



International Journal of Environment and Geoinformatics (IJEGEO) is an international, multidisciplinary, peer reviewed, open access journal.

Comparative Analysis on Deep Learning based Pan-Sharpening of Very High-Resolution Satellite Images

Peijuan WANG., Ugur ALGANCI., Elif SERTEL

Chief in Editor

Prof. Dr. Cem Gazioğlu

Co-Editors

Prof. Dr. Dursun Zafer Şeker, Prof. Dr. Şinasi Kaya,
Prof. Dr. Ayşegül Tanık and Assist. Prof. Dr. Volkan Demir

Editorial Committee (June 2021)

Assoc. Prof. Dr. Abdullah Aksu (TR), Assit. Prof. Dr. Uğur Algancı (TR), Prof. Dr. Bedri Alpar (TR), Assoc. Prof. Dr. Aslı Aslan (US), Prof. Dr. Levent Bat (TR), Prof. Dr. Paul Bates (UK), İrşad Bayırhan (TR), Prof. Dr. Bülent Bayram (TR), Prof. Dr. Luis M. Botana (ES), Prof. Dr. Nuray Çağlar (TR), Prof. Dr. Sukanta Dash (IN), Dr. Soofia T. Elias (UK), Prof. Dr. A. Evren Erginal (TR), Assoc. Prof. Dr. Cüneyt Erenoğlu (TR), Dr. Dieter Fritsch (DE), Prof. Dr. Çiğdem Göksel (TR), Prof. Dr. Lena Halounova (CZ), Prof. Dr. Manik Kalubarme (IN), Dr. Hakan Kaya (TR), Assist. Prof. Dr. Serkan Kükreler (TR), Assoc. Prof. Dr. Maged Marghany (MY), Prof. Dr. Michael Meadows (ZA), Prof. Dr. Nebiye Musaoğlu (TR), Prof. Dr. Masafumi Nakagawa (JP), Prof. Dr. Hasan Özdemir (TR), Prof. Dr. Chryssy Potsiou (GR), Prof. Dr. Erol Sarı (TR), Prof. Dr. Maria Paradiso (IT), Prof. Dr. Petros Patias (GR), Prof. Dr. Elif Sertel (TR), Prof. Dr. Nüket Sivri (TR), Prof. Dr. Füsün Balık Şanlı (TR), Prof. Dr. Uğur Şanlı (TR), Duygu Ülker (TR), Prof. Dr. Seyfettin Taş (TR), Assoc. Prof. Dr. Ömer Suat Taşkın (TR), Assist. Prof. Dr. Tuba Ünsal (TR), Dr. Manousos Valyrakis (UK), Dr. İnese Varna (LV), Dr. Petra Visser (NL), Prof. Dr. Selma Ünlü (TR), Assoc. Prof. Dr. Oral Yağcı (TR), Prof. Dr. Murat Yakar (TR), Assoc. Prof. Dr. İ. Noyan Yılmaz (AU); Assit. Prof. Dr. Sibel Zeki (TR)

Abstracting and Indexing: TR DIZIN, DOAJ, Index Copernicus, OAJI, Scientific Indexing Services, International Scientific Indexing, Journal Factor, Google Scholar, Ulrich's Periodicals Directory, WorldCat, DRJI, ResearchBib, SOBIAD

Comparative Analysis on Deep Learning based Pan-sharpening of Very High-Resolution Satellite Images

Peijuan Wang ¹, Ugur Alganci ^{2*}, Elif Sertel ²

¹ Istanbul Technical University, Institute of Informatics, Istanbul, Turkey

² Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey

* Corresponding author: Ugur Alganci
E-mail: alganci@itu.edu.tr

Received 30 Nov 2020

Accepted 09 Dec 2020

How to cite: Wang et al., (2021). Comparative Analysis on Deep Learning based Pan-sharpening of Very High-Resolution Satellite Images, *International Journal of Environment and Geoinformatics (IJEgeo)*, 8(2):150-165. doi. 10.30897/ijegeo.834760

Abstract

Pan-sharpening is a fundamental task of remote sensing, aiming to produce a synthetic image having high spatial and spectral resolution of original panchromatic and multispectral images. In recent years, as in other tasks of the remote sensing field, deep learning based approaches have been developed for this task. In this research, a detailed comparative analysis was conducted to evaluate the performance and visual quality of pan-sharpening results from traditional algorithms and deep learning-based models. For this purpose, the deep learning based methods that are CNN based pan-sharpening (PNN), Multiscale and multi-depth convolutional neural networks (MSDCNN) and Pan-sharpened Generative Adversarial Networks (PSGAN) and traditional methods that are Brovey, PCA, HIS, Indusion and PRACS were applied. Analysis was performed on regions with different land cover characteristics to evaluate the stability of the methods. In addition, effects of the filter size, spectral indices, activation and loss functions on the pan-sharpening were investigated. For the accuracy assessment, commonly used with-reference and without-reference quality metrics were computed in addition to visual quality evaluations. According to results, the deep learning-based methods provided promising results in both the reduced resolution and full resolution experiments, while PRACS method outperformed other traditional algorithms in most of the experimental configurations.

Keywords: Pan-sharpening, Deep Learning, Convolutional Neural Network (CNN), Satellite Images

Introduction

The current very high resolution (VHR) optical satellites provide multi-resolution sensor data, where the multi-spectral (MS) sensor acquires a higher degree of spectral resolution image than the panchromatic (PAN) sensor, whereas the PAN image has a higher spatial resolution than the MS image, to fulfil the high-resolution requirements in both domains. The pan-sharpening process enables a sensor fusion between the MS image and PAN image, which provides both high spectral resolution and high spatial resolution (Pohl and Van Genderen, 1998; Çelik and Gazioglu, 2020; Tombul and Kavzoglu, 2020; Ozcelik et al., 2020). Thus, pan-sharpening is quite significant for the remote sensing image interpretation and frequently used as the pre-processing step for image analysis such as feature extraction, segmentation, and classification. Therefore, it has been the focus and hotspot of remote sensing research.

In the last decades, several methods were developed for the pan-sharpening process. The research performed by Vivone et al. (Vivone et al., 2015; Erdem and Avdan, 2020; Oztürk et al., 2020), presented a detailed review of the traditional pan-sharpening methods. These methods can be divided into three main categories: (1) Component Substitution (CS) based methods; (2) Detail

Injection based methods and (3) Sparse Representation based methods. In addition, model based methods gained growing attention in the recent decade (Kang et al., 2014) (Palsson et al., 2020).

One of the basic CS based pan-sharpening method is based on intensity-hue-saturation (IHS) transform, which can only be applied on three spectral bands (Shettigara, 1992). The IHS method includes the RGB-IHS transformation model and a generalized IHS (GIHS) image fusion procedure with the colour distortion problem (Tu et al., 2001). Principal component analysis (PCA) (Chavez Jr. and Yaw Kwarteng, 1989), Brovey transform (BT) (Gillespie et al., 1987), Gram-Schmidt (GS) spectral sharpening (Laben and Brower, 2000), partial replacement (PRACS) (Choi et al., 2011) and band-dependent spatial-detail (BDSD) (Zhong et al., 2017) are among the commonly used CS based methods for the pan-sharpening purpose.

To deal with the colour distortion problem encountered in the CS methods and provide a better spectral fidelity, detail injection-based methods were proposed. The concept of detail injection is to extract the spatial information (high-frequency detail) from the PAN image by wavelet transform, Laplacian pyramid, etc., in the first step and then inject it to the up-sampled MS images to generate the fused image (Aiuzzi et al., 2002). Several

methods were proposed for this category, such as à Trouis Wavelet Transform (ATWT) and curvelet transform based fusion algorithms (Aiazzi et al., 2002) (Nencini et al., 2007). Khan et al. (Khan et al., 2008), proposed a new approach, named Indusion method, by using the induction scaling technique instead of bicubic interpolation to obtain sharper, better correlated and better co-registered images in an upscaled domain. To reduce the distortion, approaches are developed by tuning the filters to match the modulation transfer functions (MTFs) of the sensors (Aiazzi et al., 2003). The smoothing-filter-based intensity modulation (SFIM) is one of the famous methods among them (J. G. Liu, 2010).

Besides, several researches devoted to the sparse representation-based methods. In the study of Zhu and Bamler (Zhu and Bamler, 2013), a pan-sharpening method named Sparse Fusion of Image (SparseFI) was introduced. This method is based on the compressive sensing theory and asserted give higher spatial resolution with relatively less spectral distortion (Cheng et al., 2014). Yang et al. (Yang et al., 2018), proposed a novel sparse representation-based pan-sharpening method that includes three steps-dictionary construction, PAN image decomposition, and high spatial resolution multispectral image reconstruction. The sparse representation methods require the presence of high resolution MS image, which is not practically available. Most of the algorithms rely on assumptions for high resolution MS dictionary, such as considering the original MS as the degraded version of high resolution MS, which limits the performance of these approaches (Yin, 2015).

As a summary, traditional algorithms face several difficulties such as spectral distortions, spatial detail injection limitations, and limitations based on theoretical assumptions. Moreover, performance of these methods is variable across different sensor data, land cover characteristics and acquisition geometry configurations. With developments in machine learning (ML) and deep learning (DL) in the last decades, these technologies started to be widely used in image processing, such as image classification, image segmentation, object detection super-resolution, pan-sharpening and reconstruction. As an initial CNN based super-resolution study, researchers proposed a super-resolution reconstruction method by using a three-layer convolutional neural network named as SRCNN (Dong et al., 2016). This approach was implemented to the task of pan-sharpening in remote sensing, which is named as PNN and was proved that it is a more promising approach (Masi et al., 2016). However, the design of PNN is relatively simple, thus still needs to be improved. The well-known drawback of the deeper neural networks is the difficulty of the training phase. In 2016, He et al. (He et al., 2015), demonstrated the depth problem of the networks and proposed a remarkable solution, which has since allowed the training of over 2000 layers with increasing accuracy. With the development of the residual neural networks, more researchers have contributed to the pan-sharpening task. Wei and Yuan (Wei and Yuan, 2017) proposed deep residual learning

for remote sensing imagery; however, it only tested on the reduced resolution dataset. Another example for the pan-sharpening includes transfer learning of the residual network in addition to fine-tuning by changing the loss function, and the test part was applied both in reduced resolution and the full-resolution dataset (Scarpa et al., 2018). More recently, a novel network specifically designed for the pan-sharpening task named as the multiscale and multi-depth convolutional neural network (MSDCNN) was proposed (Yuan et al., 2018). In this design, the PNN was used as the shallow part and a deeper multiscale feature extraction layer was used as the deep part to generate a deep architecture.

