Deep Learning Based Pansharpening Using a Laplacian Pyramid

Doyoung Jeong (1), Yongil Kim (1)

[1] Seoul National Univ., 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea
Email: tmits37@snu.ac.kr; yik@snu.ac.kr;

KEY WORDS: Pansharpening, Deep learning, Remote sensing, Very high spatial resolution imagery

ABSTRACT: Pansharpening is an image fusion technique aimed at acquiring high-resolution multispectral images by fusing a multispectral (MS) image and high-resolution panchromatic (PAN) images. Various types of pansharpening algorithm have been developed for the last decades, and in recent years, with remarkable growth in the field of deep learning, deep learning based methods have been presented. Because the scale of MS and PAN is different, MS needs to be interpolated with the scale of PAN as inputs for most pansharpening networks.

The interpolated MS has the same information as the original MS, but has a limitation in that the computation cost and the inherent error in the satellite image can be overestimated during the deep learning process. We propose a learning based approach to synthesize high-resolution MS based on convolutional neural networks. The proposed network consists of extracting and synthesizing high spatial frequency features through a separate network, rather than interpolated MS and PAN in an integrated network. The proposed method showed good performance in preserving the spatial characteristics but limited in preserving the spectral characteristics.

1. INTRODUCTION

Due to the trade-off relationship between the spatial resolution and the spectral resolution of the sensor limits the use of sensors that meet both condition. Most commercial optical very high spatial resolution satellites provide low resolution multi-spectral (LRMS) images and a high resolution panchromatic (PAN) images. Pansharpening is a kind of image fusion technique that combines the spatial information of PAN and the spectral information of LRMS to synthesize images with high spatial resolution and high spectral resolution. Pansharpening is widely used as preprocessing process for the application such as object detection, classification.

During the recent decades, various pansharpening algorithms have been developed. Most pansharpening algorithms can be divided into four categories according to the problem approach. (1) Component substitution (CS) method, (2) multiresolution analysis (MRA) method, (3) model-based optimization (MBO) based methods, (4) deep-learning based methods. The CS methods replace the spatial information of interpolated LRMS with the spatial information of PAN. Algorithms belonging to the CS method include Gram-Schmidt (GS), principal component analysis etc., which improve the spatial information of LRMS but could cause spectral distortion. The MRA method refers to a method of replacing missing information of LRMS with high frequency contents of PAN image. Examples of MRA method include Laplacian pyramids, modulation transfer function (MTF), and focus on injecting PAN information into decomposed LRMS information. The MRA method maintains high spectral consistency, but some spatial distortion may occur. The MBO based methods describe the assumed burring, downsampling, and noise using the objective functions.

All above-mentioned methods assume a linear model for the synthesis of LRMS and PAN images. Significant advances in deep learning can overcome the limitations of linear models through the introduction of nonlinear models using the convolutional neural networks (CNN).

2. RELATED WORK

2.1.Deep learning based Super-Resolution

Super-resolution refers to the process of reconstructing a HR image from a LR image. It improves the resolution of the images and applies it to real-world applications such as medical imaging, security among others (Zhang, 2010). In general, the problem is an ill-posed problem in that there are multiple HR images corresponding to a single LR image. Rapid development of deep learning techniques in recent year has led to the exploration of deep learning based super-resolution techniques and achieve the improvement of performance. Since Super Resolution CNN (SRCNN), which first introduced the early CNN in super-resolution, numerous CNN-based algorithms have been developed (Dong, 2014). Unlike the conventional methods, deep learning-based super-resolution reconstructs HR images directly from LR images through the networks. It is a technique to recover high spatial frequency information based on numerous training data. Because the performance of the network depends on learning processes, the number and quality of datasets could directly affect the results. Increasing the depth of the CNN allows the model to learn more complex and general relationships between LR images and HR images.

