Generative Adversarial Network for Pansharpening with Spectral and Spatial Discriminators

Anaïs Gastineau, Jean-François Aujol, Yannick Berthoumieu, and Christian Germain [*][†]

Abstract

The pansharpening problem amounts to fusing a high-resolution panchromatic image with a low-resolution multispectral image so as to obtain a high-resolution multispectral image. So the preservation of the spatial resolution of the panchromatic image and the spectral resolution of the multispectral image are of key importance for the pansharpening problem. To cope with it, we propose a new method based on a bi-discriminator in a Generative Adversarial Network (GAN) framework. The first discriminator is optimized to preserve textures of images by taking as input the luminance and the near infrared band of images, and the second discriminator preserves the color by comparing the chroma components Cb and Cr. Thus, this method allows to train two discriminators, each one with a different and complementary task. Moreover, to enhance these aspects, the proposed method based on bi-discriminator, and called MDSSC-GAN SAM, considers a spatial and a spectral constraints in the loss function of the generator. We show the advantages of this new method on experiments carried out on Pléiades and World View 3 satellite images.

Index term : **Deep learning, Generative Adversarial Network, multi-discriminator, remote sensing, pansharpening.**

1 Introduction

Remote sensing is the set of techniques that, through the acquisition of images, provide information about the surface of the Earth without direct contact with it. It is the whole process of capturing and recording the energy of an emitted or reflected electromagnetic radiation, processing and analyzing the information it represents. Remote sensing methods provide relevant tools for monitoring agricultural resources, changes in biodiversity and ecosystems for land cover or oceans, natural disasters and for studying of the atmosphere.

In remote sensing, the spatial resolution is given by the surface of the ground captured by one pixel. It affects the reproduction of details in the image. The spectral resolution is given by the number of bands of the image and by the bandwidth of the signal captured by the sensors producing the images. The higher the number of channels is or the narrower the bandwidth is, the higher the spectral resolution gets.

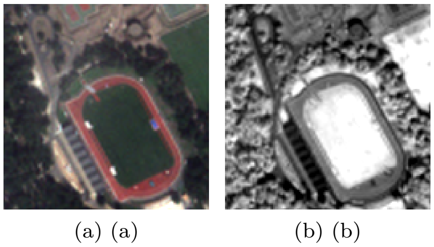
Recent satellites such as Spot 6-7, Ikonos, GeoEye or Pléiades offer multispectral and panchromatic images. The panchromatic chanel is characterized by a high spatial resolution and a low spectral resolution, whereas multispectral images have a low spatial resolution and a high spectral resolution. For Pléiades satellite, the panchromatic image is composed of one large band and the multispectral image composed of four finer bands, these bands being green, blue, red (RGB) and near infrared, respectively. More precisely, the ground sampling distance is 2.8 m for multispectral images and 0.7 m for panchromatic images.

In terms of image content, it is interesting to note: i) As the spectral bands (RGB) overlap, it means that for this channel the spatial content is obviously highly correlated. ii) Since the vegetation reflects better in the infrared, the NIR band allows to get much geometrical and texture information in vegetation areas. Indeed, the spectral reflectance is the signature of healthy vegetation and it is specific. The reflectance in the visible spectrum is low because the majority of the visible light absorbed during the photosynthesis to create chlorophyll. Besides, the reflectance is much higher in the near infrared (NIR) region since the NIR radiation are not absorbed by the vegetation for the

[*] A. Gastineau and J.-F. Aujol are with Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400 Talence, France, (e-mail: anais.gastineau@u-bordeaux.fr, jean-francois.aujol@math.u-bordeaux.fr)

[†] Y. Berthoumieu and C. Germain are with Univ. Bordeaux, Bordeaux INP, CNRS, IMS, UMR 5218, F-33400 Talence, France, (e-mail: yannick.berthoumieu@ims-bordeaux.fr, christian.germain@ims-bordeaux.fr)

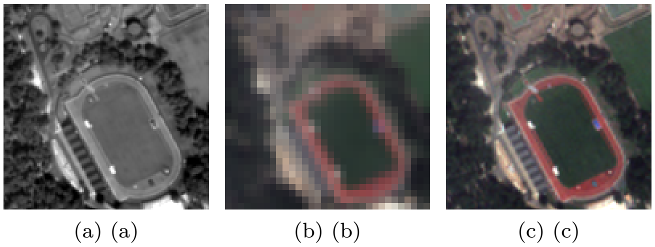
photosynthesis. So healthy vegetation can be easily identified thanks to the NIR band. Fig. 1 shows an example of it: the intensity of pixels is higher where the vegetation is healthy or dense.


(a) (a)

(b) (b)

Figure 1: Illustration of the vegetation reflectance on optical satellite images. A vegetation cover on RGB bands (a) and on the near infrared (NIR) band. The NIR band appears more textured than the RGB bands for vegetation area.

The pansharpening problem consists in fusing a panchromatic and a multispectral image in order to reconstruct a high spatial resolution multispectral image. For Pléiades images, this imposes a resolution factor setting to four. Fig. 2 gives an example of RGB bands of the pansharpening problem. The panchromatic image gives high frequency information (in particular about geometry and texture), whereas the multispectral image gives information about the spectral diversity.


(a) (a)

(b) (b)

(c) (c)

Figure 2: Example of Pléiades satellite images with the panchromatic image (a), the RGB bands of the low-resolution multispectral image (b) and the RGB bands of the high-resolution multispectral image (c). This example illustrates the problem of pansharpening focused on reconstruction of the high-resolution multispectral image.

Generally, this problem is formulated as follows:

$$\left\{ \begin{array}{lcl} y^k & = & SH^k u^k + B^k, \quad \forall\, k \leq K \\ P & = & \displaystyle\sum_{k \leq K} \alpha_k u^k \end{array} \right. , \tag{1}$$

where $y \in \mathbb{R}^{m \times n \times K}$ is the observed low-resolution multispectral image, $P \in \mathbb{R}^{M \times N}$ the high-resolution panchromatic image and $u \in \mathbb{R}^{M \times N \times K}$ the fused high-resolution image with a factor s between the low and the high spatial resolution such as $M = m \times s$ and $N = n \times s$ and K the number of spectral bands. In most cases, $s = 4$ for satellite images. In this model (1), S is the subsampling operator, H the blur operator and B a gaussian noise. The second equation supposes that P can be approached by the linear combination of the bands of u. The coefficients $(\alpha_k)_{k \leq K}$ are specific to the sensors of each satellite.

Several approaches have been proposed to solve the pansharpening problem. They can be grouped into four main classes:

- Component substitution methods that use a linear transformation such as PCA on the multispectral image to dissociate spectral from spatial details in order to replace the spatial details by the panchromatic image [1, 2].
- Multiresolution analysis methods that decompose panchromatic and multispectral images into a pyramid or a sequence of signals with a decreasing information content. Then the high frequencies of the panchromatic image are added to the multispectral image [3, 4].

- Variational and bayesian methods giving an a priori on the solution by solving an inverse problem [5, 6, 7].
- Learning methods modeling the relationship between multispectral and panchromatic images through different level of features, without the need of a model such as (1).

Recently, approaches based on deep learning give the best state of the art results. Masi *et al.* [8] adapt a CNN, initially proposed for the super-resolution problem [9], to the pansharpening problem. This CNN mimics the behavior of a sparse representation in three convolutional layers. Palsson *et al.* [10] propose to use a 3D CNN by considering multispectral images as 3D images: two spatial dimensions and one spectral dimension. This makes it possible to better model the inter bands spectral correlation. Guo *et al.* [11] propose a four layers CNN robust to the inconsistencies across satellites with dilated multilevel blocks. These dilated multilevel blocks allow to make full use of features extracted by the convolutional layers. Moreover, to avoid overfitting, they propose to use a l_2-regularization term on the weights in the loss function.