Another important breakthrough in the DL field is Generative Adversarial Networks (GAN) proposed by Goodfellow et al. in 2014 (Goodfellow and Pouget-Abadie, 2014). GANs try to train a generator to generate the fake image and train a discriminator to distinguish whether the output image is real or fake. The goal is to minimize the difference between the generated and real images. Ledig et al. (Christian Ledig Ferenc Huszar, 2017) successfully applied the GAN to the super-resolution task. In other researches, a Generative adversarial network (PSGAN) was proposed for remote sensing image pan-sharpening (X. Liu et al., 2018a) (X. Liu et al., 2018b). More recently, the PanColorGAN method was introduced for the pan-sharpening of satellite images, which reconstructs the problem by colorization of the panchromatic image and achieved high level of spatial detail conservation (Ozcelik et al., 2020).

This research aims to perform a comparative evaluation of DL based approaches and traditional methods by investigating the performance of three DL methods that are PNN as a representative of shallow network, a modified version of MSDCNN as a representative of deep network and PSGAN as a representative of adversarial network, while taking the Brovey, PCA, HIS, Indusion and PRACS as the representatives of traditional methods. Within the scope of the research, satellite images from different VHR sensors that represent different land cover characteristics and acquisition conditions were used for evaluating the pan-sharpening performances of the methods over broad coverages, different land cover configurations and different sensors. In addition, effects of different hyper parameter and input data configurations on the DL based methods were investigated.

Materials and Methods

Subsections provide information about traditional and deep learning based methods applied on this research, details of satellite images and pre-processing of the images, and lastly the details of the quantitative evaluation metrics. The processing steps and workflow of this research is provided in Figure 1.

Traditional Pan-sharpening Methods

Traditional methods that are Brovey (Gillespie et al., 1987), PCA (Chavez Jr. and Yaw Kwarteng, 1989), IHS

(Tu et al., 2001), Indusion (Khan et al., 2008), and PRACS (Choi et al., 2011) were applied on the dataset for comparative evaluation. Some of the implementation

codes of these traditional methods provided by Vivone et al. (Vivone et al., 2015) and available online (*Open Remote Sensing*, 2016).

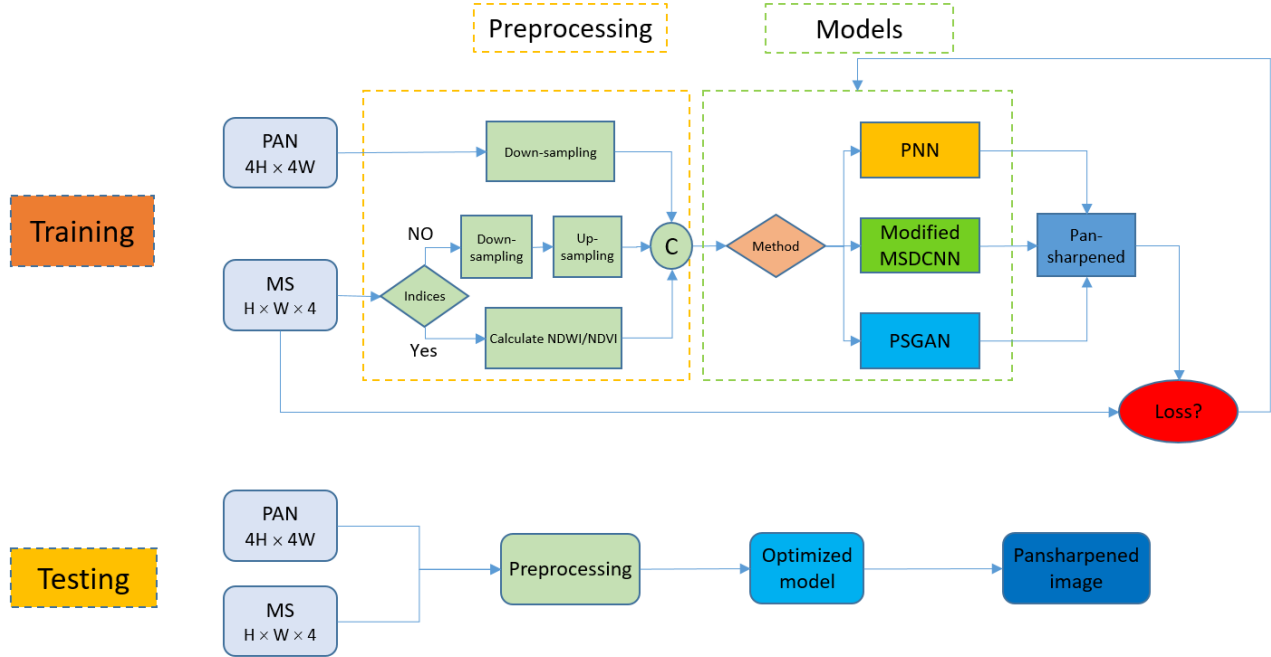


Figure 1. Workflow of the processing steps and methods used

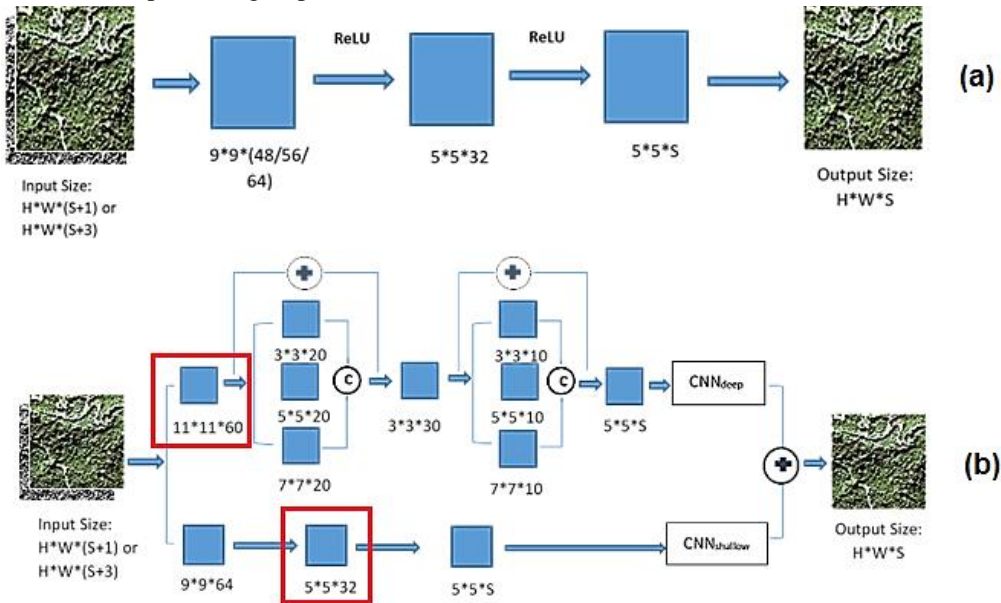


Figure 2. The architecture of, (a) the PNN, (b) improved MSDCNN method (red boxes indicates modified kernels).

Overview of the Deep Learning Network Architectures

For the traditional fully connected networks, each neuron takes the outputs of all neurons from the previous layer as the input and feeds its output to all the neurons of the next layer. Thus, the network includes a large number of parameters (weights and biases), which results in a computationally expensive training process. The Convolutional Neural Networks (CNN) overcome this problem as in CNN each neuron only has a limited receptive field by using the filters (also called the kernel, such as 5x5 or 9x9) to slide over the image spatially and to compute the dot products. This structure enables sharing the weights and bias (parameters of the filter) in

the same layer, thus, greatly reduces the computational overloads.

CNN-based pan-sharpening (PNN)

The initial PNN pan-sharpening method was motivated by the image super-resolution network, which is named as SRCNN (Dong et al., 2016). For the convolutional neural networks inside the architecture, the Rectified Linear Units (ReLU) is used as the activation function, which can be described as:

$$f(x) = \max(0, x) \quad (1)$$

The ReLU function can be activated only when the units are positive and zero. For the negative units, called the dead neurons, the gradient will be always in the off state as zero. Thus, Leaky ReLU was proposed. Leaky ReLU can be described as:

$$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (2)$$

Where, α corresponds to a small negative slope constant (0.01) that fixes off state problem of the RELU.

The CNN-based pan-sharpening (PNN) method uses the same three-layer convolutional networks as presented in the study of Dong et al. (Dong et al., 2016), but replaces the 1x1 kernel of the middle layer with 5x5 kernel. Thus, the architecture of PNN can be described as in Figure 2(a).

Improved MSDCNN pan-sharpening

Motivated by the application of residual network, a method with skip connection, which is named as Multiscale and multi-depth convolutional neural networks (MSDCNN), was proposed (Yuan et al., 2018). The MSDCNN model includes two subnetworks: a shallow three-layer CNN with the same architecture as in the study of Cheng et al. (Cheng et al., 2014), and a deep CNN with two multiscale convolutional layer blocks. In the original structure of MSDCNN, the shallow subnetwork consists three convolutional layers with 64 kernels in the first layer and the deep subnetwork contains the multiscale blocks and the skip connection. In Yuan et al. (Yuan et al., 2018), three different architectures compared and the above-mentioned version proved to be the most effective one.

