A Long-time-series Spatio-Temporal-Spectral Fusion Method via Multi-task Learning

Mingyuan Peng[1], Canhai Li[1], Xiaoqing Zhou[1], Guoyuan Li[1]

[1]Land Satellite Remote Sensing Application Center, China

Keywords: spatio-temporal-spectral fusion, hyperspectral data, multispectral data, long-time-seires.

Abstract

Due to the limitations of sensor hardware, clouds and fog, and data transmission limitations, it is difficult for the data obtained by spaceborne remote sensing imager to achieve high temporal, spatial and spectral resolution at the same time, which limits its application in long-time-series high-frequency monitoring. At present, there are several spatio-temporal-spectral algorithms that can realize the fusion of temporal, spatial and spectral resolution, but most of them are based on one to two discrete images, and the integrated fusion at the multi-dimensional level has not yet been realized. There is currently no research on the spatio-temporal-spectral fusion method based on LONG-TIME-SERIES multi-scene remote sensing data. Aiming at solving the bottleneck of spatio-tempora-spectral resolution of remote sensing data, this study proposes a new long-time-series spatio-temporal-spectral fusion method based on multi-task learning to realize the multi-dimensional optimization of multi-source remote sensing data resolutions. Experiments used simulated and real datasets, both of which contain 4 images of 10m ZY1-02D multispectral data, 7 images of 16m GF-6 multispectral data and 4 images of 30m ZY1-02D hyperspectral data, and obtained 7 images of 10m hyperspectral data. The experiments The results show that our method performs the best compared to other methods. This method can provide effective data support for applications based on long-time series remote sensing data.

1. Introduction

With the rapid development of remote sensing satellites and earth observation technology, earth observation satellites have shown an explosive development with an increasing number of in-orbit satellites. To this day, humans have obtained a massive amount of remote sensing datasets. The accumulation of a large amount of historical remote sensing data has made it possible for applications related to long-time-series remote sensing data monitoring.

However, the potentials of long-time-series hyperspectral data has not been fully tapped. Due to the bottleneck constraints of sensors, orbital height and data transmission capability (Wang and Wang, 2010), there is currently no spaceborne remote sensing satellite that can obtain high temporal, spatial and spectral resolution at the same time. Different data has advantages in one or two resolutions, but has inferior in other indicators, which has become a major limiting factor in the application of long-time-series remote sensing data monitoring. In practical applications, for the single scene coverage area within the research area, the actual temporal resolution often differs significantly from the theoretical time resolution capability. Due to the fact that the influence of cloud cover, the temporal resolution of the dataset is also difficult to reach the nominal revisit period. At present, the improvement of temporal resolution can be achieved through satellite networking and data fusion. However, data obtained from different camera may differ in irradiance characterization, and different imaging time may bring more deviation.

Thus, this study aims to solve the bottleneck problem of mutual constraints on the temporal-spatial-spectral resolution of hyperspectral satellites that hinders the applications of long-time-series hyperspectral datasets. We proposed a long-time-series spatio-temporal-spectral fusion method . On the basis of constructing discrete multi temporal remote sensing data into long-term MDD multidimensional remote sensing data (MDD), a multi-task learning model is used to perform joint extraction of multidimensional spatio-temporal-spectral information. The model contains three branches of convolutional nerual networks which extracts mappings of spatio, temporal and spectral information seperately, and learns them jointly via multi-task structure. This method can achieve multi-dimensional integration of spatio-temporal-spectral information and improve the resolution of hyperspectral data.