An important point for the pansharpening problem is to preserve the geometry and the color of images. To do so, Yang *et al.* [12] propose the PanNet network that trains in the high frequency (HF) domain to preserve the structure and the geometry of the panchromatic image. The spectral information, i.e. the color, is preserved by propagating the multispectral image thanks to residual connections in a ResNet architecture.

More recently, there is the emergence of GANs for the generative problem. Generative Adversarial Networks (GANs), introduced by Goodfellow *et al.* [13], are a class of unsupervised learning algorithms. This type of network mimics any data distribution. Usually, GANs are generative models where two networks are competing each other. The first network, the generator G_θ, generates samples, while its adversary, the discriminator network D_η tries to detect if it is a real sample or if it is the result of the generator.

For example, Liu *et al.* [14] propose a GAN based method. Their method, named PSGAN, considers a simple architecture, i.e. a stack of multiple layers for the generator and the discriminator. But more recently, they propose an extension [15] by changing the architecture of the generator. They choose a residual auto-encoder architecture for the generator with two sub-networks allowing to extract complementary features of the panchromatic and the multispectral images. Zhang *et al.* [16] propose an architecture using the deep learning module Spatial Feature Transform (SFT) initially designed by Wang *et al.* [17] for the super-resolution problem. This structure allows to reproduce the spatial characteristics of the panchromatic image within the multispectral images. In the same spirit, our recent paper [18] builds on a RDGAN (Residual Dense Generative Adversarial Network) method. In our RDGAN approach, we consider a Residual Dense architecture for the generator with a geometrical constraint in the non adversarial loss function to restore the spatial resolution of images. He *et al.* [19] preserve the spectral resolution by adding a regularisation term based on the Spectral Angle Map (SAM) measure in the generator in an adversarial autoencoder framework. This measure computes the spectral distortion between two images and they propose to minimize this quantity in the loss function.

Since GANs are efficients for image reconstruction, many improvements in the GAN algorithm have been proposed, especially by considering multiple discriminators. This multi-discriminator framework is mainly used for the super-resolution problem. For example, Zhu *et al.* [20] consider three discriminators. The first one is a pixel discriminator, commonly used in GAN, the second one compares colors with the low-resolution images and the third one compares edges and textures by considering the grayscale images. For the generator, they optimize a loss function considering a l_2 term between the generated and the target images and two feature terms comparing output of a pre-trained VGG network in order to preserve details. Lee *et al.* [21] keep the pixel discriminator of traditional GAN approaches and consider two other discriminators. One avoids checkerboard artifacts by comparing the discrete cosine transform and the other one restores high frequencies by comparing the histograms of the gradient magnitude of the input images. Park *et al.* [22] propose a method based on multiple discriminators in a CycleGAN context for image enhancement. In addition of the conventional discriminator, they propose to use a feature discriminator. It consists in using, as input of the discriminator, the output of one layer of a pre-trained VGG network.

In this paper, we propose a different GAN architecture exploiting jointly the spatial and spectral information sources. On the one hand, we seek to see how best to condition the discriminator part. On the other hand, we also study how we can introduce pertinent levels of information considering the non adversarial loss of the generator part. We propose a new bi-discriminator GAN based approach for the pansharpening problem with spectral and spatial constraints leading to the general architecture presented in Fig. 3.

The preservation of the spatial resolution of the panchromatic image and the spectral resolution of the multispectral image are of key importance for the pansharpening problem. In order to cope with it, we propose to separate these two tasks by considering two "orthogonal" discriminators. The first one is optimized to preserve texture and geometry of images by taking as input the luminance and the NIR band of images. The second one preserves the color

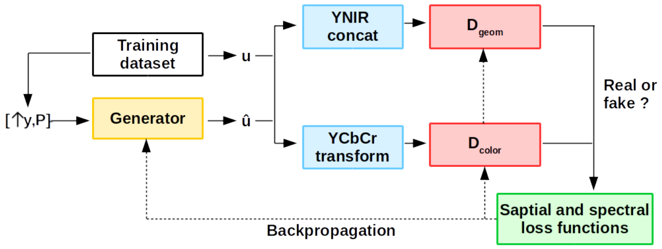


Figure 3: General framework of a bi-discriminator GAN, where the input $[\uparrow y, P]$ is the concatenation of the high-resolution panchromatic image P and $\uparrow y$ the low-resolution multispectral image y upsampled at the resolution of P, u is the target image, $\hat{u}$ is the generated image, 'YNIR concat' concatenate the luminance and the near infrared band, 'YCbCr transform' concatenate the chroma components Cb and Cr. D_{color} and D_{geom} are the two discriminators.

and the spectral resolution of images by comparing the chroma components Cb and Cr. Thus, this method allows us to train two discriminators, each one associated to a different and complementary task. Moreover, to enhance these aspects, we propose a method based on multi-discriminators, called MDSSC-GAN SAM, which considers a spatial and a spectral constraints for designing the generator.

Very recently, Ma *et al.* [**?**] proposed a method called Pan-GAN, based on multi-discriminators, for the pansharpening problem. This method considers a spatial and a spectral discriminator, and it adds constraints into the loss function. However, many differences have to be underlined with respect to our approach:

- The Pan-GAN method uses a 3 layer network with skip connections, whereas we take advantage of a residual dense architecture for the generator. Moreover, Ma *et al.* only consider convolutional layers in the discriminator while we use both convolutional and dense layers.
- Contrary to the Pan-GAN approach, we work in a residual framework, meaning we dedicate the modeling to the learning of spectral-spatial high-frequency missing features. It is one of the strengths of our approach in the context of pansharpening.
- As opposite to the Pan-GAN method, we develop a selective policy in the choice of channels supplying our both discriminators. To reconstruct the spectral resolution, Ma *et al.* use the whole set of low-resolution multispectral channels, without discernment, and the downsampled pansharpened image to focus on the intensity. To preserve the spatial resolution, the Pan-GAN network works with the panchromatic image and the luminance of the multispectral image. In our method, motivated by the spatial reconstruction in vegetation areas, we select the NIR band and the luminance to focus on the spatial reconstruction. For the spectral discriminator, we select the CbCr channels because these components are less correlated than RGB bands and so they allow us to get another color representation of the data.
- Finally, the non-adversial constraints used are very different. In the Pan-GAN approach, constraints follow the task of the discriminators. Because the spectral discriminator takes in input the low-resolution multispectral image and the downsampled pansharpened image, the associated constraint is the l_2 norm of the difference. In our method, we propose a constraint slightly different from the task of the discriminator. It allows us to re-inforce the spectral reconstruction, but with another point of view than the discriminator by minimizing the SAM metric, which is an illumination invariant spectral distorsion measure. For the spatial discriminator, we work identically. While the Pan-GAN approach minimizes the Frobenius norm of the differences between the spatial gradient fields, we choose to minimize the absolute value of the inner product of the gradient vector fields. Our choice is motivated by the fact that, as already shown in previous work [5], the absolute value of the scalar product is more efficient for aligning the local direction of the spatial gradient vector field.

2 Proposed method

In the literature based on GANs, see e.g. [14, 16], the following loss functions are classically used:

$$\begin{aligned}\mathcal{L}(G_\theta) = \sum_{i \leq N_b} \alpha log(D_\eta(G_\theta(\uparrow y, P))) \\ + \delta ||u - G_\theta(\uparrow y, P)||_1\end{aligned} \tag{2}$$

for the generator G_θ, where the first term is the cross entropy term used in adversarial learning and the second one is the l_1 norm between the target and the generated images. And

$$\mathcal{L}(D_\eta) = \sum_{i \leq N_b} log(1 - D_\eta(G_\theta(\uparrow y, P))) + log(D_\eta(u)) \tag{3}$$

is the loss function for the discriminator D_η. Here, N_b is the batch size, $[\uparrow y, P]$ is the input of the network G_θ with $\uparrow y$ corresponding to the low-resolution multispectral y image upsampled with a bicubic interpolation to the size of the panchromatic image P, $G_\theta(\uparrow y, P)$ is the output of the generator, θ and η the parameters of the generator G_θ and the discriminator D_η to optimize and u is the target image. The goal is to optimize θ and η.

