Deep Learning based technique for fusion of Remote Sensing Images

Project Report

Submitted in fulfillment of the requirements of
BITS F221 Practice School - I
at
Indian Institute of Remote Sensing, Dehradun (IIRS-ISRO)

By

Vaibhav Ganatra & Vidhani Hiten Jitendra
(2019A7PS0010G) & (2019B4A70812H)

Under the supervision of:
Dr. Manish Bhatt (BITS Goa)
&
Dr. Hari Shanker Srivastava (Indian Institute of Remote Sensing, Dehradun)

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI
July 2021

Contents

List of Figures

List of Tables

Abbreviations

PAN	**PAN**chromatic
HS	**H**yper**S**pectral
MS	**M**ulti**S**pectral
CS	**C**omponent **S**ubstitution
MRA	**M**ulti**R**esolution **A**nalysis
PCA	**P**rincipal **C**omponent **A**nalysis
SAR	**S**ynthetic **A**perture **R**adar
UIQI	**U**niversal **I**mage **Q**uality **I**ndex

Chapter 1

Introduction to Remote Sensing

1.1 Remote Sensing

Remote Sensing is the act of obtaining information about objects remotely, i.e. without any physical contact with the object. The data collected can take different forms depending on the type of observations being made: variations in acoustic wave distributions (sonar), variations in force distributions (gravity meter), variations in electromagnetic energy distributions (e.g., eye) etc. A daily example of remote sensing is the act of "seeing" by humans. A remote sensing apparatus essentially consists of: a. An energy source that sends out radiation in the direction of the target region to be observed b. A sensor which senses the radiation reflected by the target c. Data Processing setup that processes the data recorded by the sensor

A schematic of the steps involved in remote sensing is shown in Fig. 1.1.

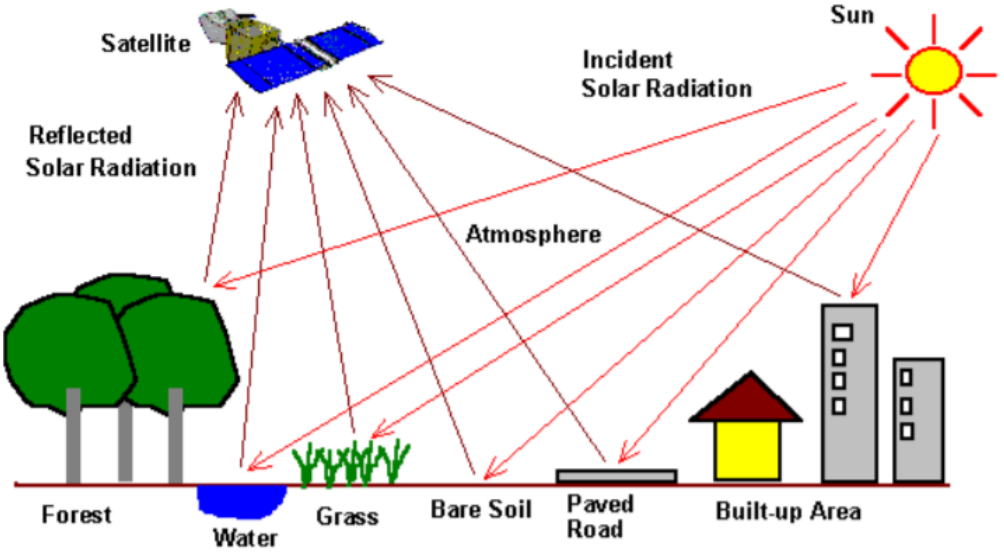


FIGURE 1.1: Basic Steps in Remote Sensing

1.2 Types of Remote Sensing

Remote Sensing can be of different types based on several classification parameters. Here, we specify some of the common classifications of remote sensing.

1.2.1 Based on the Source of Energy:

1. Active Remote Sensing: The energy is sent out from the platform that is sensing the reflected radiation
2. Passive Remote Sensing: The energy radiated by the sun is observed by the sensors on the platform.

1.2.2 Based on the Frequency of Radiation Used

Energy could be observed in different portions of the Electromagnetic Spectrum to record different types of information. Based on which portion of the spectrum is used for sensing, remote sensing can be classified as optical remote sensing (using the visible portion of the spectrum), microwave remote sensing (using microwaves), infrared remote sensing, etc

1.2.3 Based on the height of the observing platform

1. Ground Level Remote Sensing: Sensors are placed very close to the ground
2. Aerial Remote Sensing: Sensors are placed on aerial vehicles
3. Space Shuttle Remote Sensing: Sensors placed on Space Shuttles, Satellites, etc

1.3 Ideal vs Real Remote Sensing Systems

Ideally, a remote sensing system consists of a uniform energy source that sends out energy equally in all wavelengths, a non-interfering atmosphere, a super sensor, that senses all the wavelengths of radiation equally well, etc. However, in real life, such setups are hard to find. Hence, practical remote sensing is limited by factors such as the atmosphere decreasing the signal strength, a sensor sensing some frequencies of radiation better than others, etc. These must be taken into account when designing a remote sensing application to be applied in the real world

Chapter 2

Introduction to Image Fusion in Remote Sensing

2.1 Introduction

The goal of remote sensing is to acquire information about different locations on the surface of the earth without physically coming in contact with it. This is done by capturing electromagnetic radiation scattered by the region through sensors located at multiple locations (ground/aerial/space). The accuracy of the information retrieved from the recorded data is highest when the spectral, spatial and radiometric resolutions of the sensed images are high. A high spectral resolution helps in class differentiation of the surface, whereas a higher space resolution conveys physical information about the shape, and structure of the Earth's surface. However, practically, sensors can resolve only one of these to such an extent that accurate conclusions may be drawn. It is often not possible to resolve all the above simultaneously in an image and different sensors record complementary information about the Earth surface. Hence, there is a need for "fusion" of images containing this complementary information into an image that has high spectral, spatial and radiometric resolutions, which improves the accuracy of the information extracted from it. Fusion of images also makes it convenient to share the recorded information for further research.

