End-to-End Spectral-Temporal Fusion Using Convolutional Neural Network

Tayeb Benzenati[1,2(✉)], Abdelaziz Kallel[1,2], and Yousri Kessentini[1,3]

[1] Digital Research Center of Sfax, 3021 Sfax, Tunisia
[2] Advanced Technologies for Image and Signal Processing, University of Sfax, Sfax, Tunisia
[3] MIRACL Laboratory, University of Sfax, Sfax, Tunisia

Abstract. In the last few years, Earth Observation sensors received a large development, offering, therefore, various types of data with different temporal, spatial, spectral, and radiometric resolutions. However, due to physical and budget limitations, the acquisition of images with the best characteristics is not feasible. Image fusion becomes a valuable technique to deal with some specific applications. In particular, the vegetation area, which needs a high spectral resolution and frequent coverage. In this paper, we present a novel fusion technique based on Convolutional Neural Networks (CNN) to combine two kinds of remote sensing data with different but complement spectral and temporal characteristics, to produce one high spectral and temporal resolution product. To the best of our knowledge, this is the first attempt to deal with the spectral-temporal fusion problem. The feasibility of the proposed method is evaluated via Sentinel-2 data. The experimental results show that the proposed technique can achieve substantial gains in terms of fusion performance.

Keywords: Image fusion · Remote sensing · Convolutional neural networks · Spectral enhancement · Temporal enhancement

1 Introduction

Remote sensing (RS) generally comprises all techniques for measuring electromagnetic energy emitted from areas or objects on surface or subsurface of Earth without the need to a physical contact. In the last years, Optical Remote sensing has received a significant development thanks to the recent advancement in electronics field, allowing the acquisition of a large variety of data with a different spatial, spectral, radiometric, and temporal resolution, depending on the sensor proprieties and missions. The spatial resolution represents the size of the smallest area (pixel), the spectral resolution indicates the capability of the sensor to detect fine wavelength bands and the number of spectral channels, the temporal resolution is the period to take for revisiting the same region, and the radiometric resolution refers to the sensor's sensitivity to distinguish between smallest brightness values while capturing an image [6]. Currently, satellites offer

C. Djeddi et al. (Eds.): MedPRAI 2020, CCIS 1322, pp. 60–72, 2021.
https://doi.org/10.1007/978-3-030-71804-6_5

precious data about the Earth that help for detecting and monitoring in a multitude of applications. For example, vegetation monitoring [10], change detection [8], military target detection [20], and mineral exploration [21]. However, due a physical and budget limitations of satellite platforms, the acquisition of data with high characteristics for all aspects is not possible. When the technologies tackle a specific issue, researchers handle the latter with an appropriate method. Remote Sensing image fusion is a successful technique, which aims to combine one or multiple satellite images, in order to produce a fused product that includes more knowledge than each of the inputs [2]. Fusion techniques are advantageous and increase the performance in several remote sensing tasks, including object identification, segmentation, change detection, etc. The RS fusion products are in progressive increase, thanks to the growing requests from reputable companies such as Google Earth and Microsoft Visual Earth, which require very high resolution images in their commercial products, which could be done using the appropriate fusion technique.