Since we work with a residual framework, the output of the generator $G_\theta(\uparrow y, P)$ is a residual image. This residue contains information about spatial and spectral details. The final reconstructed image is obtained by adding this residue image with the low-resolution multispectral image $\uparrow y$. So the equations (3) and (2) are formulated in a residual way as follows:

$$\begin{aligned}\mathcal{L}(G_\theta) = \sum_{i \leq N_b} & \alpha log(D_\eta(G_\theta(\uparrow y, P) + \uparrow y)) \\ & + \delta ||u - G_\theta(\uparrow y, P) - \uparrow y||_1\end{aligned} \tag{4}$$

and

$$\begin{aligned}\mathcal{L}(D_\eta) = \sum_{i \leq N_b} & log(1 - D_\eta(G_\theta(\uparrow y, P) + \uparrow y)) \\ & + log(D_\eta(u)).\end{aligned} \tag{5}$$

For better understanding, in the rest of the paper, we set $\hat{u} = G_\theta(\uparrow y, P) + \uparrow y$ the generated image.

2.1 Bi-Discriminator

2.1.1 First discriminator

Motivated by texture preservation, the first discriminator considered D_{geom} takes in input the luminance and the NIR bands. Indeed, grayscale images allow to highlight texture and geometry. Moreover, as vegetation reflects in the NIR band, this band is very important to get information of texture and geometry in vegetation areas. So, the luminance and the NIR band are concatenated in order to get images with two bands and then used in input of this discriminator. The luminance is obtained by computing a linear combination of the red, blue and green bands.

Then, the loss function of the discriminator is:

$$\mathcal{L}_{D_{geom}} = \sum_{i \leq N_b} log(1 - D_{\eta_g}(\hat{u}_{YIR})) + log(D_{\eta_g}(u_{YIR})), \tag{6}$$

where $\hat{u}_{YIR}$ and u_{YIR} are the concatenation of the luminance Y and the NIR bands of $\hat{u}$ and u respectively and η_g the parameters of D_{geom} to optimize.

2.1.2 Second discriminator

Next, to preserve the spectral resolution and the color in images, we consider a second discriminator comparing the color of the target image and the fused image by using the chroma components Cb and Cr from a conventional YCbCr transformation. Indeed, the RGB inter-channel dependance is usually higher than the YCbCr inter-channel dependance. Consequently, to preserve the color, the Cb and Cr components are used. This amounts to minimizing:

$$\mathcal{L}_{D_{color}} = \sum_{i \leq N_b} log(1 - D_{\eta_c}(\hat{u}_{CbCr})) + log(D_{\eta_c}(u_{CbCr})), \tag{7}$$

where u_{CbCr} and $\hat{u}_{CbCr}$ correspond to the concatenation of the chroma components Cb and Cr for u and $\hat{u}$ respectively and η_c the parameters of D_{color} to optimize.

2.1.3 Architecture of discriminators

The architecture of each discriminator, presented in Fig. 4, is composed of seven convolutional layers with a number of feature maps increasing from 32 to 1024. These convolutional layers are used to extract enough features and capture the representation of data in a space. Then, to classify, two dense layers are added. This type of layers learns a function in the data space detecting whether the generated image is real or fake. Indeed, while convolutional layers extract features from a small receptive field, dense layers provide learning features from all the combinations of the features of the previous layer. So this type of layer allows to classify the information by taking into account all the previous features.

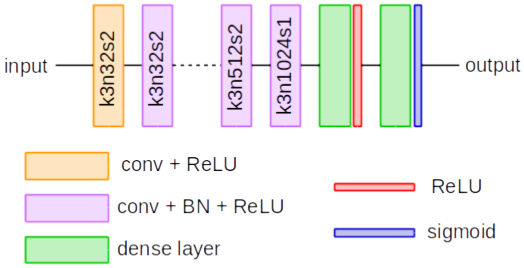


Figure 4: Architecture of each discriminator, where k is the kernel size, n the number of filters and s the stride of the convolutional layers.

2.2 Generator

2.2.1 Loss function of the MDSSC-GAN SAM method

The proposed loss function takes into account two types of constraints: a geometrical and a spectral constraint. Indeed, He *et al.* [19] propose to preserve the spectral resolution by adding a constraint based on the Spectral Angle Map (SAM) measure. They show that this term allows to improve visual results but also evaluation metrics. So, to preserve the spectral resolution, we propose to add this term in the loss function. In this case, the following loss function is minimized:

$$\begin{aligned}\mathcal{L}(G_\theta) = \sum_{i \leq N_b} & \alpha_g log(D_{\eta_g}(\hat{u}_{YIR})) + \alpha_c log(D_{\eta_c}(\hat{u}_{CbCr}) + \\ & \alpha_{l1} ||\hat{u} - u||_1 + \alpha_t \sum_{x \in \Omega} |\nabla u(x)^{\perp} . \nabla \hat{u}(x)| + \\ & \alpha_{sam} \sum_{x \in \Omega} arccos\left(\frac{u(x).\hat{u}(x)}{||u(x)||_2 ||\hat{u}(x)||_2} \right), \end{aligned} \tag{8}$$

where $\nabla(.)$ is the gradient operator, $\perp$ is the orthogonal vector, . is the inner product, Ω the image domain and $\alpha_{g,c,l1,t,sam}$ are weights. The first and the second term are the cross entropy terms associated to each discriminator. The third term is the l_1 norm between the target and the fused image. The fourth term is the geometrical constraint, it forces the alignment of gradients of each band of the solution with each band of the target image. This term, used in the RDGAN approach [18], allows us to improve the reconstruction of edges in images. The last one is the SAM measure computing the absolute value of the angle between two vectors composed of pixel values of the target and the generated image at each point of the domain. A SAM measure equal to zero indicates no spectral distortion but radiometric distortions can be present. It means that vectors are parallel but with a different wavelength.

2.2.2 Architecture

We consider the same architecture as in our RDGAN [18] approach; a residual dense architecture as in Fig. 5 for the generator.

Residual Dense architecture keep advantages of dense [23] and residual [24] connections. These types of architecture were introduced in order to solve the vanishing gradient problem faced during the training process of a deep network.

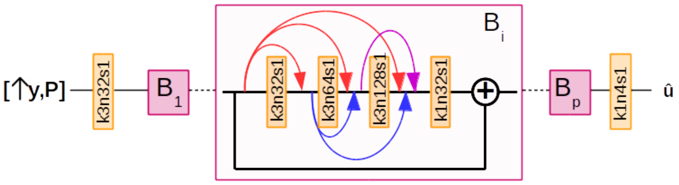


Figure 5: General architecture of the generator. The input $[P, \uparrow y]$ is the concatenation $[.]$ of the panchromatic image P and $\uparrow y$ the multispectral image $y = (y^1, ..., y^N)$ resized to the size of P, blocks B_i, $i \leq p$, are residual dense blocks, k is the kernel size of the convolution, n the number of filter and s the stride. Each convolutional layer is followed by a ReLU, except the last layer of the network. Arrows represent the dense connections and the + residual connection.

The considered architecture in Fig. 5 takes in input the concatenation of the panchromatic image P and $\uparrow y$ the multispectral image $y = (y^1, ..., y^N)$ resized to the size of P and is composed of several residual dense blocks.

3 Experiments

3.1 Dataset and quality evaluation

Pléiades satellite images and World View 3 satellite images are used to train and test the networks. Both images were acquired in the Bordeaux area, the Pléiades in August 2012 and the World View 3 in July 2018.

The World View 3 database is composed of panchromatic images and multispectral images with 4 bands (blue, green, red and near infrared). The wavelength of the blue band is between 450 nm and 510 nm, 510 nm and 580 nm for the green band, 630 nm and 690 nm for the red band and 770 nm and 895 nm for the near infrared band. Finally, the spectral range of the panchromatic images is 450 nm to 800 nm. In addition, the ground sampling distance is 0.31 m for the panchromatic images and 1.24 m for the multispectral images.

For the Pléiades satellite, the ground sampling distance is 2.8 m for multispectral images and 0.7 m for panchromatic images. The wavelength of the blue band is between 430 nm and 550 nm, 490 nm and 610 nm for the green one, 600 nm and 720 nm for the red one and 750 nm and 950 nm for the near infrared one. The panchromatic channel is sensitive to a wide range of wavelengths of visible light between 480 nm and 830 nm.

Thus, the satellites have different spatial and spectral resolution.

Satellite images are cropped into patches of size 128 × 128 to train and test. Finally, the Pléiades database is composed of 3173 samples for training and 356 for testing. The World View 3 database is composed of 5615 samples for training and 623 samples for testing.

To evaluate the results, we use the SAM metric presented in the loss function in Equation (8). In addition, to measure the quality of generated images, several criteria are considered. We note X and Y two images, X^i with $i \leq K$ the i^{th} band of X, X_j with $j \leq |\Omega|$ the j^{th} pixel of X, Ω the image domain and L the number of bands of the images.

3.1.1 The Cross Correlation coefficient (CC)

It evaluates the spatial distortion between the fused image and the target image by computing the intraband and interband correlations,

$$CC(X,Y) = \frac{1}{K} \sum_{i \leq K} \frac{\sum_{j \in \Omega} (X_j^i - \mu_X)(Y_j^i - \mu_Y)}{\sqrt{\sum_{j \in \Omega} (X_j^i - \mu_X)^2 \sum_{j \in \Omega} (Y_j^i - \mu_Y)^2}}, \tag{9}$$

where μ_X and μ_Y are the mean of X and Y respectively. The CC coefficient is between -1 and 1 with an ideal value of 1. Indeed, a coefficient equal to 1 indicates images are strongly positively correlated, i.e. images share a lot of spatial information.

3.1.2 The Peak Signal to Noise Ratio (PSNR)

It measures the reconstruction quality of the fused image,

$$PSNR(X,Y) = 10log_{10}\left(\frac{d^2}{MSE(X,Y)}\right), \tag{10}$$

where $MSE(X,Y)$ is the mean square error between X and Y with d the peak value, corresponding to the maximum fluctuation in the image.

3.1.3 The PSNR on high frequencies

Because texture and geometry are more visible in high frequencies, we propose to compute the PSNR on high pass filtered images. It is equivalent to consider:

$$PSNR_h(X,Y) = PSNR(X*h, Y*h), \tag{11}$$

where h is a Butterworth filter.

3.1.4 The Erreur Relative Global Adimentionnelle de Synthèse (ERGAS)

It gives a global measure of the quality of the fused image,

$$ERGAS(X,Y) = \frac{100}{d}\sqrt{\frac{1}{L}\sum_{l\leq L}\left(\frac{RMSE_l}{\mu_l}\right)^2}, \tag{12}$$

whith μ_l the mean of the l^{th} band, and d the resolution factor. The ideal value of this metric is 0.

Since the previous criteria require a target image, the Wald's protocol [25] is used to train the networks. It consists in reducing the spatial resolution of observed images in order to use the original multispectral image as target image. The spatial resolution is reduced using a gaussian kernel and a subsampling operator. This protocol offers a convenient way to evaluate results and to compare methods. It is commonly used when dealing with satellite images.

3.2 Details of implementation

The proposed method is implemented with Tensorflow 1.2 and it uses ADAM algorithm to optimize weights of networks with an initial learning rate of 0.0002 and a momentum of 0.5. Finally, the batch size is adjusted to 19 for the Pléiades database and 17 for the World View database.

Parameters $\alpha_{g,c,l_1,t,sam}$ are optimized to get the best balance between all the metrics. Moreover, we put the same weight on each discriminator so that they have the same importance.

3.3 Results

3.3.1 Ablation study

First, to underline the advantages of the proposed method, we compare our method with several methods. The first and the second one consist in turning off each discriminator individually. The third one consists in considering one discriminator taking in input the concatenation of the luminance, the near infrared band and the chroma components. This method allows to highlight the advantage of the bi-discriminator aspect. Finally, the last method is the proposed bi-discriminator method without the spectral constraint in the generator loss function. This last method allows to underline the contribution of the spectral constraint of our proposed method. So it amounts to compare five methods, summarized in Tab. 1:

Method	Number of discriminators	inputs of discriminators	constraints in the loss function of the generator
'MDSSC-GAN concat'	1	[Y, NIR, Cb, Cr]	spatial
'MDSSC-GAN texture'	1	[Y, NIR]	spatial
'MDSSC-GAN color'	1	[Cb, Cr]	spatial
'MDSSC-GAN'	2	[Y,NIR] and [Cb, Cr]	spatial
'MDSSC-GAN SAM'	2	[Y,NIR] and [Cb, Cr]	spatial and spectral

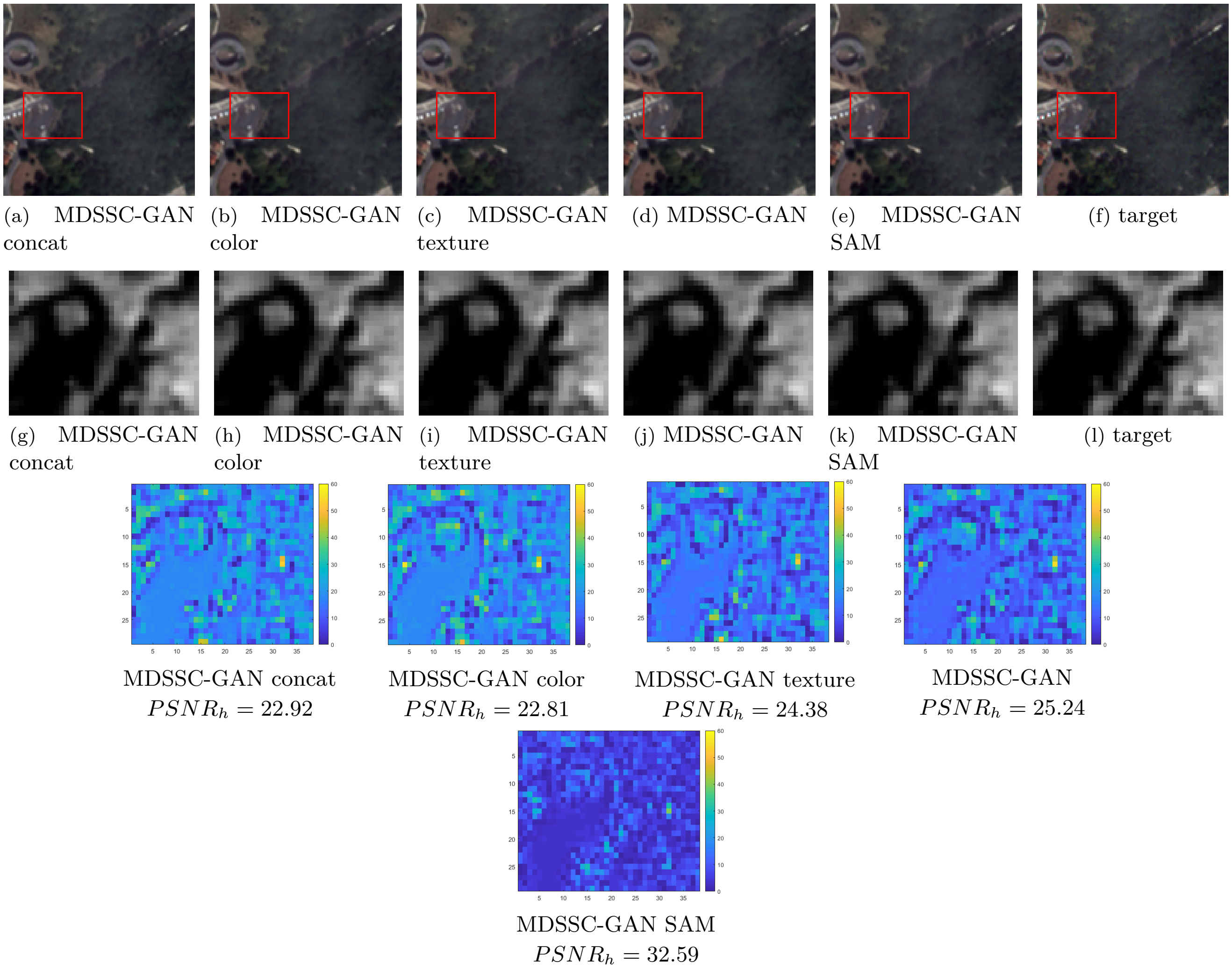
Table 1: Summary of compared methods, where [.] represents the concatenation, Y the luminance, NIR the near infrared band and Cb and Cr the chroma components.

Quantitative results are presented in Tab. 2.

Method	CC	SAM	PSNR	$PSNR_h$	ERGAS
ideal value	1	0	max	max	0
MDSSC-GAN concat	<u>0.969</u>	<u>0.141</u>	29.416	<u>27.224</u>	<u>3.888</u>
MDSSC-GAN color	**0.970**	0.139	29.492	27.406	**3.875**
MDSSC-GAN texture	<u>0.969</u>	0.140	<u>29.394</u>	**27.5015**	3.887
MDSSC-GAN	**0.970**	0.138	**29.493**	27.361	3.884
MDSSC-GAN SAM	**0.970**	**0.137**	29.455	27.455	3.881

Table 2: Quantitative results when comparing performances of each discriminator individually with the proposed methods on the Pléiades database. 'MDSSC-GAN color' considers only the color discriminator, 'MDSSC-GAN texture' considers only the geometrical/texture discriminator, 'MDSSC-GAN concat' considers only one discriminator taking in input the concatenation of the luminance, the near infrared band and the chroma components Cb and Cr, 'MDSSC-GAN' and 'MDSSC-GAN SAM' the proposed methods, the last one adding a spectral constraint to the loss function. Best results are in bold and worst results are underlined.

First, we can note that the bi-discriminator aspect improves substantially quantitative results (Tab. 2). Indeed, the 'MDSSC-GAN concat' method shows that if one just modifies the inputs while keeping only one discriminator is not sufficient to improve results. Then, when only the geometrical discriminator is considered, results are not very convincing, except for the high frequencies measure (PSNR$_h$). While considering only the color discriminator, the obtained results are equivalent to the proposed bi-discriminator methods except for the PSNR$_h$ metrics which is not as well as the 'MDSSC-GAN SAM' method. Tab. 2 shows that the geometrical discriminator reconstructs better the high frequencies and the color discriminators improves the other metrics in general. Finally, when comparing the 'MDSSC-GAN' method with the other proposed 'MDSSC-GAN SAM' method, we can see that adding a spectral constraint in the loss function of the generator improves the SAM and the PSNR$_h$ measure while degrading the PSNR metric.



(a) MDSSC-GAN concat (b) MDSSC-GAN color (c) MDSSC-GAN texture (d) MDSSC-GAN (e) MDSSC-GAN SAM (f) target

(g) MDSSC-GAN concat (h) MDSSC-GAN color (i) MDSSC-GAN texture (j) MDSSC-GAN (k) MDSSC-GAN SAM (l) target

MDSSC-GAN concat $PSNR_h = 22.92$ MDSSC-GAN color $PSNR_h = 22.81$ MDSSC-GAN texture $PSNR_h = 24.38$ MDSSC-GAN $PSNR_h = 25.24$

MDSSC-GAN SAM $PSNR_h = 32.59$

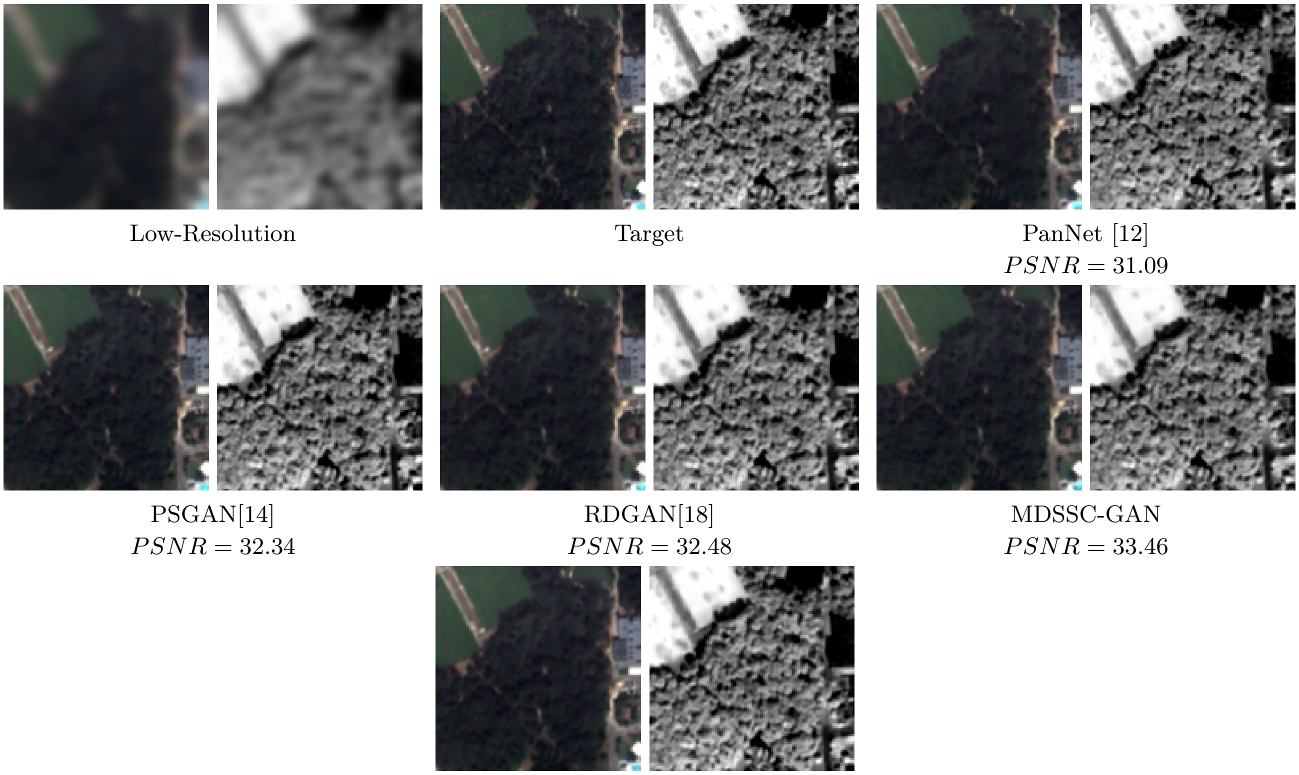
Figure 6: Visual results of a Pléiades sample on vegetation areas when comparing performances of each discriminator individually with the proposed MDSSC-GAN method. MDSSC-GAN color considers only the color discriminator, MDSSC-GAN texture considers only the geometrical discriminator, MDSSC-GAN concat considers only one discriminator taking in input the concatenation of the luminance, the near infrared band and the chroma components Cb and Cr, MDSSC-GAN andMDSSC-GAN SAM are the proposed methods, the last one adding a spectral constraint to the loss function. The first line is RGB images, the second line near infrared bands and the third line differences between the high frequencies of target images with the high frequencies of fused images. We can see that the best results is given by the proposed MDSSC-GAN SAM method.

Visual results in Fig. 6 show an example on vegetation areas. As vegetation reflects better in infrared, we choose to display results of the near infrared band. This example shows that the proposed method with a spectral constraint reconstructs high frequencies much better than the other one.

So, quantitative and visual results show advantages to use several discriminators compared to only one discriminator, by an improvement of metrics and a better high frequencies reconstruction.

3.3.2 Comparison with state-of-the-art

In a second time, we compare our proposed method with some recent state-of-the-art pansharpening methods on the Pléiades and the World View 3 databases. The coefficient injection methods GLP [27] and GSA [26] and the PanNet [12], PSGAN [14] and RDGAN [18] networks. For a more coherent and relevant comparison, all networks are trained with the same Pléiades dataset. The training time is about 24 hours for the 'PanNet' and the proposed methods based bi-discriminator 'MDSSC-GAN' and 'MDSSC-GAN SAM', 15 hours for the 'PSGAN' method and 12 hours for the 'RDGAN' method.



Low-Resolution

Target

PanNet [12]
$PSNR = 31.09$

PSGAN[14]
$PSNR = 32.34$

RDGAN[18]
$PSNR = 32.48$

MDSSC-GAN
$PSNR = 33.46$

MDSSC-GAN SAM
$PSNR = 33.76$

Figure 7: Visual results for state-of-the-art methods on a Pléiades sample. For each method is displayed the RGB image on the left and the near infrared (NIR) band on the right. A significant improvement if the PSNR is observed with our new method MDSSC-GAN SAM.

Tab. 3 presents a comparison of quantitative results between state-of-the-art methods with the Pléiades database. In a first time, we can see that deep learning approaches give better results than GLP and GSA methods. But it is important to note that the coefficient injection methods GLP and GSA need the coefficients α_k, $k \leq K$, of the Equation (1) for the fusion and we do not have these coefficients. So it may be possible to get slightly better results with these methods. Then, the proposed method gives best quantitative results for most metrics. We can note an improvement of the SAM, the CC and the PSNR$_h$ metrics meaning a better preservation of the spectral and spatial resolution and a better reconstruction of high frequencies with the 'MDSSC-GAN SAM' method, but a better PSNR with the 'MDSSC-GAN' method.

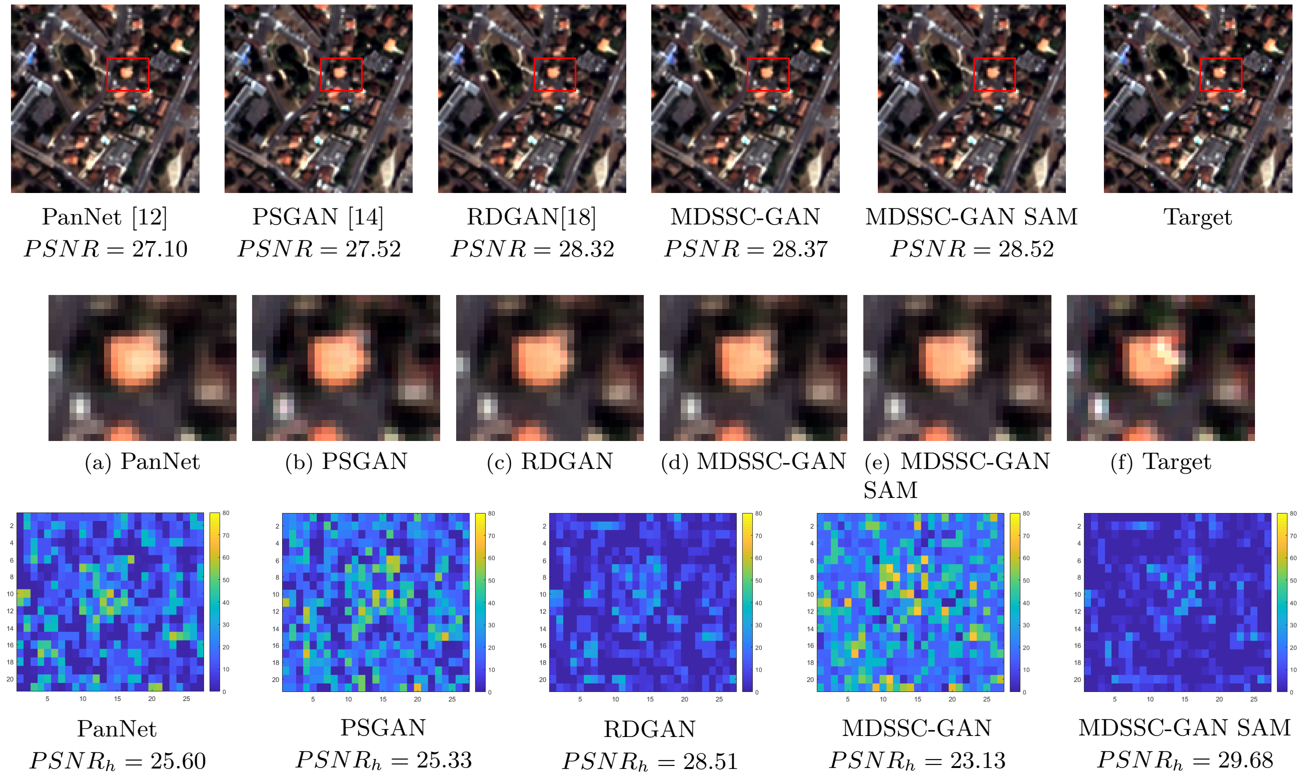
Method	CC	SAM	PSNR	$PSNR_h$	ERGAS
ideal value	1	0	max	max	0
GSA [26]	0.871	0.237	21.94	22.41	10.74
GLP [27]	0.866	0.242	21.74	23.05	10.91
PanNet [12]	0.950	0.157	28.36	26.60	8.77
PSGAN [14]	0.952	0.155	26.59	26.96	4.28
RDGAN [18]	0.969	0.138	29.38	27.23	3.94
MDSSC-GAN	**0.970**	0.138	**29.49**	27.36	**3.88**
MDSSC-GAN SAM	**0.970**	**0.137**	29.45	**27.45**	**3.88**

Table 3: Quantitative results when comparing performances of state of the art methods with the Pléiades database. Best results are in bold and worst results underlined.

First, Fig. 7 is a general example of image fusion of all the state-of-the-art methods. It is difficult to see a major visual difference in this example, but we note a significant improvement of the PSNR with our proposed method

’MDSSC-GAN SAM’. Indeed, we improve the PSNR about 0.3 dB on this image face to the best state-of-the-art method.

Fig. 8 gives an example of the high frequencies reconstruction for each method. In fact, this figure displays the RGB image and the difference between high frequencies of the target image and high frequencies of the fused image for each method. We can see that the best visual result is for the proposed ’MDSSC-GAN SAM’ method. In this example, the best PSNR does not give the best high frequencies reconstruction. Indeed, we can see that the worst high frequencies reconstruction is for the ’MDMD-GAN’ method, but the PSNR of this image is one of the best. On the contrary, the proposed ’MDSSC-GAN SAM’ method has a similar PSNR, but the high frequencies are much better reconstructed.



PanNet [12] $PSNR = 27.10$

PSGAN [14] $PSNR = 27.52$

RDGAN[18] $PSNR = 28.32$

MDSSC-GAN $PSNR = 28.37$

MDSSC-GAN SAM $PSNR = 28.52$

Target

(a) PanNet

(b) PSGAN

(c) RDGAN

(d) MDSSC-GAN

(e) MDSSC-GAN SAM

(f) Target

PanNet $PSNR_h = 25.60$

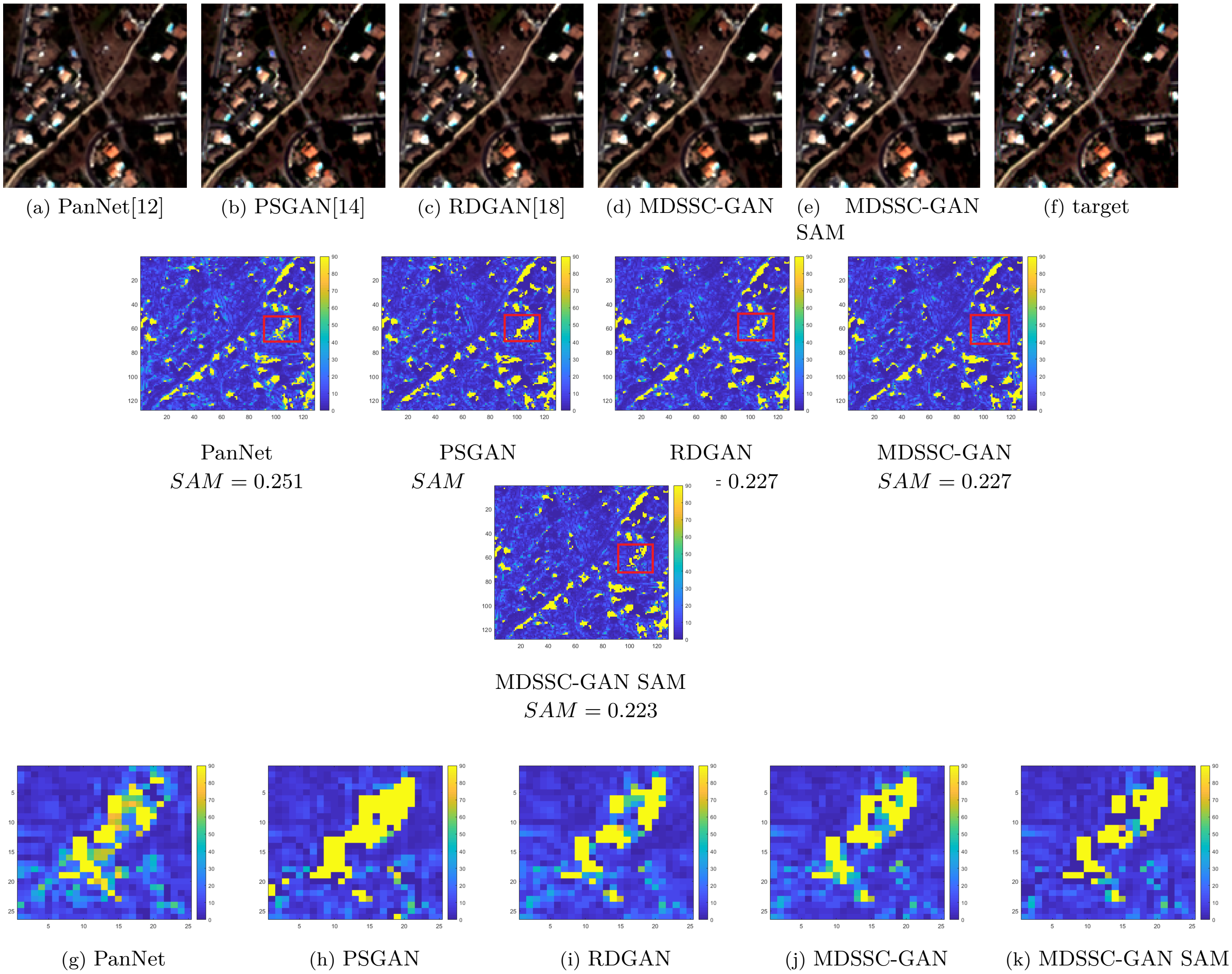
PSGAN $PSNR_h = 25.33$

RDGAN $PSNR_h = 28.51$

MDSSC-GAN $PSNR_h = 23.13$

MDSSC-GAN SAM $PSNR_h = 29.68$

Figure 8: Visual results with Pléiades database of state of the art methods on urban areas. The first line shows the RGB images. The second line displays some zoom part of the RGB image. The third line presents differences between high frequencies of the target image and high frequencies of the fused image in the zoom part. The best high frequency reconstruction is given by the proposed MDSSC-GAN SAM method.


(a) PanNet[12] (b) PSGAN[14] (c) RDGAN[18] (d) MDSSC-GAN (e) MDSSC-GAN SAM (f) target

PanNet $SAM = 0.251$ PSGAN SAM RDGAN $= 0.227$ MDSSC-GAN $SAM = 0.227$

MDSSC-GAN SAM $SAM = 0.223$

(g) PanNet (h) PSGAN (i) RDGAN (j) MDSSC-GAN (k) MDSSC-GAN SAM

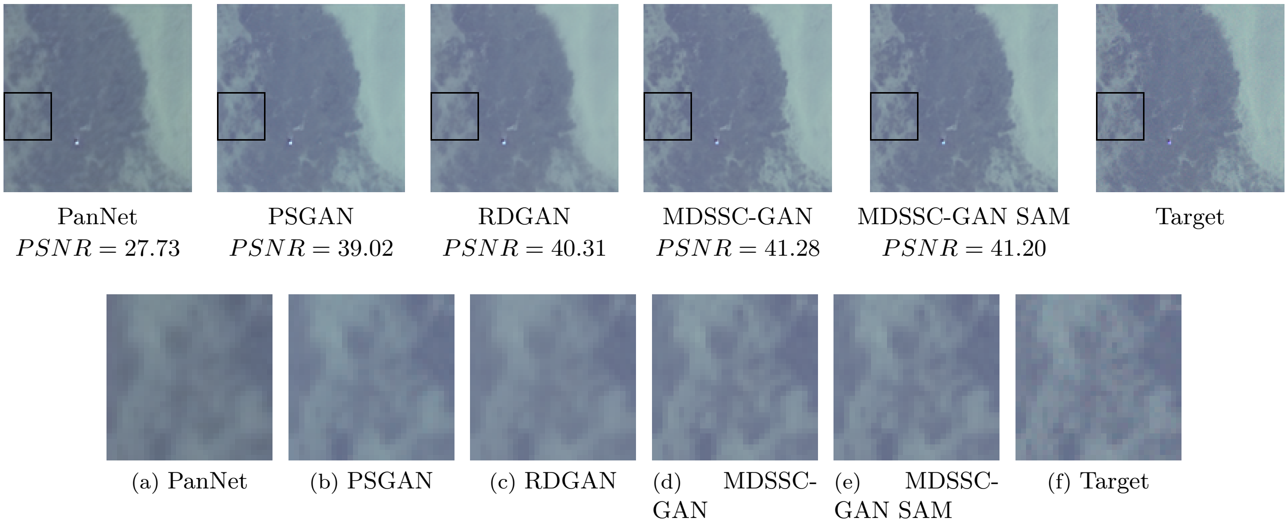
Figure 9: Visual results of state of the art methods on a Pléaides sample. The first line shows the RGB images. The second line displays the SAM map, i.e. the angle of distortion between the fused image and the target at each pixel. The third line presents a zoom of the SAM map. The blue color indicates an angle of distortion of 0 and the yellow one an angle of 90 degrees. The best result is obtained by the proposed MDSSC-GAN SAM method.

Then, Fig. 9 shows an example of the preservation of the spectral resolution. This Fig. displays the SAM metrics at each pixel, i.e. the angle of distortion between the target image and the fused image at each pixel. A pixel with zero distortions is displayed in blue and a pixel with a high distortion in yellow. In this example, it is more difficult to see a substantial difference on the whole image but in the zoomed part, we can see that adding the SAM term in the loss function of the generator ('MDSSC-GAN SAM') allows to get better results.

Method	CC	SAM	PSNR	$PSNR_h$	ERGAS
ideal value	1	0	max	max	0
GSA [26]	0.879	0.184	22.67	23.14	19.94
GLP [27]	0.880	0.179	22.67	23.07	14.53
PanNet [12]	0.930	0.135	23.22	23.93	8.01
PSGAN [14]	0.966	0.110	29.30	26.82	6.33
RDGAN [18]	0.966	0.107	29.39	26.20	6.58
MDSSC-GAN	0.967	0.106	29.44	26.59	6.48
MDSSC-GAN SAM	**0.968**	**0.104**	**29.62**	**26.82**	**6.31**

Table 4: Quantitative results when comparing performances of state of the art methods with the World View 3 database. Best results are in bold and worst results underlined.

Finally, on the World View 3 database, quantitative results presented in Tab. 4 show the same improvement with our proposed method. Compared to the results with the Pléiades database, we can see a better SAM metric for all methods, but other metrics are of the same order. The best approach is once again our 'MDSSC-GAN SAM' method.


PanNet $PSNR = 27.73$

PSGAN $PSNR = 39.02$

RDGAN $PSNR = 40.31$

MDSSC-GAN $PSNR = 41.28$

MDSSC-GAN SAM $PSNR = 41.20$

Target

(a) PanNet

(b) PSGAN

(c) RDGAN

(d) MDSSC-GAN

(e) MDSSC-GAN SAM

(f) Target

Figure 10: Visual results of state of the art methods of a World View 3 sample on vegetation area. The first line shows the RGB images. The second line displays a zoom of the RGB images. We can see a better reconstruction of texture when considering multiple discriminators. Our proposed methods 'MDSSC-GAN' and 'MDSSC-GAN SAM' give better results.

With the World View 3 database, both quantitative results presented in Tab. 4 and visual results presented in Fig. 10 show that the 'MDSSC-GAN' and 'MDSSC-GAN SAM' methods give better results. Indeed, on Fig. 10 (which is an example of vegetation area), the texture of the forest is better reconstructed when considering multiple discriminators.

4 Conclusion

To conclude, we propose a method, named 'MDSSC-GAN SAM', for the pansharpening problem. This method considers bi-discriminator in a Generative Adversarial Network framework, by training two discriminators with a different and complementary task. The first one is to improve the spatial resolution and the second one to improve the spectral resolution. Moreover, both geometrical and spectral constraints in the generator loss function are added to enhance these aspects.

Experiments on Pléiades and World View 3 satellite images have shown that the proposed method 'MDSSC-GAN SAM' gives better quantitative and visual results. Quantitative results show an improvement for all spatial and spectral metrics and visual results confirm it by reconstructing better high frequencies and spectral contents in images.

Acknowledgment

The authors would like to thank the team of the thematic recipe users of the program ORFEO of accompaniment Pléiades for the Pleiades images (©CNES (2012,2013), distribution Airbus DS/Spot Image) used within the framework of this study. The World View 3 satellite image was provided by DigitalGlobe, Inc. under Creative Commons license. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - SysNum (ANR-10-IDEX-03-02). The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-18-CE92-0050 SUPREMATIM.

References

[1] M. Gonzalez-Audicana, J.L. Saleta, R. Garcia Catalan, and R. Garcia, "Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA Mergers based on Wavelet Decomposition," *IEEE TGRS, vol. 42, No. 6, pp.1291-1299*, 2004.

[2] W. Carper, T. Lillesand, and R. Kiefer, "The use of Intensity-Hue-Saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm Eng Remote Sensing, vol. 56, No. 4, pp. 459-467*, 1990.

[3] M. Gonzalez-Audicana, X. Otazu, O. Fors, and A. Seco, "Comparison between Mallat's and the 'A trous' discrete wavelet transform algorithms for the fusion of multispectral and panchromatic images," *Int J Remote Sens, vol. 26, no. 3, pp. 595-614*, 2005.

[4] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on communications, vol. 31, No. 4, pp.532-540*, 1983.

[5] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A Variational Model for P+XS Image Fusion," *IJCV, vol. 69, no. 1, pp 43-59*, 2006.

[6] J. Duran, A. Buades, B. Coll, and C. Sbert, "A non local variational model for pansharpening image fusion," *SIAM, vol. 7, no. 2, pp. 761-796*, 2015.

[7] F. Palsson, J. Sveinsson, M. Ulfarsson, and J. Benediktsson, "A New Pansharpening Method Using an Explicit Image Formation Model Regularized via Total Variation," *IEEE International Geoscience and Remote Sensing Symposium*, 2012.

[8] G. Masi, D. Cozzolino, L. Verdolina, and G. Scarpa, "Pansharpening by Convolutional Neural Network," *Remote Sensing, vol. 8, no. 7, pp. 594-616*, 2016.

[9] C. Dong, L.K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on pattern analysis and machine intelligence*, 2016.

[10] F. Palsson, J. Sveinsson, and M. Ulfrasson, "Multispectral and Hyperspectral Image Fusion Using 3D Convolutional Neural Network," *IEEE Goescience and remote sensing*, 2017.

[11] Y. Guo, F. Ye, and H. Gong, "Learning an Efficient Convolution Neural Network for Pansharpening," *Algorithms, vol. 12, pp. 16*, 2019.

[12] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A Deep Network Architecture for Pan-Sharpening," *IEEE International Conference on Computer Vision*, October 2017.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Neural Information Processing Systems Proceedings*, 2014.

[14] X. Liu, Y. Wang, and Q. Liu, "PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-sharpening," *IEEE International Conference on Image Processing*, 2018.

[15] Q. Liu, H. Zhou, Q. Xu, X. Liu, and Y. Wang, "PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-sharpening," *IEEE Transactions on Geoscience and Remote Sensing, vol. 14, no. 8*, 2020.

[16] Y. Zhang, X. Li, and J. Zhou, "SFTGAN : a generative adversarial network for pan-sharpening equipped with spatial feature transform layers," *Journal of Applied Remote Sensing, vol.13, no. 2*, 2019.

[17] X. Wang, K. Yu, C. Dong, and C.-C. Loy, “Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform,” *IEEE Conference on Computer Vision Pattern Recognition, pp. 606-615*, 2018.

[18] A. Gastineau, J.-F. Aujol, Y. Berthoumieu, and C. Germain, “A Residual Dense Generative Adversarial Network for Pansharpening with Geometrical Constraints,” *IEEE International Conference on Image Processing*, 2020.

[19] G. He, J. Zhong, J. Lei, Y. Li, and W. Xie, “Hyperspectral Pansharpening based on Spectral Constrained Adversarial Autoencoder,” *Remote Sensing*, 2019.

[20] X. Zhu, Y. Cheng, J. Peng, M. Wang, R. Le, and X. Liu, “Super-Resolved Image Peceptual Quality Improvement via Multi-Feature Discrimiantors,” *CoRR, abs/1904.10654*, 2019.

[21] O.-Y. Lee, Y.-H. Shin, and J.-O. Kim, “Multi-Perspective Discriminators-Based Generative Adversarial Network for Image Super Resolution,” *IEEE*, 2019.

[22] J. Park, D. Han, and H. Ko, “Adaptive Weighted Multi-Discriminator CycleGAN for Underwater Image Enhancement,” *Journal of MArine Science and Engineering*, 2019.

[23] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberer, “Densely connected convolutional networks,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[25] L. Wald, T. Ranchin, and M. Mangolini, “Fusion of satellite images of different spatial resolution: Assessing the quality of resulting images,” *Photogramm. Eng. Remote Sens., vol. 63, no. 6*, 1997.

[26] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, “Context-adaptive Pansharpening Based on Image Segmentation,” *IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 2*, 2017.

[27] G. Vivone, R. Restaino, and J. Chanussot, “Full Scale Regression-based Injection Coefficients for Panchromatic Sharpening,” *IEEE Transactions on Image Processing, vol. 27, no. 7*, 2018.