GENERATING ARTIFICIAL NEAR INFRARED SPECTRAL BAND FROM RGB IMAGE USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

Xiangtian Yuan[1],[*] Jiaojiao Tian[1], Peter Reinartz[1]

[1]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany -
(Xiangtian.yuan, Jiaojiao.tian, Peter.reinartz)@dlr.de

KEY WORDS: Near-infrared, RGB, Gerative adversarial networks, Robust loss function, Conditional GAN

ABSTRACT:

Near infrared bands (NIR) provide rich information for many remote sensing applications. In addition to deriving useful indices to delineate water and vegetation, near infrared channels could also be used to facilitate image pre-processing. However, synthesizing bands from RGB spectrum is not an easy task. The inter-correlations between bands are not clearly identified in physical models. Generative adversarial networks (GAN) have been used in many tasks such as generating photorealistic images, monocular depth estimation and Digital Surface Model (DSM) refinement etc. Conditional GAN is different in that it observes some data as a condition. In this paper, we explore a cGAN network structure to generate a NIR spectral band that is conditioned on the input RGB image. We test different discriminators and loss functions, and evaluate results using various metrics. The best simulated NIR channel has a mean absolute error of around 5 percent in Sentinel-2 dataset. In addition, the simulated NIR image can correctly distinguish between various classes of landcover.

1. INTRODUCTION

In remote sensing, near-infrared bands (NIR) have been playing important roles in many aspects. They exhibit additional potential for representing ground objects in comparison to RGB bands, especially in representing vegetations. For example, indices involving NIR have been developed and used for tasks such as landcover classification. These indices includes Normalized Vegetation Index (NDVI) and Normalized water index (NDWI), which have been proven to be effective in highlighting vegetation and open water feature in remote sensing imagery (McFeeters, 1996). In addition to identifying vegetation and water, NIR band is also capable of discerning materials such as plastic, minerals, sea foams, trace gases, and the health problems of trees. In data-hungry machine learning or deep learning methods for landcover classification, these characteristics enable NIR bands to be used to improve coarse ground truth, and correct wrong labels with their capability of distinguishing between classes with subtle difference in spectral signature. Moreover, NIR-derived indices have also been used in some tasks such as atmospheric correction (Kaufman, Sendra, 1988).

But NIR bands are not always available in every sensor. Some low cost satellite might not be equipped with sensor capable of capturing NIR bands. Some airborne systems also only consist of cameras capturing RGB bands. Moreover, sometimes when doing landcover change detection, data from old sensors might not provide NIR bands as the newer ones, thus hindering the accuracy of change detection. Therefore, synthesizing NIR bands from RGB is of practical values.

The generation of NIR band from RGB can be regarded as a nonlinear mapping from RGB to NIR. Neural networks have been proven to be effective in nonlinear mapping. For example, one paper (Fu et al., 2018) proposed a network structure for hyperspectral image reconstruction from RGB bands. The network consists of a spectral sub network, which performs the

spectral nonlinear mapping, and a spatial sub network, which models the spatial correlation. Then hyperspectral bands are generated by minimizing a mean squared error between generated bands and real bands.

In recent years, generative adversarial networks (GANs) have been extensively used in remote sensing community to tackle various tasks. For example, GAN and its variants are capable or refining Digital Surface Models (DSMs) derived from stereo matching (Bittner et al., 2019). In addition, GANs are applied in hyperspectral image classification (Zhan et al., 2017), PAN-sharpening (Liu et al., 2018) and super resolution (Ledig et al., 2017) tasks.

Due to the versatility of GANs, we want to test if GANs are capable of generating realistic NIR band reflectance. The generated NIR bands should keep the original image textures, as well as the physical radiometric properties. To this purpose, GAN in conditional setting is more suitable, meaning that the generated NIR bands will be conditioned on the visible spectra (red, green and blue). This conditional setting ensures that the generated NIR bands are not only realistic, but also close to RGB input in terms of information content. To this end, additional loss functions such as L1 or L2 are added to the GAN loss to ensure that the output is close to the ground truth (Isola et al., 2017). However, such losses are susceptible to outliers. Some robust loss functions are able to handle outliers by putting less sensitivity to large error. A single robust loss function proposed by (Barron, 2019) encompasses several common robust loss functions. This robust loss function is controlled by a single continuous-valued parameter that can also be optimized when training neural networks.