Recently, the unavailability of products with high spatial and temporal resolution dues to the spatial-temporal limitation of optical satellites has been addressed by the reputable commercial company Planet lab [17]. Planet lab exploits more than 175 dove satellites in orbit to provide daily images from the whole Earth. Doves are nano-satellites that weight approximately 4 kg, which can acquire data at high spatial resolution of 3 m with four spectral channels (Red, Green, Blue, and near-infrared). Thanks to these satellites and its ability to offer a series of daily observation at high resolution, the need of spatio-temporal fusion has decreased. Especially, for applications that require high spatial and temporal resolutions such as the detection of rapid changes, object extraction, and object identification. Unfortunately, these satellites are characterized by a low spectral resolution as they include a limited number of bands with broad wavelengths. Besides, several agriculture applications such as monitoring vegetation seasonal dynamics, detection of leaf disease, and water consumption monitoring, need a higher number of bands with fine wavelengths, especially, in the near-infrared range. In contrast, Sentinel-2 satellite can provide a higher number of bands (13 bands), which make it the most suitable for vegetation monitoring tasks thanks to its valuable fine red-edge wavelength bands at 10 and 20 m. These bands are principally utilized to derive the vegetation indices via red-edge bands, which are significant to study the health of vegetation. However, Sentinel-2 is characterized by a low temporal resolution, as it has a long revisit circle of 5 days. Indeed, Planet and Sentinel-2 provide complementary characteristics, i.e., Planet provides a high spatial-temporal resolution but low spectral resolution, contrary to Sentinel-2, which produces images with high spectral resolution but low temporal resolution.

In this paper, we introduce a spectral-temporal fusion technique based on CNN to tackle the physical limitation of widely used satellites. To the best of our knowledge, it is the pioneering approach addressing the spectral-temporal image fusion problem. To study the feasibility of the fusion technique, only Sentinel-2 satellite is considered as a simulation tool. We assume that a low temporal

resolution characterizes the fine wavelength bands at 20 m, while the broad ones at 10 m have a higher temporal resolution. The main objective of this paper is to integrate Sentinel-2 broad and fine wavelength bands on a prior date $(t-1)$, a broad wavelengths bands on prediction date (t), to produce a product at 20 m with high spectral resolution unavailable due to the technical limitation.

2 Related Works

The literature of RS fusion is rich with works proposed to address some particular image fusion problems, and the number of papers increases continuously every year. Ones of the most common RS image fusion approaches are: pansharpening [1,3] and spatio-temporal fusion [11]. Indeed, optical satellites (e.g., WorldView-3 and QuickBird) have the ability to capture two kinds of images, a low spatial Multispectral (MS) image with several fine wavelength intervals and a high spatial Panchromatic (PAN) image in one broadband (unique band). Due to a Signal-to-noise ration (SNR) constraint, the acquisition of a MS image with high spatial resolution is not realizable. This key issue has principally motivated RS researchers to develop several sophisticated pansharpening fusion techniques to deal with this system limitation. Pansharpening technique becomes a fundamental preprocessing for many RS tasks [5,12].

The majority of optical sensors that have an acceptable spatial resolution (from 5 to 60 m), like Landsat TM and Sentinel-2, offer a long revisit time (16 days for Landat and 5 days for Sentinel-2), in addition to other factors like the cloud contamination and other atmospheric conditions, which can further increase the revisit periods. Even if their fine spatial resolution is appropriate for tasks like change detection and cover mapping, the detection of rapid land cover changes at the right time becomes limited using this type of satellites. In contrast, another type of satellites, such as Sentinel-3 and MODIS, offer opposite properties. In other words, they produce images with a coarse spatial resolution, with a very short revisit time (less than 2 days). However, this coarse resolution is inappropriate for monitoring small land cover changes. As a satellite with high spatial resolution and daily observation does not exist, several works on fusion techniques have been developed, aiming to combine the two different types of satellites images, that have complementary characteristics, and fuse the latter to offer a high spatial and temporal resolution product. Therefore, the ability for monitoring and detecting land surface changes will be improved.

In the last few years, Convolutional Neural Networks (CNN) have received large attention from the computer science community in many fields, including RS image fusion [3,16], thanks to their great success in many image processing tasks, especially, super-resolution (SR) [9,13]. SR aims to reconstruct a high-resolution image from a unique low-resolution image. Recently, the authors proposed a super resolution technique based on CNN known as SRCNN [9]. SRCNN is a pioneering model that proved very competitive results compared with traditional techniques. Inspired by its great success, plenty of works adapted the latter as prepossessing in many fields that need a high spatial resolution, including pansharpening and spatial-temporal fusion techniques, as they all aim to enhance