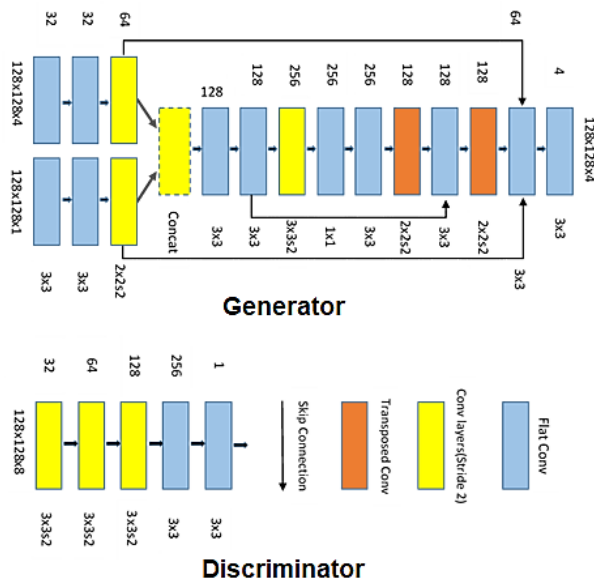


Figure 3. The architecture of PSGAN method.

This research proposes an improved version of MSDCNN architecture to make it suitable with the Pleiades satellite images. The modified architecture keeps using a similar design as the PNN model for shallow subnetwork by replacing the 1x1 kernels in the middle layer with the 5x5 kernels. In the deep

subnetwork, the first 7x7 kernel is replaced with an 11x11 kernel. Moreover, NDVI and NDWI index data are added to the dataset in order to observe the effect of additional data on the pan-sharpening process (Yang et al., 2018). The architecture of the implementation is given in Figure 2(b).

PSGAN pan-sharpening

Motivated by the application of the Generative Adversarial Networks (GANs) introduced by Goodfellow et al. (Goodfellow and Pouget-Abadie, 2014), a PSGAN model that applies the GAN to the pan-sharpening task was proposed in Liu et al. (X. Liu et al., 2018a). The architecture of the PSGAN is presented in Figure 3.

The generator that takes down-sampled MS image and PAN image as the input and produces the fake pan-sharpened images and the original multispectral images are regarded as the real images, which is compatible with the dataset produced with Wald protocol. Discriminator detects whether the processed images are from the Generator or real images. The interpolation and normalization methods were modified in the implementation of PSGAN to perform a comparative evaluation with the other DL models.

Data and pre-processing

In this research, very high-resolution (VHR) satellite images acquired by the Pleiades satellite were used in training and testing phases. Pleiades image provides 0.5 m spatial resolution for panchromatic image and 2.0 m spatial resolution for the multispectral image with 4 bands in the visible and NIR portions of the electromagnetic spectrum with 12-bit radiometric resolution. For additional independent testing, Worldview 2 and Worldview 3 satellite images were used. Worldview 2 provides 0.4 m and 1.6 m spatial resolution for panchromatic image and multispectral image, while Worldview 3 provides 0.3 m and 1.2 m spatial resolutions respectively. Both sensors have an 11-bit radiometric resolution (AIRBUS, 2020) (DigitalGlobe, 2020).

Table 1 provides the details of the image dataset used for training and test processes. Two Pleiades scenes covering different regions of Turkey were used as the training datasets. One scene covers Istanbul, Turkey, which has a size of 10160×10540 pixels at the MS scale. The other one covers Iznik, Turkey, which has a size of 10036×10136 pixels at the MS scale. For the training, the PAN image and MS image were split into 57602 patches separately. For the pre-processing, the normalization was applied to make sure the data range is between 0 and 1.

In addition, spectral indices added to the input data to see if it can improve the resulting performance (Nouri et al., 2013). In this research the NDVI and NDWI indices were applied on the training dataset.

Table 1. Detailed information about dataset used in training and test procedures.

	Mode	Sensor	Area	Patches	Input size		Output size
					MS	PAN	HRMS
Training	Reduced resolution	Pleiades	Istanbul Izник	57602	60x60x4	240x240	60x60x4
	Reduced resolution	Pleiades	Bursa	200	200x200x4	800x800	200x200x4
Testing	Full resolution	Pleiades	Bursa				
	Full resolution	Worldview2/ Worldview3	Washington DC Tripoli	100/65/65	200x200x4	800x800	800x800x4

*Pleiades satellite image data provided by ITU – CSCRS under the licence of AIRBUS Defence and Space, while the Worldview satellite image data is open access by Maxar and can be downloaded from: <http://www.digitalglobe.com/samples?search=Imagery>

Another Pleiades image scene that covers Bursa, Turkey was used as the test image and is completely independent of the training dataset. For the reduced resolution experiments, 200 patches with a size of 200×200×4 pixels for the MS image and 800×800 pixels for the PAN image were used. For the full resolution experiments (full resolution experiment), 100 patches from Pleiades image, 65 patches from Worldview 2 and 65 patches from Worldview 3 were used with a patch size of 200×200×4 pixels for the MS images and 800×800 pixels for the PAN images. The patches were generated with use of tiling tool of the GDAL library.

Quantitative Evaluation Metrics

Metrics for Reduced Resolution Experiments

Wald et al. (Wald et al., 1997), provided important properties to verify the quality of the synthesis images. For instance, (a) once the synthesis image is degraded to its original resolution, it should be as identical as possible to the original image; (b) the synthesis image should be as identical as possible to the high-resolution multispectral image if it could be acquired by the sensor.

One of the important assessment metrics to evaluate the spectral quality of the pan-sharpened image is Correlation Coefficient (CC), which presents the similarity between structures of the original MS images and the pan-sharpened image (Panchal and Thakker, 2015). CC can be defined as in Eq. 3:

$$CC(R, F) = \frac{\sum_{mn}(R_{mn}-\bar{R})(F_{mn}-\bar{F})}{\sqrt{(\sum_{mn}(R_{mn}-\bar{R})^2)(\sum_{mn}(F_{mn}-\bar{F})^2)}} \quad (3)$$

where \bar{F} and \bar{R} are the mean of the images that are fused (F) and reference (R). The value of CC ranges from 0 to 1, where higher values represent the better correlation.

Several full-reference metrics were proposed for the assessment of the spatial and spectral distortions of the fused images for reduced resolution images. These can be listed as the Universal Image Quality Index (Q-index) (Zhou and Bovik, 2002), Root Mean Square Error (RMSE), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) (J. G. Liu, 2000) and the spectral angle mapper (SAM) (Yuhas et al., 1992).

Metrics for Full Resolution Experiments

For evaluating the full-resolution images, a single index named as QNR (Quality with No Reference) was

proposed. QNR index is the product of the complements of the spatial and spectral distortion indices and it varies in the range of 0 and 1, where the optimal value is 1. QNR can be written as follows:

$$QNR \triangleq (1 - D_\lambda)^\alpha (1 - D_s)^\beta \quad (4)$$

where D_λ is the Spectral Distortion Index derived from the difference of inter-band Q values described in the previous part from the pan-sharpened MS bands (written as $\{\hat{G}_I\}_{I=1}^L$) and the low-resolution MS bands (written as $\{\tilde{G}_I\}_{I=1}^L$). Thus D_λ is defined as in Eq. 5:

$$D_\lambda \triangleq \sqrt[P]{\frac{1}{L(L-1)} \sum_{I=1}^L \sum_{\substack{r=1 \\ r \neq I}}^L |Q(\hat{G}_I, \hat{G}_r) - Q(\tilde{G}_I, \tilde{G}_r)|^P} \quad (5)$$

D_s is the Spatial Distortion Index defined as in Eq. 6:

$$D_s \triangleq \sqrt[q]{\frac{1}{L} \sum_{I=1}^L |Q(\hat{G}_I, P) - Q(\tilde{G}_I, \tilde{P})|^q} \quad (6)$$

Where, P is the PAN image, and \tilde{P} is a spatially degraded PAN image at the size of the original MS image. Both D_s and D_λ alters in [0, 1] and obtains the best value zero when two vectors are identical.

Results

Reduced Resolution Experiments

Pan-sharpening requires high spatial resolution PAN band and high spectral resolution MS image to generate a fused image to achieve high resolution both in spectral and spatial domains. Due to the nature of the problem, the reference original pan-sharpened image does not exist, thus the Wald protocol is applied to the dataset to produce reduced resolution dataset for the training of the CNN architectures and to use the full resolution MS images as the reference data (Wald et al., 1997). As the first step of the Wald protocol, down sampling is performed according to the spatial resolution ratio of the PAN and MS images. Secondly, the down-sampled MS image is up sampled to the original size of the MS image. In the next step, the down sampled PAN and 4-band MS images are stacked together to obtain 5-band data. In this step, the spectral index products can be

added to the dataset as in the study of (Masi et al., 2016), such as the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) to increase the amount of input data for the CNN (Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W.; Harlan., 1974) (Gao, 1996). The original MS image is used as the reference image throughout the training.

The training process applied with the use of reduced resolution Pleiades dataset and tests were performed on the independent reduced resolution datasets with an optimization of kernel size and activation functions. Besides, spectral indices such as Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) that computed from the multispectral images were added to data layers in order

to investigate the effect of indices on the pan-sharpening task. In addition, the effects of the L1 and L2 loss functions on training process tested in the scope of this research.

Table 2. Parameters used in model training step.

Network	Learning rate	Epochs	Optimizer
PNN_indices_L1	0.001	100	SGD
PNN_indices_L2	0.001	300	SGD
PNN_noindices_L1	0.001	200	SGD
PNN_noindices_L2	0.001	300	SGD
MSDCNN_indices_L1	0.001	50	SGD
MSDCNN_indices_L2	0.001	100	SGD
MSDCNN_noindices_L1	0.001	100	SGD
MSDCNN_noindices_L2	0.001	150	SGD
PSGAN_L1	0.0002	200	Adam
PSGAN_L2	0.0002	50	Adam

Table 3. Accuracy metric and correlation coefficient results of reduced resolution experiments with Pleiades dataset.