2. Related Work

Spatio-temporal-spectral fusion method can be traced back to the work of Huang et al. in 2013. Using the maximum posterior probability (MAP) criterion, based on two Landsat and MODIS sensors, the authors explored the integrated fusion method on multi-temporal spatial-spectral images. By using 19 bands of two scenes, MODIS images with 250 m/500 m/100 m spatial resolution and 6 bands of one scene Landsat image fusion with 30m spatial resolution produces a reconstructed image with MODIS spectral resolution and Landsat spatial resolution that is missing temporally, realizing the combination of spatio-temporal fusion and spatial spectral fusion (Huang et al., 2013). Shen et al. utilized MAP to construct a multi-source sensor spatio-temporal-spectral integrated fusion framework (Shen et al., 2016), which is commonly used for multi view image spatial fusion, spatiotemporal fusion, spatio-temporal fusion, and spatio-temporal-spectral fusion. The performance of spatio-temporal-spectral fusion method was validated using MODIS, Landsat, and SPOT satellites. Jiang et al. constructed a heterogeneous spatiotemporal spectral integrated fusion framework using Deep Residual CycleGAN (Deep Residual CycleGAN)(Jiang et al., 2021). Peng et al., based on the semi-coupled sparse tensor factorization method, and built an integrated spatio-temporal-spectral fusion model on the basis of four-dimensional tensors. In addition to verifying spatiotemporal fusion and spatial-spectrum fusion, they also used three groups of Hyperion data to simulate low spaial, low temporal, high spectral resolution data and high spatial, high tempoal, low spectral resolution data for spatio-temporal-spectral integration fusion (Peng et al., 2021). Wei et al. designed a spatio-temporal-spectral fusion method for serial

panchromatic image sharpening and spatiotemporal fusion using the same platform camera data of Gaofen-1, aiming to construct a 2-meter multispectral image with high temporal resolution. The method was experimentally validated using three temporal phases of Gaofen-1 panchromatic, multispectral, and wideband data, However, the integration of spatio-temporal-spectral has not yet been achieved (Wei et al., 2021).

All of the above achievements have made outstanding contributions to the improvement of multi-source remote sensing data resolution, but most of them are research oriented to two to three scene remote sensing image fusion, and there is no research on spatio-temporal-spectral fusion method oriented to long time series or multi scene remote sensing data which involves datasets more than 4 time phrases.