2.2 Need for fusion of Images

Through the improvement in the technology of sensors, it is now possible to record scattered radiation from multiple regions of the electromagnetic spectrum. Consequently, we have multiple types of images, Panchromatic (PAN) Images, Multispectral (MS) and Hyperspectral (HS) Images, Microwave (SAR) Images,etc. Each of these have their characteristic spectral and spatial

resolutions. For example, PAN images have high spatial resolution, but have poor spectral resolution. On the other hand, Multispectral and Hyperspectral Images have higher spectral resolution, but lack details about the physical structure of the area. Hence, multiple techniques have been experimented with, to incorporate the spatial data of the PAN image into the MS Image. This procedure is called PANSharpening of the MS Image. The outcome of this process is a fused image, that is rich in spatial as well as spectral information about the area. Hence, any further processing done on the fused image (like Land Cover Classification) yields a higher accuracy than when done on any of the source images.[Ghassemian, 2016] Hence, fusion of source images containing complementary data is vital to improving the utility of the remote sensing process.

Recently, techniques have been suggested for the fusion of SAR-Images with MS Images. SAR-Images record the backscattered energy (which is dependent on the surface roughness, moisture content, dielectric properties, etc) in the microwave region of the spectrum by the Synthetic Aperture Radar (SAR) Sensor. These images have high spatial resolution (like PAN Images) and are available irrespective of environmental conditions such as cloud cover, haze, dust, etc (which is not the case with PAN Images, which are sensed in the visible region). Hence, fusion of MS Images with SAR Images boosts the accuracy of further processes even more when carried out on the fused images. It must be noted here that any sort of fusion of optical and SAR-images requires pre-processing of the image, where they are georeferenced and co-registered.

2.3 Types of Fusion Techniques

2.3.1 Pixel-Level Fusion Techniques

2.3.1.1 Component Substitution (CS)

In the Component Substitution method, the low spatial resolution MS Image is projected onto a different space such that the spectral and the spatial information are separated. Next, the SAR Image (which has high spatial information) replaces the spatial component of the projected MS image. This technique works best when there is a high correlation between the spatial component of the MS Image and the SAR Image. Hence, histogram matching might need to be conducted between the SAR Image and the spatial component of the projected MS image.

There are multiple techniques through which the separation of the spectral and spatial components might be done. Example: IHS Transform, Principal Component Analysis transform, Brovery transform, etc Transformation of the MS image into IHS form, separates the spectral properties into the Hue and the Saturation component, whereas the spatial component is largely separated into the Intensity component. Hence, after performing histogram-matching between the Intensity

component and the SAR-Image, the SAR-Image replaces the intensity component. Similarly, in the PCA transform, the spectral information is distributed between multiple components, whereas the spatial information is largely contained in the first principal component, which is then replaced by the SAR-Image.[Kulkarni and Rege, 2020]

2.3.1.2 Multiresolution Analysis (MRA)

The steps involved in this method are, firstly, with the help of wavelet/pyramid transform we decompose the input images into several layers in decreasing level of resolution(image with the highest resolution is at the bottom). Secondly, all the layers corresponding to the first image type are fused with the second image type. Thirdly, we apply the inverse transform to merge all the layers in the pyramid, to obtain the fused image. This method requires more time and space complexity than the Component Substitution method, as we do the similar process of the CS method in several layers of the pyramid. But, this method has been found to perform better than the CS method. There are several pyramids that can be used for decomposition such as Gaussian pyramid, Laplacian pyramid, etc. The other transforms which can be used in the place of wavelet transforms are curvelet and contourlet transforms. The major advantage of the wavelet transform is that, in the subsequent layers, resolution decreases but the size of the image remains the same.

2.3.1.3 Hybrid Methods

These methods combine both CS and MRA methods to use the advantages of both of them. A study on fusion of SAR and optical images tells us that IHS based fusion preserves the spatial details but the disadvantage is that it gives rise to spectral distortion. To solve this problem of spectral distortion we can use the IHS transform combined with non-subsampled wavelet transform, which is also known as "à Trous wavelet i.e. AWT" and empirical mode decomposition(EMD).

2.3.1.4 Model based Methods

Two major models come under model based methods and they are, variational models and sparse representation-based models. In the sparse representation-based model the process of image fusion is considered as image restoration, so now the aim becomes to generate high resolution fused images. Variational models aim to minimize the energy function which helps in fusion of color information(from optical image) and spatial information(from SAR image).

2.3.2 Feature Level Fusion Techniques

To carry out the feature level fusion process, first, we extract the features from the input images using image segmentation techniques. The features are shape, size, edges. Similar features from various sources are classified and fused for further processing. Basically, the features arising in multiple input images are merged into a single feature set. The major advantage of this technique over pixel level fusion techniques(for fusion of MS and PAN images) is that it doesn't require resampling of images. This can be performed for any given aspect ratio of the pixels of PAN and MS images.

2.3.3 Decision Level Fusion Techniques

This technique involves highest data processing. In this technique, initially each source is classified with the help of feature level technique, the final output is based on the combined decision of several classifiers. It is the process of fusing the decision by selecting a hypothesis from M hypotheses, given the decision of N sensors.