In this work, we will present a method to generate NIR band from RGB bands, which applies a robust loss function in conditional GAN setting. We tested the method on Sentinel-2 Dataset and analysed the applicability of the proposed method. The contribution of our work is twofold: we tested a conditional GAN for task that not only requires perceptive realness but also

---

*Corresponding author

NIR band with meaningful radiometric properties; we also adopted a robust loss function that contributes to better learning in the generative model.

The paper is structured as follows. In chapter two we describe in detail the concepts and methodology involved; in chapter three, dataset and experiment settings are detailed. Results are analysed in chapter four, followed by conclusions in chapter five.

2. METHODOLOGY

GANs are built on game theory (Goodfellow et al., 2014) and have been used in multitudes of tasks in computer vision. In remote sensing, GANs have been proven to be effective in many applications and have achieved good results. The characteristics of cGAN will be briefly described in this chapter.

2.1 Conditional GAN

GAN comprises generator and discriminator. The generator tries to produce output while the discriminator tries to classify if the output is fake or real (Goodfellow et al., 2014). The input of GAN is usually random noise vector, and the output is image that is similar to realistic images. Different from conventional GANs, conditional GANs (Mirza, Osindero, 2014) observe input data. In our case, the network should generate NIR band while observing RGB bands. Then the discriminator tries to distinguish between the real and the fake image from generator until it can not distinguish anymore. In cGAN, the discriminator is also conditioned on the input RGB bands similar to the generator. Therefore, NIR band corresponding to RGB bands can be generated from the cGAN.

2.1.1 Generator Generating realistic NIR band from RGB bands can be regarded as a mapping from input to output of the same spatial resolution. As the input and output are representation of the same ground objects, they should match in structure, texture and have same semantics. A number of GAN generators adopt encoder-decoder structures that first reduce the spatial resolution of input and gradually recover it. This structure loses the low level information from previous stages, resulting in lack of details. Therefore, encoder-decoder network with skip connection is more suitable for this task. This structure is capable of retaining information from different stages in the network, which is popularly known as U-Net structure (Ronneberger et al., 2015). This generator is adopted in image-to-image translation model Pix2pix (Isola et al., 2017). The U-net in our experiment consists of 8 blocks in both encoder and decoder. In encoder each block encompasses convolution, batch normalization and LeakyReLU of slope 0.2. In decoder each block comprises transposed convolution, batch normalization and ReLU layers. The convolution has a filter size of 4 and stride of 2 in both encoder and decoder. In some conditional GANs, Gaussian noise z is provided to generator as input to avoid deterministic results matching delta function (Isola et al., 2017). Different from this approach, Pix2pix model employs drop out in generator during both training and testing phase. Although this approach results in reduced stochasticity, it is still suitable for our task as our task does not need much randomness as other computer vision tasks such as image translation.

2.1.2 Discriminator As for discriminator, various options are available depending on the task. One choice is the Markovian discriminator, which is also termed as PatchGAN (Isola et al.,

2017). It classifies whether a $N \times N$ patch in the input image is real or fake, and average all the patches in the image. The discriminator is made of several blocks consisting of 2D convolution, batch normalization and leaky ReLU layers. The stride of all convolutions are 2 except for the last and second last convolutions. The size of receptive field of previous block is calculated as:

$$(outputsize - 1) \times stride + kernelsize \qquad (1)$$

It should be noted that the patch size of the patch discriminator is defined as the size of receptive field in input that corresponds to one output pixel. Therefore, the deeper the discriminator, the larger the patch size. The detail of a 3-layer (excluding the last two layers) PatchGAN discriminator is shown in Figure 1. Ignoring padding, the patch size is 70×70 for such 3-layer patch discriminator. It could be understood as a form of texture loss (Isola et al., 2017).