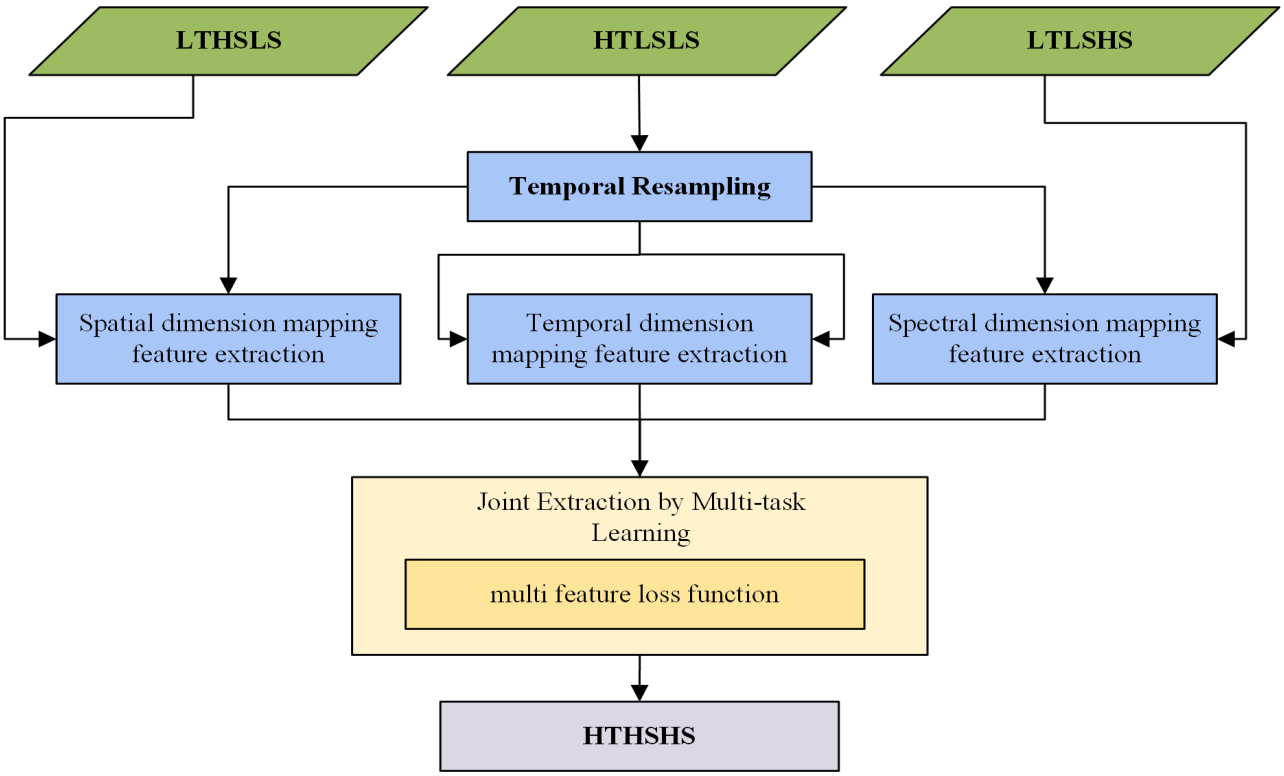


Figure 1. Overall Framework

3. Proposed Method

3.1 Spatio-Temporal-Spectral Fusion Framework

The irradiance is converted into a continuous time-varying optical signal through the detection element, which is then converted into an electronic signal. After being amplified and processed by the element, it is sampled and quantified into pixel values (DN) in an analog-to-digital converter over time. The specific values of the signal are obtained by integrating the sensor at certain time intervals, wavelength intervals, and spatial intervals based on the characteristics of the component. It can be said that the signal is obtained by convolution of input radiation in spectral, spatial, and temporal dimensions with the instrument response function.

$$o(z) = i(z) * r(z)$$

in which represents the output of the instrument on the channel, represents the input of the instrument on, represents convolution operation, and represents the response function of the instrument on.

Therefore, the multidimensional data integration image process under the assumption of sensor nonlinearity is

$$X_{4D}^{N} = f_N(Z_{4D})$$

in which represents the long-term multidimensional observation dataset obtained by sensors, represents the real (fusion reconstruction) long-term multidimensional dataset, and represents the nonlinear mapping relationship of comprehensive temporal resampling, spatial resampling, and spectral resampling factors.

Based on formula (4), a multi-source and multi-dimensional remote sensing data fusion model framework can be obtained, namely

$$Z_{4D} = f(X_{4D}^{1}, X_{4D}^{2}, ..., X_{4D}^{N})$$

in which represents the multi-source and multidimensional remote sensing data obtained after fusion reconstruction, and represents the integrated mapping relationship based on the multidimensional remote sensing dataset.

Combining the multidimensional dataset fusion framework, construct a feature information extraction model for multi-source remote sensing data in temporal, spatial, and spectral dimensions, namely

$$\delta_{4D}^{n} = f_N(X_{4D}^{n})$$

in which represents the first multidimensional remote sensing dataset to be fused, represents the feature information extraction model of the time, space, or spectral dimensions, and represents the corresponding feature information.

Spatial dimension mapping feature extraction utilizes the relationship between high spatial resolution panchromatic datasets and time encoded resampled high temporal resolution multispectral datasets for extraction, i.e

$$\delta_{4D}^{spatial} = f_{spatial}(X_{4D}^{pan}, X_{4D}^{mul'})$$

Time dimension mapping feature extraction utilizes the relationship between the original high-resolution multispectral

dataset and the time encoded resampled high-resolution multispectral dataset, i.e

$$\delta_{4D}^{temporal} = f_{temporal}(X_{4D}^{mul}, X_{4D}^{mul'})$$

Due to the presence of band noise and data concatenation issues in the shortwave infrared spectrum of the hyperspectral data of the ZY1-02D/E satellite, it is proposed to perform principal component transformation on the hyperspectral data, using the transformed data to extract spectral dimension mapping features to filter out data noise. Extracting high temporal resolution multispectral datasets and their relationships using time encoded resampling, i.e

$$\delta_{4D}^{spectral} = f_{spectral}(X_{4D}^{mul'}, X_{4D}^{hy'})$$

Due to the different scales of multi-source and multidimensional remote sensing data in spatial, spectral, and temporal dimensions, the design of mapping feature extraction networks focuses on achieving multi-scale information extraction. We plan to perform multi-scale feature extraction with different filter operator sizes on the basis of stacking residual blocks to achieve feature extraction at multiple spatial, temporal, and spectral scales. For each residual block, its input and output can be simplified as

$$y = x + \sum_{i=1}^{C} \mathcal{T}_i(x)$$

in which , is the output of the residual block, is the input of the residual block, is the number of group convolutions, that is, the cardinality of the network, and is the feature extraction network of the th scale. This approach is easy to achieve convergence while increasing the cardinality of the network, which can increase the accuracy of the network without increasing its depth. The architecture of the mapping feature extraction network residual block is proposed to be designed in Figure 2.