the spatial resolution of one or more low resolution images. With regard to pansharpening, in [16], the authors developed a SRCNN to deal with pansharpening task known as PNN. PNN aims to generate a high resolution MS image from the PAN and the MS images, that can be captured by the satellite. PNN achieved a consistent gain in performance compared with the state-of-the-art techniques. As another example, in [4], we proposed a pansharpening CNN based on Multiresolution Analysis called GIP-CNN, which aims to estimate the injection gains to control the amount of details need to be inserted into the MS image. GIP-CNN can improve the fusion quality with respect to the literature approaches, as well as the CNN-based techniques. In addition, SR success has influenced the spatial-temporal fusion. One can cite, the authors in [18] implemented an approach based on SRCNN capable to perform a complex mapping from a coarse image (captured by MODIS) to fine image (captured by Landsat) sensor, and utilize the pretrained model to predict the unavailable fine image due to the low temporal resolution of the sensor from its associated coarse image. Similarly, in [15], the authors proposed two-stream spatio-temporal fusion technique based on SRCNN. The latter focuses on learning a mapping between the corresponding fine residual image as a target and corresponding coarse images, captured in neighboring dates, ending up with two pretrained models. These models are used to produce two similar fine images; then, the outcomes are combined to generate a unique fine resolution product by applying a local weighting strategy.

3 Background

3.1 Sentinel-2 Satellite

Sentinel-2 was launched in 2015 by the European Space Association (ESA), as part of the European Copernicus program [19]. Sentinel-2 represents two identical satellites: Sentinel-2A and Sentinel-2B. They provide multispectral images with 13 bands covering the visible and near-infrared, in addition to the shortwave-infrared (SWIR) spectral region, incorporating three new spectral bands in the red-edge region (vegetation red-edge). The new bands are very useful on vegetation applications. For instance, to retrieve the canopy chlorophyll content [7]. The availability of 4 vegetation red-edge bands (B5, B6, B7, B8A) at 20 m is a unique characteristic that discriminates Sentinel-2 from other optical satellite sensors. Table 1 describes the spectral bands of Sentinel-2. In this paper, we consider the bands: B5, B6, B7, and B8a as high spectral resolution bands with low temporal resolution, and B7 and B8, as lower spectral resolution bands and higher temporal resolution.

3.2 Super-Resolution CNN

Super-Resolution is a technique to reconstruct a high-resolution (HR) images from one or more low spatial resolution (LR) images. After the considerable success of CNNs in many computer vision applications, particularly, image enhancement, such as image denoising, image inpainting, image dehazing, and super-resolution. Recently, single image super-resolution technique based on CNN

Table 1. Specifications of Sentinel-2 spectral bands

Spectral band	Central wavelength (nm)	Band width (nm)	Resolution (m)
B1	443	20	60
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20
B8	842	115	10
B8a	865	20	20
B9	940	20	60
B10	1375	30	60
B11	1610	90	20
B12	2190	180	20

(SRCNN) has been proposed [9], which aims to learn a mapping from the low resolution input to the high resolution image. Even the shallowness and the simplicity of the architecture, SRCNN can achieve superior performance over the state-of-the-art techniques. SRCNN includes only three convolutional layers. As a preprocessing, the low resolution image must be interpolated to the same size of the high resolution one in order to train the network. In the next section, we propose a modified version of SRCNN adapted for the spectral-temporal fusion problem. Therefore, the inputs of the network are substituted by the concatenation of two images (multiple inputs) with different spectral and temporal characteristics instead of single input, and the output provides an improved product that combines the best features of each input (high spectral and temporal resolution).