Since image super-resolution is an ill-posed problem, how to perform upsampling is the key problem (Wang, 2019). (Dong, 2014) first adopted the pre-upsampling SR frameworks. They proposed SRCNN to learn an end-to-end mapping from interpolated LR images to HR images. LR images is interpolated to the desired size using conventional methods (e.g., bicubic interpolation) as the size of HR, so they proposed a network with the same input and output size. The advantage of this framework is that the upsampling process is done in predefined methods, thus reducing the learning difficulties. However, pre-upsampling methods have been introduced some side effects such as noise amplification and blurring. Besides, since most operations are performed in high-dimensional space, the cost of time and space is much higher than other frameworks (Wang, 2019).

Post-upsampling methods is proposed to solve the problem of computational efficiency of pre-upsampling. The methods perform most of the nonlinear mapping in low-dimensional space, the images are upsampled at the end of the network by introducing learnable upsampling layers. Through post-upsampling methods, the computation complexity and spatial complexity is much reduced, it improves the training speed and inference speed. For this reason, the framework has emerged as the one of the mainstream framework in super-resolution.

2.2.Deep Learning based pan-sharpening

Pansharpening is a special form of super-resolution (Zhang, 2019). After excellent achievement in super-resolution field, several deep learning-based pansharpening techniques have been develop that apply SR’s methodology. Pansharpening Neural Networks (PNN) was the first attempt to solve the pansharpening problem by introducing the network as SRCNN (Yang, 2017). PNN was made by stacking three CNN layers, which, like SRCNN, had shallow networks and had limitations, leaving room for improvements. Residual learning proposed by ResNet has shown great performance and used to alleviate the degradation problem by increasing network depths and improve learning ability. Multi-Scale and Multi-Depth Convolutional Neural Network (MSDCNN) is a residual learning based model to solve the pansharpening problem. MSDCNN proposed by (Yuan, 2018) was able to improve the fusion performance by enabling more complex nonlinear mapping through residual learning. More recently, (Guo, 2019) proposed pansharpening network based on a dilated multilevel block to use of the extracted features and enlarge the receptive field. The pansharpening techniques described above are all end-to-end mapping from interpolated LRMS to HRMS. In the preparation phase, LRMS needs to upsample to the size of PAN image using bicubic interpolation. then, the upsampled LRMS images is concatenated with the PAN image to compromise input. The output of the network is HRMS images with the same spatial resolution as the PAN images. It is known that noise amplification is introduced in pre-upsampling in super-resolution fields. Deep learning based pansharpening networks use satellite images as input with a lower signal-to-noise ratio (SNR) than the dataset of super-resolution, thereby maximizing the problem caused by noise amplification. Pre-upsampling process has advantages of simplifying the network learning, but suffer from the computation complexity and the cost of time.

3. METHODOLOGY

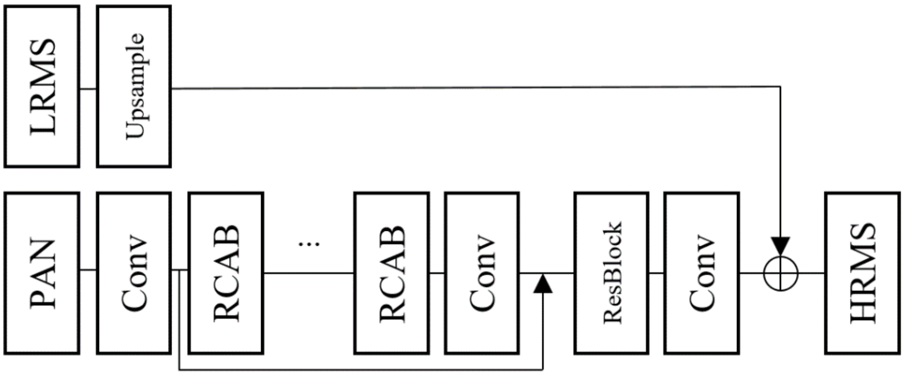

Figure 1. Structure of the proposed network

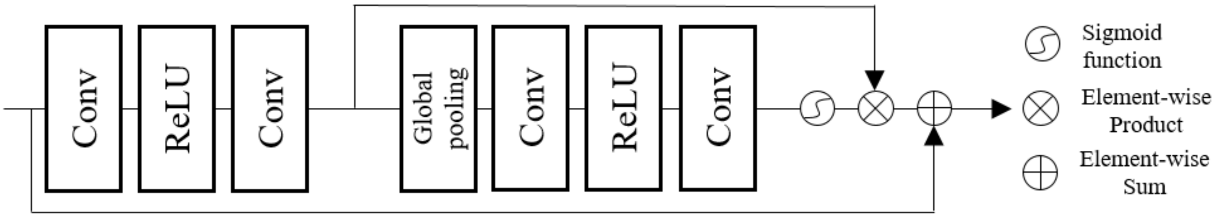

Figure 2. Residual Channel Attention Block (RCAB)