Method	With Reference Metrics				Correlation Coefficient				
	Q↑	SAM↓	ERGAS↓	RMSE↓	Red	Green	Blue	NIR	Average
PRACS	0.968	2.266	2.552	63.875	0.799	0.859	0.836	0.655	0.788
PCA	0.895	6.042	5.146	162.641	0.948	0.944	0.940	0.958	0.948
Indusion	0.943	2.953	3.878	104.463	0.885	0.891	0.883	0.883	0.885
IHS	0.972	2.936	3.863	76.828	0.890	0.842	0.791	0.957	0.870
Brovey	0.971	2.093	3.233	78.774	0.932	0.887	0.825	0.937	0.895
PNN_indices_L1	0.980	1.935	1.653	44.900	0.976	0.978	0.966	0.976	0.974
PNN_indices_L2	0.977	2.203	1.825	48.354	0.970	0.974	0.952	0.973	0.967
PNN_noindices_L1	0.979	1.945	1.643	45.292	0.975	0.978	0.969	0.975	0.974
PNN_noindices_L2	0.976	2.297	1.895	49.687	0.968	0.972	0.952	0.972	0.966
MSDCNN_indices_L1	0.981	1.923	1.653	44.468	0.975	0.978	0.966	0.976	0.974
MSDCNN_indices_L2	0.976	2.236	1.878	49.218	0.964	0.971	0.947	0.972	0.964
MSDCNN_noindices_L1	0.980	1.931	1.638	45.036	0.976	0.977	0.969	0.975	0.974
MSDCNN_noindices_L2	0.976	2.304	1.899	49.696	0.968	0.972	0.952	0.972	0.966
PSGAN_L1	0.981	2.012	2.226	54.805	0.954	0.954	0.937	0.978	0.956
PSGAN_L2	0.956	2.772	4.222	57.689	0.952	0.962	0.940	0.974	0.957

All of the DL based pan-sharpening models were trained in the Tensorflow DL framework (Tensorflow, 2015) over reduced resolution training dataset. With the Tensorflow supported by a GPU (Graphics Processing Unit, NVIDIA GeForce GTX 1060 3GB), the training process costs almost 10 hours for the PNN model, 14 hours for the improved MSDCNN model and 10 hours for PSGAN model. The learning rate, epoch number and optimizer information of the PNN, MSDCNN and PSGAN training process are provided in Table 2.

Quantitative assessment

The performance of the DL based and traditional pan-sharpening algorithms was evaluated on 200 patches for the reduced resolution experiments. According to Table 3, DL based methods provided better results than traditional methods. Although traditional methods provided quite closer accuracy with DL based methods in terms of Q metric, the accuracy improvement is significant in favour of DL based methods for SAM, ERGAS and RMSE metrics.

Moreover, the DL based methods trained with L1 loss function provided better performance than the same methods trained with L2 loss function in almost all cases. For the PNN and MSDCNN models, implementations with NDVI and NDWI indices perform better than the standard implementations for the Q, SAM, and RMSE metrics, whereas for the ERGAS metric, the standard implementation provided better results than the indices added implementation. PSGAN with L1 provided the highest performance in terms of Q metric. For SAM and RMSE, MSDCNN with indices and L1 loss function performed the best. Table 3 also presents the CC metric that calculated for spectral bands of the results with reference to the bands of the reference MS images. According to CC metric MSDCNN-noindices-L1, PNN-noindices-L1, MSDCNN-noindices-L1, PNN-noindices-L1, PSGAN-L1 combinations provided the highest correlation for red, green, blue and NIR bands, respectively. The highest average correlation derived from all bands were obtained by PNN and MSDCNN implementations with no indices and L1 loss

function. These results indicate the superiority of DL based methods over the traditional ones.

Visual interpretation and CC based assessment

In addition to metric based evaluation, the results obtained from the pan-sharpening algorithms evaluated visually as the colour representation and object sharpness are important considerations, which should be judged by human perception. For this purpose, three patches were selected to perform a visual and statistical evaluation in a reduced resolution dataset, which represents industrial areas, agricultural and vegetated lands and a heterogeneous area with mixed land cover characteristics respectively. Moreover, the spectral correlation between analysis results and the original MS reference image was tested with the use of CC metric for each band for these patches. The CC metric has the advantage of being effective in case of bias between pan-sharpening results and reference data.

In the first patch, the traditional methods injected the spatial details well especially for the rooftops of the industrial buildings, while DL based methods present closer looks to the original MS data, which is the reference for the expected result (Figure 4). The deep methods with L2 loss function resulted in blur effects when compared to L1. In addition, data set with no indices provided a more realistic colour representation. According to CC results, the DL methods provided a higher correlation with the reference MS image when compared to traditional methods for all bands, where the

PSGAN L1 combination reached the highest score (Table 4).

In the second patch that covers agricultural and vegetated lands, the Indusion, IHS and Brovey methods completely injected the spatial details from the PAN image and they provide sharpest visual results, while there is an obvious blur problem in PCA and PRACS method results (Figure 5 (c-g)). The MSDCNN L1 method provides the best visual match with reference MS followed by the PNN L1 method. For this patch, PSGAN provided sharper results but distortions in the colour tones are observable (Figure 5 (h-q)). According to CC analysis results, there is an obvious increment in correlation with DL methods. MSDCNN L1 algorithms provided highest scores followed by PNN L1, which is in line with visual inspections (Table 4).

The last patch consists of a heterogeneous region with small housings, roads, vegetated areas, and bare lands (Figure 6). For the traditional methods, spatial detail injection is high except the PRACS method, which provides similar visual quality as the reference data (Figure 6 (d)) and higher CC value when compared to other methods (Table 4). The DL methods outperformed the traditional methods with noticeable improvement in CC values for all bands and PSGAN L1 method provided the highest CC value and a balanced spectral similarity – spatial detail injection performance (Figure 6 (p)). PSGAN method is followed by MSDCNN noindices L1 and PNN noindices L1.

Table 4. Correlation Coefficient results of reduced resolution Pleiades experiments for different land cover schemas.

Land Cover Type	Industrial				Urban				Agriculture and Vegetation			
	Red	Green	Blue	NIR	Red	Green	Blue	NIR	Red	Green	Blue	NIR
PCA	0.947	0.946	0.945	0.946	0.868	0.856	0.856	0.901	0.895	0.930	0.907	0.615
PRACS	0.970	0.968	0.969	0.944	0.917	0.907	0.907	0.917	0.940	0.937	0.938	0.963
Indusion	0.953	0.952	0.951	0.931	0.852	0.851	0.844	0.828	0.840	0.845	0.842	0.868
HIS	0.970	0.963	0.957	0.919	0.897	0.856	0.823	0.894	0.860	0.791	0.752	0.966
Brovey	0.973	0.968	0.964	0.871	0.909	0.878	0.846	0.855	0.917	0.860	0.805	0.944
PNN_indices_L1	0.996	0.996	0.993	0.957	0.985	0.979	0.967	0.949	0.969	0.974	0.968	0.978
PNN_indices_L2	0.994	0.994	0.991	0.950	0.981	0.974	0.960	0.944	0.962	0.968	0.950	0.974
PNN_noindices_L1	0.996	0.996	0.995	0.957	0.985	0.980	0.972	0.948	0.969	0.975	0.968	0.978
PNN_noindices_L2	0.993	0.994	0.992	0.950	0.979	0.975	0.964	0.945	0.961	0.969	0.950	0.974
MSDCNN_indices_L1	0.996	0.995	0.993	0.959	0.985	0.979	0.967	0.953	0.969	0.975	0.968	0.978
MSDCNN_indices_L2	0.993	0.994	0.991	0.949	0.978	0.972	0.958	0.941	0.959	0.964	0.944	0.973
MSDCNN_noindices_L1	0.996	0.996	0.995	0.957	0.986	0.980	0.972	0.948	0.969	0.974	0.970	0.978
MSDCNN_noindices_L2	0.993	0.995	0.992	0.951	0.980	0.976	0.964	0.945	0.961	0.970	0.950	0.974
PSGAN_L1	0.999	0.999	0.998	0.986	0.990	0.989	0.985	0.972	0.919	0.925	0.915	0.965
PSGAN_L2	0.997	0.997	0.996	0.978	0.987	0.984	0.979	0.964	0.927	0.952	0.921	0.963



Figure 4. Reduced resolution experiment results for industrial area. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

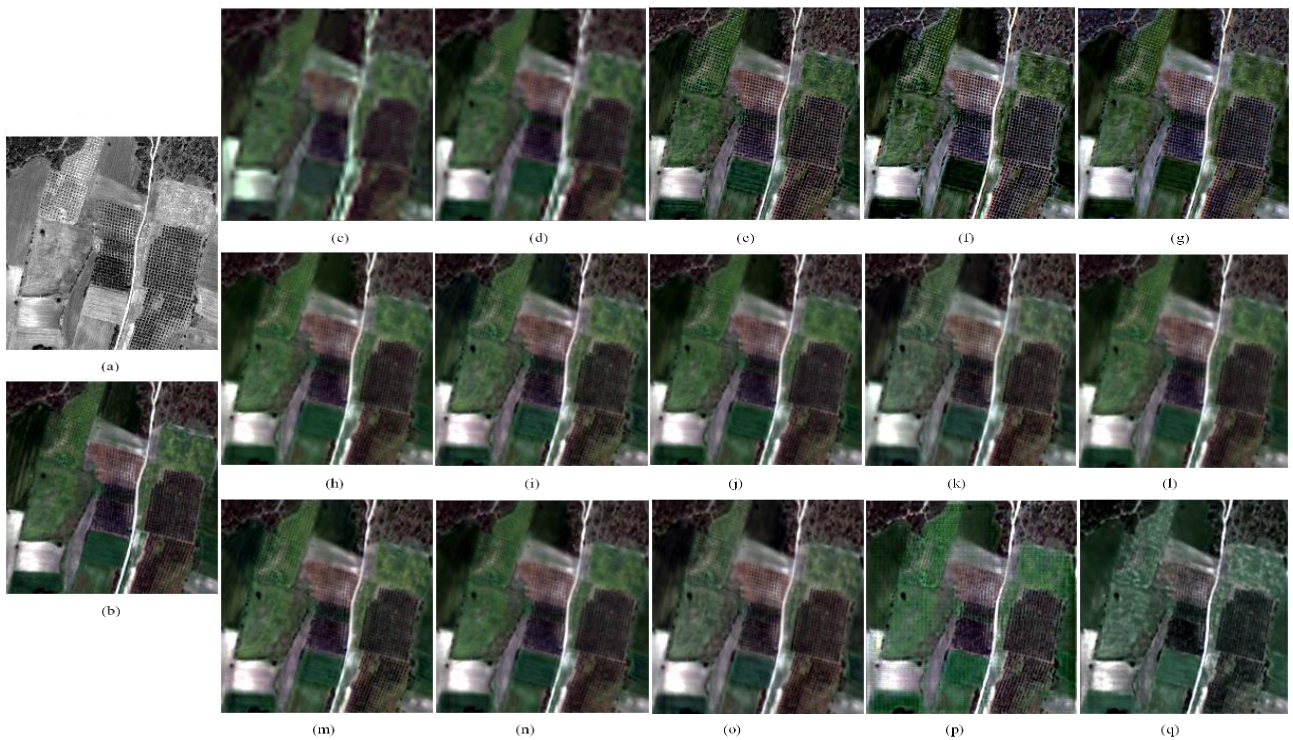


Figure 5. Reduced resolution experiment results for agricultural lands and vegetated areas. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