4 The Spectral-Temporal Fusion Technique

The objective of the paper is to estimate a high spectral high temporal (HSHT) product S_{20}^{t} captured at date t, using the associated low spectral high temporal (LSHT) image S_{10}^{t} captured at the same date t, accompanied with a high spectral low temporal (HSLT) image S_{20}^{t-1} captured at a previous date $t-1$. The process results in a high spectral resolution product with higher temporal resolution (HSHT). In this work, we assume that the Sentinel bands: B5, B6, B7, and B8a constitute a HSLT image, whereas the bands: B4 and B8 are considered as LSHT resolution bands. As the considered HS and LS bands are characterized by different wavelength, the selection of LS input is based on the correlation with target HS one. For example, B4 (wavelength 650–680 nm) is used to estimate

B5 (wavelength 698–713 nm) as their spectral ranges are close to each other. Similarly, B8 (wavelength 785–900 nm) is utilized to estimate B6, B7, and B8a (wavelengths are 733–748 nm, 773–793 nm, and 855–875 nm, respectively).

The proposed technique includes a training phase and a prediction phase. Each phase is performed for each band separately respecting the same conditions.

Inspired by the great success of SRCNN in image super-resolution [9], the proposed method aims to learn an end-to-end mapping function ($F(.)$) from HSLT and LSHT images to HSHT one. As SRCNN require the same size of input and target images, the LS image (S^t_{10}) input is downsampled by a factor of 2 to the same size of HS S^t_{20} target image (Y), producing ($\hat{S}^t_{10}$); then, the latter is concatenated with the associated S^{t-1}_{20} as a single input (X). The main goal of the training phase is to optimize the following Mean Squared Error (l_2) loss function:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|F(X_i;\theta) - Y_i\|^2 \tag{1}$$

where N is the total number of the training samples, θ indicates the network weights need to be optimized, i represents the image index, and F is estimated high spectral resolution image at t. The objective function is minimized using Adam [14] with a learning initialized to 10^{-4}. To guarantee a successful training of the network, the whole image is cropped into 10000 sub-images 64×64 pixels, divided into training and validation parts. The former includes 80 % sub-images, and the latter comprises the rest. Figure 1 illustrates the framework of the proposed fusion technique.

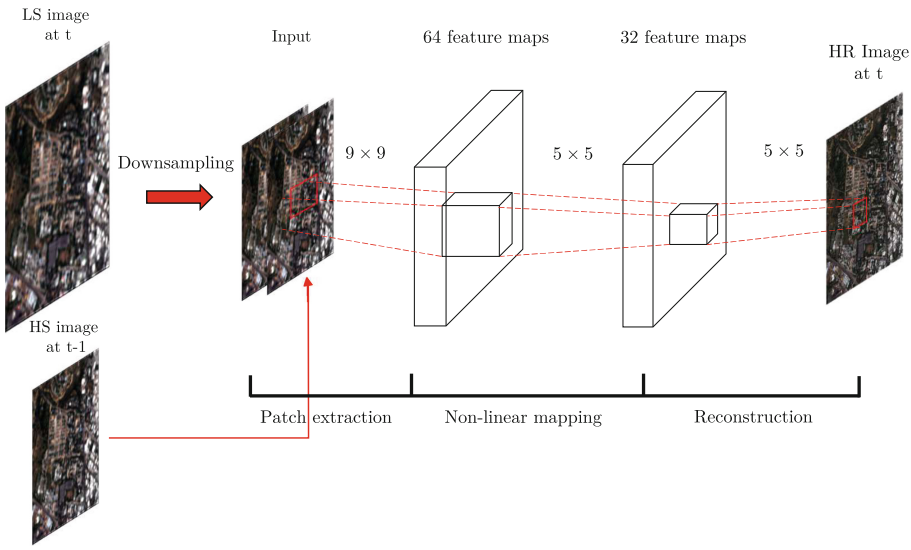
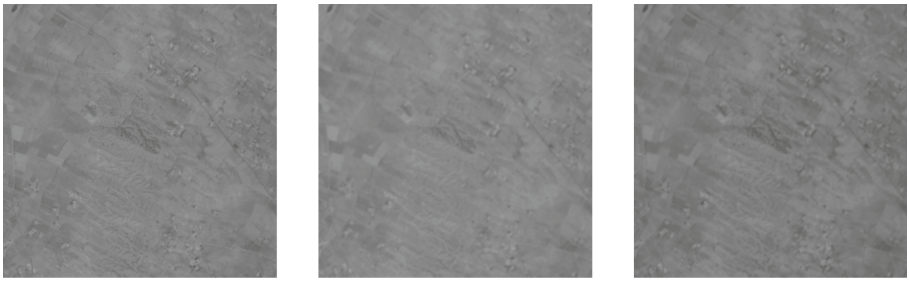


Fig. 1. The framework of the proposed fusion technique.