In most earth observation satellites, the spatial resolution of LRMS is four times that of the PAN image. For the design of the end to end network, most deep-learning based pansharpening algorithms need to Upsample LRMS images to the scale of the PAN image and use them as input. But, conventional upsampling the MS images definitely can lead to severe errors such as noise amplification. the LRMS image is noisy compared to PAN due to higher MTF and lower SNR than the PAN image. Comparing the simulated PAN image with the original LRMS image with the same spatial resolution can confirm that the simulated PAN image contains noticeably high spatial frequency information. The MRA method also generate HRMS through the synthesis of bicubic interpolated LRMS and modified PAN information. Therefore, the main purpose of the proposed network is to extract the PAN information to inject into the interpolated LRMS.

The acquired LRMS (m) and PAN (p) are assumed as degraded observations of the high-resolution multispectral imagery (HRMS, x), the degradation process can be modelled as follows:

$$\begin{cases} m = (x * k) \downarrow_4 + \varepsilon_{MS} \\ p = x * H + \varepsilon_{PAN} \end{cases} \tag{1}$$

where, $x * k$ represents the convolution between HRMS and a blurring kernel k, $\downarrow_4$ means the process of downsampling with a scale of 4; H is a spectral response matrix, ε_{MS} and ε_{PAN} are additive noise. For the MRA based method, the composition of PAN to LRMS is derived via linear decomposition, and the general form of the MRA based pansharpening algorithms is defined as:

$$\hat{x} = y_k + w_k(p - p * h) \tag{2}$$

where $\hat{x}$ and $\mathrm{y_k}$ are the k-th band of HRMS and LRMS, and $\mathrm{w_k}$ represents the injection gain of the k-th band, h is the corresponding decomposition operator.

Since the spatial resolution of PAN and LRMS is different, the two images are designed to pass through different network passes. Similar with MRA-based methods, the proposed network aims to extract the information to be injected by training the PAN images into the network alone. LRMS upsamples with the scale of PAN using sub-pixel convolution, a learned upsampler, and serves as a guide for color information by adding it to PAN's high frequency extraction network. (Zhang, 2018)

Inspires by (Zhang, 2018), a similar type of network to residual channel attention networks (RCAN) was established in the process of extraction of high spatial frequency of the PAN images. The core of RCAN is the introduction of residual channel attention blocks). The high-frequency components would usually be regions, being full of edges, textures, and other details (Zhang, 2018). RCAB ensures to fully capture channel-wise dependencies, making it useful to learn nonlinear interactions between channels. RCAM was selected to simulate the nonlinear process of the decomposition of a PAN image and to train $\mathrm{w_k}$ at each step. To extract the PAN information to be injected into the MS, we combine the shallow information passed through the initial CNN and the deep information passed through the RCAB. Finally, HRMS is reconstructed by adding the extracted PAN information to the upsampled LRMS.

4. EXPERIMENTAL RESULTS

4.1 Input data

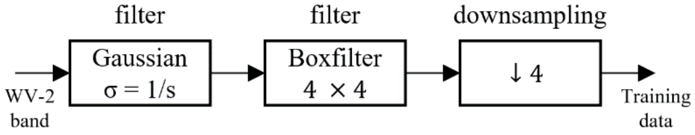

Figure 3. The downsampling process used for simulating the data

We use very high resolution remote sensing data from WorldView-2 (WV-2). WV-2 satellite sensor launched October 8, 2009 provides a high resolution panchromatic at 0.46m and 1.84m 8 bands multispectral imagery. The imagery will be made available commercially by resampling as 0.5m and 2.0m. We can acquire satellite imagery from DigitalGlobe corresponding the two product levels, Standard and Ortho-standard. Standard products are 16bit images with radiometrically and sensor-corrected and georectified processing. Ortho-standard products are 8bit images that processed with oorthorectified in addition to processing of standard products. We used 6 images of WV-2 for training process. Direct training of the CNN model is not possible due to the absence of the HRMS image which is a true image to use as a label for training. We followed Wald's protocol for network training and experiment simulation (Wald, 1997).