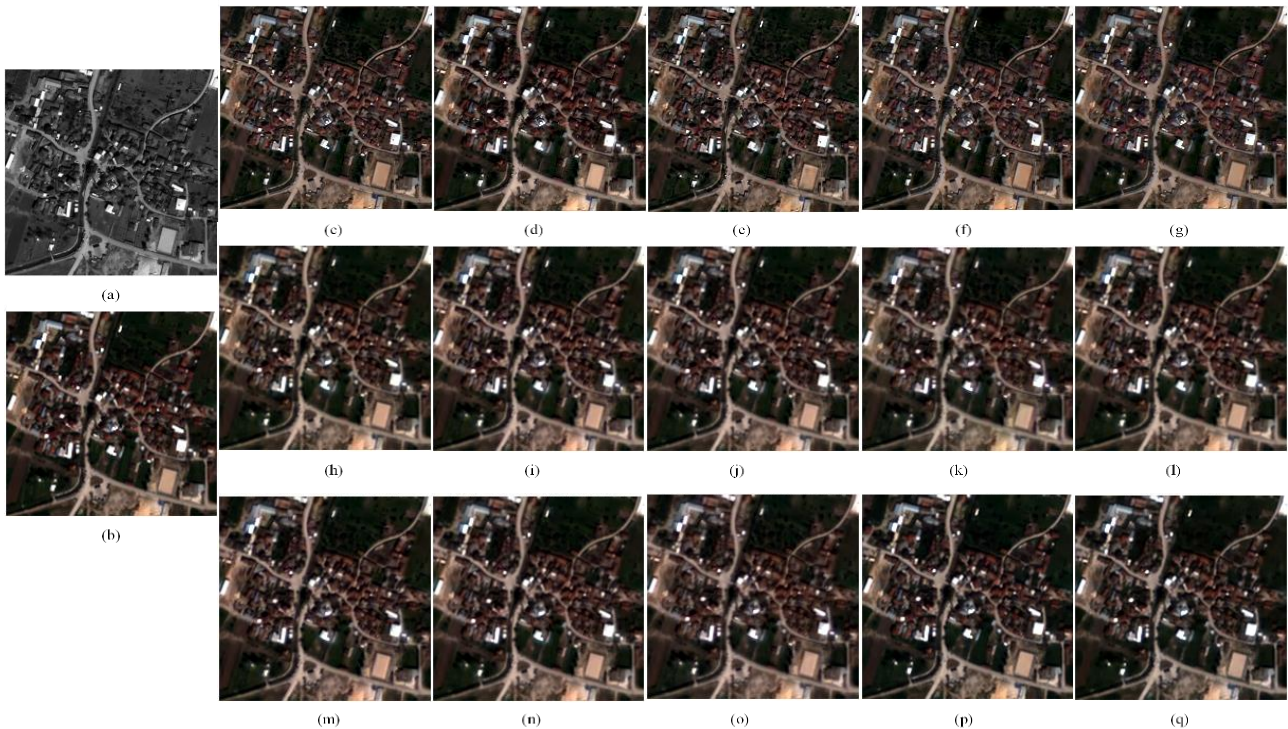


Figure 6. Reduced resolution experiment results for urban areas. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

Full Resolution Experiments

The full resolution experiments were performed with the full-resolution images where the PAN and MS image has 0.5m and 2m spatial resolution respectively. The 100 patches that belong to the Bursa region, which is independent of the training region. This experiment is important to evaluate the performance of the pan-sharpening models in original resolution conditions,

which is the main purpose. Moreover, Worldview 2 and Worldview 3 datasets were evaluated in full resolution mode to test the transferability of the approaches across different sensors.

Quantitative assessment

The quantitative assessment results of Pleiades images are presented in Table 5.

Table 5. Accuracy metric and correlation coefficient results of full resolution experiments with Pleiades dataset.

	Without Reference Metrics			Correlation Coefficient				Average
	QNR \uparrow	D $_s$ \downarrow	D $_{\lambda}$ \downarrow	Red	Green	Blue	NIR	
PRACS	0.885	0.092	0.034	0.944	0.957	0.958	0.709	0.892
PCA	0.805	0.150	0.056	0.991	0.983	0.984	0.992	0.988
Indusion	0.879	0.063	0.056	0.872	0.866	0.870	0.871	0.870
IHS	0.608	0.280	0.174	0.833	0.765	0.661	0.989	0.812
Brovey	0.671	0.223	0.149	0.929	0.841	0.703	0.962	0.859
PNN_indices_L1	0.811	0.086	0.113	0.965	0.966	0.962	0.982	0.969
PNN_indices_L2	0.827	0.079	0.105	0.956	0.957	0.906	0.986	0.951
PNN_noindices_L1	0.821	0.077	0.111	0.956	0.953	0.951	0.981	0.960
PNN_noindices_L2	0.808	0.060	0.142	0.950	0.942	0.913	0.988	0.948
MSDCNN_indices_L1	0.810	0.085	0.115	0.959	0.956	0.953	0.981	0.962
MSDCNN_indices_L2	0.830	0.076	0.104	0.955	0.960	0.902	0.987	0.951
MSDCNN_noindices_L1	0.821	0.079	0.109	0.955	0.951	0.951	0.981	0.959
MSDCNN_noindices_L2	0.806	0.085	0.115	0.948	0.941	0.913	0.988	0.948
PSGAN_L1	0.872	0.074	0.059	0.954	0.951	0.931	0.977	0.953
PSGAN_L2	0.769	0.104	0.146	0.902	0.904	0.874	0.967	0.912

Table 6. Accuracy metric results of full resolution experiments with Worldview-2 and Worldview-3 image.

	Worldview-2				Worldview-3			
	QNR↑	D_s↓	D_lambda↓	CC↑	QNR↑	D_s↓	D_lambda↓	CC↑
PCA	0.806	0.167	0.033	0.952	0.737	0.249	0.021	0.949
PRACS	0.891	0.096	0.014	0.872	0.789	0.196	0.019	0.967
Indusion	0.789	0.134	0.089	0.890	0.790	0.166	0.054	0.883
IHS	0.782	0.192	0.033	0.954	0.718	0.266	0.027	0.944
Brovey	0.769	0.169	0.074	0.770	0.645	0.242	0.146	0.592
PNN_indices_L1	0.850	0.061	0.095	0.935	0.864	0.065	0.077	0.946
PNN_indices_L2	0.844	0.062	0.100	0.931	0.878	0.066	0.061	0.941
PNN_noindices_L1	0.845	0.092	0.069	0.954	0.884	0.072	0.048	0.952
PNN_noindices_L2	0.846	0.085	0.075	0.952	0.855	0.086	0.065	0.948
MSDCNN_indices_L1	0.848	0.060	0.097	0.941	0.875	0.059	0.071	0.946
MSDCNN_indices_L2	0.843	0.063	0.101	0.929	0.884	0.059	0.061	0.941
MSDCNN_noindices_L1	0.852	0.091	0.063	0.953	0.891	0.064	0.048	0.951
MSDCNN_noindices_L2	0.835	0.087	0.085	0.952	0.855	0.082	0.068	0.949
PSGAN_L1	0.880	0.048	0.076	0.936	0.869	0.038	0.096	0.937
PSGAN_L2	0.815	0.064	0.129	0.942	0.869	0.038	0.097	0.944

Table 7. Correlation Coefficient results of Worldview 2 and Worldview 3 dataset for different land cover schemas.

Methods/ Bands	Agriculture andVegetation				Urban				Water			
	Red	Green	Blue	NIR	Red	Green	Blue	NIR	Red	Green	Blue	NIR
PCA	0.987	0.999	0.995	0.703	0.972	0.941	0.960	0.598	0.985	0.985	0.992	0.875
PRACS	0.999	0.998	0.998	0.997	0.960	0.949	0.956	0.989	0.996	0.995	0.995	0.989
Indusion	0.799	0.775	0.791	0.772	0.892	0.892	0.893	0.889	0.940	0.939	0.939	0.958
IHS	0.934	0.871	0.841	0.995	0.897	0.846	0.812	0.962	0.880	0.838	0.794	0.996
Brovey	0.979	0.925	0.879	0.979	0.921	0.877	0.844	0.916	0.937	0.884	0.819	0.987
PNN_indices_L1	0.986	0.982	0.984	0.994	0.963	0.961	0.975	0.961	0.965	0.964	0.967	0.994
PNN_indices_L2	0.983	0.981	0.959	0.995	0.972	0.967	0.968	0.972	0.969	0.966	0.956	0.995
PNN_noindices_L1	0.984	0.980	0.985	0.993	0.961	0.961	0.965	0.964	0.967	0.967	0.968	0.994
PNN_noindices_L2	0.978	0.969	0.963	0.996	0.975	0.966	0.966	0.970	0.970	0.963	0.959	0.996
MSDCNN_indices_L1	0.986	0.981	0.983	0.993	0.964	0.960	0.976	0.961	0.968	0.965	0.973	0.994
MSDCNN_indices_L2	0.984	0.983	0.962	0.995	0.975	0.969	0.969	0.975	0.967	0.969	0.953	0.996
MSDCNN_noindices_L1	0.984	0.977	0.985	0.993	0.962	0.961	0.967	0.963	0.967	0.965	0.969	0.994
MSDCNN_noindices_L2	0.977	0.970	0.963	0.996	0.973	0.967	0.966	0.970	0.970	0.963	0.959	0.996
PSGAN_L1	0.955	0.956	0.927	0.991	0.980	0.976	0.981	0.986	0.977	0.972	0.974	0.994
PSGAN_L2	0.882	0.913	0.877	0.964	0.957	0.949	0.958	0.970	0.963	0.963	0.963	0.994