Another option is a pixel level discriminator, which only classifies real or fake on pixel level. Different from PatchGAN, the kernel size and stride equal to 1. Therefore, the feature map size remains unchanged across the network and no texture information is considered by discriminator. An illustration of the pixel discriminator is shown in Figure 2.

Both discriminators' last layer is a binary cross entropy layer that classifies if the generated image is true or false. The result is averaged over the whole image.
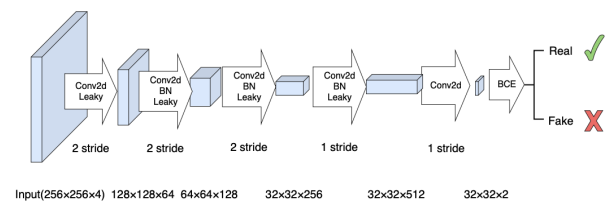

Figure 1. Illustration of PatchGAN Discriminator. The first block has no batch normalization. The first three convolutions have a filter size of 4×4 and stride of 2. The last two convolution layers has stride of 1, therefore retaining the spatial resolution. The output is passed on to a binary cross entropy function. The output is a score for whole image.
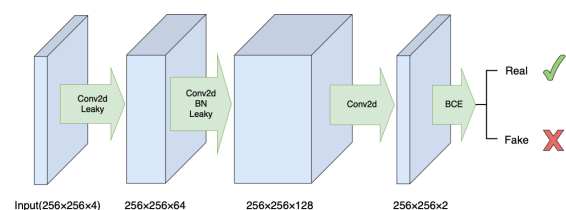

Figure 2. Illustration of Pixel Discriminator. The convolution filters have a size of 1×1 and stride of 1. The classification of real and fake is only on pixel level, without any contextual information. The result after binary cross entropy (BCE) is averaged.

2.2 Loss Function Formulation

2.2.1 GAN Loss In GANs, random noise z conforming to certain probability distribution is mapped to the desired output y by generator G. Conditional GAN, on the other hand, learns a mapping not solely from random noise z, but from both random

noise z and input image x. $G : \{x, z\} \to y$. Discriminator D is trained adversarially against generator G to distinguish between real image and generated image. The objective function of conditional GAN can be expressed as:

$$L_{cGAN}(G, D) = \mathbb{E}_{x,y}[LogD(x, y)] \\ + \mathbb{E}_{x,z}[log(1 - D(G(x, z)))] \quad (2)$$

The loss of the unconditional GAN can be written as:

$$L_{GAN}(G, D) = \mathbb{E}_y[LogD(y)] \\ + \mathbb{E}_z[log(1 - D(G(z)))] \quad (3)$$

2.2.2 Traditional Loss It has been found beneficial to combine GAN loss with traditional loss functions such as L1 or L2 (Isola et al., 2017). In our task, the generated NIR band should not only be distinguishable from the real NIR band, but also has to be close to the real NIR band numerically. Therefore, traditional loss is helpful in enforcing results to be close to ground truth. Compared with L2, L1 loss encourages less blurring (Isola et al., 2017). L1 loss is calculated as:

$$L_{L1}(G) = \mathbb{E}_{x,y,z}[\|z - G(x, z)\|_1] \quad (4)$$

The final loss can be expressed as:

$$arg \min_G \max_D = L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (5)$$

2.2.3 Robust Loss L1 and L2 losses suffer from the problem of outliers, meaning that outlier contributes equally to loss as inlier. The ability to handle outliers is termed robustness in machine learning. Robustness is a crucial property that is desired in machine learning models. There are several robust loss functions that have reduced sensitivity to large errors, such as Cauchy/Lorentzian (Black, Anandan, 1996), Geman-McClure (Geman, McClure, 1985), Welsch (Dennis Jr, Welsch, 1978), Charbonnier (Charbonnier et al., 1994) and generalized Charbonnier (Sun et al., 2010). These loss functions have saturating or even reduced gradient when the loss is large. A robust loss function proposed by (Barron, 2019) is the superset of many common robust loss functions mentioned above. It is able to adjust its robustness as a continuous parameter during training. The loss function is defined as:

$$f(x, \alpha, c) = \frac{|\alpha - 2|}{\alpha} \left(\left(\frac{(x/c)^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} - 1 \right) \quad (6)$$

