

FUSING PANCHROMATIC AND SWIR BANDS BASED ON CNN – A PRELIMINARY STUDY OVER WORLDVIEW-3 DATASETS

Mengxiao Guo¹, Haiyi Ma¹, Ying Bao¹, Leiguang Wang^{2*}

¹ School of Forestry, Southwest Forestry University, Kunming, China, 650224

² Institutes of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming, 650024, China – wlgain@126.com

KEY WORDS: Pan-sharpening, Convolutional Neural Network, Short-wave Infrared, Deep Learning, Image fusion, Remote Sensing

ABSTRACT: The traditional fusion methods are based on the fact that the spectral ranges of the Panchromatic (PAN) and multispectral bands (MS) are almost overlapping. In this paper, we propose a new pan-sharpening method for the fusion of PAN and SWIR (short-wave infrared) bands, whose spectral coverages are not overlapping. This problem is addressed with a convolutional neural network (CNN), which is trained by WorldView-3 dataset. CNN can learn the complex relationship among bands, and thus alleviate spectral distortion. Consequently, in our network, we use the simple three-layer basic architecture with 16×16 kernels to conduct the experiment. Every layer use different receptive field. The first two layers compute 512 feature maps by using the 16×16 and 1×1 receptive field respectively and the third layer with a 8×8 receptive field. The fusion results are optimized by continuous training. As for assessment, four evaluation indexes including Entropy, CC, SAM and UIQI are selected built on subjective visual effect and quantitative evaluation. The preliminary experimental results demonstrate that the fusion algorithms can effectively enhance the spatial information. Unfortunately, the fusion image has spectral distortion, it cannot maintain the spectral information of the SWIR image.

1. INTRODUCTION

Remotely-sensed images have exhibited explosive growth trends in multi-sensor, multi-temporal, and multi-resolution characteristics (Bin, 2017). Duo to technical limitations (Liang, 2004) of multi-sensor technology, remote sensing sensors cannot obtain images with both high spatial and high spectral properties. Remote sensing image fusion, also known as image pan-sharpening, becomes a crucial procedure to combine the features of multi-sensors.

A large collection of literature has explored all kinds of fusion methods. These methods can be roughly divided into two families, i.e., the component substitution (CS) (Shettigara, 1992) and the multi-resolution analysis (MRA) (Yocky, 1996). The first family relies on the substitution of a component with the PAN image. This class includes intensity-hue-saturation (IHS) (Carper, 1990), principal component analysis (PCA) (Shah, 2008), Gram-Schmidt (GS) spectral sharpening methods (Laben, 2000), etc. The MRA approach extracts multi-scale spatial

details through filtering or decomposing the Panchromatic image, and then injects these details into the resampled MS bands (Vivone, 2015). This class includes high-pass filtering (HPF) (Psjr, 1991) based technique, smoothing filter-based intensity modulation (SFIM) (Schowengerdt, 1997) method, decimated wavelet transform (DWT) (Mallat, 1989), and Laplacian pyramid (LP) method (Burt, 2003).

These methods assume that the spectral coverage of panchromatic (Pan) band is overlapped with that of spectral bands, and the spectral bands are highly related to the panchromatic band. However, many newly-launched satellites, such as worldview-3, besides some typical green, blue, red and near infrared (NIR) bands, also can provide some short-wave infrared bands (SWIRs) (Pacifi, 2016). The spatial resolution of these SWIRs is lower than the Pan bands, meanwhile the spectral ranges are not overlapped by that of the Pan band. Therefore, the assumption that holds by state-of-the-arts methods is false under this situation. How to combine complementary information from the panchromatic and SWIR bands become the key issue for fusing Pan and SWIR band.