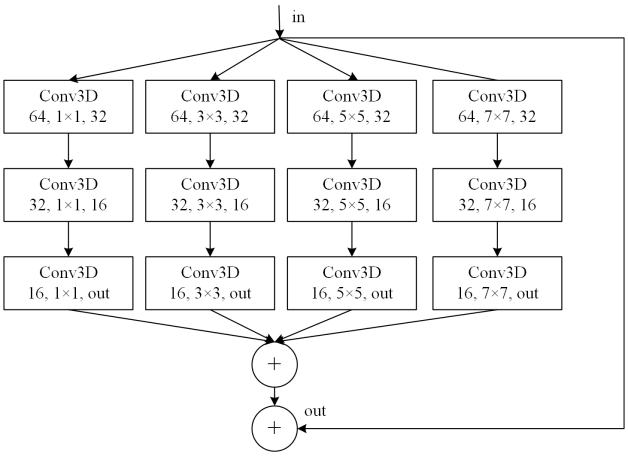


Figure 2. Residual Block

3.2 Joint Extraction Based on Multitask Learning

Through the design of multi feature loss function based on multi task learning, the objective loss function of all single task feature network branches is optimally combined to achieve joint feature extraction and fusion of multi task learning. Assuming that the learnable parameter set of a multitask learning network is, which represents all weight parameters in the network. The input set of the multi task learning network is, the label set is, and all predicted label sets are. For each single task characteristic network branch, its loss function is, where. The total loss function of the multi task learning network is designed as a linear combination of each single task characteristic network branch, namely

$$L_{comb}(x, y_{\mathcal{T}}, y_{\mathcal{T}}'; W_{\mathcal{T}}) = \sum_{\tau \in \mathcal{T}} L_\tau(x, y_\tau, y_\tau'; w_\tau) \cdot c_\tau$$

in which, is the weight of the loss function of the branch of the task characteristic network. The most direct way is to manually assign values, but the model is sensitive to weight parameters and may affect the final accuracy results. Therefore, a learnable parameter set will be added, namely

$$W_{\mathcal{T}} = (\theta_{\mathcal{T}}, c_{\mathcal{T}})$$

To force a positive value to avoid a negative value when it drops to, the regularization term is designed as

$$R_{pos}(c_{\mathcal{T}}) = \ln(1 + c_{\mathcal{T}}^2)$$

Therefore, the total loss function of multi task learning is finally designed as

$$L_{\mathcal{T}}(x, y_{\mathcal{T}}, y_{\mathcal{T}}'; W_{\mathcal{T}}) = \sum_{\tau \in \mathcal{T}} \frac{1}{2 \cdot c_{\mathcal{T}}^2} L_\tau(x, y_\tau, y_\tau'; w_\tau) + \ln(1 + c_{\mathcal{T}}^2)$$

Perform multi-dimensional fusion and reconstruction of the feature information extracted from the model to obtain a spatiotemporal spectral fusion dataset, achieving the following process

$$Z_{4D} = g(\delta_{4D}^1, \delta_{4D}^2, \ldots, \delta_{4D}^n)$$

By concatenating and fusing the mapping feature extraction network, a spatiotemporal spectral fusion dataset can be obtained.

4. Experiments and Results

4.1 Datasets and Experimental Settings

The technical goal is to construct a long-term spatiotemporal spectrum fusion method based on the idea of multitasking learning, and use ZY1 02D/E multispectral data, hyperspectral data, and GF-6 wide-swath multispectral data to construct a long-term dataset fusion reconstruction, obtaining the spatial resolution of ZY1-02D/E multispectral data high temporal resolution remote sensing data set of time resolution of GF 6 and spectral resolution of ZY-1-02D/E hyperspectral data. The comparison of theoretical target resolution indicators before and after fusion is shown in Table 1.

Two datasets are used in this study. Dataset I is the simulated dataset simulated using 7 dates of hyperspectral images obtained from ZY1-02D. The images are sensed during Dec.13th, 2021 to Jun.25th, 2022 at N37.5 E116.7 in China. Dataset I mainly contains buildings and crop fields. The hyperspectral data were simulated to 4 images of 90m hyperspectral data, 7 images of 60m multispectral data and 30m multispectral data. The spatial simulation was done using bilinear interpolation and the spectral simulation used spectral response functions of GF-6 wide-swath camera and ZY1-02D visible/near-infrared camera. Dataset II is the real dataset which contains 4 images sensed by ZY1-02D hyperspectral camera and multispectral camera as well as Gaofen-6 wide-swath camera. The images are sensed during Sep.7th, 2020 to Jan. 27th,2021. The dataset were are preprocessed with geometric correction, geospatial registration and normalized radiometric correction.

