high dimensional data. More specifically inspired by the tremendous progress achieved by incorporating the benefits of CNN in image processing and other computer vision tasks, in this manuscript we have presented a novel two stream dual attention based pansharpening method. It mainly consists of three modules feature extraction, feature fusion and image reconstruction. The feature extraction module has two separate networks for both images, works like encoder and combining the advantages of attention mechanism to recalibrate the obtained features. The fusion module fuses the obtained features and forms more compact representation of these. This fused feature map is fed to the reconstruction network for creation of final pansharped image. As proved from the work of many researchers, the adaptation of residual block for propagating the low level details along with high level details and ℓ_1

loss function instead of ℓ_2, also significantly improve the results and overall performance of the proposed architecture.
For the evaluation of the proposed architecture, experiments are carried out using the Pleiades dataset and results are compared with the most recently proposed approach. Images of two different locations of same sensor are used for training and testing respectively. The computed quality measures confirms that the proposed model can efficiently learns the degradation process of the corresponding sensor and can effectively produces a pansharped image having high spatial details while maintaining the spectral fidelity of underlying image.

5.2 Future Work

Due to the limited availability of sensors' data, exhausted experiments cannot be performed. In future we will explore the transferability of our proposed network among different sensors. This will help us in understanding and demonstrating the connection between the network architecture and sensors specifications.
The incorporation of fine-tuning step can also significantly improve the performance of proposed approach on the cost of acceptable computational complexity. This step will not require any involvement from the user. Alteration of a loss function can also be explored that is specifically designed and work for the task of pansharpening.
With technical advancements HS images are now available and being used by some remote sensing applications. High spectral resolution of HS image ensures the presence of spectral signature in them. Libraries are also available containing the spectral signatures of different objects; these can be utilized after extracting spectral signatures from high spectral resolution images for classification or identification of different objects. These are not very common because of their high computational cost and low spatial resolution. But In future we will try to extend our work to HS images.
Furthermore, we will explore the GANs architecture, that have been recently adopted in many domains. In this our proposed network will act as generator of GAN and we will devise a corresponding discriminator network. This will also significantly improve the performance of proposed approach.

References

[1] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio–temporal–spectral fusion of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7135–7148, 2016.

[2] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *European conference on computer vision*, pp. 63–78, Springer, 2014.

[3] N. Akhtar, F. Shafait, and A. Mian, "Hierarchical beta process with gaussian process prior for hyperspectral image super resolution," in *European Conference on Computer Vision*, pp. 103–120, Springer, 2016.

[4] C. Wang, Y. Liu, X. Bai, W. Tang, P. Lei, and J. Zhou, "Deep residual convolutional neural network for hyperspectral image super-resolution," in *International conference on image and graphics*, pp. 370–380, Springer, 2017.

[5] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE transactions on neural networks and learning systems*, no. 99, pp. 1–11, 2018.

[6] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison of pansharpening algorithms," in *2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 191–194, IEEE, 2014.

[7] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[8] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments,"

in *Signal and Image Processing for Remote Sensing*, pp. 552–599, CRC Press, 2012.

[9] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pan-sharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Information Fusion*, vol. 46, pp. 102–113, 2019.

[10] W. CARPER, T. LILLESAND, and R. KIEFER, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogrammetric Engineering and remote sensing*, vol. 56, no. 4, pp. 459–467, 1990.

[11] P. Chavez, S. C. Sides, J. A. Anderson, *et al.*, "Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic," *Photogrammetric Engineering and remote sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[12] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at ihs-like image fusion methods," *Information fusion*, vol. 2, no. 3, pp. 177–186, 2001.

[13] P. Kwarteng and A. Chavez, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens*, vol. 55, no. 339-348, p. 1, 1989.

[14] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Engineering and remote sensing*, vol. 58, no. 5, pp. 561–567, 1992.

[15] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive pca approach and contourlets," *IEEE transactions on geoscience and remote sensing*, vol. 46, no. 5, pp. 1323–1335, 2008.

[16] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood, "Worldview-2 pan-sharpening," in *Proceedings of the ASPRS 2010 Annual Conference, San Diego, CA, USA*, vol. 2630, pp. 1–14, 2010.

[17] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000. US Patent 6,011,875.

[18] T. Maurer, "How to pan-sharpen images using the gram-schmidt pan-sharpen method-a recipe," *International archives of the photogrammetry, remote sensing and spatial information sciences*, vol. 1, p. W1, 2013.

[19] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of ms + pan data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007.

[20] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 446–453, 2011.

[21] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of ms + pan data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007.

[22] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Mtf-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 5, pp. 591–596, 2006.

[23] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 930–934, 2013.

[24] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[25] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "Hyperspectral image superresolution: An edge-preserving convex formulation," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4166–4170, IEEE, 2014.

[26] J. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000.

[27] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[28] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and statistics*, pp. 281–299, Springer, 1995.

[29] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," *IEEE Transactions on signal processing*, vol. 40, no. 10, pp. 2464–2482, 1992.

[30] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.

[31] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[32] J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE transactions on image processing*, vol. 16, no. 2, pp. 297–309, 2007.