2. PROPOSED PANSARHPENING METHOD

2.1 Convolutional Neural Network (CNN)

Traditional methods are often suffers from high spectral and spatial distortion, because Pan and SWIR components are not acquired in overlapped spectral ranges. In contrast, Convolutional Neural Network (CNN), a breakthrough method in the machine learning field, is a promising solution to the problem due to the following properties. Firstly, CNN can learn features automatically by using the multi-layer nonlinear mapping. The model can learn the optimal deep support value filter at various levels. Secondly, the multi-scale support value filter can extract the high frequency information of the images at all levels effectively. The high frequency information of the Pan image is injected into the SWIR, which avoids the space and spectral distortion simultaneously. Finally, via two or more

deep learning networks, the network can find the ideal solution in the training process quickly. Figure 1 shows a basic flowchart describing fusion process of the CNN. The input of the CNN model is the SWIR and PAN band contents. The output is a SWIR image with the same spatial resolution as the PAN. We use the simple three-layer basic architecture with 16×16 kernels. The parameters and weight adjustment of the network need to be trained with the desired high resolution SWIR image as the reference data. Unfortunately, such reference data does not exist. Therefore, before entering the network, we create a down-sampled reference data following the Wald protocol (Wald 2009). In detail, the SWIR image at the target panchromatic resolution is employed as reference, original SWIR and PAN image is employed as input. The space and spectral features are extracted by the operation of multiple convolution layers for image reconstruction. The output of the network is the final fusion result.

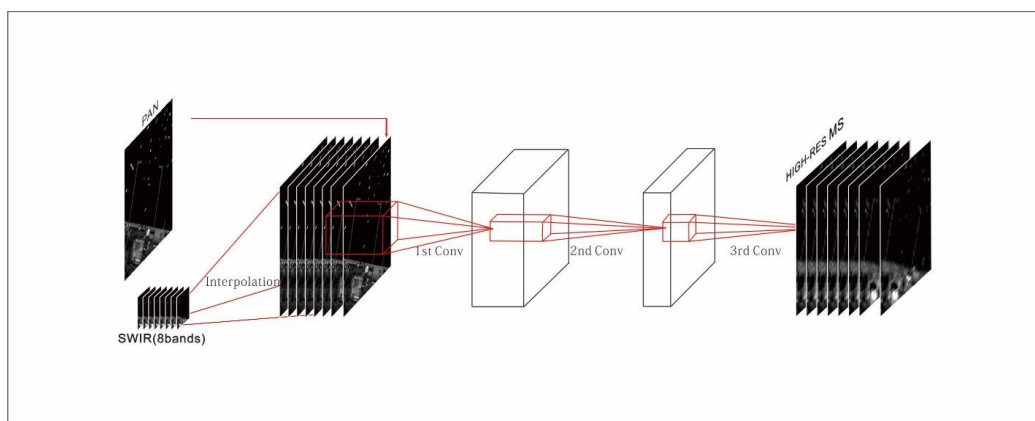


Figure 1. The CNN structure of for pansharpening

Figure 2 shows the training process based on the Wald training protocol. In the training process, the input of the CNN is down-sampled PAN and SWIR images (by ratio 4), i.e., training data. And the output of the network is original SWIR component, i.e., reference data. Hence, the first layer using a 16×16 receptive field and a ReLU nonlinearity computes 512 feature maps. The second step also computes 512 feature maps but using a 1×1 receptive field and ReLU. In the end, the third

layer, with a 8×8 receptive field and a identity activation function, provides the hypothetical pansharpened SWIR image. During learning, it is based on the stochastic gradient descent of backpropagation. As updating iteration, it refers to a set of 10 input tiles selected at random from the training data. We use the mean square error to represent the loss function. Moreover, in the test procedure, the input of the network is original PAN and resampled SWIR image.