Data		Spatial resolution	Temporal resolution	Spectral resolution
ZY1-02D	Hyperspectral camera	30m	55d	5nm/10nm
	Visible /Near-infrared	10m	55d	1 band
GF-6	Wide-swath camera	16m	6d	8 bands

Table 1. Resolutions of Different Instruments

4.2 Results

The experiments of different kernels' effect were conducted on the two datasets. The kernels are set as 1, 1/3, 1/3/5, as different cardinalities. The results for the two datasets are shown as Table 2 and Table 3. As we can see that for the two datasets, 1 and 1/3 perform the best overall.

For the spectral indicators SAM, smaller kernel shows better performance as results of kernel 1 are the smallest among three groups. For the overall indicators CC and PSNR and the spatial indicators SSMI, 1-3 performs best for the simulated datasets yet 1 performs best for the real dataset. This may due to the fact that simulation was conducted on hyperspectral data, which has blur effect than the finer multispectral data, and 1/3 kernel can better capture the blur effect. When it comes to real dataset with no blur effect on fine images, smaller kernel has better capability to capture fine textures. The running times shows that smaller kernel costs less time. Table 4 and Table 5 show the results of different dates on both dataset on relatively best average result.

Size	CC	SAM	PSNR	SSMI	Time/s
1	0.7935	**2.9558**	29.5757	0.6505	**680**
1/3	**0.8362**	3.4243	**29.6338**	**0.6977**	1473
1/3/5	0.7650	3.3315	28.9650	0.6798	3114

Table 2. Quantitative Results of Different Kernel Size on Dataset I

From the result below we can come to the conclusion that, for real datasets with fine multispectral data, models with only 1 kernel performs the best.

TABLE II QUANTITATIVE RESULTS OF DIFFERENT KERNEL SIZE ON REAL DATASET

Size	CC	SAM	PSNR	SSMI	Time/s
1	**0.7760**	**5.8687**	**24.9184**	**0.5321**	**535**
1/3	0.7141	6.1855	22.3987	0.4073	1023
1/3/5	0.7371	9.5547	22.3015	0.4833	2986

Table 3. Quantitative Results of Different Kernel Size on Dataset II

In order to show the effectiveness of our method, we compare the results of real dataset with combination of spatiotemporal and spatial-spectral methods. We chose traditional methods of ESTARFM(Zhu et al., 2010) and CNMF(Yokoya et al., 2012), as well as deep learning methods of STFDCNN(Peng et al., 2020) and SRECNN(Peng et al., 2019). Among them, STFDCNN is a integrated spatio-temporal methods which can process multi-temporal images of all time. The two-stage methods are combined by each spatiotemporal and spatial-spectral methods. The results are shown in Table III below. We can see that compared to the two-stage methods, our methods perform the best in all quantitative indices as well as cost time. This shows that our integrated method can utilize spatio-temporal-spectral information better compared to two-stage methods. Deep learning methods cost less time, and integrated methods such as STFDCNN can effectively reduce processing time.

Size	CC	SAM	PSNR	SSMI	Time
Ours	**0.7207**	**4.5935**	**25.6058**	**0.5012**	**1473**
ESTARFM +CNMF	0.7099	6.2116	21.6568	0.4333	48972
ESTARFM +SRECNN	0.2849	22.9863	22.4711	0.2502	20060
STFDCNN +CNMF	0.5613	12.3754	20.0720	0.3095	34082
STFDCNN +SRECNN	0.4191	14.2738	14.6536	0.1530	4666

Table 4. Quantitative Results of Different Methods on Dataset I

Size	CC	SAM	PSNR	SSMI	Time
Ours	**0.5375**	**9.5249**	**20.5763**	**0.3399**	**1023**
ESTARFM +CNMF	0.4378	15.8413	16.2998	0.1978	11681
ESTARFM +SRECNN	0.4465	15.7589	16.4954	0.2025	7352
STFDCNN +CNMF	0.5613	12.3754	20.0720	0.3095	5376
STFDCNN +SRECNN	0.4191	14.2738	14.6536	0.1530	1553