It is a generalisation of many losses. In Equation 6, α controls the robustness of the loss; $c > 0$ is the scale parameter that controls the size of quadratic bowl nean $x = 0$. A general probability distribution can be constructed from the robust loss, so that the log-likelihood of the probability density is a shifted version of the robust loss function. The distribution is defined as:

$$p(x|\mu, \alpha, c) = \frac{1}{cZ(\alpha)} exp(-f(x - \mu, \alpha, c)) \quad (7)$$

In this equation, $Z(\alpha)$ is a partition function:

$$Z(\alpha) = \int_{-\infty}^{\infty} exp(-f(x, \alpha, 1)) \quad (8)$$

The logarithmic of the partition function can be approximated using the cubic Hermit spline. The negative log likelihood of the distribution can avoid skewing towards ignoring outlier by forcing extra penalty for small errors. The details can be found in paper (Barron, 2019). Therefore, the final objective for cGAN with robust loss function can be expressed as:

$$arg \min_G \max_D = L_{cGAN}(G, D) + \lambda L_{Robust}(G) \quad (9)$$

3. EXPERIMENT

We use the multispectral images from SEN12MS dataset based on Sentinel-1 and Sentinel-2 dataset (Schmitt et al., 2019). The Sentinel-2 data from SEN12MS is level 1-C Top of Atmosphere reflectance (TOA) product. The images have in total 13 band with spatial resolution from 10m to 60 m. In our experiment, we selected the red (R), green (G), blue (B) and near-infrared (NIR) bands with 10 m resolution. The dataset encompasses areas including desert, field, forests, urban areas, water bodies etc. Example images are shown in Figure.4. The images in SEN12MS are distributed across the world as can be seen from Figure.3. It shows that the data are distributed globally, with varying latitudes and climate conditions. The landcover type also varies drastically in different locations. The dataset is categorised by seasons. In this paper, we used data acquired in summer for training and testing to avoid problems incurred by properties of multi-seasonal dataset. Details of band information can be seen in Table.1.

Figure 3. Visualization of locations of images in the experiment. Blue dots denote the location of clusters of smaller images.

Bands	Wavelength(nm)		Bandwidth	Resolution
	S2A	S2B	(nm)	(m)
Red	664.6	664.9	31	10
Green	559.8	559.0	36	10
Blue	492.4	492.1	66	10
NIR	832.8	832.9	106	10

Table 1. Data Description from (European Space Agency, 2015). S2A and S2B are the two satellites respectively. Wavelength is the central wavelength

3.1 Data Pre-processing

We randomly selected 30000 images from the summer scenes for training and 300 images for testing. In Sentinel-2 Level-1C data, the digital number (DN) is TOA reflectance multiplied by 10000. We therefore converted the DN to physically meaningful reflectance and zero-centered the pixel value for training. In

(a) Image with mostly high vegetation   (b) Image with mostly low vegetation and urban area   (c) Image with mostly barren landscape   (d) Image with around half water and half high vegetation
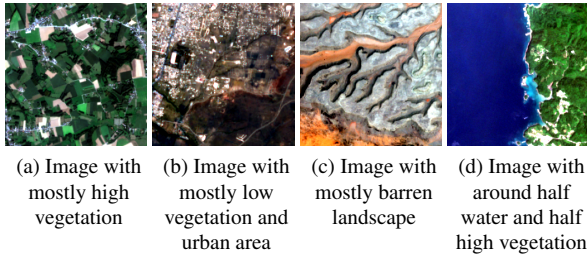
Figure 4. Example images from different regions with various landcover types. As can be seen from images, landcover is very distinct from region to region.

generative models, data pre-processing is very crucial for learning, and we find this pre-processing strategy effective.