Inspired by (Lanaras, 2018), we downsample the original WV-2 data, by first blurring it with a Gaussian filter of standard deviation σ, emulating the modulation transfer function (MTF) of WV-2. From (Poli, 2015), we get a range of 0.46~0.49 for the point spread function (PSF) of MS bands and of 0.57~0.63 for that of the PAN, which is given the relation $\mathrm{psf} = \sqrt{-2\log(mtf)/\pi^2}$. Then, we downsample by averaging over s x s windows with s = 4. In this way, we obtain two datasets for training, validation and testing.

4.2. Implementation details

We sample 9 representative scenes of WV-2 standard products, 7 for training and 2 for testing. For training, we sample 8,000 random patches per training image. Out of theses patches to 90% are used for training the weights, and remaining 10% for validation process. To test the network, we run on 2 images. Following Wald Protocol, simulated LRMS and PAN images were used as inputs, and original LRMS images were used as reference images.

The network is implemented with PyTorch as back-end, and training is run on a NVIDIA GTX 1080Ti. The patch size of MS images and PAN images for training were 32 x 32 and 144 x 144 pixels, the mini-batch size of ADAM is set to 32, the learning rate lr =1e-4,

4.3. Results and Comparison

Table 1. Evaluation Metrics for performance comparison of Pansharpening algorithms

	Evaluation Metrics
QNR	$\mathrm{QNR} = (1 - \mathrm{D}_\lambda)(1 - D_s)$ $\mathrm{D}_\lambda = \frac{1}{B(B-1)} \sum_{b=1}^{B} \sum_{l=1(b\neq 1)}^{B} \lvert Q(x_b, x_l) - Q(\widetilde{x_b}, \tilde{x}_b) \rvert$ $\mathrm{D}_\mathrm{s} = \frac{1}{B} \sum_{b=1}^{B} \lvert Q(x_b, P) - Q(\widetilde{x_b}, \tilde{P}) \rvert$
ERGAS	$\mathrm{ERGAS100} \triangleq \frac{d_h}{d_l} \sqrt{\frac{1}{B} \sum_{b=1}^{B} (\frac{RMSE(b)}{\mu(b)})^{2}}$
CC	$\mathrm{CC} = \frac{\sum_{m=1}^{M} (P_m - \overline{p_m}) \times (R_m - \overline{R_m})}{\sqrt{\sum_{m=1}^{M} (P_m - \overline{p_m})^{2} \times \sum_{m=1}^{M} (R_m - \overline{R_m})^{2}}}$
SAM	$\mathrm{SAM}(\mathrm{v}, \hat{v}) \triangleq \arccos(\frac{(\mathrm{v}, \hat{v})}{\lVert v \rVert_2 \times \lVert \hat{v} \rVert_2})$
Q4	$\mathrm{Q}_4 = \frac{\lvert \sigma_{z1}\sigma_{z2} \rvert}{\sigma_{z1}\sigma_{z2}} \times \frac{2\sigma_{z1} \times \sigma_{z2}}{\sigma_{z1}^2 + \sigma_{z2}^2} \times \frac{2\lvert \overline{z_1} \rvert \times \lvert \overline{z_2} \rvert}{\lvert \overline{z_1} \rvert^2 + \lvert \overline{z_2} \rvert^2}$

After improving spatial resolution to original resolution through application of several pansharpening algorithms, the relative performance of the algorithm was numerically compared by using original LRMS images as reference data. In this experiments, four metric shown in the table were used to validate the simulated data: the *Erreur Relative Globale Adimensionnelle de Synthese* (ERGAS), the correlation coefficient (CC), the spectral angle mapper (SAM) and the universal image quality metric (Q).