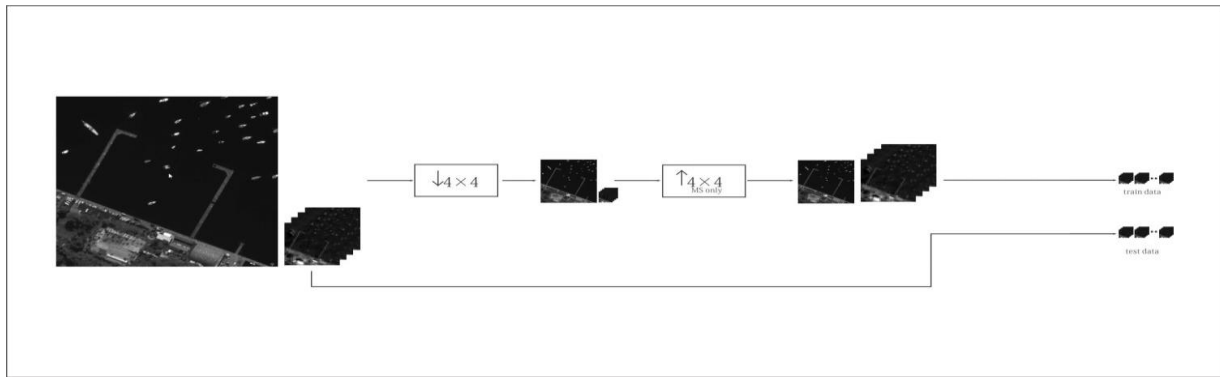


Figure 2. Training process based on the Wald protocol

2.2 Proposed Pan-Sharpening Method

In summary, as shown in Figure 3, our experiment consists of the following 5 steps:

- (1) Preprocessing. Before the fusion, the PAN and SWIR images are firstly orthographical correction to eliminate geometric distortion;
- (2) Generate low resolution images pairs. The Pan and SWIR images are down-sampled simultaneously. The pixel size of the Pan is the size of the original SWIR, SWIR image pixel size reduced with the same proportion;
- (3) Registration. The SWIR image is up-sampled and interpolated to have the same pixel size as the panchromatic image. Next, compare of various interpolation methods of

fusion results to select the appropriate interpolation method. Eventually, we up-sample the data via cubic convolution method;

- (4) Training. All bands of down-sampled low resolution images (PAN and SWIR) are used as the input to the convolutional neural network, and the original SWIR image as an output to train the network and correct the weight;
- (5) Fusion. The trained network model is applied to the original image data to finally obtain the SWIR image with panchromatic image spatial resolution.

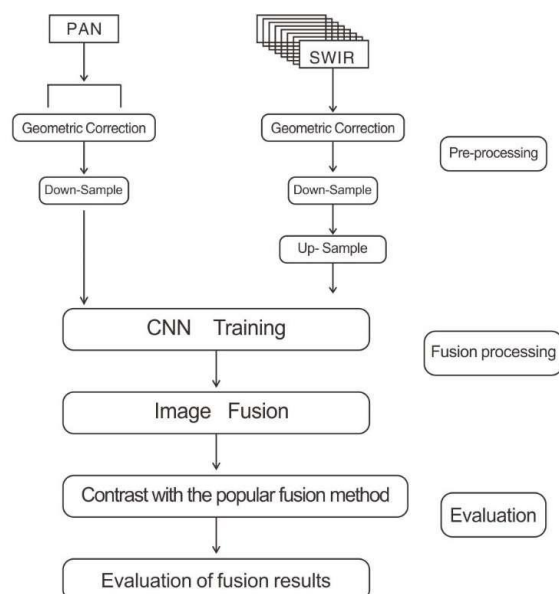


Figure 3. The processing flowchart of the proposed pan-sharpening method.

3. EXPERIMENTAL RESULTS AND COMPARISONS

3.1 Experimental Setting

To illustrate the effectiveness of the method, two datasets acquired by WorldView-3 sensors were considered. The test images were acquired over Rio de Janeiro, Brazil region. The resolution of PAN in this satellite ranges from 450nm to 800nm. And the resolution cell is $1.2\text{m} \times 1.2\text{m}$ with 1000×1000 pixels for the PAN. The datasets also comprise of 8 SWIR bands in the wavelength range from 1210nm to 2330nm with spatial resolution 7.5m and contains 1000×1000 pixels. The test scene includes building, water body and vegetation. We compare with three widely used pan-sharpening method: ATWT, IHS, PCA.

Spatial resolution	SWIR:7.5m
	PAN:1.2m
Spectral range	SWIR:1210-2330nm
	Blue:450-510nm
	Green:510-580nm
	Red:630-690nm
Land cover types	Building, water body, vegetation
Locations of images	Rio de Janeiro, Brazil

Table 1. Characteristics of the employed WV-3 datasets

3.2 Assessment Metrics

We use four widely used quantitative assessment indexes to evaluate performance: the entropy, the correlation coefficient (CC) (Soiva,2007) , the spectral angle mapper (SAM) (Yuhas,1992), the universal image quality index (UIQI) (Wang,2002).