Table 5. Quantitative Results of Different Methods on Dataset II

(a)GT on date 2

(b)Ours on date 2

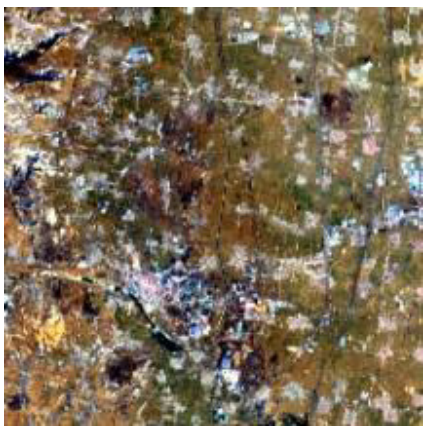
(c) GT on date 4

(a)GT on date 2

(d) Ours on date4

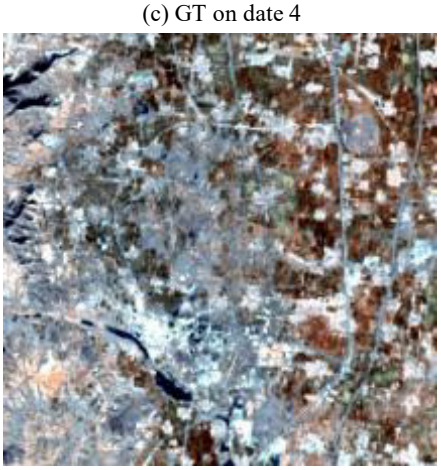
(b)Ours on date 2

(e)GT on date 6



(c) GT on date 4

(f) Ours on date 6



(d) Ours on date4

Figure 3. RGB composites of Ground Truth and Our Methods on Dataset I

(e)GT on date 6

(f) Ours on date 6

Figure 4. RGB composites of Ground Truth and Our Methods on Dataset II

5. Conclusions

This article proposed a long-time-series spatio-temporal-spectral fusion method. By constructing multi-temporal remote sensing datasets into four-dimensional dataset, temporal information can be extracted compactly. The method extract spatial information, spectral information and temporal information by three branches of convolutional neural networks, and fuse them jointly by multi-task learning structure. The method was tested on a simulated dataset and a real dataset, and was compared with two-stage fusion methods. The results show that our method performs the best in both accuracy and cost time.

References

Huang, B. O., Zhang, H., et al., 2013. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations, *Remote Sensing Letters*, 4, 561-569

Jiang, M., Shen, H., et al., 2021. An Integrated Framework for the Heterogeneous Spatio-Spectral-Temporal Fusion of Remote Sensing Images, *arXiv e-prints*

Peng, M., Zhang, L., et al.: Improving Spectral Resolution of Multispectral Data Using Convolutional Neural Network, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium,

Peng, M., Zhang, L., et al., 2020. A Fast Three-Dimensional Convolutional Neural Network-Based Spatiotemporal Fusion Method (STF3DCNN) Using a Spatial-Temporal-Spectral Dataset, *Remote Sensing*, 12, 3888, 10.3390/rs12233888

Peng, Y., Li, W., et al., 2021. Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal-spectral fusion of remote sensing images, *Inf. Fusion*, 65, 21-36

Shen, H., Meng, X., et al., 2016. An Integrated Framework for the Spatio–Temporal–Spectral Fusion of Remote Sensing Images, *IEEE Transactions on Geoscience and Remote Sensing*, 54, 7135-7148, 10.1109/tgrs.2016.2596290

Wang, J. and Wang, Y., 2010. Noise model of hyperspectral imaging system and influence on radiation sensitivity, *Journal of Remote Sensing*, 14, 607-620

Wei, J., Yang, H., et al., 2021. Spatiotemporal-Spectral Fusion for Gaofen-1 Satellite Images, *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5

Yokoya, N., Member, S., et al., 2012. Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion, *IEEE Transactions on Geoscience & Remote Sensing*, 50, 528-537

Zhu, X., Chen, J., et al., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions, *Remote Sensing of Environment*, 114, 2610-2623, 10.1016/j.rse.2010.05.032