2.4 CNN-Based Pixel Level Fusion Technique

Recently, there has been a rise of deep-learning based methods for image fusion. Here, we review a method that utilises convolutional neural networks (CNNs) for fusing Sentinel-1 and Sentinel-2 images at the pixel level.[Shakya et al., 2020] The paper uses 3 different methods in which CNNs can be used in generating a weight/feature map, which can then be decomposed by a pyramid as is done in Multi-Resolution Analysis.

The 3 methods were IVF, MEDF and MFF methods. IVF (Infrared and visible Image fusion) method is essentially a 4 step process:

1. Generation of a CNN-based feature map by applying CNN-kernels on the input image,which summarizes the features of the input image.
2. Decomposition of the feature map via Gaussian pyramid (as mentioned in the MRA method of image fusion
3. Calculation of the fusion coefficients/local energy maps and fusion via the fusion strategy.
4. Reconstruction of the fused image using the Laplacian pyramid.

MEDF, originally proposed for medical image fusion, uses multi-scale processing and adaptive fusion using images acquired through Computed Tomography (CT) and Magnetic Resonance

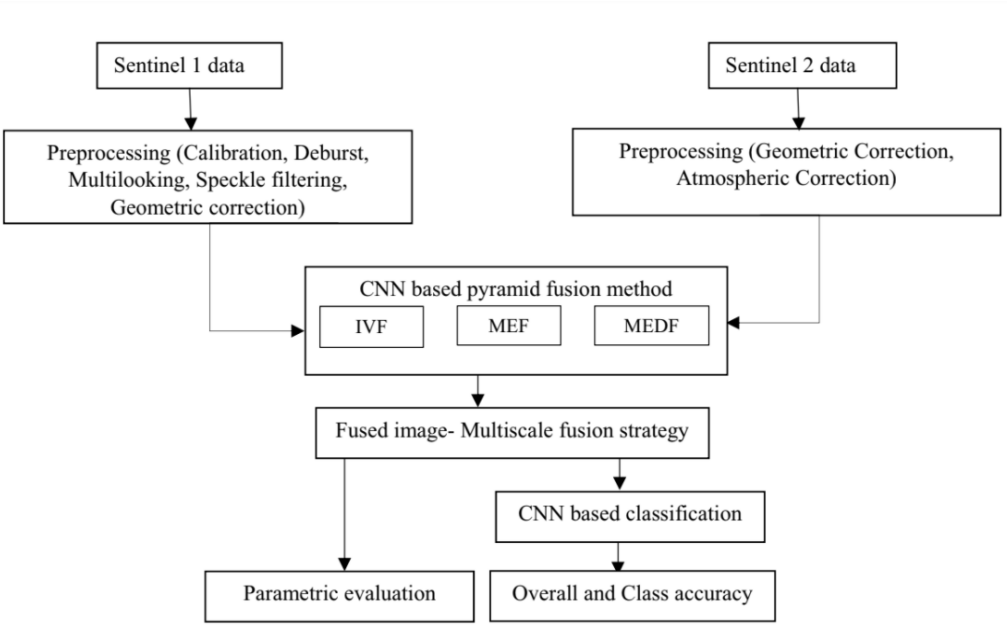


FIGURE 2.1: Steps followed in the fusion of images and then using the fused images for Land Cover Classification

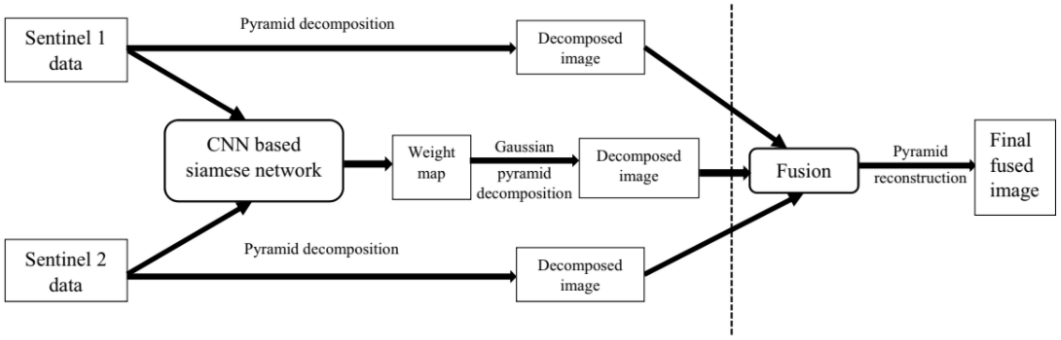


FIGURE 2.2: Details of the CNN-Based Fusion technique

(MR) imaging process. This approach generates a weight map for integrating the pixel information from two input images so as to achieve better results with the fused image. MFF is a multi-focus image fusion approach employing deep siamese CNN for direct mapping between source images and focus map using dataset acquired from ImageNet dataset. The network is trained using patches extracted from an image and blurred images for encoding to achieve better fusion performance in terms of both visual quality and the objective analysis. The above mentioned methods were used for generating a weight map which could then be decomposed by a Gaussian pyramid. The original images were also subjected to pyramid decompositions, and the 3 were fused and reconstructed using the Laplacian pyramid. After completion of fusion, the fused image was evaluated using the various evaluation metrics. It was also used for land cover classification using a 2-D CNN. The performance of the fused image was then contrasted with the performance of the individual images.

Chapter 3

Deep Learning based Image Fusion

3.1 Introduction

The source images are decomposed into their base parts and detail content by the image decomposition approach explained in Section 3.2. Then we fuse the base parts of both the image using weighted-averaging strategy to obtain the fused base part (F_b). As the detail content contains features which we want to preserve in our final image, we use deep learning network[Li et al., 2018] to obtain the fused detail content(F_d). More details about obtaining F_d are explained in Section 3.3.2. The final fused image is then obtained by fusing F_b and F_d. It must be noted here, that this technique was initially used for fusing visible and infrared images. We extend this technique to fusing optical and SAR images. Fig. 3.1 outlines the basic steps to be followed in the fusion process.