(1) The entropy is an important index to measure the information richness of images. The entropy value reflects the average amount of information contained in the image. It is defined as:

$$H = - \sum_{i=0}^{255} p_i \log_2 p_i \quad (1)$$

where *Entropy* is the entropy of image, and p_i is the probability of pixel value i in the image.

(2) CC (Soiva, 2007) measures the correlation between the SWIR image and the fused image. Its definition is given by:

$$CC = \frac{\sum_{i=1}^m \sum_{j=1}^n [M(i,j) - \bar{M}][F(i,j) - \bar{F}]}{\sqrt{\{\sum_{i=1}^m \sum_{j=1}^n M(i,j) - \bar{M}\}^2 \{\sum_{i=1}^m \sum_{j=1}^n F(i,j) - \bar{F}\}^2}} \quad (2)$$

where M and F representative the mean values of the original SWIR and fused image.

(3) SAM (Yuhas, 1992) is defined as calculating the angle between the corresponding pixels of the fused and reference image, its optimal value is 0. It is defined as follows:

$$SAM(I_{\{i\}}, J_{\{i\}}) = \arccos \left(\frac{\langle I_{\{i\}}, J_{\{i\}} \rangle}{\|I_{\{i\}}\| \|J_{\{i\}}\|} \right) \quad (3)$$

where $I_{\{n\}} = [I_{1,\{n\}}, \dots, I_{N,\{n\}}]$ is a pixel vector of the SWIR image I

with N bands. In which $\langle \cdot, \cdot \rangle$ denotes abc the scalar product,

and $\|\cdot\|$ denotes the vector ℓ_2 -norm.

(4) UIQI (Wang, 2002) can measures the distortion of two image structure and it is universal, which is suitable for evaluating different image processing processes. It is expressed as:

$$Q(I, J) = \frac{\sigma_{IJ}}{\sigma_I \sigma_J} \frac{2\bar{I}\bar{J}}{(\bar{I})^2 + (\bar{J})^2} \frac{2\sigma_I \sigma_J}{(\sigma_I^2) + (\sigma_J^2)} \quad (4)$$

where σ_{IJ} is the sample covariance of I and J, and \bar{I} is the sample mean of \bar{I} . σ_I and σ_J are the standard deviation of the fused and original images respectively.

3.3 Results

Figure 4 and 5 show the different fused results of WV-3. Every patches is mainly covered by different land cover types, such as building, water body and vegetation. Every patches with the size of 1000×1000 pixels. Table 2-3 show these quantitative scores for evaluation results. The best performance of each metric is in bold. The SWIR image has eight bands and is displayed by a combination of 1, 2, 3 bands as red, green, blue, as shown in figure 4.

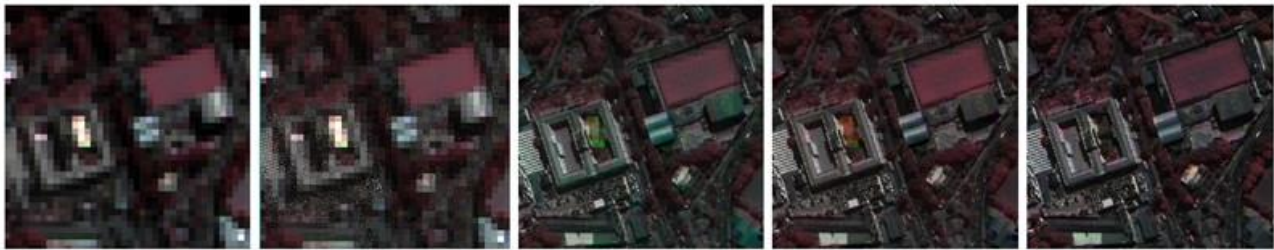


Figure 4. Results of the fusion experiment on an area of building. From left to right: Original SWIR images; ATWT; HIS; PCA and the proposed method.



Figure 5. Results of the fusion experiment on an area of building. From left to right: Original SWIR images; ATWT; HIS; PCA and the proposed method.