For the numeric and visual assessment, we will compare the trained network with other widely accepted pansharpening algorithms. We will evenly selectthe four classification algorithms of the pansharpening described earlier: the GS method (Laben, 2000), the Partial Replacement Adaptive Component Substitution (PRACS) method (Choi, 2010) belonging to CS, the MTF-based generalized Laplacian pyramid (MTF-GLP) (Aiazzi, 2006), the Indusion (Li, 2008) method belonging to the MRA method, l1/2 gradient based (l1/2) model (Zheng, 2016) belonging to the MBO method, MSDCNN, GUO belonging to deep learning based method.

Figure 4. and Table 2. shows an example of the experiments performed on the simulated data. For visualization, the red, green, and blue bands were chosen for display. Figure 4. (a)~(e) shows the results of the conventional pansharpening methods, and Figure 4. (f)~(h) shows the results of the deep-learning based pansharpening methods. The PRACS, MTF-GLP, l1/2, MSDCNN, GUO and methods preserve spectral information well. In particular, deep-learning

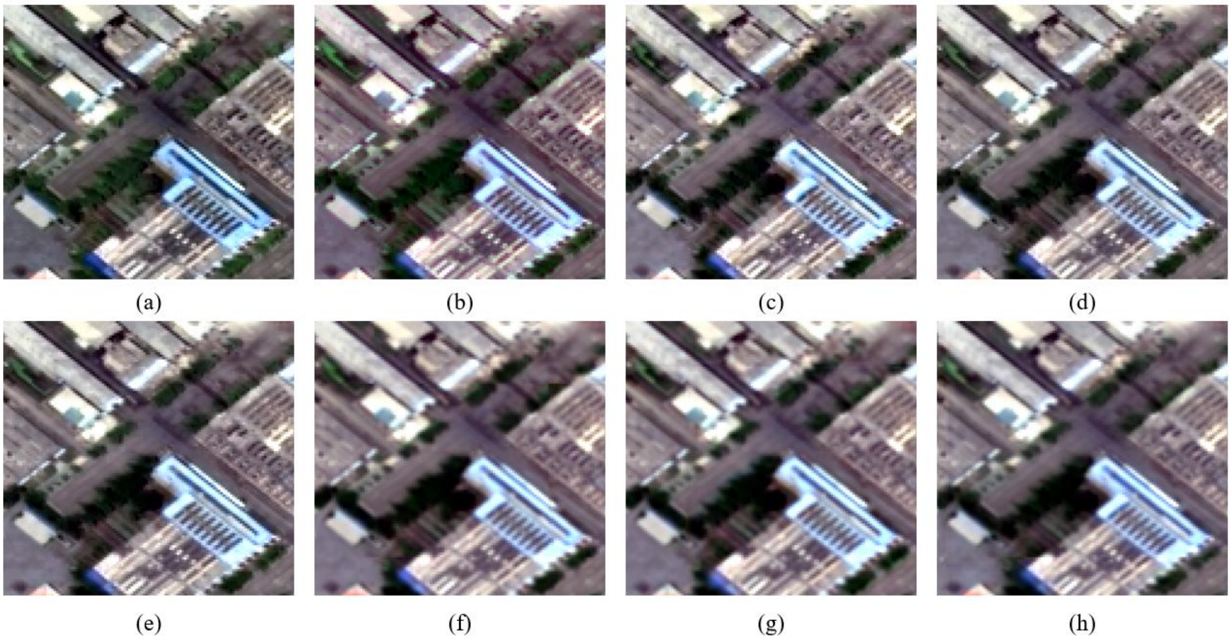

Figure 5. Resuts of the simulated experiment on city area, which was extracted from a WV-2 image of Ansan, South Korea, obtained in 2018. **(a)** GS; **(b)** PRACS; **(c)** Indusion; **(d)** MTF-GLP; **(e)** l1/2; **(f)** MSDCNN; **(g)** Guo; **(h)** Proposed.