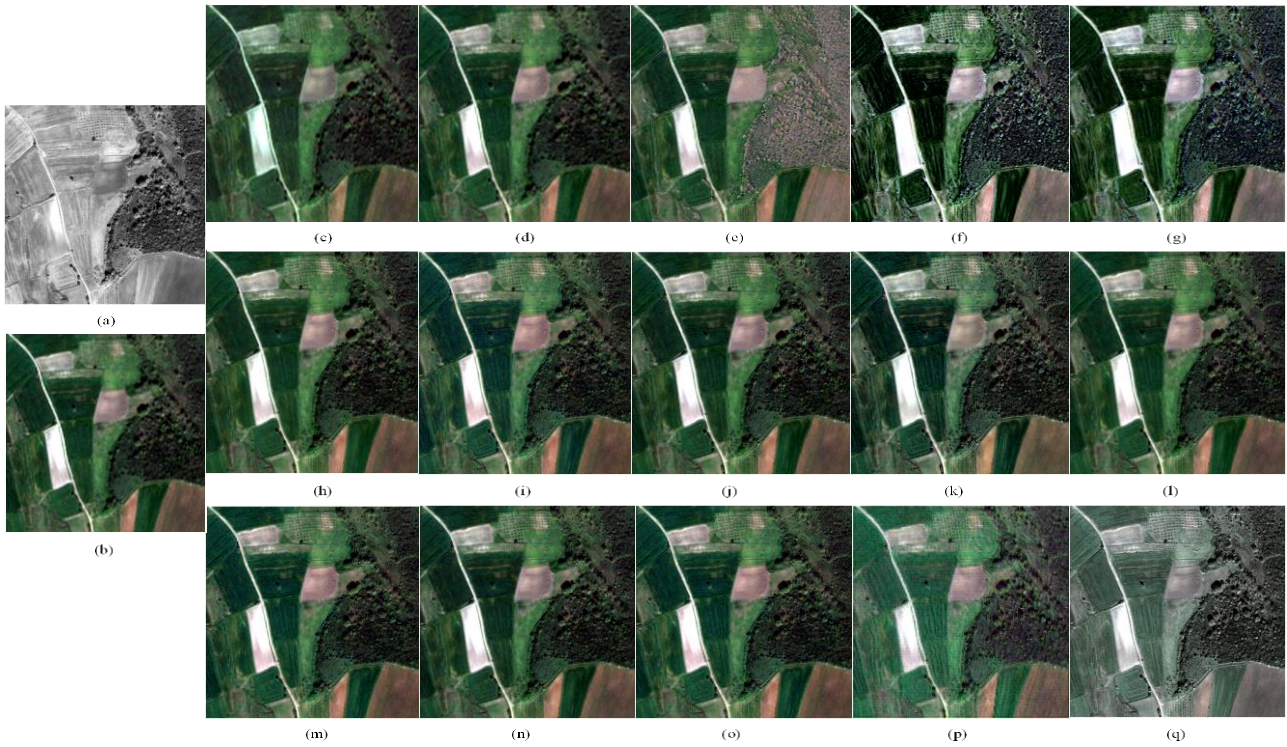


Figure 7. Full resolution experiment results for agricultural lands and vegetated areas. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

According to QNR and D_Lambda metrics PRACS, Indusion and PSGAN_L1 methods provided the highest performance with similar scores. Regarding the DL methods, each method with the L1 loss function

provided better performance than the L2 loss function for QNR and Lambda metrics. On the other hand, the L2 Loss function provided better performance than the L1 Loss function for the D_s metric.

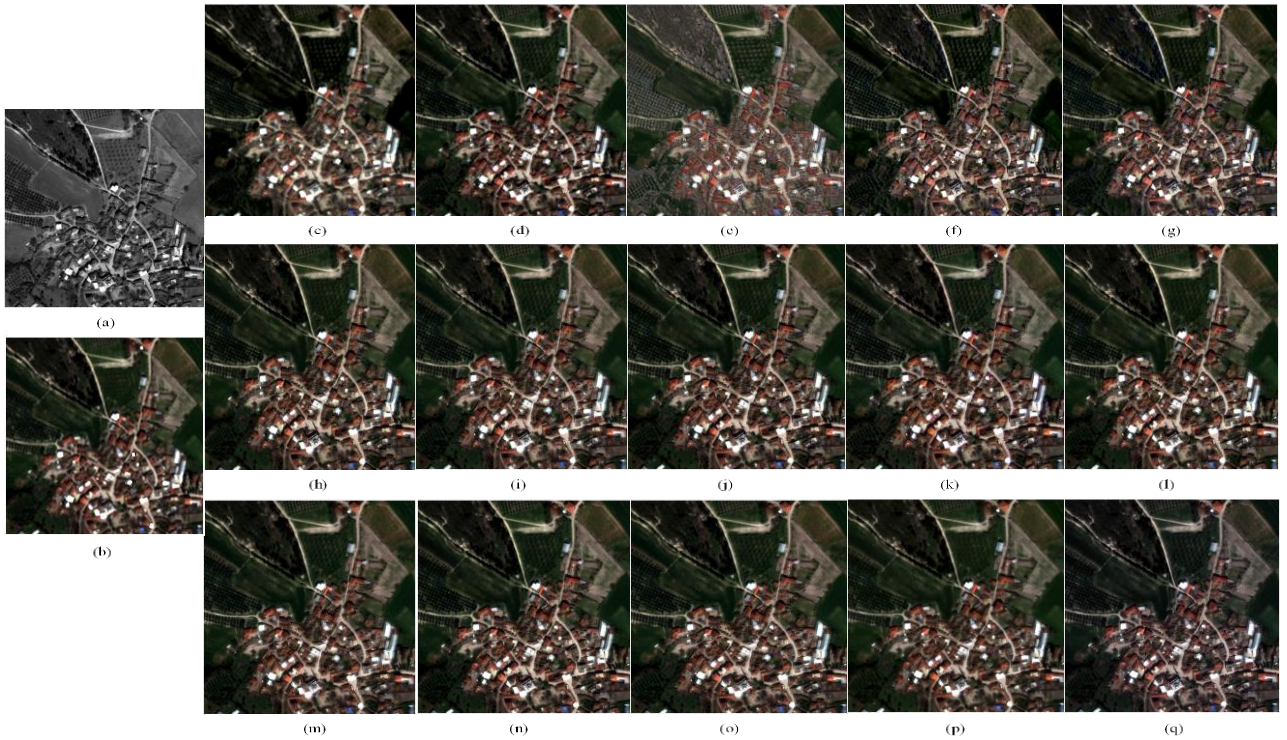


Figure 8. Full resolution experiment results for urban areas. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

Table 5 presents the CC metrics derived from all patches in the independent full resolution dataset. According to this evaluation, the PRACS method provided the highest CC value for all bands and for average. The deep learning-based methods, especially with the L1 loss function, followed the PRACS with very close CC values. The quantitative assessment results of Worldview-2 and Worldview-3 images including the average of CC metric are presented in Table 6. According to results of Worldview dataset, the PRACS method provided the highest performance for traditional methods. There is an obvious improvement with DL based methods for the QNR metric and D_s metric, while worse results achieved for D_{lambda} metric. Moreover, DL methods provided stable and high CC values around 0.95, however CC is variable across the traditional methods.

Visual interpretation and CC based assessment

Similar to the reduced resolution data experiments, three patches were selected to visually interpret the performance of methods. In addition, CC metrics were calculated for each band of the patches. The first patch covers agricultural lands and forest areas, the second patch covers a heterogeneous region with urban areas, agricultural lands and roads and the third patch cover bare lands, vegetation, small housings, and water body. For the first patch, the PRACS and PCA provided high spectral similarity with the original MS data, however, the spatial detail injection performance was lower when compared to DL methods (Figure 7 (c, d)). The remaining conventional methods failed both in spectral and spatial aspects for this patch (Figure 7 (e, f, and g)).

For DL methods, PNN and MSDCNN methods provided good performance in spectral similarity and results with L1 loss function provided sharper images. PSGAN_L1 method also provided good spatial detail injection, but colour tone deformation was evident. The CC metrics are in agreement with the visual interpretation results and PCA and PRACS provided highest scores, however, PNN, MSDCNN and PSGAN_L1 methods also provided very close values to these methods (Table 7).

For the second patch, all of the DL methods provided good balance between spatial detail injection and spectral conservation, while no indices and L1 loss function combinations provided slightly better performance especially for the textural information of the forest area and colour representation of the bare lands (Figure 8 (h-q)). According to the evaluation of traditional methods, PCA method suffered from blur effect and Indusion resulted in colour distortion, while remaining methods provided similar performance with deep learning based methods (Figure 8 (c-g)). The deep learning based methods provided high CC values where PSGAN_L1 provided highest scores in RGB bands. The PRACS provided similar CC performance with the DL methods (Table 7).

In the last patch, PRACS and PCA methods provided balanced trade-off between the spectral similarity and spatial detail injection, while Indusion method suffered from colour distortion and HIS, Brovery methods injected spatial detail in a way that unwanted fine – grained texture is evident in the visuals (Figure 9 (c-g)).