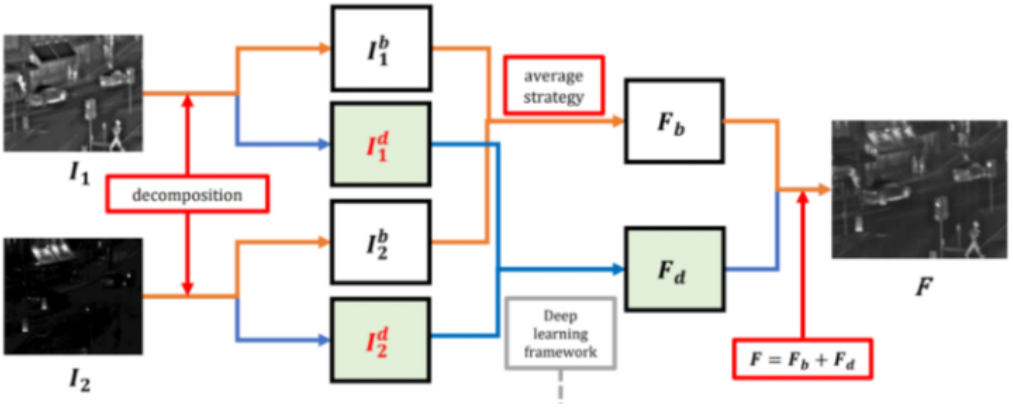


FIGURE 3.1: Steps followed in the fusion of the source images

3.2 Image Decomposition

The two source images are decomposed into a base part and a detail content via a guided filtering process[Li et al., 2013]. The process aims at separating each source image into a base layer containing the large-scale variations in intensity and a detail layer containing the small scale details. The base for the images part is obtained by solving the optimization problem

$$I_k^b = \underset{I_k^b}{argmin} ||I_k - I_k^b||_F^2 + \lambda(||g_x * I_k^b||_F^2 + ||g_y * I_k^b||_F^2). \tag{3.1}$$

Here, $g_x = [-1 \quad 1]$ and $g_y = [-1 \quad 1]^T$ are the horizontal and vertical gradient operators, respectively. The parameter λ can be varied according to our need. We use $\lambda = 5$ for obtaining our results. After getting the base part we obtain the detail content by the following equation:

$$I_k^d = I - I_k^b \tag{3.2}$$

3.3 Image Fusion

Here, we describe the steps to obtain the final fused image starting from the components of the decomposed image as obtained in Sec. 3.2

3.3.1 Fusion of Base Component

The base component I_k^b for k=1,2 is fused using a weighted average method such that

$$I_{fused}^b = \alpha_1 I_1^b + \alpha_2 I_2^b. \tag{3.3}$$

where I_{fused}^b is the base-component of the final fused image, which when added with I_{fused}^d , gives the final fused image

We use $\alpha_1 = \alpha_2 = 0.5$.. In the Results section, we have shown the results of using various different α_is

3.3.2 Fusion of Detail Content

In order to fuse the detail-content, we make use of the VGG-19 network. The overall outline of the fusion procedure is shown in Fig. 3.2
The VGG-19 architecture has 16 convolution layers and 3 fully-connected layers. As is the case with many deep CNNs, the convolutional layers are interspersed with max-pooling layers. The