3.2 Training Settings

We test cGAN networks with pixel discriminator and patch discriminator. We also test a U-net generator without cGAN setting in order to verify if the cGAN objective facilitates better learning. Among these models, we compare traditional L1 loss and robust loss in the final objective. The experiment is performed based on Pytorch framework. We used Adam optimizer (Kingma, Ba, 2014) and learning rate of 0.0002. The parameter of the robust loss function is optimized together with network parameters. Batch size is set to 16. The input patch size is 256×256 without any cropping or rotation. The parameter weights are initialized by uniform distribution between 0 to 1. The λ is set to 100 for the cGAN because the L1 loss is significantly smaller than cGAN loss. We train the network for 200 epoch. In training process, dropout is employed in generator to serve as random noise.

3.3 Inference

The discriminator is only active during training. During inference, discriminator is no longer involved. Dropout is also employed in generator to avoid deterministic results. The inference is ran one image per batch. The output image pixel values are converted back to reflectance using the pre-calculated statistics from the dataset.

4. RESULT ANALYSIS

We evaluate the generated near-infrared band based on mean absolute error (MAE), mean absolute percentage error (MAPE) and structural similarity (SSIM). We also evaluate the MAE of the resulting NDVI and NDWI. The MAPE is not calculated for NDVI and NDWI because these indices can be zero sometimes, causing undefined results. SSIM index is a method for evaluating the perceived quality of generated images (Wang et al., 2004). The SSIM index takes into consideration luminance (l), contrast(c) and structure (s), making it a more comprehensive metric. General forms of MAE, MAPE and SSIM are defined as:

$$MAE(x,y) = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \qquad (10)$$

$$MAPE(x,y) = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{x_i - y_i}{y_i} \right| \qquad (11)$$

$$SSIM(x,y) = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma \qquad (12)$$

In the above SSIM definition, α, β, γ are parameters that define the relative importance of the three components. The mean intensities are μ_x and μ_y, standard deviations are σ_x and σ_y, C_1 and C_2 are constants that are used to avoid zero denominator instability, and are related to dynamic range of pixel values. Mean intensity and standard deviation are weighted by a Gaussian weighting function of $\sigma = 1.5$. If we set α, β and γ all equal to 1, the equation becomes:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (13)$$

In addition to approximating the reflectance values, we expect the near-infrared channel to correctly reflect the characteristic of various landcover types. Specifically, the generated NDVI should also have low value at water bodies and high values at vegetation areas. We implement a simple classification rule which performs quantization on the NDVI and derives four classes that can be roughly summarized as water, baren, low vegetation, high vegetation. The classe definition is illustrated in Equation.14. It should be noted that the classification is an over generalization for all the landcover types in the dataset. But as we only want to test if fake NIR is capable of separating classes with distinctive spectral characteristics, the classification scheme is still meaningful used in our evaluation.

$$
\begin{aligned}
Pixel = Water, \quad if: \\
-1 < PixelValue < 0.1 \\
Pixel = Barren, \quad if: \\
0.1 \leq PixelValue < 0.1 \\
Pixel = LowVegetation, \quad if: \\
0.1 \leq PixelValue < 0.4 \\
Pixel = HighVegetation, \quad if: \\
0.4 \leq PixelValue \leq 1.0
\end{aligned} \qquad (14)
$$

After performing quantization on NDVI, we evaluate this 4-class classification map using Jaccard index, which is widely used in semantic segmentation tasks. It is defined as:

$$J = \frac{|x \cap y|}{|x \cup y|} \qquad (15)$$

The Jaccard Index is the area of intersection of prediction (x) and ground truth (y) divided by the area of union of prediction and ground truth. It can still give fair evaluation if class distribution is unbalanced.