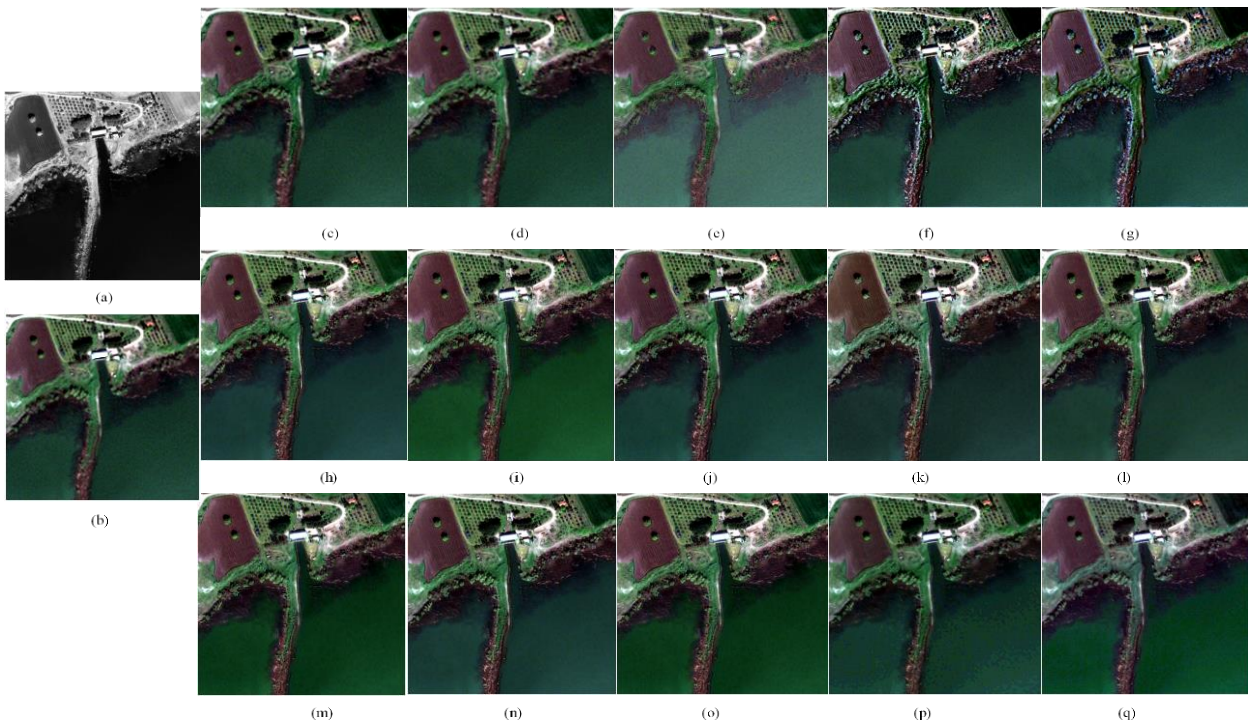


Figure 9. Full resolution experiment results for water areas. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.



Figure 10. Full resolution experiment results for Worldview-2 image patch. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

DL methods provided satisfactory results; however, there is a colour distortion with L2 loss function, which results in greenish looks (Figure 9 (h-q)). According to CC metrics, PRACS provided high values for all bands followed by PCA for only RGB bands. The Indusion method also provided very high CC values, which

contradicts with visual inspection. DL methods mostly provided high values and stable characteristics through all bands (Table 7). When the results obtained from Worldview dataset are investigated, it can be stated that PNN and MSDCNN architectures provided similar performances when compared to Pleiades dataset.

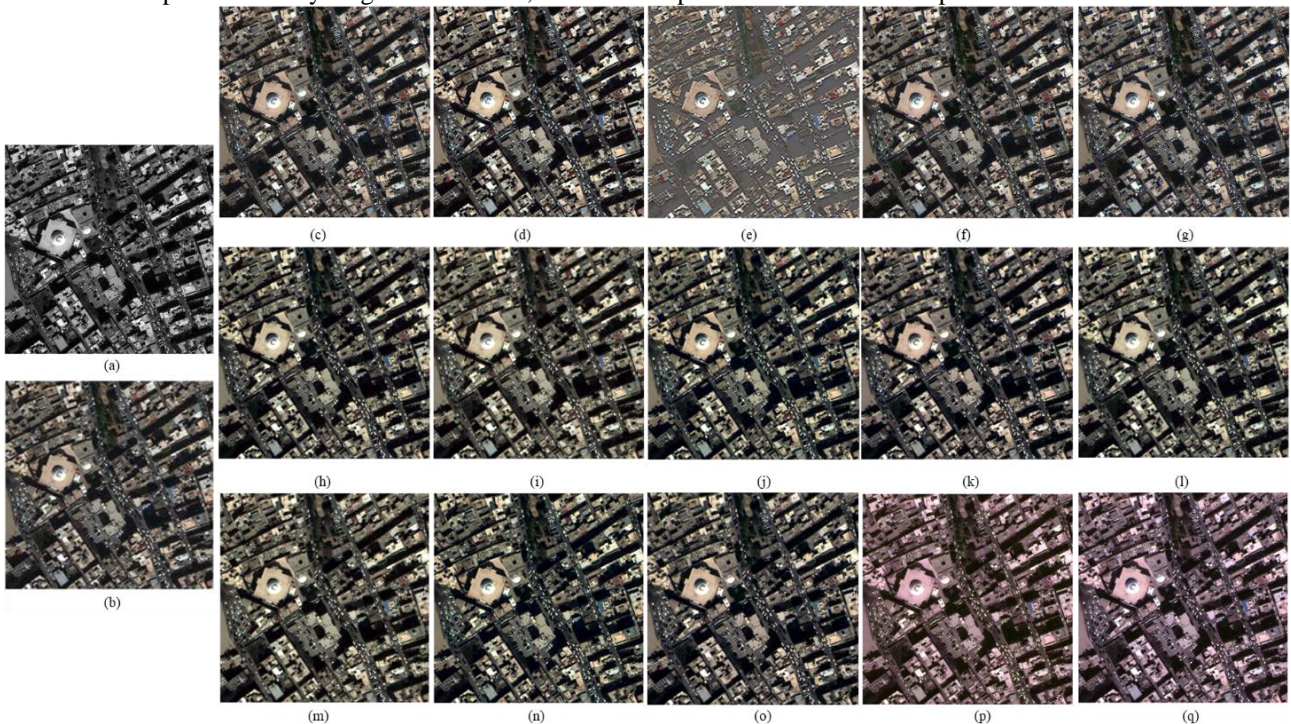


Figure 11. Full resolution experiment results for Worldview 3 image patch. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

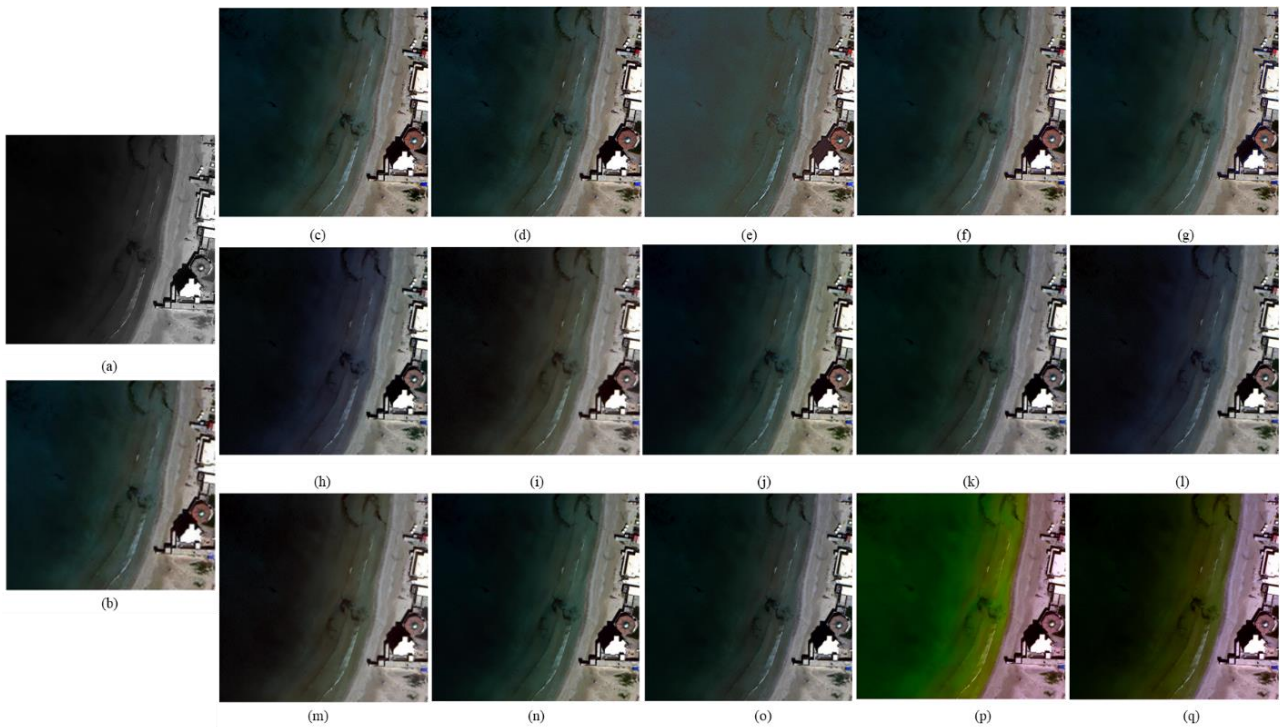


Figure 12. Full resolution experiment results for Worldview-3 image patch. (a) PAN Image; (b) Original MS; (c) PCA MS; (d) PRACS MS; (e) Indusion MS; (f) IHS MS; (g) Brovey MS; (h) PNN-indices-L1; (i) PNN-indices-L2; (j) PNN-noindices-L1; (k) PNN-noindices-L2; (l) MSDCNN-indices-L1; (m) MSDCNN-indices-L2; (n) MSDCNN-noindices-L1; (o) MSDCNN-noindices-L2; (p) PSGAN-L1; (q) PSGAN-L2.

Thus, it provided good level of transferability to the new sensor dataset with different spatial resolutions and spectral characteristics (Figures 10, 11 and 12). On the other hand, PSGAN method suffered from colour distortion, which indicates the lack of transferability to a new sensor data with current training schema.