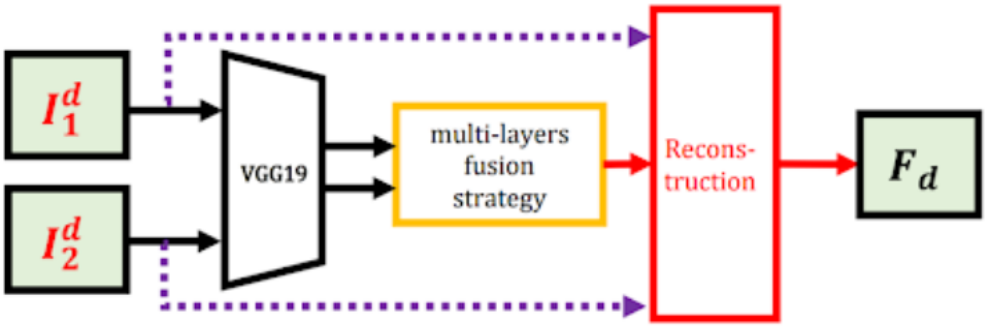


FIGURE 3.2: Steps for obtaining F_d

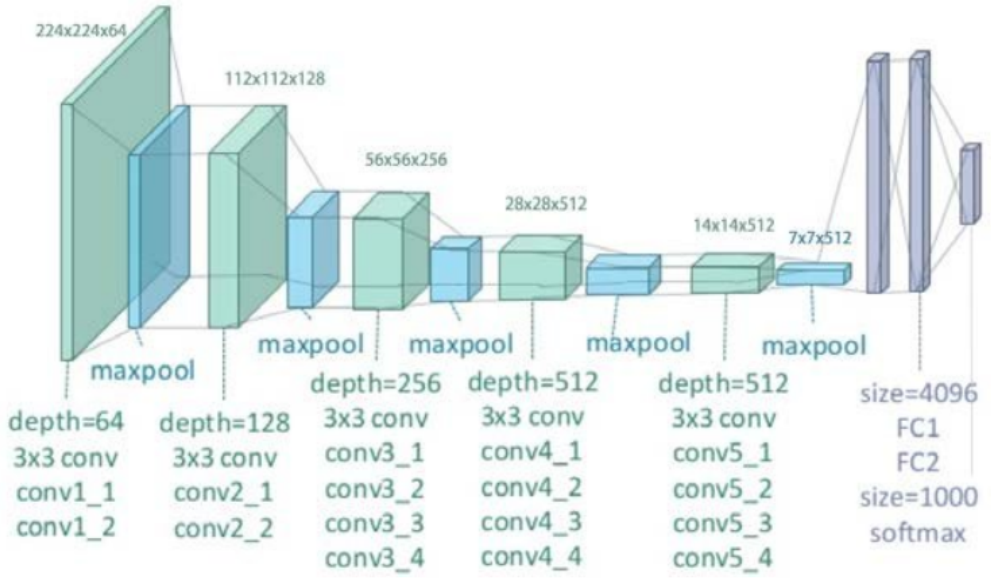


FIGURE 3.3: VGG-19 architecture

VGG-19 consists of max-pooling layers. The VGG-19 architecture is shown in Fig. 3.3
It is commonly known that the filters of the VGG-19 learn parameters that detect different features in the source image. In order to fuse the detail-content of both the source images into a single fused detail-content, we will make use of these features learnt by the filters. We extract the feature-maps from the VGG-19 just before the first 4 max-pooling layers. This gives us 4 pairs of 3-D feature volumes which we will now process in order to produce the final fused detail-content. The steps of processing the feature-volumes extracted from the VGG:19 are as follows:

(a) Flattening the feature-volumes in 2-dimensional feature maps

(b) Calculation of the activity level maps

(c) Upscaling the activity level maps to the size of the source image

(d) Calculation of the initial fusion map for each pair of feature-maps

(e) Calculation of the final fused detail-content from the initial fusion maps.

3.3.2.1 Flattening the feature-volumes in 2-dimensional feature maps

The feature volumes extracted from the VGG:19 network now need to be flattened into 2-Dimensional feature maps. This is done by taking the L1-Norm of the features across the 3rd dimension.
Given an input volume of size (a x a x b), we can flatten it into an (a x a) feature map, by the above described technique.

$$\hat{C}_k^i(x,y) = \sum_{z=0}^{z=b} |C(x,y,z)| \tag{3.4}$$

where C represents the input feature-volume of size ($axaxb$), and $\hat{C}$ represents the output flattened 2-D feature map.
This flattening is done for each of the 4 pairs of feature-volumes extracted from the VGG-19 network. The output of the flattening process is 4 pairs of 2-D feature maps of various dimensions.

3.3.2.2 Calculation of the activity level maps

For each pair of flattened feature maps, the values are normalized to [0,1] by calculating the activity level maps. This is done by calculating the proportion of each feature-map (corresponding to each source image) in the pair.

$$W_k^i(x,y) = \frac{\hat{C}_k^i(x,y)}{\sum_{k=1}^{2} \hat{C}_k^i(x,y)} \tag{3.5}$$

The output of this step is the activity level map for each of the 4 pairs of input feature-maps.

3.3.2.3 Upscaling the activity level maps to the size of the source image

The activity-level maps obtained from the previous step are of different sizes corresponding to the output shape obtained from the VGG-19 network. These are scaled to the size of the input image, by using the upscaling operator:

$$\hat{W}_k^i(x+p, y+q) = W_k^i(x,y) \tag{3.6}$$

where $p, q = 0, 1, 2...2^{i-1}$ and 2^{i-1} is the scaling factor The output of this step is 4-pairs of activity-level maps of the same size as the input image.

3.3.2.4 Calculation of the initial fusion map for each pair of feature-maps

With the upscaled activity level maps, we are ready to obtain the initial fusion map for each of the 4-feature maps extracted from the VGG-19. The activity-leveL maps essentially denote the level of contribution of their corresponding source image in the features denoted by the feature-map. Hence, the initial fusion map is calculated by weighted addition of the source images, with the activity-level being the corresponding weights.

$$F^i_{fused}(x,y) = \sum_{k=1}^{2} \hat{W}^i_k(x,y) * I^d_k \tag{3.7}$$

3.3.2.5 Calculation of the final fused detail-content from the initial fusion maps

The final fused detail-content of the fused image is calculated using the 4 initial fusion maps generated in the previous step by using a max-selection strategy.

$$F^d_{fused} = max(F^i_{fused}) \tag{3.8}$$

3.3.3 Obtaining the Final Fused Image

From the above steps, the final fused detail content Fdfused has been generated. The fused base content Fbfused had been generated using the weighted average technique described earlier. Finally the fused image is calculated by adding the 2 parts together:

$$F_{fused} = F^b_{fused} + F^d_{fused} \tag{3.9}$$

3.3.4 Fusion of optical(RGB) and SAR images using the above technique

The technique described above is applicable to fuse two different 2-D images into a fused 2-D image. We extend the technique to fuse multi-channel images. We treat the data in separate channels as an independent 2-D image and fuse that channel data with the SAR image. Once all the individual channels have been merged, we reconstruct the multi-channel fused image using the "fused-channels".

We then assess the quality of the fused images using the Universal Image Quality Index. Here, image distortion is modeled as a combination of three factors namely, loss of correlation, luminance distortion and contrast distortion. Its value ranges between -1 to 1. It is desirable to have this value as high as possible. It is given by the following equation

$$Q = \frac{4\sigma_{rf}\bar{r}\bar{f}}{(\sigma_r^2 + \sigma_f^2)(\bar{r}^2 + \bar{f}^2)} \tag{3.10}$$

where r and f are reference and fused images, respectively. σrf is covariance between the reference and fused image and $\hat{r}$ an $\hat{f}$ are mean values of the reference and the fused image respectively. Initially, local Q index is calculated over a 16 $\times$ 16 window and finally all local Q indices are averaged to get a global index.