The results for MAE, MAPE and SSIM of generated NIR band are shown in Table.2. In all 3 models the robust loss function show some improvement over L1 loss. It should be noted that patch discriminator might not be suitable for NIR generation task where fine grained information should be retained. The MAE of NDWI and NDVI, as well as the 4-class classification mIoU is illustrated in Table.3. Among all the network structures, the cGAN with Pixel Discriminator acquires the lowest MAE and MAPE, as well as the highest SSIM. It also achieves

the highest mIoU scores for NDVI classification. The cGAN-PixelD model also acquires the lowest MAE and MAPE for both NDWI and NDVI. As for loss function, robust loss shows improvement over L1 loss function in every model variation. On the other side, cGAN-PatchD with L1 loss function results the worst performance in all the evaluation. In this model variation, robust loss has demonstrated the biggest improvement over L1 loss.

Network	Loss	MAE ($\times 10^{-3}$)	MAPE (%)	SSIM
U-Net	L1	10.70	5.54	93.08
	Robust	10.60	5.46	93.13
cGAN-PatchD	L1	19.95	14.33	79.89
	Robust	10.85	5.19	92.28
cGAN-PixelD	L1	9.99	4.82	93.43
	Robust	9.67	4.73	93.63

Table 2. The result comparison of different methods. MAE, MAPE and SSIM of generated NIR band is calculated.

Network	Loss	MAE ($\times 10^{-3}$) NDWI	MAE ($\times 10^{-3}$) NDVI	mIoU (%)
U-net	L1	21.85	20.76	95.39
	Robust	21.79	20.70	95.34
cGAN-PatchD	L1	55.56	58.81	87.43
	Robust	21.80	20.11	95.41
cGAN-PixelD	L1	21.97	18.85	95.68
	Robust	18.90	17.61	95.79

Table 3. The MAE results for NDWI and NDVI respectively, and the mIoU based on NDVI classification. The mIoU score is the average among all four classes.

In Figure. 5, 6 and 7 we present some example results from various models. We find that the lowest MAE is always achieved in noisy images due to the relative low absolute reflectance values. These NIR bands have even discrete values. Therefore we exclude noisy images in visualization. We select some random images to show the performance in different landcover types, including water, forest, mountain, field and urban areas. We present the result in false color for better visualization. NDVI indices are shown in jet color map for visual comparison. We plot the histograms of fake and real NIR bands to compare the probability distributions. The blue one denotes the original NIR while orange one is that of fake NIR band. Except for some noisy images, almost all the fake NIR images demonstrate high level of realness compared with the real ones. The cGAN-PixelD model acquired best results in all evaluation metrics. As can be seen in figure.5e and figure.5j, the histograms of the generated NIR bands in general match reasonably well with the histograms of the real NIR bands. Patch discriminator cGAN model, on the other hand, shows decreased performance. It is the only combination that has mIoU below 90 percent. As we can see in figure.6e and figure.6j, the distributions show big differences and shifts from that of the real NIR bands. The NDVI values also have large difference in some specific areas. The U-net model can also yield reasonable results without cGAN objectives, but the result is not as good as cGAN-PixelD model.

However, the best results still show some degree of information loss compared with the original NIR band. Specifically, the generated images are relatively blurry in edges compared with the original NIR band. The reduced texture is possibly caused by the convolution operation and down sampling.

5. CONCLUSION

NIR band is important in remote sensing applications as it provides additional information about landcover types. They have also shown significance in many interdisciplinary researches. However, not every equipment is capable of capturing NIR bands, or sometimes NIR bands are missing in time series dataset. Generating NIR band from RGB bands has very important practical uses. Generative adversarial models have been proven to have good performance in many generative tasks. They have been transferred to many applications in remote sensing field such as DSM refinement and monocular depth estimation. In this paper, we have shown that cGAN is a viable method to be used in NIR band generation from RGB bands. The quality of the generated NIR band is numerically good, with around 5 percent of mean absolute percentage error. In addition, it can in general reflect accurate land cover class information through various NIR-based indices. However, the model still suffers from some texture loss that could potentially harm the usability of generated data. In the future, we will try to design a better generator structure that can retain more texture information, and a stronger discriminator that can distinguish between more subtle differences. In addition, imagery from different seasons will later be tested in following research to verify if the method is applicable when atmospheric property is distinct. Training and testing on multi-sensor dataset will also be experimented.