[33] W. Liao, X. Huang, F. Van Coillie, S. Gautama, A. Pižurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, *et al.*, "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 ieee grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2015.

[34] N. Yokoya, Y. Takehisa, and I. Akira, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, 2012.

[35] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1174–1184, 2004.

[36] M. Simoes, J. M. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via

subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[37] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, 2015.

[38] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multiband images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1117–1127, 2015.

[39] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3176–3180, IEEE, 2014.

[40] Y. Zhang, A. Duijster, and P. Scheunders, "A bayesian restoration approach for hyperspectral images," *IEEE transactions on geoscience and remote sensing*, vol. 50, no. 9, pp. 3453–3462, 2012.

[41] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 7, no. 5, pp. 1792–1805, 2013.

[42] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for p+ xs image fusion," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[43] A. Huck, F. de Vieilleville, P. Weiss, and M. Grizonnet, "Hyperspectral pan-sharpening: a variational convex constrained formulation to impose parallel level lines, solved with admm," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, IEEE, 2014.

[44] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, "A variational approach for sharpening high dimensional images," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 150–178, 2012.

[45] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views

of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[46] T. N. Sainath, A.-r. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for lvcsr," in *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 8614–8618, IEEE, 2013.

[47] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 8599–8603, IEEE, 2013.

[48] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of machine learning research*, vol. 12, no. ARTICLE, pp. 2493–2537, 2011.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[51] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[52] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, no. Icwrcoe, pp. 133–142, 2015.

[53] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sensing*, vol. 8, no. 7, p. 594, 2016.

[54] L. He, Y. Rao, J. Li, J. Chanussot, A. Plaza, J. Zhu, and B. Li, "Pansharpening via detail injection based convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1188–1204, 2019.

[55] Y. Guo, F. Ye, and H. Gong, "Learning an efficient convolution neural network for pansharpening," *Algorithms*, vol. 12, no. 1, p. 16, 2019.

[56] X. Liu, C. Deng, B. Zhao, and J. Chanussot, "Feature-level loss for multispectral pan-sharpening with machine learning," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 8062–8065, IEEE, 2018.

[57] S. Eghbalian and H. Ghassemian, "Multi spectral image fusion by deep convolutional neural network and new spectral loss function," *International Journal of Remote Sensing*, vol. 39, no. 12, pp. 3983–4002, 2018.

[58] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 978–989, 2018.

[59] W. Huang, X. Fei, J. Yin, and Y. Liu, "A multi-direction subbands and deep neural networks bassed pan-sharpening method," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5139–5142, IEEE, 2018.

[60] G. He, S. Xing, Z. Xia, Q. Huang, and J. Fan, "Panchromatic and multi-spectral image fusion for new satellites based on multi-channel deep model," *Machine Vision and Applications*, vol. 29, no. 6, pp. 933–946, 2018.

[61] Y. Chen, M. Zhang, W. Li, and Q. Du, "Joint feature extraction for multispectral and panchromatic images based on convolutional neural network," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5005–5008, IEEE, 2018.

[62] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE transactions on neural networks and learning systems*, no. 99, pp. 1–11, 2018.

[63] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1585–1594, 2019.

[64] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network,"

IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 639–643, 2017.

[65] W. Cai, Y. Xu, Z. Wu, H. Liu, L. Qian, and Z. Wei, "Pan-sharpening based on multilevel coupled deep network," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7046–7049, IEEE, 2018.

[66] Y. Xing, M. Wang, S. Yang, and L. Jiao, "Pan-sharpening via deep metric learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 165–183, 2018.

[67] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1795–1799, 2017.

[68] S. Vitale, G. Ferraioli, and G. Scarpa, "A cnn-based model for pansharpening of worldview-3 images," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5108–5111, IEEE, 2018.

[69] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial–spectral fusion by combining deep learning and variational model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 6169–6181, 2019.

[70] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive cnn-based pansharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5443–5457, 2018.

[71] R. Gaetano, D. Ienco, K. Ose, and R. Cresson, "Mrfusion: A deep learning architecture to fuse pan and ms imagery for land cover mapping," *arXiv preprint arXiv:1806.11452*, 2018.

[72] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Information Fusion*, vol. 55, pp. 1–15, 2020.

[73] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 11, no. 5, pp. 1656–1669, 2018.

[74] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, 2017.

[75] X. Mei, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, and J. Ma, "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 8, p. 963, 2019.

[76] Z. Yan, W. Liu, S. Wen, and Y. Yang, "Multi-label image classification by feature attention network," *IEEE Access*, vol. 7, pp. 98005–98013, 2019.

[77] L. Mou and X. X. Zhu, "Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 110–122, 2019.

[78] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7314–7323, 2019.

[79] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing.," in *AAAI*, pp. 11908–11915, 2020.

[80] S. Yin, Y. Wang, and Y.-H. Yang, "A novel image-dehazing network with a parallel attention block," *Pattern Recognition*, vol. 102, p. 107255, 2020.

[81] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, "Attention branch network: Learning of attention mechanism for visual explanation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10705–10714, 2019.