Chapter 4

Results from Image Fusion

4.1 Results

In this chapter, we demonstrate the results obtained from the application of the Image-Fusion technique described in Chapter-3. We first show a fused image obtained by fusing an optical image with infrared image. Next, we present images obtained by fusing optical and SAR-images. These images are obtained from the QXS-SAROPT dataset.[Huang et al., 2021] We evaluate the fusion accuracy through the Universal Image Quality Index (UIQI). Table. 4.1 shows the UIQI values for the different fused images that we have obtained.

Image	UIQI with Optical Reference	UIQI with SAR Reference
Fig. 4.2(c)	0.8	0.45
Fig. 4.3(a)	0.95	0.343
Fig. 4.3(c)	0.56	0.623
Fig. 4.4(c)	0.92	0.74
Fig. 4.5(c)	0.944	0.29

TABLE 4.1: Evaluation metric - UIQI values for the subsequent fused images

4.2 Conclusion

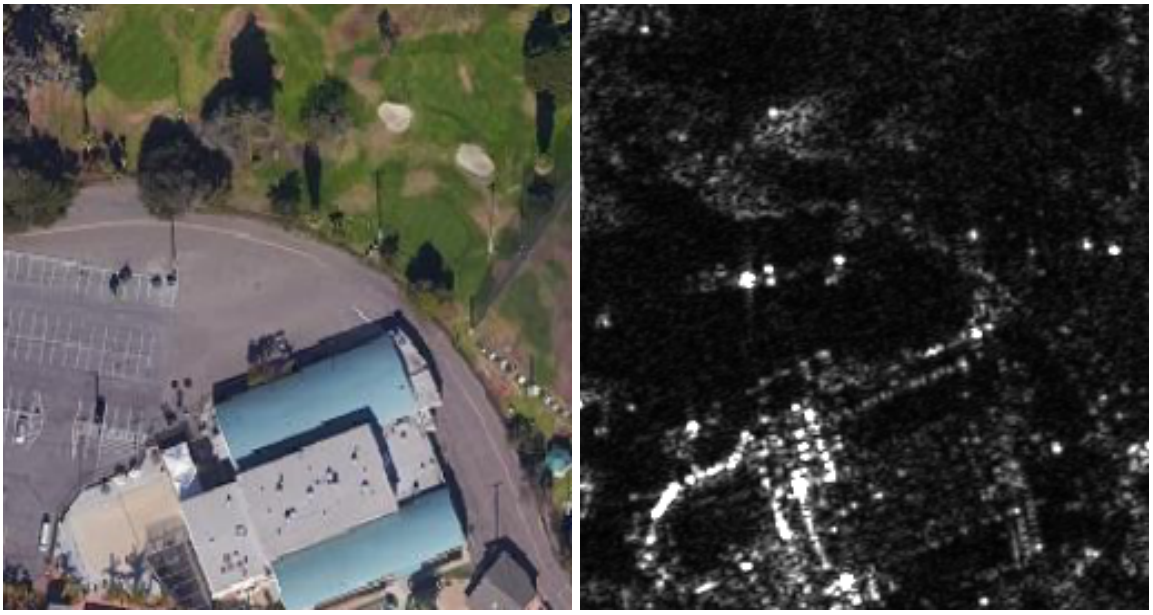
In this project, we have extended the method initially devised for fusion of infrared and visible images to fuse visible and SAR-images. The evaluation metric used here, UIQI, shows the accuracy of the fusion process. Varying different parameters of fusion leads to different values of the metric. However, the usefulness of the fused images needs to be evaluated through other metrics of image fusion and the use of the fused images for further applications such as Land Cover Classification, etc.

(a)

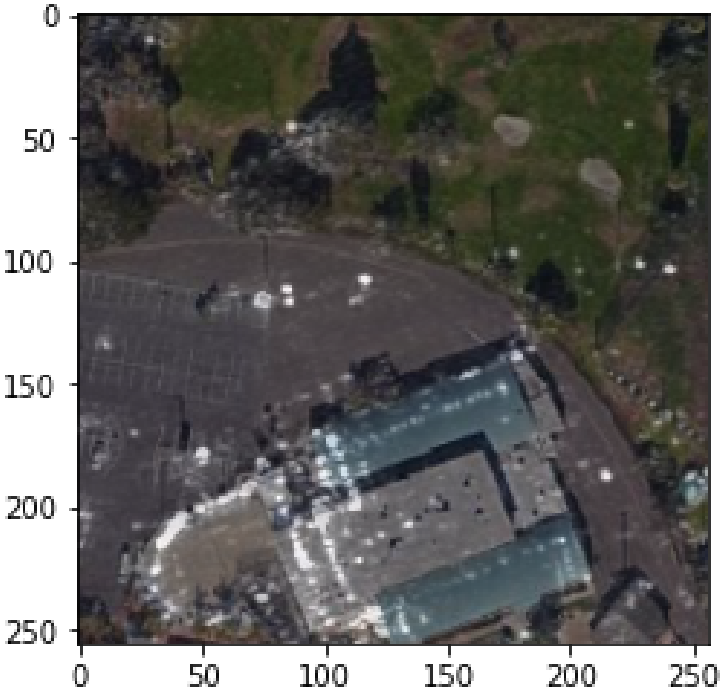
(b)

(c)

FIGURE 4.1: Fused Image (c) obtained upon fusing the source visible image (a) with the source infrared image (b)

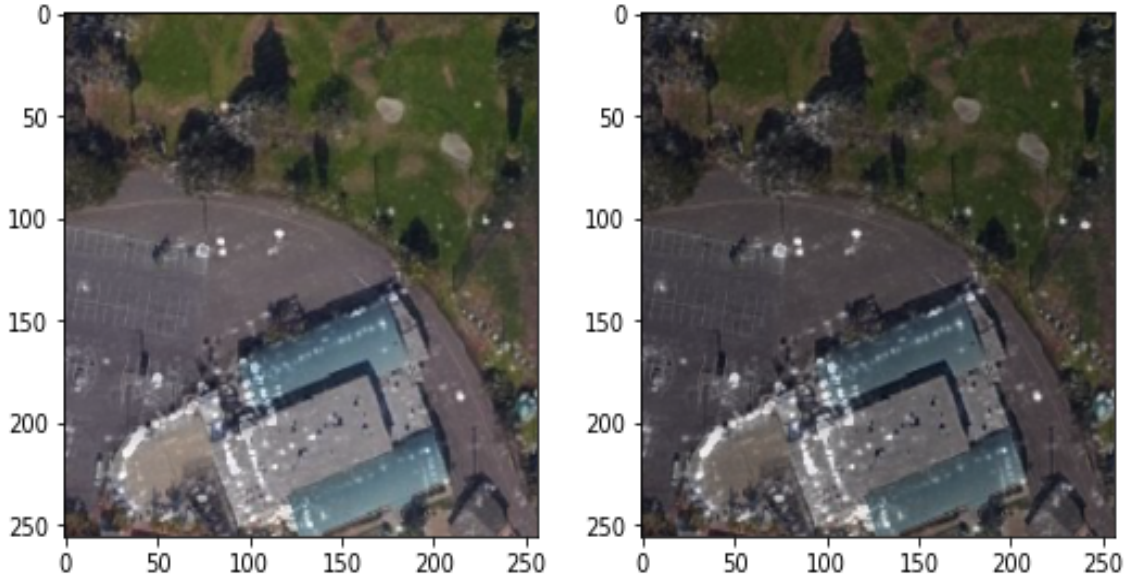


(a) (b)



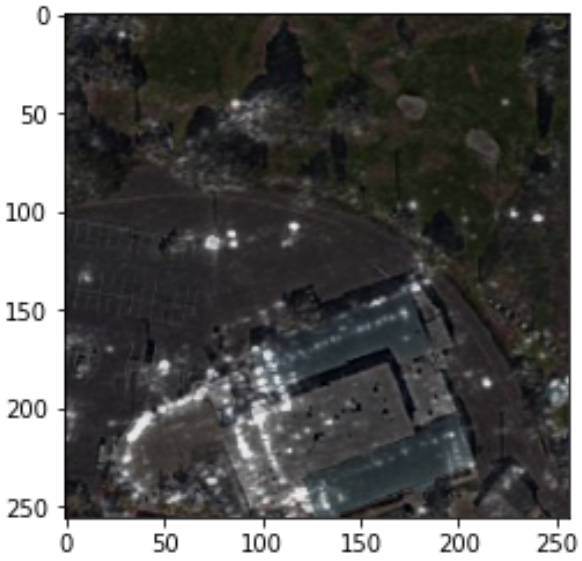
(c)

FIGURE 4.2: Fused Image (c) obtained by fusing the Source Optical Image(a) with the Source SAR Image(b)



(a)

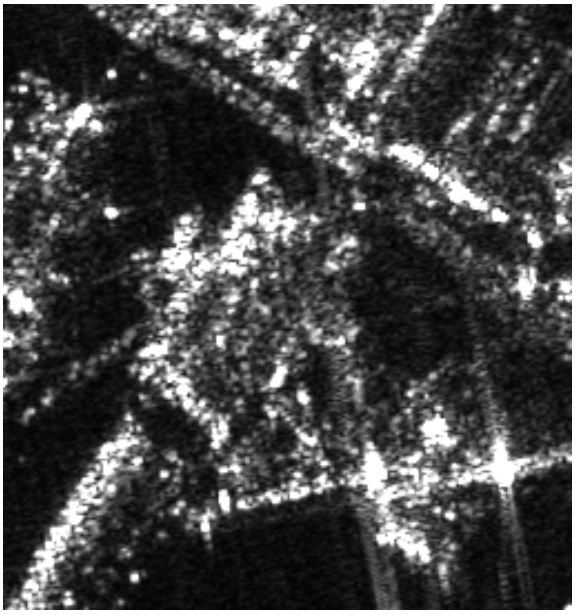
(b)



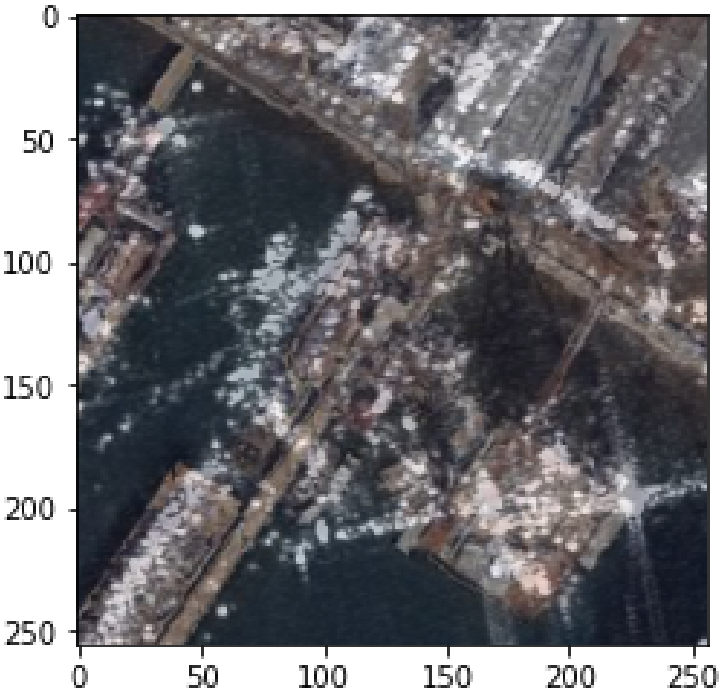
(c)