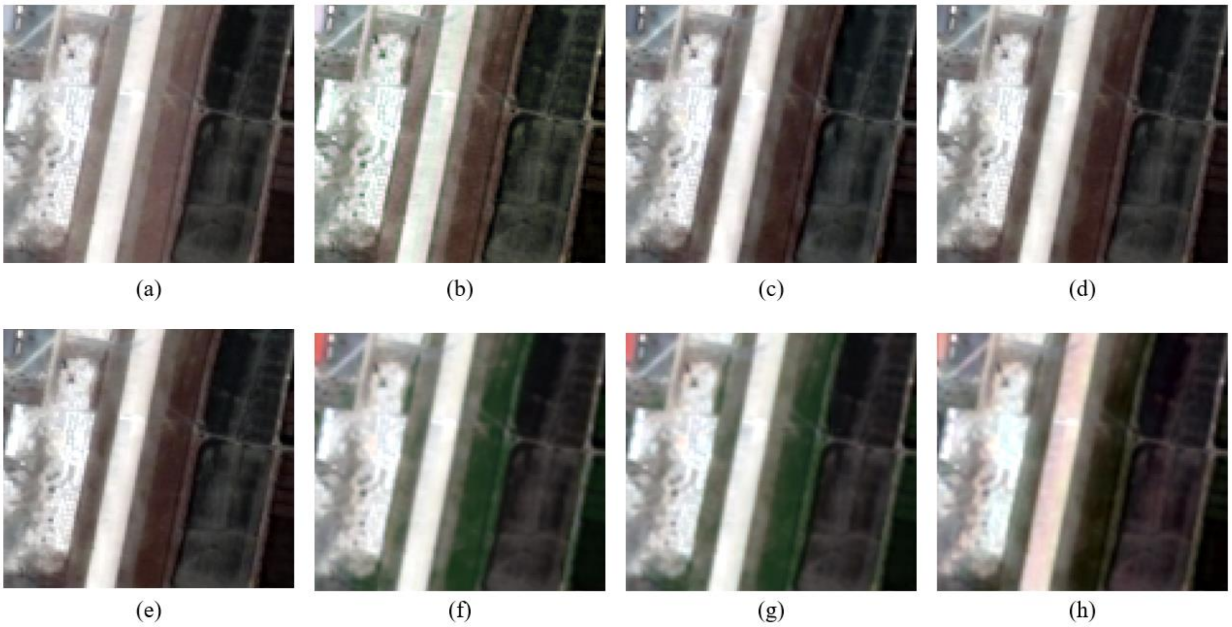

Figure 4. Resuts of the full-resolution experiment on city area, which was extracted from a WV-2 image of Bucheon, South Korea, obtained in 2010. **(a)** GS; **(b)** PRACS; **(c)** Indusion; **(d)** MTF-GLP; **(e)** l1/2; **(f)** MSDCNN; **(g)** Guo; **(h)** Proposed.

based methods have shown a common strength in preserving spectral information, but produce different levels of blurring effects. The proposed method achieved the best CC, which appears to have been able to successfully extract spatial information from the PAN. However, the metrics that represent the spectral characteristics such as ERGAS, SAM, Q4 are lower than other deep-learning based methods. This is because the process of injecting bicubic interpolated LRMS, a common characteristic of other networks was omitted. A fall in spectrometric indicators means a lack of variables in the injection process of MS and PAN. It is considered that the injection process of the current network may improve the spectral characteristics if the patches are considered in

Table 2. Performance indicators of simulated data

	ERGAS	CC	SAM	Q4
GS	4.8896	0.9876	0.7066	0.6207
PRACS	3.6122	0.9922	**0.7767**	0.6604
Induison	3.7790	0.9928	0.5878	0.6312
MTF-GLP	3.1565	0.9952	0.5446	0.6616
l1/2	3.2029	0.9951	0.5454	0.6595
MSDCNN	**2.8067**	0.9960	0.7185	**0.7245**
Guo	2.9444	0.9956	0.7353	0.7051
Proposed	3.1069	**0.9964**	0.6120	0.6978

Table 3. Performance indicators of real data

	QNR	D_s	D_λ
GS	0.8052	0.1554	0.0467
PRACS	**0.8753**	0.0948	**0.0330**
Induison	0.8667	**0.0309**	0.0530
MTF-GLP	0.8717	0.0685	0.0642
l1/2	0.8732	0.0677	0.0634
MSDCNN	0.8686	0.0696	0.0464
Guo	0.8666	0.0691	0.0691
Proposed	0.8587	0.0624	0.0842

the injection process through stacking of the CNN layers.