After the completion of the first phase, the prediction phase consists on estimating the HSHT images on another dates using the pretrained model. To this end, the training images at t and $t-1$ by different dates $\hat{t}$ and $\hat{t}-1$, which are previously unseen by the network, respectively, on the same region of study.

4.1 Datasets

In the experiments, we assume that the Sentinel-2 bands: B4, B8, are characterized by a low spectral resolution and high temporal resolution. In contrast, the bands: B5, B6, B7, and B8a, are considered as high spectral and low temporal resolution. Therefore, the principal objective is to estimate the high spectral bands at date t, which are assumed unavailable due to physical limitations. Figure 2 shows some considered broad and fine wavelength bands provided by Sentinel-2.

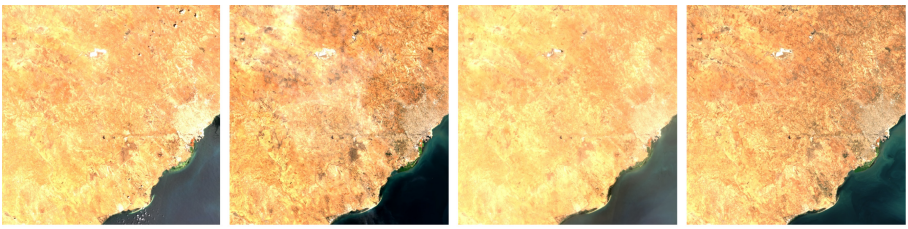


(a) B8 (b) B8a (c) B6

Fig. 2. Illustration of some Sentinel-2's bands.

To ensure an effective training and evaluation process of the proposed fusion technique, we considered two datasets of size 2600×2600 pixels captured in different dates by Sentinel-2 sensor from the city of Sfax (35°06 N, 10°54 E). The datasets include an agriculture area from Jebiniana town, located about thirty kilometers north of Sfax. The first dataset is used to train the network. It was captured on June 6, 2017 $(t-1)$ and December 12, 2017 (t). The second one is used to evaluate the fusion performance of the proposed technique. It was acquired over the same region on June 30, 2018 $(\hat{t}-1)$ and December 17, 2018 $(\hat{t})$.

A period of 6 months is chosen between observations on both datasets, as the latter can guarantee the apparition of more change in the region of interest. Consequently, the fusion process becomes more challenging. Figure 3 illustrates the train and test images.



(a) (b) (c) (d)

Fig. 3. Illustration of train and test datasets. (a) and (b) represent the train images on June 6, 2017 ($t-1$) and December 12, 2017, respectively. (c) and (d) are the test images on June 30, 2018 and December 17, 2018.

5 Experimental Results and Discussion

The proposed approach was compared with the state-of-the-art fusion technique: the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) [11]. STARFM was adapted to deal with the spectral-temporal fusion problem.

In order to assess the performance of the fusion techniques, the fused images were compared with the ground-truth images at full-reference manner through the three highly used metrics: Root Mean Square Error (RMSE), Spectral Angle Mapper (SAM), and Structural Similarity (SSIM) index.

5.1 Root Mean Square Error

Root Mean Square Error (RMSE) measures the fusion distortion between the target product (A), and fused image (B). It is defined as follows:

$$RMSE(A,B) = \sqrt{E\left[(A-B)^2\right]} \tag{2}$$

where E is the expectation value. The lower RMSE, the lower distortion. Its ideal value is 0.