ACKNOWLEDGEMENTS

REFERENCES

Barron, J. T., 2019. A general and adaptive robust loss function. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4331–4339.

Bittner, K., Körner, M., Fraundorfer, F., Reinartz, P., 2019. Multi-Task cGAN for Simultaneous Spaceborne DSM Refinement and Roof-Type Classification. *Remote Sensing*, 11(11), 1262.

Black, M. J., Anandan, P., 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1), 75–104.

Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M., 1994. Two deterministic half-quadratic regularization algorithms for computed imaging. *Proceedings of 1st International Conference on Image Processing*, 2, IEEE, 168–172.

Dennis Jr, J. E., Welsch, R. E., 1978. Techniques for nonlinear least squares and robust regression. *Communications in Statistics-simulation and Computation*, 7(4), 345–359.
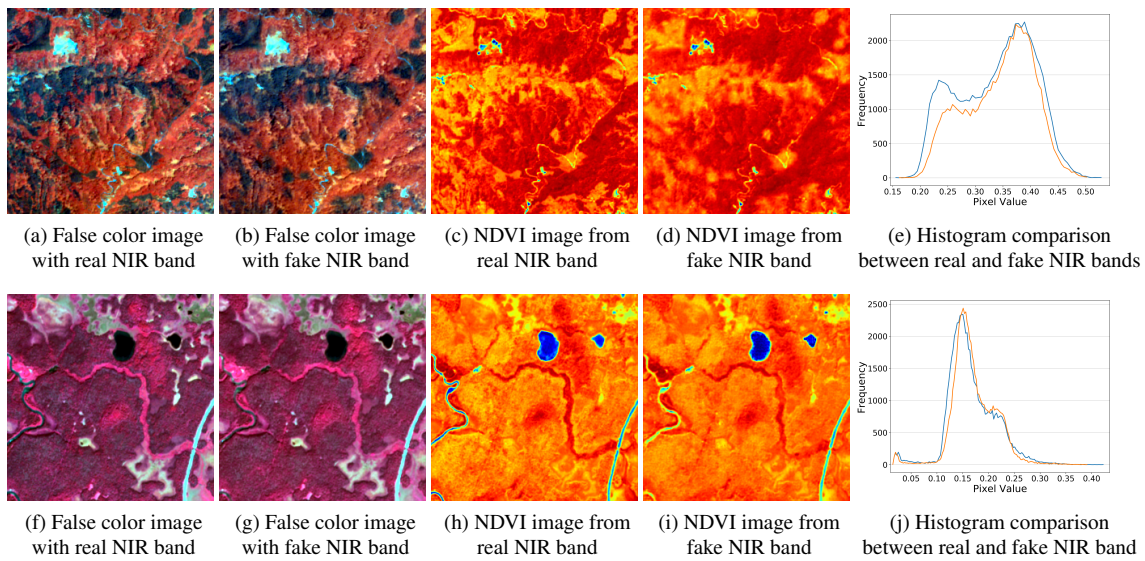
(a) False color image
with real NIR band

(b) False color image
with fake NIR band

(c) NDVI image from
real NIR band

(d) NDVI image from
fake NIR band

(e) Histogram comparison
between real and fake NIR bands

(f) False color image
with real NIR band

(g) False color image
with fake NIR band

(h) NDVI image from
real NIR band

(i) NDVI image from
fake NIR band

(j) Histogram comparison
between real and fake NIR band

Figure 5. The results of U-net with l1 loss.

(a) False color image
with real NIR band

(b) False color image
with fake NIR band

(c) NDVI image from
real NIR band

(d) NDVI image from
fake NIR band

(e) Histogram comparison
between real and fake NIR bands

(f) False color image
with real NIR band

(g) False color image
with fake NIR band

(h) NDVI image from
real NIR band

(i) NDVI image from
fake NIR band

(j) Histogram comparison
between real and fake NIR band

Figure 6. The result of cGAN-PatchD with l1 loss.