FIGURE 4.3: Fused Images (a,b,c) obtained by variation of α_1 and α_2
(a) $\alpha_1 = 0.75, \alpha_2 = 0.25$
(b) $\alpha_1 = 0.5, \alpha_2 = 0.5$
(c) $\alpha_1 = 0.25, \alpha_2 = 0.75$

(a)

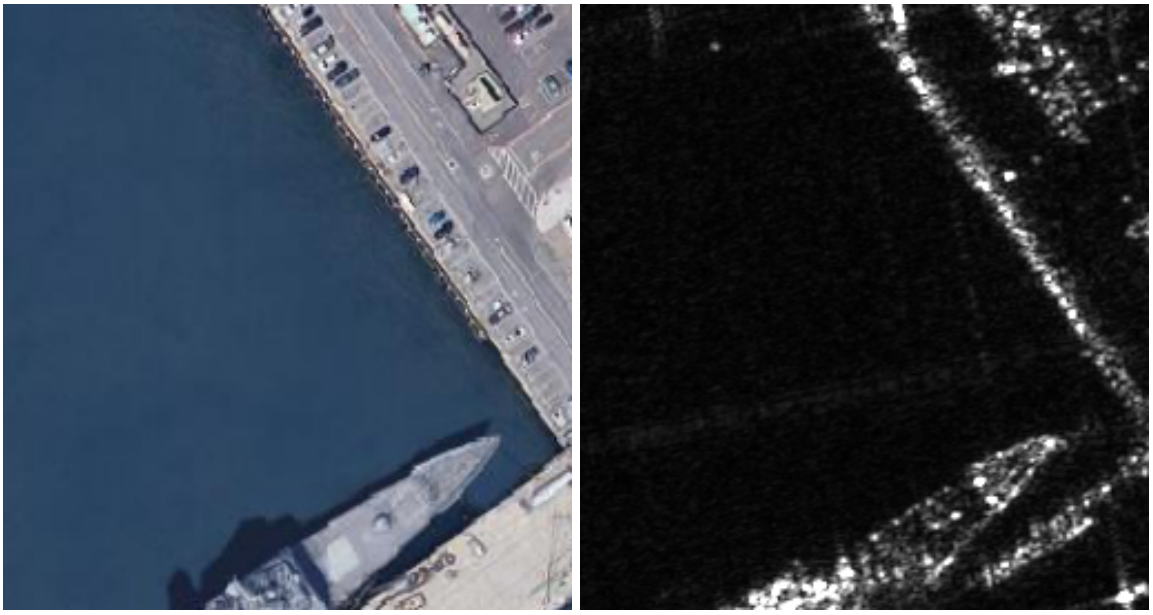


(b)

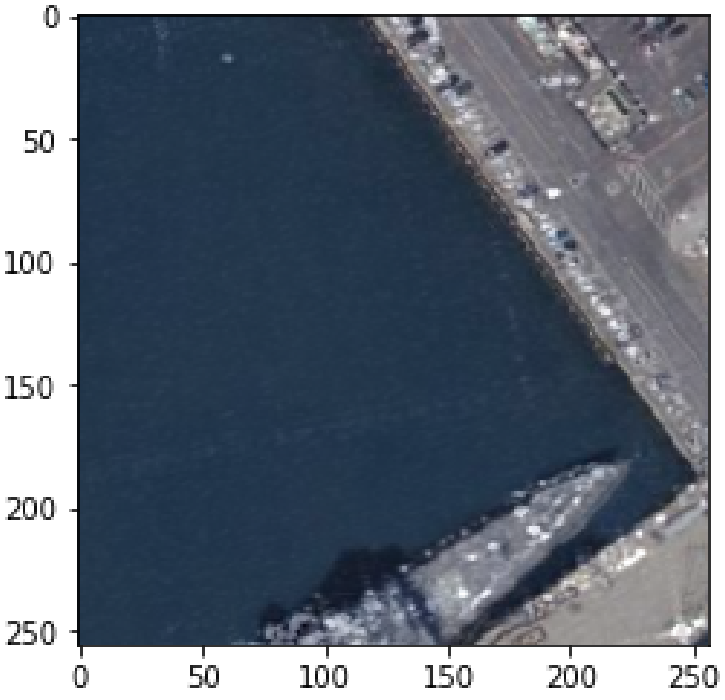


(c)

FIGURE 4.4: Fused Image (c) obtained by fusing the Source Optical Image(a) with the Source SAR Image(b)



(a) (b)



(c)

FIGURE 4.5: Fused Image (c) obtained by fusing the Source Optical Image(a) with the Source SAR Image(b)

References

[Ghassemian, 2016] Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32:75–89.

[Huang et al., 2021] Huang, M., Xu, Y., Qian, L., Shi, W., Zhang, Y., Bao, W., Wang, N., Liu, X., and Xiang, X. (2021). The qxs-saropt dataset for deep learning in sar-optical data fusion.

[Kulkarni and Rege, 2020] Kulkarni, S. C. and Rege, P. P. (2020). Pixel level fusion techniques for sar and optical images: A review. *Information Fusion*, 59:13–29.

[Li et al., 2018] Li, H., Wu, X.-J., and Kittler, J. (2018). Infrared and visible image fusion using a deep learning framework. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 2705–2710.

[Li et al., 2013] Li, S., Kang, X., and Hu, J. (2013). Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875.

[Shakya et al., 2020] Shakya, A., Biswas, M., and Pal, M. (2020). Cnn-based fusion and classification of sar and optical data. *International Journal of Remote Sensing*, 41(22):8839–8861.