With a careful visual inspection of the results, we can find that all of these results have serious spectral distortion, especially in vegetation. This may be due to the CS method has higher spectral overlap requirement. As for the CNN, it may attribute to our network's structure. However, the spatial quality of the image is greatly improved.

In order to further evaluate the effect of fused images, quantitative indices were used to evaluate the spectral fidelity and spatial details of the fused images. The entropy can measure the information of the fusion image. The higher the entropy is, the better the quality of the result will be. The CC, UIQI and SAM can evaluate the similarity of the fusion image and

original short-wave infrared image. Especially the SAM, it can measure the spectral similarity of the two, the smaller the SAM is, the better the spectral fidelity of the result is. In table 2 and 3, the entropy of the proposed method is significantly higher than the original SWIR image. This demonstrates that this methods are effective to increase the spatial resolution of the SWIR images, this is consistent with the visual inspection. From the visual interpretation, although there are spectral distortions in all four methods, the CNN method has the smallest spectral angle mapper indicating that its spectral fidelity is better than the other methods. Unfortunately, the CC and UIQI are lower than the ATWT. In general, the reason for this result is that the number of samples we use to train the network is insufficient.

Method	Entropy	CC	UIQI	SAM
ATWT	0.141	0.942	0.927	4.416
PCA	0.270	0.388	0.385	7.413
IHS	0.266	0.384	0.381	8.584
Proposed	0.272	0.376	0.293	2.459

Table 2. Quality evaluation of fused images: Building (corresponding to Figure 4). The best performance of each metric is in bold.

Method	Entropy	CC	UIQI	SAM
ATWT	0.172	0.952	0.940	4.061
PCA	0.876	0.416	0.404	6.068
IHS	0.785	0.350	0.346	8.446
Proposed	1.058	0.408	0.352	0.637

Table 3. Quality evaluation of fused images: water body and vegetation (corresponding to Figure 5). The best performance of each metric is in bold.

4 CONCLUSION

In this paper, a novel image fusion algorithm for SWIR and Panchromatic images based on CNN domain is proposed, and applying it to WV-3 images. The experiments have achieved good results and can provide an effective means for the fuse of Pan and SWIR bands. The experimental results demonstrate the effectiveness of the proposed fusion algorithms. Though it can enhance the spatial information, there are also spectral distortion of the SWIR images. Furthermore, how to eliminate the resolution of fusion image and realize adaptive fusion based on sensor is the direction of our future research.

5 ACKNOWLEDGEMENTS

The data in this paper was provided by DigitalGlobe. The support is gratefully acknowledged.

REFERENCES

- Burt PJ, Adelson EH. 2003. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4), pp. 532-540.
- Laben CA, Brower BV. 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening: US, US6011875.
- Lecun Y, Bottou L, Bengio Y, et al. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp. 2278-2324.
- Mallat S G. 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Computer Society*.
- Pacifici F. 2016. On the predictive value of the WorldView3 VNIR and SWIR spectral bands. *Geoscience and Remote Sensing Symposium. IEEE*, pp. 898-901.
- Psjr C, Sides S C, Anderson J A. 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering & Remote Sensing*, 57(3), pp. 265-303.
- Schowengerdt R A. 1997. Models and Methods for Image Processing.
- Shettigara VK. 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogrammetric Engineering & Remote Sensing*, 58(5), pp.561-567.
- Soiva, Keijo. 2007. Image Fusion Using the Ehlers Spectral Characteristics Preservation Algorithm. *Mapping Sciences & Remote Sensing*, 44(2), pp. 93-116.
- Vivone G, Alparone L, Chanussot J. 2015. A Critical Comparison Among Pansharpening Algorithms. *IEEE Transactions on Geoscience & Remote Sensing*, 53(5), pp. 2565-2586.
- Wald L, Ranchin T, Mangolini M. 2009. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering & Remote Sensing*, 63(6), pp.691-699.
- Wang Z, Bovik AC. 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9(3), pp. 81-84.
- Yocky DA. 1996. Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT panchromatic data. *Photogrammetric Engineering & Remote Sensing*, 62(9), pp. 1067-1074.
- Yuhas R, Goetz A F H, Boardman J W. 1992. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm.