International Journal of
Geo-Information

Article

Comparative analysis of remote sensing image fusion based on deep convolutional neural network with different loss functions

Abstract: Spatial resolution is an important attribute of remote sensing images. Remote sensing image fusion (also known as pan-sharpening), which is of great significance to information extraction and analysis, is an important method to improve the spatial resolution of existing images. Recently, pan-sharpening based on convolutional neural network has advanced rapidly. However, comparison between their loss functions is lacking. Therefore, this paper compares ×2 and ×4 pan-sharpening performance of the four widely used loss functions, namely, L1 loss, L2 loss, the mixed loss of L1 loss and adversarial loss (L1+ad), and the mixed loss of L2 loss and adversarial loss (L2+ad). Experiments on images acquired from Landsat8-OLI demonstrate that L1 loss yields better image reconstruction effect with the smallest distortion, whereas mixed losses exhibit better performance in terms of visual effect.

Keywords: loss function, pan-sharpening, deep convolutional network, remote sensing

1. Introduction

Remote sensing has been widely used in environment monitoring, city planning, and military reconnaissance. Accordingly, it is of great significance to environmental protection, economic development, and national defense construction. High spatial resolution remote sensing images are important data sources for accurate ground object recognition and classification. However, improving the spatial resolution by developing high-resolution satellite sensors is time consuming and costly. Image fusion algorithm is an effective and economical way to improve image spatial resolution. Two kinds of images, namely, low-resolution multi-spectral image (LRMS) and high-resolution panchromatic image (PAN), can be acquired simultaneously by most of the optical imaging sensors. Pan-sharpening is an image fusion approach that uses LRMS and PAN to generate a high-resolution multi-spectral image (HRMS).

Pan-sharpening algorithm gained considerable attention because of its efficiency and economy and the fact that the fusion image considers spatial and spectral information. The traditional algorithms for pan-sharpening can be divided into component substitution [1], compressed sensing theory [2], and multi-resolution analysis [3]. Component substitution mainly includes HIS transform [4], Brovey transform [5], PCA transform [6], and Gram–Schmidt [7]. Recently, pan-sharpening algorithm based on deep learning has also been proposed. Experiments show that these algorithms outperform traditional ones [8]. Since then, pan-sharpening algorithm based on deep learning gained considerable attention, and a large number of algorithms, such as PNN [8] and PanNet [9], have been proposed.

Loss function is an important part of deep learning algorithms. The value of loss is generated by the loss function as a metric. Then, the deep neural network minimizes this loss through gradient descent and obtains its targeted network parameters. Loss function can be selected according to different tasks. For example, L1 loss and L2 loss are always used in regression tasks, and cross entropy is often used in classification tasks.

As a special regression task, L1 loss and L2 loss are still used in many works of remote sensing image fusion. Inspired by SRCNN[10], Giuseppi et al. proposed a convolutional neural network (CNN) called PNN for remote sensing image fusion[8] using L2 loss as the loss function. Rao et al. proposed a residual CNN for pan-sharpening (RCNNP) for image fusion of Landsat7-ETM+, which also used L2 loss[11]. Yang et al. proposed PanNet and used the squared Frobenius-norm as the loss function but was in fact equivalent to L2 loss[9]. Lanaras et al. proposed DSen2 model for the image fusion of Sentinal-2 satellite images. They used L1 loss and pointed out that L1 loss led to a sharpened edge and better convergence[12]. Scarpa et al. found that L1 loss had better effect than L2 loss in pan-sharpening tasks[13].

Generative adversarial network (GAN) is a framework that obtains the model parameters by optimizing the generator and the discriminator alternatively [14]. Adversarial loss is generated by discriminator to optimize the two neural networks. GAN has been applied to image processing fields, such as image generation[15], image-to-image translation[16], and image super resolution[17,18]. In the fields of remote sensing image fusion, the PSGAN model proposed by Liu et al.[19] and RED-cGAN proposed by Shao et al.[20] adopted the GAN framework, in which a combination of the L1 loss and adversarial loss was used as the loss function. The experiments showed that the model had better effect than previous models.

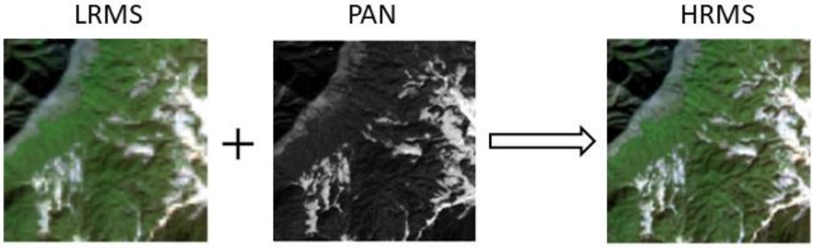


Figure 1. The process of pan-sharpening

In summary, L1 loss, L2 loss, the mixed loss of L1 or L2 loss with adversarial loss are widely used for pan-sharpening networks. Considering that many neural networks have been proposed, the performance of this kind of network has been very outstanding, and improving it is very difficult[13]. However, with regard to the selection of loss functions for pan-sharpening, experiment evidence is lacking. For this reason, this paper does not focus on the slight improvement of the networks but on the selection of loss functions. The performance of the four loss functions on ×2 and ×4 pan-sharpening, which are two frequently-used image fusion ratios, are compared. The main contributions of this paper are presented as follows:

- The paper discusses principle of the losses in the background of pan-sharpening and compares the performance of pan-sharpening networks under different loss functions. And we also provide advice on their research and engineering.
- For the sake of comparing loss functions, we summarized the proposed models and designed a universal model for pan-sharpening. By using the same model, the influence of differences in models can be avoided.
- In this paper, loss functions in different bands, ground objects, and fusion ratios are compared. Hence, the experiments were very ample.

The paper is organized as follows: Section 2 introduces the theory of loss functions and the structure of the neural network used in the experiments. Section 3 presents the experiment details, results, and discussions. Section 4 concludes this paper.

2. Materials and Methods

2.1 Pan-sharpening Networks

Inspired by PanNet, a simple deep CNN for pan-sharpening, we designed a fundamental reconstructed network. Several kinds of pan-sharpening networks are referred.

According to the different ratios of up sampling, the proposed network was classified into ×2 and ×4 networks.

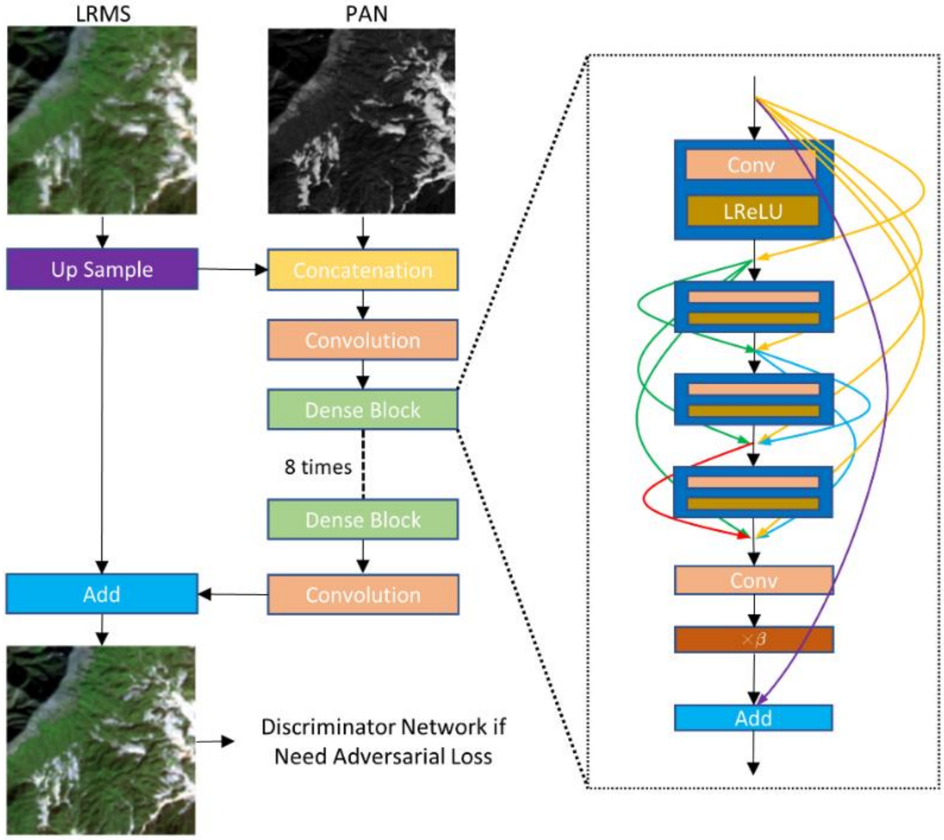


Figure 2. The structure of the generator

The network has two inputs, PAN and LRMS. After up sampling, LRMS and PAN are concatenated. Then, the dimension of the feature becomes 8. After a convolutional layer, the feature was mapped to 64 dimensions. Subsequently, 8 times of Dense Block[21] are used to learn the nonlinear mapping. The Dense Block is shown in Figure 2. The input of Dense Block was convoluted and activated by LReLu for 4 times and every convolutional layer mapped the feature to 32 dimensions. After each activation, the current features, the output of previous convolutional layers, and the input of the Dense Block are connected. Next, a convolutional layer is used to map the dimension to the input dimension. Finally, it was scaled and added to the input as the output of the Dense Block. The obtained scaling factor β in the experiments is 0.2. After 8 Dense Blocks in a row, a convolutional layer is used to map the feature to 7 dimensions, same as the HRMS. In the end, considering the idea of residual learning, the output of the last convolutional layer is added to the up sampled LRMS, and then outputted. In fact, the blocks are not limited to Dense Block. Many choices, such as Residual Block[22] and RRDB[18], are available for the blocks. In this experiment, for the sake of comparison, the blocks are all set to Dense Block. Moreover, because the Batch Normalization (BN) layer[23,24] cannot enhance the performance of pan-sharpening networks and leads to calculation overload[25], it was not used in the experiments.

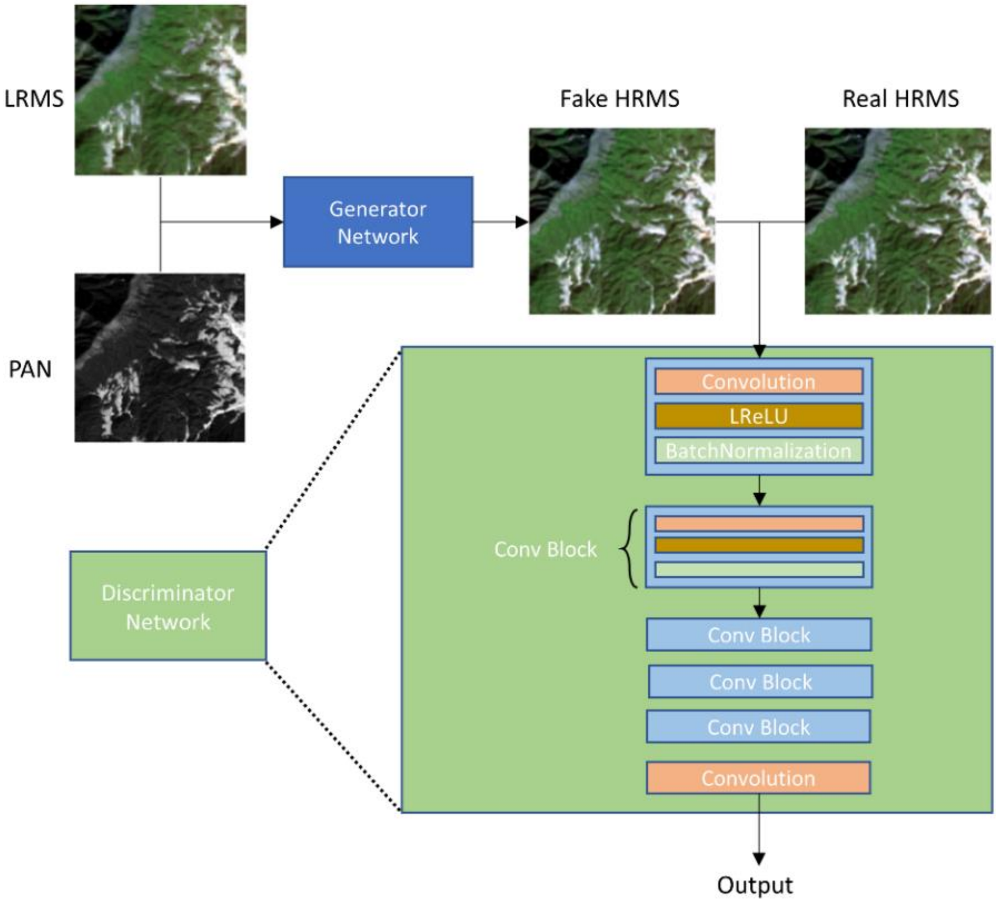


Figure 3. The structure of the discriminator

In this paper, the discriminator is another deep CNN for generating adversarial loss. The structure of the discriminator is shown in Figure 3. The input of the model is the fake HRMS generated by the generator or the real HRMS in the dataset. Five convolutional layers are used every time the convolution maps the feature to 64 dimensions, and strides are 2. Then, a convolutional layer is used to map the features to one dimension as the output of the discriminator.

2.2 Loss Functions

When processing the images by using neural networks, we aim to minimize the distance between the outputs and the labels. When calculating the distance on single image, the formula is expressed as follows:

$$L = \frac{1}{N}\varepsilon(f(X_P, X_M;\theta), Y)\,, \tag{1}$$

where $f(X_P, X_M;\theta)$ is the output of the neural network, X_P is the panchromatic image, X_M is the LRMS image, θ represents the parameters of the neural network, N is the number of all pixels, Y is the label image, and ε is a measure of the distance.

The measure ε could be L_p norm. For pan-sharpening, L_p norm takes a formula as follows:

$$L_p = (\sum_{\mathrm{n}=1}^{N} \left| f(X_P, X_M;\theta)_n - Y_n \right|^p)^{\frac{1}{p}}\,. \tag{2}$$

$p = 1$ indicates L1 loss. When $p = 2$, the squared L_2 norm is used to define L2 loss [26].

2.2.1. L2 loss

For the pan-sharpening task, the most popular loss function is L2 loss. It is calculated as follows:

$$L=\frac{1}{N}\sum_{n=1}^{N}(f(X_P,X_M;\theta)_n-Y_n)^2 . \quad (3)$$

An equivalent form mentioned by Yang et al.[9] is presented as follows:

$$L=\|f(X_P,X_M;\theta)-Y\|_F^2 , \quad (4)$$

where $\|\bullet\|_F$ represents the Frobenius-norm.

If we use y to represent the pixel value of one pixel and use $f(X_P,X_M;\theta)_p$ to represent the value predicted by the neural network at the same pixel, then when given the LRMS and the PAN, the pixel value is assumed to be a Gaussian random variable, and the prediction value of the neural network is the mean of the Gaussian distribution, that is, $y\sim N(f(X_P,X_M;\theta)_{\mathrm{p}},\sigma^2)$. If maximum likelihood estimation is performed, then:

$$\begin{aligned}\theta_{\mathrm{ML}}&=\arg\max_{\theta}\prod_{n=1}^{N}p(y_n\mid X_P,X_M;\theta)\\&=\arg\max_{\theta}\prod_{n=1}^{N}N(y_n\mid f(X_P,X_M;\theta)_n,\sigma^2)\end{aligned} . \quad (5)$$

Furthermore, the log-likelihood is maximized. The formula is equivalent to the following formula:

$$\theta_{\mathrm{ML}}=\arg\max_{\theta}-\frac{1}{N}\sum_{n=1}^{N}(f(X_P,X_M;\theta)_n-Y_n)^2 . \quad (6)$$

This formula is equivalent to L2 loss. Thus, the use of L2 loss assumes that the predicted values are Gaussian random variables or Gaussian noise is present on the prediction of the pixels. When L2 loss is used, the neural network provides a mean of all possible solutions, which leads to a blurring effect. Smooth results are shown in many works, such as SRCNN[10] and SRGAN[17].

In addition, the derivative of the L2 loss for each pixel in the predicted values is:

$$\frac{\partial L}{\partial f(X_P,X_M;\theta)_n}=\frac{1}{N}\cdot 2(f(X_P,X_M;\theta)_n-Y_n) . \quad (7)$$

In some literature, $\frac{1}{2}$ is added to L2 loss for convenience, but it is equivalent to our formula. The value of the derivative depends on the difference between the prediction and the label, thereby leading to a bigger gradient at the beginning of training. However, this phenomenon also makes L2 loss more tolerant to small errors. Furthermore, L2 loss is easy to be influenced by outliers, thereby reducing the stability of the training.

2.2.2 L1 Loss

The L1 loss for pan-sharpening is defined as follows:

$$L=\frac{1}{N}\sum_{n=1}^{N}|f(X_P,X_M;\theta)_n-Y_n| . \quad (8)$$

Same as L2 loss, when using L1 loss, the predicted values that obey the Laplace distribution are assumed, that is, $y\sim \mathit{Laplace}(f(X_P,X_M;\theta)_p,\gamma)$. The L1 loss enables the neural network to provide the median of all solutions.

Compared with L2 loss, L1 loss has greater robustness for outliers. The derivative of L1 loss for each pixel of the pan-sharpened image is:

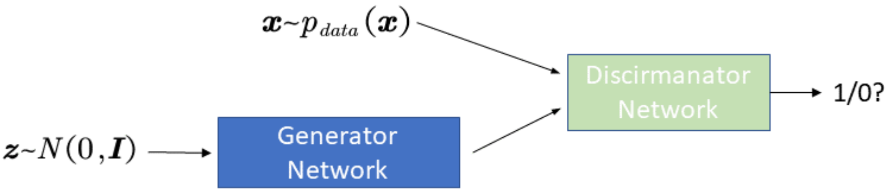
$$\frac{\partial L}{\partial f(X_P, X_M;\theta)_n} = \frac{1}{N} sign(f(X_P, X_M;\theta)_n - Y_n) \, . \qquad (9)$$

For the predicted values deviant from the label values, the same derivatives are provided by L1 loss as those close to the label values. In other words, the training process is very stable, either in the early stage or the late stage of training. Outliers are common for remote sensing images. Therefore, L1 loss seems to be more suitable for remote sensing images with many outliers.

2.2.3 Adversarial Loss

The generative adversarial net is a framework that uses the idea of two-player max–min game. The mapping from latent variable to the random variable of real data distribution is learned by neural networks. The framework consists of two networks, namely, the generator and the discriminator. As shown in Figure 4, for a sample z from a distribution, the generator aims to map it to a sample from the real data distribution as the output. The discriminator receives the output of the generator or the real data x and distinguishes if the data are the real data. It outputs the probability of whether the data are the real data. The goal could be written as follows:

$$\begin{aligned}\min_G \max_D V(D,G) = {} & E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] \\ & + E_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))] \, .\end{aligned} \qquad (10)$$



$\boldsymbol{x}\sim p_{data}(\boldsymbol{x})$
$\boldsymbol{z}\sim N(0,\boldsymbol{I})$
Generator Network
Discirmanator Network
1/0?

Figure 4. The structure of generative adversarial nets

In actual training, the generator and the discriminator are optimized alternately. For the generator, only the second term needs to be optimized. Considering the gradient of the log function, the actual optimization goal is:

$$\min E_{z\sim p_z(z)}[-\log D(G(z))] \, . \qquad (11)$$

The optimized goal of the discriminator when the parameters of the generator are fixed is:

$$\max E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + E_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}[1 - \log(D(G(\boldsymbol{z}))] \, . \qquad (12)$$

A new kind of loss called adversarial loss is induced by the adversarial generative network. For image generation, the use of adversarial loss aims to enable the network to learn the statistics of nature images and reconstruct more realistic images[27]. However, training the neural network by only using adversarial loss is not enough because it controls the distortion weakly. Therefore, combining L1/L2 loss with adversarial loss is natural. In this way, distortion and perception are taken care of.

For pan-sharpening, the LRMS and PAN images are samples from a joint distribution, and the adversarial loss could be written as:

$$L_G = -E_{X_p, X_M \sim p_{data}(X_P, X_M)}(\log \sigma(D_{\theta_G}(X_P, X_M))) \, , \qquad (13)$$

where σ is a sigmoid function.

The adversarial loss for the discriminator is written as:

$$\begin{aligned} L_D = & -E_{Y\sim p(Y)}(\log\sigma(D_{\theta_D}(Y))) \\ & -E_{X_P,X_M\sim p_{data}(X_P,X_M)}(\log(1-\sigma(D_{\theta_D}(G_{\theta_G}(X_P,X_M))))) \end{aligned} . \quad (14)$$

Different from PSGAN, we express the LRMS and PAN images as samples from a joint distribution instead of two distributions. The reason is that if the PAN image is given, the corresponding LRMS image should be a sample from a conditional distribution.

Oher methods, including the relativistic average discriminator used in RaGAN[28] and the Wasserstan-GAN proposed by Arjovsky et al.[29], can be used to generate adversarial loss. The standard GAN is used in our experiment to generate standard adversarial loss.

Combining the adversarial loss with L1/L2 loss, the total loss of generator could be written as the sum of the L1/L2 loss and the adversarial loss:

$$Loss_G = Loss_{L1/L2} + \lambda L_G , \quad (15)$$

where λ is the hyper-parameter for controlling the ratio between two losses. For L1 and L2 loss, we set λ to 0.01 and 0.0001 in experiments.

3. Results and Discussion

3.1 Dataset

Landsat8 is a medium resolution satellite launched in 2013. The OLI sensor of this satellite has 9 bands in VIR and NIR, of which the B8 is the panchromatic band with 15 m spatial resolution and the rest are the 30 m bands. The data have fine quality and are widely used. The training dataset in this paper is obtained from three selected images from Landsat-8 OLI, covering most kinds of ground objects, such as waters, cities, and vegetation. The images are divided into two groups. One contains 15 m panchromatic band with patch size of $256\times256\times1$. The other contains 30 m bands (without Cirrus band) with patch size of $128\times128\times7$. For numerical stability, the data are processed as follows:

$$newvalue = \frac{originvalue}{2000} . \quad (16)$$

Table 1. The Parameters of Landsat-8 OLI Sensor

Band	Band Name	Bandwidth	Resolution
B1	Coastal	0.43-0.45	30
B2	Blue	0.45-0.51	30
B3	Green	0.53-0.59	30
B4	Red	0.64-0.67	30
B5	NIR	0.85-0.88	30
B6	SWIR1	2.11-2.29	30
B7	SWIR2	0.50-0.68	30
B8	Pan	0.50-0.68	15
B9	Cirrus	1.36-1.38	30

The ratio of resolution of PAN to that of LRMS images has two different conditions. One is ×2 for Landsat-8, Landsat-7, and SPOT-4 so on. The other is ×4 for Quickbird and Wordview. Thus, before inputting the data into the network, the images are down-sampled with different scaling factors of ×2 and ×4 to simulate different data. The testing dataset is from another remote sensing image, and 600 images are chosen for testing.

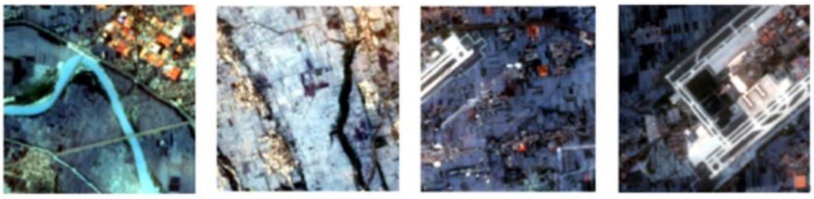


Figure 5. Sample Images in Dataset

3.2 Training details

The networks were implemented and trained in TensorFlow 2.0 framework. The batch size was set to 2, and the learning rate is 0.0001. The networks were trained by Adam [30]. Glorot Uniform is used as the initialization method [31]. To be more representative, standard discriminator is used to generate the adversarial loss.

3.3 Evaluation metrics

In this paper, five widely used metrics are used to evaluate the performance of the networks, including RMSE, ERGAS[32], SAM[33], SRE, and AG.

(1) Root mean square error (RMSE): For single band, the RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Y_n - Y_n)^2}\ , \tag{17}$$

where Y is the pan-sharpened image, and Y is the reference image, N is the total number of reference image pixels. When calculating RMSE for the multi-band image, the sum of single band RMSE is used. The RMSE measures the distance between pan-sharpened image and reference image. For ideal pan-sharpened image, the RMSE should be 0.

(2) ERGAS: The ERGAS index is defined as:

$$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\frac{RMSE(B_i)}{M_i(B_i)})^2}\ , \tag{18}$$

where h and l are the spatial resolutions of PAN images and the LRMS images, respectively; $RMSE(B_i)$ is the RMSE of ith band of pan-sharpened image and the reference image; and M_i is the mean of ith band of B_i. ERGAS reflects the reconstruction quality, the ideal value is 0.

(3) Spectral angle mapper (SAM): The SAM measures the angle of spectral vector from the same pixel of pan-sharpened and reference images. It is calculated by:

$$SAM(x,y) = \arccos(\frac{x \cdot y}{\|x\| \cdot \|y\|})\ , \tag{19}$$

where x and y are the spectral vectors of the same pixel of pan-sharpened image and the reference image, respectively. When calculating the SAM for whole image, the mean of SAM of all pixels is used. The ideal value of SAM is 0. The SAM calculated using the formula above is in radiant measure, which usually needs to be changed to degree measure. We also did that.

(4) Signal-to-reconstruction error (SRE): The SRE index is defined as:

$$SRE = 10\log_{10}\frac{\mu_y^2}{\|x-y\|^2 / n}\ , \tag{20}$$

where x and y are the vectorized pan-sharpened image and reference image, respectively; and μ_y is the mean of y. When calculating SRE for multi-band images, the mean of all bands is used. The higher SRE is, the better the reconstruction quality is.

(5) Average gradient (AG): The single band AG is defined as:

$$AG = \frac{1}{(W-1)(H-1)} \sum_{i=1}^{W-1} \sum_{j=1}^{H-1} \sqrt{\frac{(Y(i+1,j)-Y(i,j))^2 + (Y(i+1,j)-Y(i,j))^2}{2}}, \quad (21)$$

where W and H are the width and height of the pan-sharpened image, respectively; and Y is the pan-sharpened image and indicates the clarity of the reconstructed image.

3.4 Experimental Results and Discussion

In the experiments, the performance of L1 loss, L2 loss, "L1+ad" loss, and "L2+ad" loss in ×2 and ×4 pan-sharpening tasks are compared.

3.4.1 Numerical Results

From the results in Table 2, L1 loss has the best performance in nearly all evaluation metrics, except AG. In the ×2 task, compared with L2 loss, L1 loss has a slight increase in all metrics. But the AG of two losses is very close. In the ×4 task, L1 loss has a greater increase than L2 loss in nearly all metrics, except on AG. The L2 loss has better AG than L1 loss in the ×4 task. In general, L1 loss performs smaller distortion, L2 loss performs more clarity.

Table 2. Pan-Sharpening Results Under Different Loss Functions

		RMSE×100 ↓	**ERGAS** ↓	**SAM** ↓	**SRE** ↑	**AG** ↑
×2	**L1**	**9.6**	**0.74**	**0.60**	**43.9**	0.0281
	L2	10.2	0.77	0.63	43.4	0.0279
	L1+ad	13.0	0.97	0.86	42.8	**0.0314**
	L2+ad	10.6	0.80	0.66	43.2	0.0286
×4	**L1**	**14.8**	**0.57**	**0.89**	**42.5**	0.0263
	L2	15.7	0.60	0.95	41.9	0.0274
	L1+ad	16.2	0.62	0.99	42.4	**0.0291**
	L2+ad	16.1	0.61	0.97	41.6	0.0275

When discussing the conditions wherein adversarial loss is added, the comparison between "L1+ad" loss and "L2+ad" loss is not considered, because the hyper-parameter λ also influences the performance. The effect of adding adversarial loss to L1 loss and L2 loss is focused on. In Table 2, "L1+ad" loss has a greater increase in AG, which can be observed in both ×2 and ×4 pan-sharpening. However, the addition of adversarial loss leads to a worse distortion and then reduces the RMSE, ERGAS, SAM, and SRE. The results of L2 loss and "L2+ad" loss are the same.

3.4.2 Image results

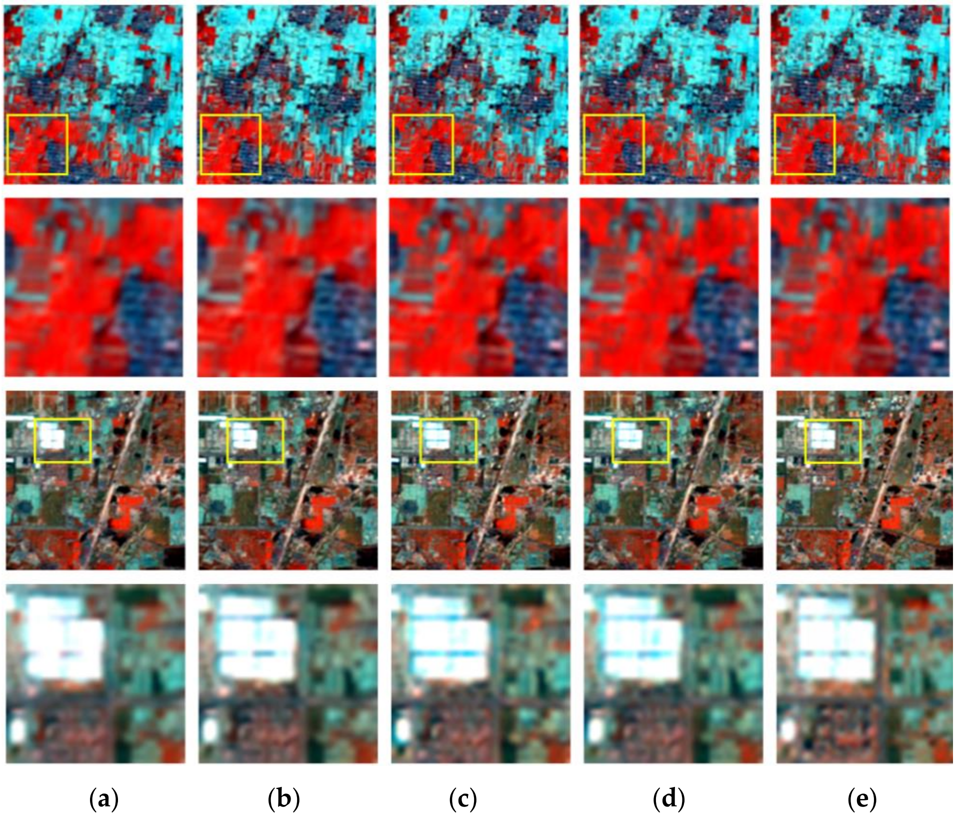
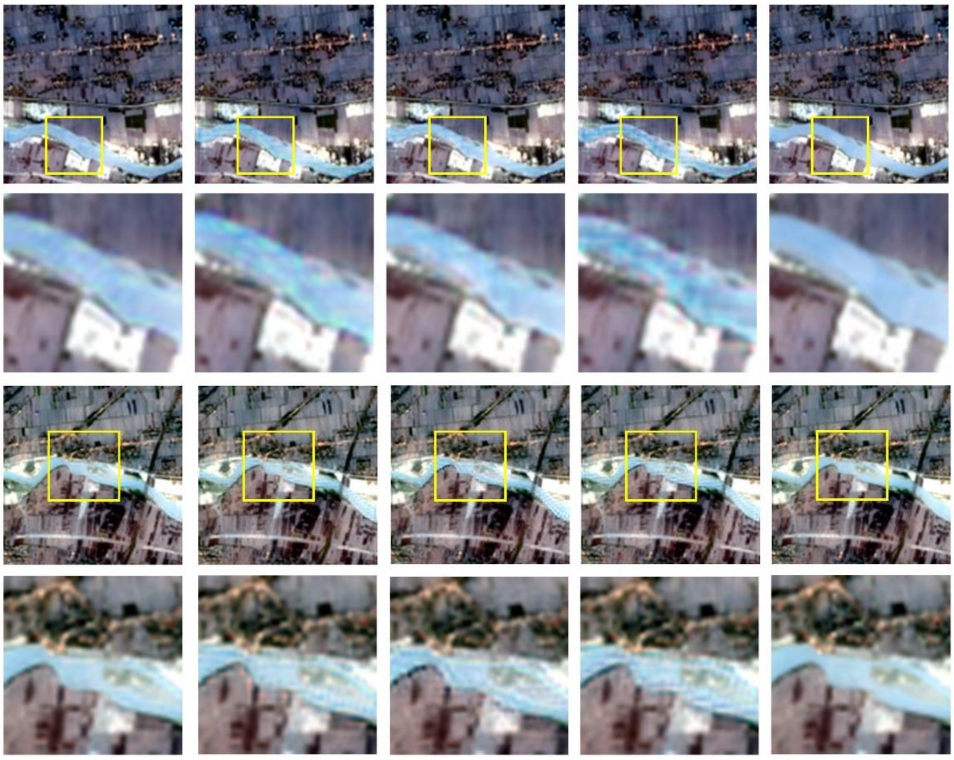


Figure 6. Result pan-sharpened images with their local details (R, G, B = B7, B6, B5). (**a**) L1 loss; (**b**) L2 loss; (**c**) L1+ad loss; (**d**) L2+ad loss; (**e**) HRMS.

For image details, the difference among these four loss functions is not significant in ×2 pan-sharpening. Thus, the results are not presented in this paper. In×4 pan-sharpening, when true color (R, G, B=B4, B3, B2) is used, the images show slight distortion. The differences are hardly noticeable to the naked eyes. However, when R, G, B = B7, B6, B5, the difference is significant, which is consistent with the above numerical values. The pan-sharpened images based on adversarial loss show more details and sharpened edges, which leads to a better perceptive effect. This phenomenon also proves why the adversarial loss function achieves a greater AG in terms of the numerical values. More blurring results are shown by L1 and L2 losses. However, these details or textures are not necessarily real, and that results have more distortion.

Significantly, in some uniform ground, such as rivers, lakes, and bright zigzagging roads, L1-based losses (L1 and "L1+ad") are observed to lead to a coherent result, whereas L2- based losses lead to a lot of artifacts in our experiments. The L1-based losses have the effect of reducing the artifacts. This effect makes L1-based losses perform better on pan-sharpening for uniform ground.

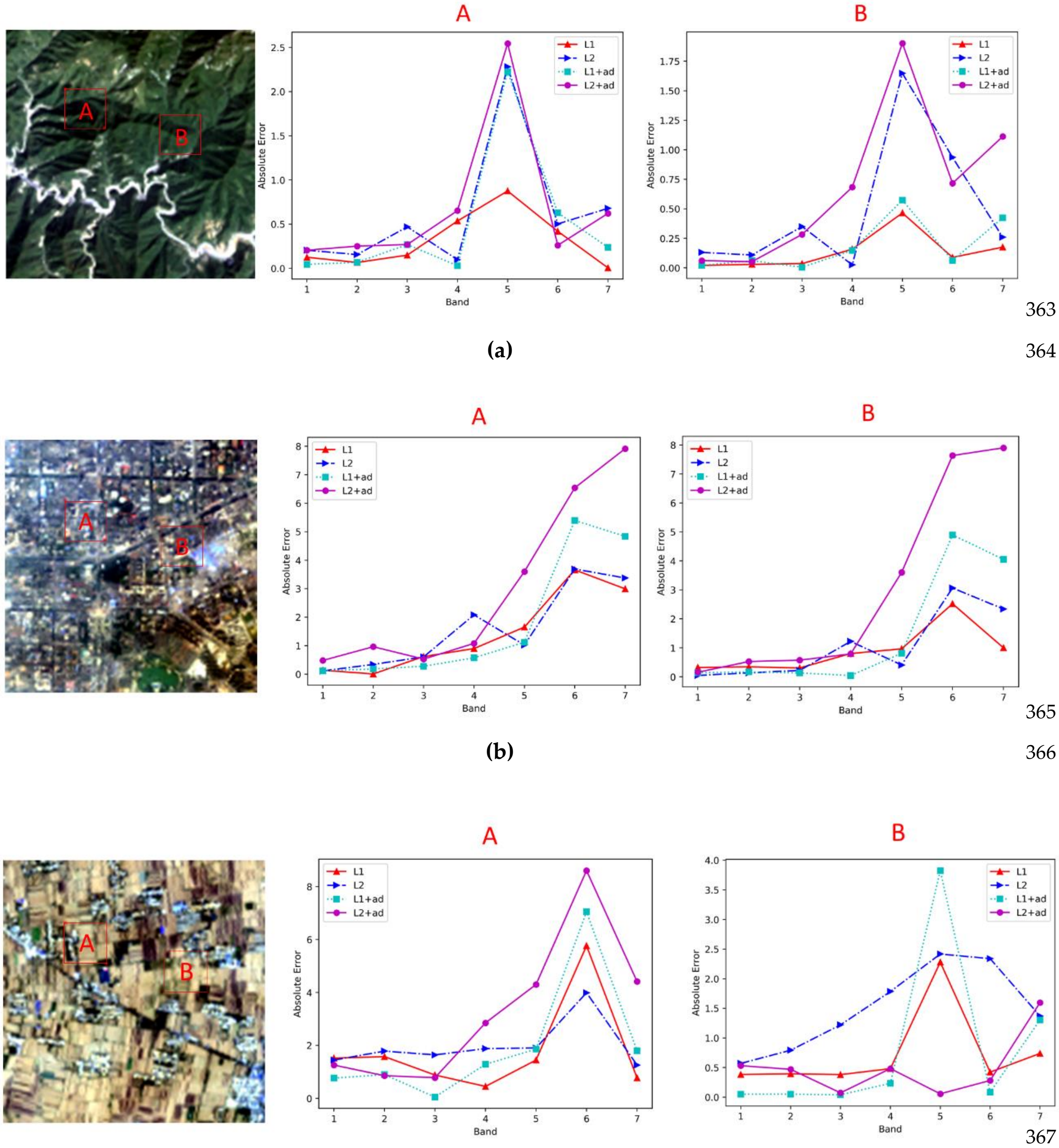


(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 7. Result Pan-Sharpened Images (notice the artifacts at the river) (R, G, B = B4, B3, B2).

3.4.3 Specific Ground Objects

In this paper, the spectral curves of specific ground objects, including vegetation, buildings, and crops, are extracted. To reduce the influence of randomness, the spectral on one pixel is not analyzed, but the mean of an area. For accuracy, the uniformity of the ground objects is kept as far as possible.



(a)

(b)

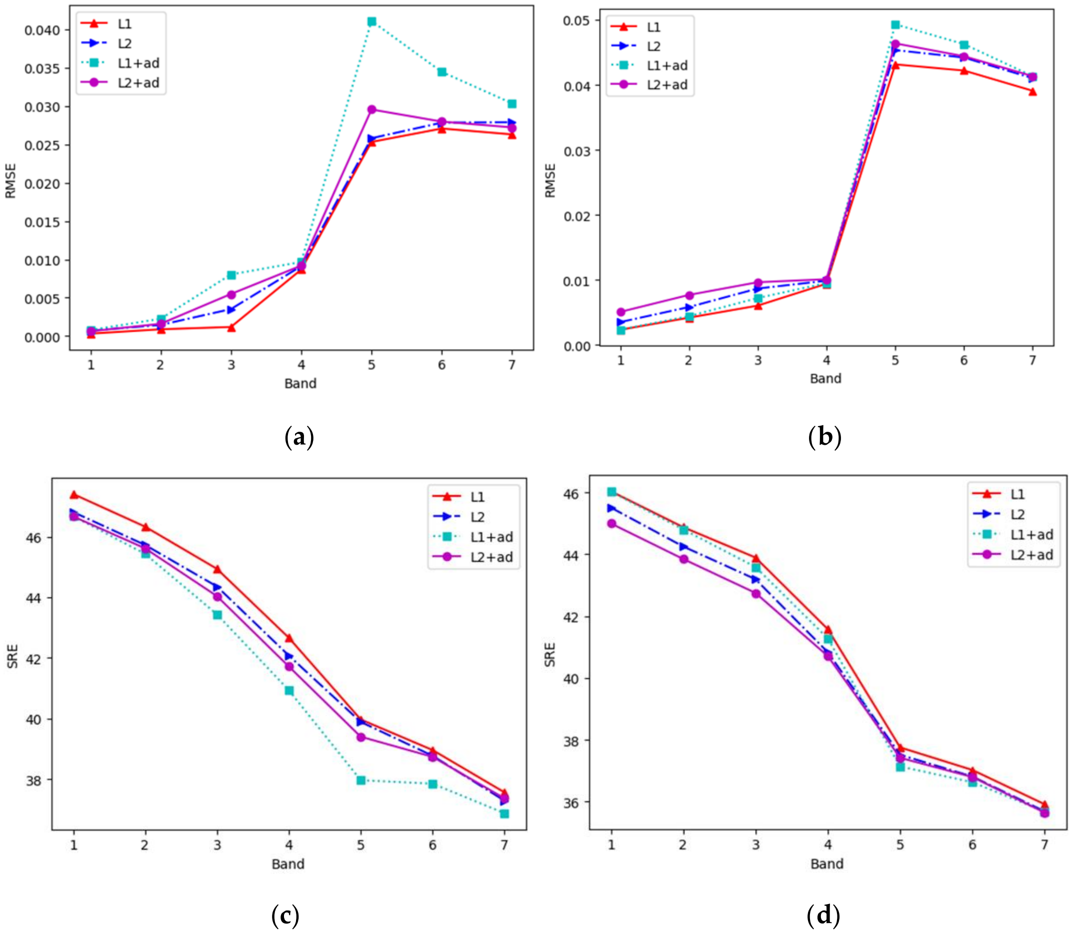
(c)

Figure 7. Spectral absolute error in different distortion (The mean absolute error between the spectrum of pan-sharpened image and reference image are calculated). (**a**) Absolute error in vegetation; (**b**) Absolute error in buildings; (**c**) Absolute error in crops.

In Figure 7, the quality of spectral reconstruction in small areas still shows some randomness. The same results are not observed as the total error in the whole test set. For example, in numerical results, L2 loss always shows better performance than “L2+ad” loss. However, when focusing on a small area of specific ground objects, such as vegetation, the L2 loss does not show an absolute advantage over “L2+ad” loss. In some bands, L2 loss has even worse performance than “L2+ad” loss. In area “A” of crops, L2 loss has lower distortion than L1 loss in B6. However, in the other areas, L2 loss has worse performance than L1 loss in all bands. Thus, the numerical results are considered more of an average manifestation, instead of an inevitable result in each pixel. However, although L1 loss does not always have better performance than the others, it always leads to a relatively small error and shows minimal fluctuation. This phenomenon explains why L1 has a better performance statistically, which is also in agreement with numerical results.

3.4.3 Results in each band

By observing the RMSE in each band in Figure 8 and Table 3, we can find that from B1 to B4, all loss functions lead to small distortion. Therefore, when observing the respective influence of the four loss functions on true color images, it is actually difficult to tell the differences. This observation is true for ×2 and ×4 pan-sharpening. However, from B5 to B7, the quality of pan-sharpening declines because less information is provided by panchromatic band to these bands. The wavelength of these bands is very different from that of panchromatic band. Thus, the statistical information is not similar. Therefore, when B5, B6, and B7 are used to show the images, many huge differences between these bands can be observed. The richer texture features are displayed by adversarial-loss based pan-sharpening. Considering that B2, B3, and B4 have low RMSE, they have high SRE accordingly. From B5 to B7, RMSE increases and SRE decreases. For every band, L1 loss still performs better than the others.



(**a**) (**b**)

(**c**) (**d**)

Figure 8. The Values Curve of Metrics in Each Band. (a) RMSE in ×2 pan-sharpening; (b) RMSE in ×4 pan-sharpening; (c) SRE in ×2 pan-sharpening; (d) SRE in ×4 pan-sharpening.

Table 3. The Values of Metrics in Each Band

		B1	B2	B3	B4	B5	B6	B7
					RMSE×1000 ↓			
×2	**L1**	0.32	0.88	1.18	8.67	25.32	27.06	26.30
	L2	0.70	1.45	3.51	9.11	25.80	27.83	27.88
	L1+ad	0.78	2.25	8.03	9.67	41.10	34.45	30.35
	L2+ad	0.62	1.60	5.48	9.18	29.55	27.97	27.21
					SRE↑			
	L1	47.41	46.33	44.94	42.66	39.97	38.96	37.56
	L2	46.81	45.73	44.36	42.08	39.89	38.77	37.28
	L1+ad	46.68	45.44	43.44	40.93	37.96	37.85	36.87
	L2+ad	46.67	45.62	44.04	41.72	39.40	38.73	37.36
					RMSE×1000 ↓			
×4	**L1**	2.32	4.16	6.01	9.38	43.13	42.18	39.06
	L2	3.48	5.80	8.65	9.85	45.33	44.18	41.00
	L1+ad	2.32	4.40	7.18	9.5	49.30	46.25	41.31
	L2+ad	5.05	7.67	9.61	10.08	46.37	44.38	41.31
					SRE↑			
	L1	46.04	44.88	43.89	41.57	37.75	37.01	35.91
	L2	45.51	44.25	43.19	40.83	37.51	36.81	35.68
	L1+ad	46.02	44.80	43.59	41.28	37.13	36.62	35.67
	L2+ad	44.50	43.85	42.75	40.71	37.41	36.79	35.64

3.5 Advice for Applications

When using the losses in applications, more attention is paid to the distortion of reflectance of single pixels. Therefore, we collected statistics on the distortion at pixel wise in reflectance unit. First, images in 3.4.3 is used and pan-sharpened. Subsequently, the radiometric calibration is applied to obtain reflectance. Then, the mean and the max are collected to observed the distortion at each pixel. Absolute error is used to measure the distortion of pixels.

Table 4. The Values of Metrics for ×2 Pan-Sharpening in Reflectance Unit ($\times 10^4$)

	B1	B2	B3	B4	B5	B6	B7
				L1			
mean	0.86	0.94	1.18	1.93	7.66	7.79	7.48
max	15.55	18.72	16.33	24.04	71.90	94.91	138.55
				L2			
mean	1.00	1.10	1.36	2.23	7.83	8.19	8.04
max	16.38	19.69	18.22	26.57	72.80	97.78	136.30
				L1+ad			
mean	1.02	1.16	1.70	2.94	12.32	10.14	8.76
max	18.25	23.00	21.96	33.76	112.24	127.47	157.78
				L2+ad			
mean	1.05	1.14	1.49	2.47	8.79	8.24	7.84
max	16.60	20.34	18.76	27.63	76.14	99.79	133.58

In ×2 pan-sharpening, the distortion of reflectance is in the order of 10^{-4} from B1 to B4, whereas from B5 to B7, it is near the order of 10^{-3}. For max distortion, B5, B6, and B7 have approximately 1% reflectance. The L1 loss showed the best control of distortion, and "L1+ad" has the worst performance.

Table 5. The Values of Metrics for ×4 Pan-Sharpening in Reflectance Unit ($\times 10^4$)

metric	B1	B2	B3	B4	B5	B6	B7
				L1			
mean	1.09	1.22	1.49	2.48	12.23	11.71	10.40
max	29.70	35.95	22.91	33.24	137.63	188.56	263.94
				L2			
mean	1.29	1.50	1.81	3.07	13.19	12.52	11.12
max	28.82	35.33	24.31	34.55	140.12	183.78	255.61
				L1+ad			
mean	1.10	1.25	1.61	2.66	14.44	12.87	10.99
max	29.36	35.65	24.02	34.81	157.55	201.33	264.32
				L2+ad			
mean	1.50	1.70	2.02	3.12	13.66	12.53	11.22
max	29.28	35.73	25.46	35.71	139.27	183.29	252.95

Due to the increasing difficulty of tasks, ×4 pan-sharpening shows worse values in metrics. From B1 to B4, the mean of the distortion is controlled at the order of 10^{-4}. From B5 to B7, the mean of distortion increases to the order of 10^{-3}. The max values increase by one order of magnitude. B5, B6, and B7 deviate 1%-2% from the real reflectance, the values increase from 8% to 43%. Various loss functions lead to very different results. The L1 loss shows the best performance, whereas "L2+ad" loss shows the worst performance in distortion.

In applications, our purpose must be considered. Quantitative remote sensing uses the remote sensing images to retrieve ground and atmosphere parameters. For example, Kaufman et al.[34,35], Holben et al.[36], and Liang et al.[37] used the remote sensing images to retrieve the aerosol optical depth. Gordan et al.[38] and Nechad et al.[39] used images to obtain the total suspended matter. High precision and strict control of distortion are required by this kind of applications. L1 loss is recommended if we want to use pan-sharpening in these applications. In addition, bands near panchromatic band have low distortion. In Tables 4 and 5, the maximal distortion is higher than 1% in B5, B6, and B7, which is far from panchromatic band. In that case, results of quantitative remote sensing will be unacceptable. Therefore, the bands near panchromatic band are highly recommended.

By contrast, qualitative remote sensing uses the remote sensing images to evaluate the ecological systems. For example, remote sensing images have been used to evaluate the ecology of forests, grasslands, and urban areas by Ochoa-Gaona et al.[40], Sullivan et al. [41], and Xu et al. [42], respectively. In these applications, requirements for precision are relatively low. Thus, adversarial losses could be chosen to improve the visual effect, especially if the bands far from panchromatic band are used.

4. Conclusion

In this paper, four popular loss functions in pan-sharpening are compared. The experiment results show that L1 loss performs better distortion than the others and reduces the artifacts. More details and textures are shown by the "L1+ad" loss and "L2+ad" loss, especially in bands that are far from the panchromatic band. During the training, we also find that the training of the generative adversarial nets is difficult and needs more time because of the adjustment of the hyper-parameters. Finally, practical applications are focused on, and advices for the use of the loss functions are offered. L1 loss is recommended in quantitative applications, and adversarial loss is recommended in qualitative applications.

References

1. Shettigara, V.K. A Generalized Component Substitution Technique for Spatial Enhancement of Multispectral Images Using a Higher Resolution Data Set. *Photogrammetric Engineering & Remote Sensing* **1992**, *58*, 561–567.
2. Zhu, X.X.; Bamler, R. A Sparse Image Fusion Algorithm With Application to Pan-Sharpening. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 2827–2836, doi:10.1109/TGRS.2012.2213604.
3. Ranchin, T.; Wald, L. Fusion of High Spatial and Spectral Resolution Images: The ARSIS Concept and Its Implementation. *Photogrammetric engineering and remote sensing* **2000**, *66*, 49.
4. Tu, T.-M.; Su, S.-C.; Shyu, H.-C.; Huang, P.S. A New Look at IHS-like Image Fusion Methods. *Information Fusion* **2001**, *2*, 177–186, doi:https://doi.org/10.1016/S1566-2535(01)00036-7.
5. Gillespie, A.R.; Kahle, A.B.; Walker, R.E. Color Enhancement of Highly Correlated Images. II. Channel Ratio and "Chromaticity" Transformation Techniques. *Remote Sensing of Environment* **1987**, *22*, 343–365, doi:10.1016/0034-4257(87)90088-5.
6. Chavez, P.S.; Kwarteng, A.Y. Extracting Spectral Contrast in Landsat Thematic Mapper Image Data Using Selective Principal Component Analysis. *PHOTOGRAM ENG REMOTE SENS* **1989**, *55*, 339–348.
7. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. *United States Patent 6* **2000**, *11*.
8. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by Convolutional Neural Networks. *Remote Sensing* **2016**, *8*, 594, doi:10.3390/rs8070594.
9. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A Deep Network Architecture for Pan-Sharpening. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); IEEE: Venice, October 2017; pp. 1753–1761.
10. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans Pattern Anal Mach Intell* **2016**, *38*, 295–307, doi:10.1109/TPAMI.2015.2439281.
11. Rao, Y.; He, L.; Zhu, J. A Residual Convolutional Neural Network for Pan-Shaprening. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP); 2017; pp. 1–4.
12. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *146*, 305–319, doi:10.1016/j.isprsjprs.2018.09.018.
13. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sensing* **2018**, *56*, 5443–5457, doi:10.1109/TGRS.2018.2817393.
14. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]* **2014**.
15. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv e-prints* **2015**, *1511*, arXiv:1511.06434.
16. Zhu, J.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); October 2017; pp. 2242–2251.

17. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *arXiv e-prints* **2016**, *1609*, arXiv:1609.04802.
18. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the Computer Vision – ECCV 2018 Workshops; Leal-Taixé, L., Roth, S., Eds.; Springer International Publishing: Cham, 2019; pp. 63–79.
19. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. *IEEE Transactions on Geoscience and Remote Sensing* **2020**, *PP*, 1–16.
20. Shao, Z.; Lu, Z.; Ran, M.; Fang, L.; Zhou, J.; Zhang, Y. Residual Encoder–Decoder Conditional Generative Adversarial Network for Pansharpening. *IEEE Geosci. Remote Sensing Lett.* **2020**, *17*, 1573–1577, doi:10.1109/LGRS.2019.2949745.
21. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv:1608.06993 [cs]* **2018**, 2261–2269, doi:10.1109/CVPR.2017.243.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; pp. 770–778.
23. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37; JMLR.org: Lille, France, July 6 2015; pp. 448–456.
24. Santurkar, S.; Tsipras, D.; Ilyas, A.; Mądry, A. How Does Batch Normalization Help Optimization? In Proceedings of the Proceedings of the 32nd International Conference on Neural Information Processing Systems; Curran Associates Inc.: Red Hook, NY, USA, December 3 2018; pp. 2488–2498.
25. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); July 2017; pp. 1132–1140.
26. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Transactions on Computational Imaging* **2017**, *3*, 47–57, doi:10.1109/TCI.2016.2644865.
27. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM Challenge on Perceptual Image Super-Resolution. In Proceedings of the Computer Vision – ECCV 2018 Workshops; Leal-Taixé, L., Roth, S., Eds.; Springer International Publishing: Cham, 2019; pp. 334–355.
28. Jolicoeur-Martineau, A. The Relativistic Discriminator: A Key Element Missing from Standard GAN. *arXiv:1807.00734 [cs, stat]* **2018**.
29. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv e-prints* **2017**, *1701*, arXiv:1701.07875.
30. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv e-prints* **2014**, arXiv:1412.6980.
31. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; JMLR Workshop and Conference Proceedings, March 31 2010; pp. 249–256.
32. Wald, L. Quality of High Resolution Synthesised Images: Is There a Simple Criterion ? In Proceedings of the Third conference ”Fusion of Earth data: merging point measurements, raster maps and remotely sensed images”; Ranchin, T., Wald, L., Eds.; SEE/URISCA: Sophia Antipolis, France, January 2000; pp. 99–103.
33. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Descrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. *Summaries of the Third Annual JPL Airborne Geoscience Workshop, JPL Publ. 92–14, Vol. 1* **1992**, 147–149.

34. Kaufman, Y.J.; Sendra, C. Algorithm for Automatic Atmospheric Corrections to Visible and Near-IR Satellite Imagery. *International Journal of Remote Sensing* **1988**, *9*, 1357–1381, doi:10.1080/01431168808954942.
35. Kaufman, Y.J.; Tanré, D.; Remer, L.A.; Vermote, E.F.; Chu, A.; Holben, B.N. Operational Remote Sensing of Tropospheric Aerosol over Land from EOS Moderate Resolution Imaging Spectroradiometer. *Journal of Geophysical Research: Atmospheres* **1997**, *102*, 17051–17067, doi:https://doi.org/10.1029/96JD03988.
36. Holben, B.; Vermote, E.; Kaufman, Y.J.; Tanre, D.; Kalb, V. Aerosol Retrieval over Land from AVHRR Data-Application for Atmospheric Correction. *IEEE Transactions on Geoscience and Remote Sensing* **1992**, *30*, 212–222, doi:10.1109/36.134072.
37. Liang, S.; Fallah-Adl, H.; Kalluri, S.; JáJá, J.; Kaufman, Y.J.; Townshend, J.R.G. An Operational Atmospheric Correction Algorithm for Landsat Thematic Mapper Imagery over the Land. *Journal of Geophysical Research: Atmospheres* **1997**, *102*, 17173–17186, doi:https://doi.org/10.1029/97JD00336.
38. Gordon, H.R.; Brown, O.B.; Evans, R.H.; Brown, J.W.; Smith, R.C.; Baker, K.S.; Clark, D.K. A Semianalytic Radiance Model of Ocean Color. *Journal of Geophysical Research: Atmospheres* **1988**, *93*, 10909–10924, doi:https://doi.org/10.1029/JD093iD09p10909.
39. Nechad, B.; Ruddick, K.G.; Park, Y. Calibration and Validation of a Generic Multisensor Algorithm for Mapping of Total Suspended Matter in Turbid Waters. *Remote Sensing of Environment* **2010**, *114*, 854–866, doi:10.1016/j.rse.2009.11.022.
40. Ochoa-Gaona, S.; Kampichler, C.; de Jong, B.H.J.; Hernández, S.; Geissen, V.; Huerta, E. A Multi-Criterion Index for the Evaluation of Local Tropical Forest Conditions in Mexico. *Forest Ecology and Management* **2010**, *260*, 618–627, doi:10.1016/j.foreco.2010.05.018.
41. Sullivan, C.A.; Skeffington, M.S.; Gormally, M.J.; Finn, J.A. The Ecological Status of Grasslands on Lowland Farmlands in Western Ireland and Implications for Grassland Classification and Nature Value Assessment. *Biological Conservation* **2010**, *143*, 1529–1539, doi:10.1016/j.biocon.2010.03.035.
42. Xu, H.; Ding, F.; Wen, X. Urban Expansion and Heat Island Dynamics in the Quanzhou Region, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2009**, *2*, 74–79, doi:10.1109/JSTARS.2009.2023088.