5.2 Spectral Angle Mapper

The Spectral Angle Mapper (SAM) [22] measures the spectral distortion between two images A and B. It is defined as follows:

$$SAM(A,B) = \arccos \frac{\langle A|B\rangle}{\|A\| \cdot \|B\|} \tag{3}$$

where $\langle .|. \rangle$ and $\|.\|$ denote the scalar product and the normalization, respectively. SAM's ideal value is 0.

5.3 Structural Similarity

The Structural Similarity (SSIM) calculates the similarity between the target image (A) and the predicted one (B) [19]. SSIM is defined as follows:

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + C_1) + (2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{4}$$

where μ_A and μ_B represent the mean of the target and the predicted images, respectively, σ_{AB} indicates the covariance between the target and the observed images, σ_A and σ_B are the variances of the target and the predict images, and C_1 and C_2 denote small constants utilized to avoid a null denominator when the means and variances are almost zero.

Table 2 describes the quantitative performance of the fusion products on the test dataset. Based on the results, the proposed approach overpasses the traditional technique STARFM on all bands in terms of spectral, as it obtained the best SAM value. In term of spatial fidelity, the proposed approach provides the best results on RMSE and SSIM metrics, which measures the spatial distortion. Furthermore, the proposed technique produces very competitive results in terms of RMSE with an error (less than 0.8%). Hence, this fusion result can be used for agriculture monitoring applications. The worst product in RMSE is related to the band B4, as its wavelength, compared with the other bands, is not included in the inputs low spectral resolution bands: B4 and B8. Consequently, the fusion task on this specific band is further complex for the CNN to learn.

Table 2. Quantitative performance of the fused products on test dataset.

Metric	Band	STARFM	Proposed
RMSE	B4	0.0250	**0.0072**
	B6	0.0277	**0.0069**
	B7	0.0287	**0.0057**
	B8a	0.0294	**0.0045**
SAM	B4	0.0797	**0.0242**
	B6	0.0736	**0.0179**
	B7	0.0701	**0.0137**
	B8a	0.0657	**0.0115**
SSIM	B4	0.7244	**0.9783**
	B6	0.7149	**0.9821**
	B7	0.7092	**0.9854**
	B8a	0.7045	**0.9881**

Figure 4 illustrates the fusion results and their associated ground-truth images for each band that have been described on Table 2. The ground-truth

bands' images are shown in the first row, The fused images produced by STRAFM are shown in the second row, and the third row shows the fused products generated by the proposed technique. The fused images of STARFM technique provide the worst fusion performance, as the latter suffers from a blurring effect compared with the ground-truth images, and it lacks of a huge amount of spatial information. For example, the region highlighted in red suffer from blurring effect, and both details and contours are not visible. The proposed technique produces the best fusion quality as the fused results are as close as possible to the ground-truth ones, and the details and the edges are well preserved. Furthermore, it is difficult to notice any type of distortion on the fused images.