(a) False color image
with real NIR band

(b) False color image
with fake NIR band

(c) NDVI image from
real NIR band

(d) NDVI image from
fake NIR band

(e) Histogram comparison
between real and fake NIR bands

(f) Real NIR band in
false color

(g) Fake NIR band in
false color

(h) NDVI image from
real NIR band

(i) NDVI image from
fake NIR band
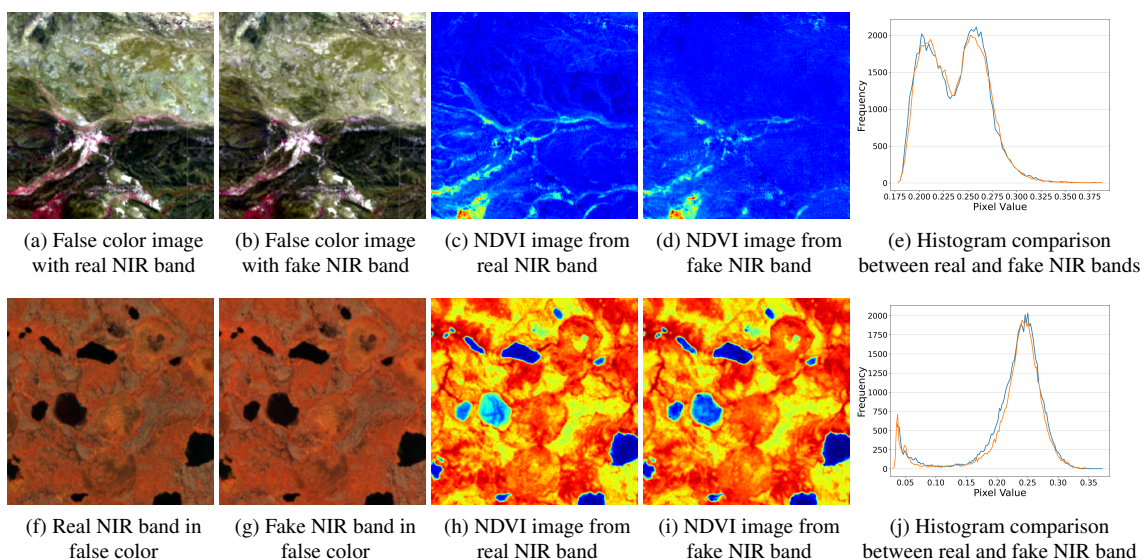
(j) Histogram comparison
between real and fake NIR band

Figure 7. The result of cGAN-PixelD with robust loss.

European Space Agency, 2015. *SENTINEL-2 User Handbook.* [Online]. Available from: `http://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook.` [Accessed 15th Jan 2020].

Fu, Y., Zhang, T., Zheng, Y., Zhang, D., Huang, H., 2018. Joint camera spectral sensitivity selection and hyperspectral image recovery. *Proceedings of the European Conference on Computer Vision (ECCV)*, 788–804.

Geman, S., McClure, D., 1985. Bayesian image analysis: An application to single photon emission tomography. *Amer. Statist. Assoc*, 12–18.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 2672–2680.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Kaufman, Y. J., Sendra, C., 1988. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *International Journal of Remote Sensing*, 9(8), 1357–1381.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Liu, X., Wang, Y., Liu, Q., 2018. Psgan: a generative adversarial network for remote sensing image pan-sharpening. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 873–877.

McFeeters, S. K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), 1425–1432.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Schmitt, M., Hughes, L. H., Qiu, C., Zhu, X. X., 2019. SEN12MS–A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *arXiv preprint arXiv:1906.07789*.

Sun, D., Roth, S., Black, M. J., 2010. Secrets of optical flow estimation and their principles. *2010 IEEE computer society conference on computer vision and pattern recognition*, IEEE, 2432–2439.

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.

Zhan, Y., Hu, D., Wang, Y., Yu, X., 2017. Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 212–216.