Discussion and Conclusion

The results of the research indicated that DL methods provide more stable and balanced results in both spectral and spatial domains. On the other hand, the performance of the convolutional methods is variable across different landscape characteristics. Secondly, the addition of the spectral indices to the training dataset did not provide remarkable improvement in the learning according to both reduced resolution and full resolution experiments. The results showed that the L1 loss function is more effective than the L2 loss function in training. The PNN and modified version of MDSCNN provided satisfactory performance in most of the experiments but faced some blur effect in few cases. The PSGAN_L1 algorithm provided satisfactory and promising results with balanced performance in different conditions and high and stable CC values in all bands of the image. However, with independent test with Worldview dataset, it faced colour distortion problems, which can be considered as lack of transferability over different sensor data. Nevertheless, when trained with data from different sensors, PSGAN can be a good candidate for pan-sharpening algorithm for satellite images that includes more than four multispectral bands, such as Landsat 8 OLI.

The deep learning-based approaches are quite promising in various image processing problems including the pan-sharpening, which plays a very important role in the remote sensing field. In this research, the most popular deep learning-based pan-sharpening approaches trained with Pleiades VHR satellite images to obtain pan-sharpened images. Besides, several traditional algorithms were applied to the dataset to perform a comparative evaluation. The accuracy analysis was performed on reduced resolution and full resolution datasets that are independent of the training dataset. The results were evaluated by use of several accuracy indices and visual interpretation. Results informed that the PNN, modified MDSCNN and PSGAN provided promising results. The advantage gained by the deep learning methods is their balanced performance in the spectral and spatial domain across different landscapes and high correlation with original multispectral data for all bands, which is not the case for conventional methods in general. PRACS method performed the best among traditional algorithms according to quantitative and visual evaluation results and provided balanced performance except some blur problems in few cases. Further studies are planned to enlarge the training dataset with images from different satellites and construct a deep learning-based pan-sharpening framework that can deal with data from various sensors.

Acknowledgements

The authors acknowledge the support of the ITU Center for Satellite Communications and Remote Sensing (ITU-CSCRS) by providing Pleiades satellite images for this research.

References

- Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A. (2002). Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing*. doi.org/10.1109/TGRS.2002.803623.
- Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., Selva, M. (2003). An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas. *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, URBAN 2003*. doi.org/10.1109/DFUA.2003.1219964
- AIRBUS. (2020). Pleiades Products. <https://www.intelligence-airbusds.com/optical-and-radar-data/#pleiades>
- Chavez Jr., P. S., Yaw Kwarteng, A. (1989). Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogrammetric Engineering and Remote Sensing*.
- Cheng, M., Wang, C., Li, J. (2014). Sparse representation based pansharpening using trained dictionary. *IEEE Geoscience and Remote Sensing Letters*. doi.org/10.1109/LGRS.2013.2256875
- Choi, J., Yu, K., Kim, Y. (2011). A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*. doi.org/10.1109/TGRS.2010.2051674
- Christian Ledig Ferenc Huszar, L. T. (2017). *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. 19.
- Çelik, O., Gazioğlu, C. (2020). Coastline Difference Measurement (CDM) Method. *International Journal of Environment and Geoinformatics*, 7(1), 1-5. doi.10.30897/ijegeo.706792.
- DigitalGlobe. (2020). Tools and Resources. <http://www.digitalglobe.com/resources>.
- Dong, C., Loy, C. C., He, K., Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi.org/10.1109/TPAMI.2015.2439281
- Erdem, F., Avdan, U. (2020). Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery. *International Journal of Environment and Geoinformatics*, 7(3), 221-227. doi.10.30897/ijegeo.684951
- Gao, B. (1996). NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sensing of Environment*, 266(April), 257–266. doi.org/10.1016/S0034-4257(96)00067-3
- Gillespie, A. R., Kahle, A. B., Walker, R. E. (1987). Color enhancement of highly correlated images. II. Channel ratio and “chromaticity” transformation techniques. *Remote Sensing of Environment*. [https://doi.org/10.1016/0034-4257\(87\)90088-5](https://doi.org/10.1016/0034-4257(87)90088-5)
- Goodfellow, I. J., and Pouget-Abadie, J. (2014). Generative Adversarial Nets. *Veterinary Immunology and Immunopathology*, 155(4), 270–275. <https://doi.org/10.1016/j.vetimm.2013.08.005>
- He, K., Zhang, X., Ren, S., Sun, J. (2015). *Deep Residual Learning for Image Recognition*. doi.org/10.1109/CVPR.2016.90
- Kang, X., Li, S., Benediktsson, J. A. (2014). Pansharpening with Matting Model. *IEEE Transactions on Geoscience and Remote Sensing*. doi.org/10.1109/TGRS.2013.2286827
- Khan, M. M., Chanussot, J., Condat, L., Montanvert, A. (2008). Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience and Remote Sensing Letters*, 5(1), 98–102. doi.org/10.1109/LGRS.2007.909934
- Laben, C. Brower, B. . (2000). Process for enhancing the spatial resolution of multispectral imagery using pansharpening. *United States Patent* 6. [https://doi.org/10.1074/JBC.274.42.30033.\(51\)](https://doi.org/10.1074/JBC.274.42.30033.(51))
- Liu, J. G. (2000). Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*. doi.org/10.1080/014311600750037499
- Liu, J. G. (2010). *Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details*. 1161. doi.org/10.1080/014311600750037499
- Liu, X., Wang, Y., and Liu, Q. (2018a). PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpener. 1–5. doi.org/10.1109/ICIP.2018.8451049
- Liu, X., Wang, Y., Liu, Q. (2018b). Remote Sensing Image Fusion Based on Two-Stream Fusion Network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10704 LNCS, 428–439. https://doi.org/10.1007/978-3-319-73603-7_35
- Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G. (2016). Pansharpening by convolutional neural networks. *Remote Sensing*. doi.org/10.3390/rs8070594
- Nencini, F., Garzelli, A., Baronti, S., Alparone, L. (2007). Remote sensing image fusion using the curvelet transform. *Information Fusion*. doi.org/10.1016/j.inffus.2006.02.001
- Nouri, H., Beecham, S., Anderson, S., and Nagler, P. (2013). High spatial resolution WorldView-2 imagery for mapping NDVI and its relationship to temporal urban landscape evapotranspiration factors. *Remote Sensing*, 6(1), 580–602. doi.org/10.3390/rs6010580
- Open Remote Sensing. (2016). A Critical Comparison among Pansharpening Algorithms. <http://openremotesensing.net/>
- Ozcelik, F., Alganci, U., Sertel, E., Unal, G. (2020). Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images via GANs. *IEEE Transactions on Geoscience and Remote Sensing*, 1–16. doi.org/10.1109/tgrs.2020.3010441
- Ozturk, O., Sarıtürk, B., Seker, DZ. (2020). Comparison of Fully Convolutional Networks (FCN) and U-Net

- for Road Segmentation from High Resolution Imageries. *International Journal of Environment and Geoinformatics*, 7(3), 272-279. doi.10.30897/ijgeo.737993
- Palsson, F., Ulfarsson, M. O., Sveinsson, J. R. (2020). Model-Based Reduced-Rank Pansharpening. *IEEE Geoscience and Remote Sensing Letters*. doi.org/10.1109/LGRS.2019.2926681
- Panchal, S., Thakker, R. (2015). Implementation and Comparative Quantitative Assessment of Different Multispectral Image Pansharpening Approaches. *Signal and Image Processing: An International Journal*, 6(5), 35–48. doi.org/10.5121/sipij.2015.6503
- Pohl, C., Van Genderen, J. L. (1998). Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. In *International Journal of Remote Sensing* (Vol. 19, Issue 5). doi.org/10.1080/014311698215748
- Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W.; Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. 1–8.
- Scarpa, G., Vitale, S., Cozzolino, D. (2018). Target-Adaptive CNN-Based Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*. doi.org/10.1109/TGRS.2018.2817393
- Shettigara, V. K. (1992). A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogrammetric Engineering and Remote Sensing*. doi.org/10.1038/050073a0
- Tensorflow. (2015). <https://www.tensorflow.org/>
- Tonbul, H., Kavzoglu, T. (2020). A Spectral Band Based Comparison of Unsupervised Segmentation Evaluation Methods for Image Segmentation Parameter Optimization. *International Journal of Environment and Geoinformatics*, 7(2), 132-139. doi.10.30897/ijgeo.641216
- Tu, T.-M., Su, S.-C., Shyu, H.-C., Huang, P. S. (2001). A new look at IHS-like image fusion methods. *Information Fusion*. doi.org/10.1016/S1566-2535(01)00036-7
- Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., Wald, L. (2015). A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*. doi.org/10.1109/TGRS.2014.2361734
- Wald, L., Ranchin, T., Mangolini, M. (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*.
- Wei, Y., Yuan, Q. (2017). Deep residual learning for remote sensed imagery pansharpening. *RSIP 2017 - International Workshop on Remote Sensing with Intelligent Processing, Proceedings*. doi.org/10.1109/RSIP.2017.7958794
- Yang, X., Jian, L., Yan, B., Liu, K., Zhang, L., Liu, Y. (2018). A sparse representation based pansharpening method. *Future Generation Computer Systems*, 88(June), 385–399. doi.org/10.1016/j.future.2018.04.096
- Yin, H. (2015). Sparse representation based pansharpening with details injection model. *Signal Processing*. doi.org/10.1016/j.sigpro.2014.12.017
- Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L. (2018). 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*. doi.org/10.1109/JSTARS.2018.2794888
- Yuhas, R., Goetz, A. F. H., Boardman, J. W. (1992). Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. *Summaries of the Third Annual JPL Airborne Geoscience Workshop, JPL Publ. 92–14, Vol. 1*.
- Zhong, S., Zhang, Y., Chen, Y., Wu, D. (2017). Combining component substitution and multiresolution analysis: A novel generalized BDSD Pansharpening algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. doi.org/10.1109/JSTARS.2017.2697445
- Zhou, W., Bovik, a C. (2002). A universal image quality index. *Signal Processing Letters, IEEE*. doi.org/10.1109/97.995823
- Zhu, X. X., Bamler, R. (2013). A sparse image fusion algorithm with application to pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2012.2213604>