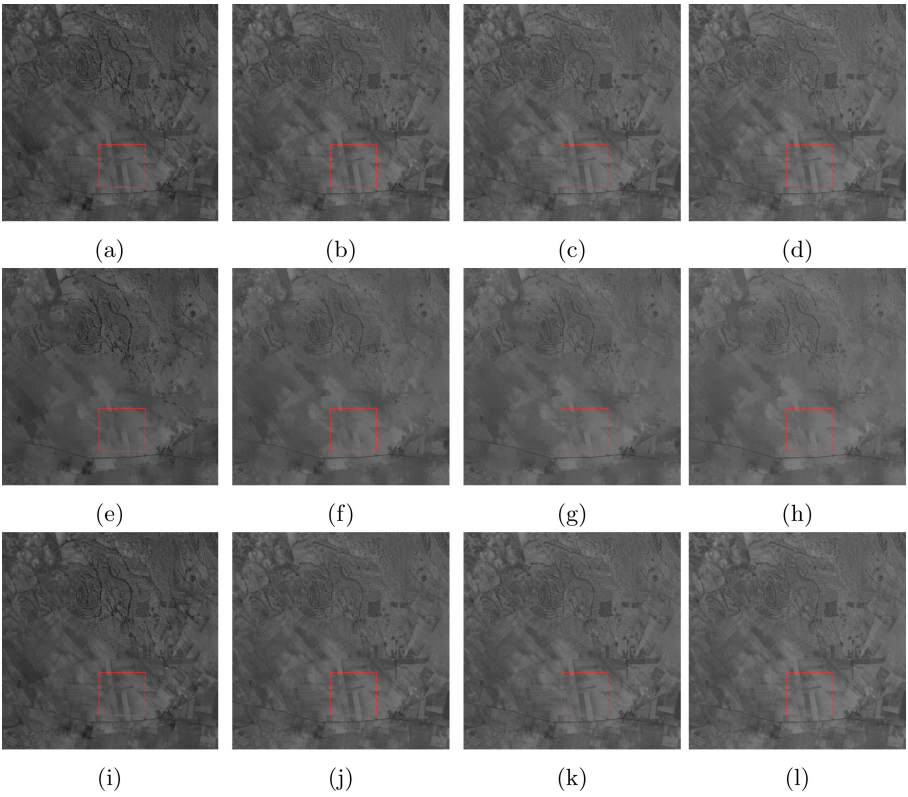


Fig. 4. Fusion products on the test dataset. The first row illustrates, from left to right, the ground-truth bands: B5, B6, B7, and B8a, respectively, the second row shows the fusion results using STARFM, and the third one shows the fusion results by the proposed technique.

Figure 5 shows the residual error images, calculated as the difference between the fused images and their associated ground-truth images for each band. The

first row illustrates the residual images of the proposed technique, and the second one shows the residual images of STARFM technique. As the residual pixel values are almost zero, they were multiplied by 100 to provide more interpretability. STARFM produces the worst results, as the errors are shown on the whole residual images of each band, and it fails to reconstruct the most of spatial details of the high spectral image. The proposed approach provides better fusions results, as it successes to conserve the spatial structures of fused products. Although, the proposed technique lacks of some spatial details on heterogeneous vegetation areas that undergo small changes.

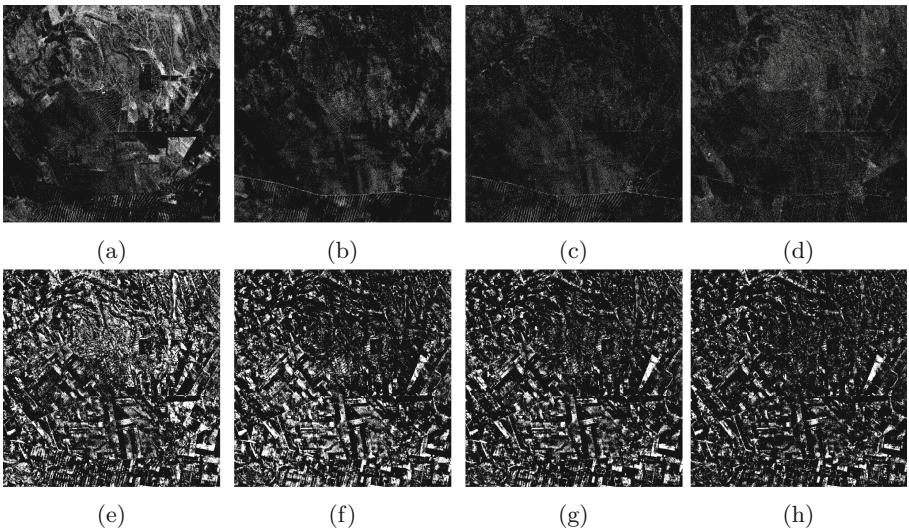


Fig. 5. Residual images on the test dataset. The first row shows, from left to right, the residual images of the proposed technique for the bands: B5, B6, B7, and B8a, respectively, the second row shows the residual images using STARFM. Brighter pixels indicate a larger error.