Figure 5. and Table 3. shows an example of the experiments performed on the real data. Fusion quality was assessed through QNR due to absence of reference data. When comparing only deep learning-based algorithms, the proposed network preserves the spatial characteristics well, but the preservation of spectral characteristics is significantly lower than other methods. As aforementioned, the same problem appears to be repeated as the injection network is shallow.

The PRACS, MTF-GLP and l1/2 shows good quality. Interestingly, the pattern of performance for real data is different from that of simulated data. Algorithms that show good performance are generally those that belong to the conventional method. This seems to be because the simulation process does not capture the actual blurring and downsampling process well. In the simulation process, only the Gaussian blurring imitating the MTF of the sensor was considered, but other noise effects such as fixed pattern noise and other random noise are affected. Other noise is recongnized as the true value of the earth irradiation because only the Gaussian blurring is assumed. Not only proposed methods but also all other deep-learning based pansharpening poses the problem. The performance for simulated data is high, but the performance for real data is low, so the actual usability of the proposed network developed can be reduced.

5. CONCLUSION

In this article, an end-to-end deep learning network is introduced to solve pre-upsampling of LRMS for preventing noise amplification etc. Inspired by MRA-based network, the proposed network consists of extracting and synthesizing high spatial frequency features through a separate network, rather than upsampled LRMS and PAN in an integrated network. The proposed method showed good performance in preserving the spatial characteristics but limited in preserving the spectral characteristics. This study is a pilot study to implement a network that computes the spectral response matrix of PAN. Our future work will improve the synthesizing process to make appropriate simulated data and increase the depth of the injection network.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2019R1I1A2A01058144)

REFERENCE

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A. and Selva, M., 2006. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering & Remote Sensing*, *72*(5), pp.591-596.

Choi, J., Yu, K. and Kim, Y., 2010. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(1), pp.295-309.

Dong, C., Loy, C.C., He, K. and Tang, X., 2014, September. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision* (pp. 184-199). Springer, Cham.

Guo, Y., Ye, F. and Gong, H., 2019. Learning an efficient convolution neural network for pansharpening. *Algorithms*, *12*(1), p.16.

Laben, C.A. and Brower, B.V., Eastman Kodak Co, 2000. *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*. U.S. Patent 6,011,875.

Li, Z. and Leung, H., 2008. Fusion of multispectral and panchromatic images using a restoration-based method. *IEEE transactions on geoscience and remote sensing*, *47*(5), pp.1482-1491.

Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E., 2018, Schindler, K. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS J. Photogramm. Remote Sens.* *146*, 305–319

Poli, D., Remondino, F., Angiuli, E., & Agugiaro, G. (2015). Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *100*, 35-47.

Wang, Z., Chen, J. and Hoi, S.C., 2019. Deep learning for image super-resolution: A survey. *arXiv preprint arXiv:1902.06068*.

Wald, L., Ranchin, T., Mangolini, M. , 1997, Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* 19 63, 691–699

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X. and Paisley, J., 2017. PanNet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5449-5457).

Yuan, Q., Wei, Y., Meng, X., Shen, H. and 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*, *11*(3), pp.978-989.

Zhang, L., Zhang, H., Shen, H. and Li, P., 2010. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, *90*(3), pp.848-859.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European*

Zhang, Y., Liu, C., Sun, M. and Ou, Y., 2019. Pan-Sharpening Using an Efficient Bidirectional Pyramid Network. *IEEE Transactions on Geoscience and Remote Sensing*. *Conference on Computer Vision (ECCV)* (pp. 286-301).

Zeng, D., Hu, Y., Huang, Y., Xu, Z. and Ding, X., 2016. Pan-sharpening with structural consistency and ℓ1/2 gradient prior. *Remote sensing letters*, *7*(12), pp.1170-1179.