[82] Y. Zhang, S. Nie, W. Liu, X. Xu, D. Zhang, and H. T. Shen, "Sequence-to-sequence domain adaptation network for robust text image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2740–2749, 2019.

[83] L. Huang, W. Wang, J. Chen, and X.-Y. Wei, "Attention on attention for image captioning," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4634–4643, 2019.

[84] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, "Deep residual attention network for spectral image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 0–0, 2018.

[85] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 11065–11074, 2019.

[86] J. Xin, N. Wang, X. Gao, and J. Li, "Residual attribute attention network for face image super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9054–9061, 2019.

[87] Y. Atoum, M. Ye, L. Ren, Y. Tai, and X. Liu, "Color-wise attention network for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 506–507, 2020.

[88] C. Zhang, Q. Yan, Y. Zhu, X. Li, J. Sun, and Y. Zhang, "Attention-based network for low-light image enhancement," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, 2020.

[89] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided cnn for image denoising," *Neural Networks*, vol. 124, pp. 117–129, 2020.

[90] M. Jiang, H. Shen, J. Li, Q. Yuan, and L. Zhang, "A differential information residual convolutional neural network for pansharpening," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 257–271, 2020.

[91] X. Li, F. Xu, X. Lyu, Y. Tong, Z. Chen, S. Li, and D. Liu, "A remote-sensing image pan-sharpening method based on multi-scale channel attention residual network," *IEEE Access*, vol. 8, pp. 27163–27177, 2020.

[92] Q. Yang, Y. Xu, Z. Wu, and Z. Wei, "Hyperspectral and multispectral image fusion based on deep attention network," in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5, IEEE, 2019.

[93] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[94] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[95] L. Ying, G. Yanfeng, and Z. Ye, "Hyperspectral feature extraction using selective pca based on genetic algorithm with subgroups," in *First International Conference on Innovative Computing, Information and Control-Volume I (ICICIC'06)*, vol. 3, pp. 652–656, IEEE, 2006.

[96] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying and Land Information Science*, vol. 62, no. 2, pp. 115–122, 2002.

[97] S. A. Robila, "Independent component analysis," in *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, pp. 109–132, Springer, 2004.

[98] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[99] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016.

[100] J. Wang, H. He, and D. V. Prokhorov, "A folded neural network autoencoder for dimensionality reduction," *Procedia Computer Science*, vol. 13, pp. 120–127, 2012.

[101] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric engineering and remote sensing*, vol. 63, no. 6, pp. 691–699, 1997.

[102] L. F. Zoran, "Quality evaluation of multiresolution remote sensing images fusion," *UPB Sci Bull Series C*, vol. 71, pp. 38–52, 2009.

[103] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Engineering & Remote Sensing*, vol. 74, no. 2, pp. 193–200, 2008.

[104] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, 2007.

[105] X.-l. Zhang, Z.-f. Liu, Y. Kou, J.-b. Dai, and Z.-m. Cheng, "Quality assessment of image fusion based on image content and structural similarity," in *2010 2nd International Conference on Information Engineering and Computer Science*, pp. 1–4, IEEE, 2010.

[106] I. Alimuddin, J. T. S. Sumantyo, H. Kuze, *et al.*, "Assessment of pansharpening methods applied to image fusion of remotely sensed multiband data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 165–175, 2012.

[107] Q. Du, N. H. Younan, R. King, and V. P. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 518–522, 2007.

[108] Y. Yusuf, J. T. Sri Sumantyo, and H. Kuze, "Spectral information analysis of image fusion data for remote sensing applications," *Geocarto international*, vol. 28, no. 4, pp. 291–310, 2013.

[109] V. Naidu, "Discrete cosine transform-based image fusion," *Defence Science Journal*, vol. 60, no. 1, p. 48, 2010.

[110] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pansharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2827–2836, 2013.

[111] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[112] V. Naidu, "Discrete cosine transform-based image fusion," *Defence Science Journal*, vol. 60, no. 1, p. 48, 2010.

[113] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376–2385, 2005.

[114] W. Wang and F. Chang, "A multi-focus image fusion method based on laplacian pyramid.," *JCP*, vol. 6, no. 12, pp. 2559–2566, 2011.

[115] B. Yang, Z.-l. Jing, and H.-t. Zhao, "Review of pixel-level image fusion," *Journal of Shanghai Jiaotong University (Science)*, vol. 15, no. 1, pp. 6–12, 2010.

[116] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 744–756, 2011.

[117] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, vol. 3, pp. III–173, IEEE, 2003.

[118] J. Zeng, A. Sayedelahl, T. Gilmore, and M. Chouikha, "Review of image fusion algorithms for unconstrained outdoor scenes," in *2006 8th international Conference on Signal Processing*, vol. 2, IEEE, 2006.

[119] A. Show and K. Xu, "Tell: Neural image caption generation with visual attention," *Kelvin Xu et. al.. arXiv Pre-Print*, vol. 23, 2015.

[120] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[121] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, 2018.

[122] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[123] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423, 2016.

[124] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[125] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, Springer, 2016.

[126] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "Pannet: A deep network architecture for pan-sharpening," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5449–5457, 2017.

[127] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

[128] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

[129] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.