6 Conclusion

In this paper, we introduced a spectral-temporal fusion technique based on CNN, which aims to integrate the spectral information of high spectral images and the temporal information of high temporal resolution images. The proposed network aims to reconstruct a high spectral and temporal products from two inputs characterized by different and complementary spectral and temporal resolutions. To the best of our knowledge, it is the first work that focuses on CNN's to deal with the spectral and temporal fusion problem.

The qualitative and quantitative experimental results conducted on Sentinel-2 satellites using simulated images show that the proposed approach achieved

comptetitive fusion quality in terms of both spectral and temporal qualities. In our future works, we intend to evaluate the proposed network to combine real data from two widely used satellites with different spectral and spatial, using deeper architecture and combining different loss functions.

Acknowledgement. We gratefully acknowledge the support of NVIDIA Corporation with the contribution of the Titan Xp GPU used for this research.

References

1. Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., Selva, M.: Twenty-five years of pansharpening: a critical review and new developments. In: Signal and Image Processing for Remote Sensing, pp. 552–599. CRC Press (2012)
2. Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A.: Remote Sensing Image Fusion. CRC Press, Boca Raton (2015)
3. Benzenati, T., Kallel, A., Kessentini, Y.: Two stages pan-sharpening details injection approach based on very deep residual networks. IEEE Trans. Geosci. Remote Sens. (2020)
4. Benzenati, T., Kessentini, Y., Kallel, A., Hallabia, H.: Generalized Laplacian pyramid pan-sharpening gain injection prediction based on CNN. IEEE Geosci. Remote Sens. Lett. **17**(4), 651–655 (2019)
5. Bovolo, F., Bruzzone, L., Capobianco, L., Garzelli, A., Marchesi, S., Nencini, F.: Analysis of the effects of pansharpening in change detection on VHR images. IEEE Geosci. Remote Sens. Lett. **7**(1), 53–57 (2009)
6. Chang, N.B., Bai, K.: Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing. CRC Press, Boca Raton (2018)
7. Clevers, J.G., Kooistra, L., Van den Brande, M.M.: Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. Remote Sens. **9**(5), 405 (2017)
8. Daudt, R.C., Le Saux, B., Boulch, A., Gousseau, Y.: Urban change detection for multispectral earth observation using convolutional neural networks. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2115–2118. IEEE (2018)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
10. Frampton, W.J., Dash, J., Watmough, G., Milton, E.J.: Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. **82**, 83–92 (2013)
11. Gao, F., Masek, J., Schwaller, M., Hall, F.: On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. **44**(8), 2207–2218 (2006)
12. Gilbertson, J.K., Kemp, J., Van Niekerk, A.: Effect of pan-sharpening multitemporal Landsat 8 imagery for crop type differentiation using different classification techniques. Comput. Electron. Agric. **134**, 151–159 (2017)
13. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

15. Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B.: StfNet: a two-stream convolutional neural network for spatiotemporal image fusion. IEEE Trans. Geosci. Remote Sens. **57**(9), 6552–6564 (2019)
16. Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G.: Pansharpening by convolutional neural networks. Remote Sens. **8**(7), 594 (2016)
17. Planet: Satellite Imagery and Archive. https://planet.com/products/planet-imagery/. Accessed 01 July 2020
18. Song, H., Liu, Q., Wang, G., Hang, R., Huang, B.: Spatiotemporal satellite image fusion using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **11**(3), 821–829 (2018)
19. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
20. Wu, H., Zhang, H., Zhang, J., Xu, F.: Typical target detection in satellite images based on convolutional neural networks. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 2956–2961. IEEE (2015)
21. Yokoya, N., Chan, J.C.W., Segl, K.: Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images. Remote Sens. **8**(3), 172 (2016)
22. Yuhas, R.H., Goetz, A.F., Boardman, J.W.: Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm (1992)