Band Reconstruction for High Resolution Satellite Images using Deep Learning

Md Aminur Hossain*, Sanjay K. Singh, Arvind K. Singh, & S. Devakanth Naidu

Space Applications Centre, ISRO, India
*Corresponding Author's email: aminur@sac.isro.gov.in

Abstract

High resolution Multispectral (MX) satellite imaging systems contain multiple bands with high spatial resolution and swath. Due to the complex mechanisms involved in image acquisition and data transmission for such systems, it is possible that the information in some of these bands is lost and even one of the bands (like Blue Band) does not exist during system de-sign to reduce system complexity. These datasets are unsuitable for end-use in such situations. Due to the missing blue band images, blue band needs to be synthesized for making a Natural Colour Composite (NCC), in order to get more insight from the data. It is essential to design methods which can faithfully reconstruct data given the available bands and a simultaneously acquired Panchromatic image (PAN). In this work, we propose a Deep Learning method to reconstruct full single-band images or a portion of a band in MX images using a dense U-Net based Wasserstein Generative Adversarial Network (WGAN). Specifically, we demonstrate the strength of our network by synthesizing blue band: a) from multispectral Red(R), Green(G), and Near Infrared (NIR) bands as inputs or b) from co-registered PAN and multispectral R, G, NIR bands as inputs. Qualitative and quantitative results obtained from experiments performed on High-Resolution Cartosat satellite images show that our method is able to reconstruct images that are spatially and spectrally accurate. Proposed method is also demonstrated for cross satellite validation i.e. if one model is trained on a particular satellite, same can be used for the similar type of other satellites without using transfer learning, overcoming the need of ground truth. Metrics like peak signal-to-noise ratio (PSNR) and Frechet Inception Distance (FID) are used to evaluate our model at different steps of training to study training convergence.

Keywords: Band Reconstruction; Deep Learning; Generative Adversarial Networks, Dense U-Net, Wasserstein Distance

Introduction

Remote sensing is the process of gathering information about an object or area from a distance, typically using sensors and instruments mounted on satellites, aircraft, drones, or ground-based platforms. It has become an essential tool in various fields such as environmental monitoring, agriculture, urban planning, disaster management, and natural resource management. Remote Sensing is a complex combination of various subsystems starting from satellite building, launching and operational onboard. In the process of launching a satellite to acquiring remotely sensed images, lot of things could occur in an unfavorable situation in, before launch or after launch.

High-resolution remote sensing data become essential for detailed analysis of ground objects. Development of high-resolution imaging sensors are becoming complex systems due to requirements of high spatial, spectral resolution and swath. Also, the advancement of sensor technology has grown in complexity to meet the demands for greater precision in spatial, spectral and temporal aspects. Modern sensors employ intricate opto-electrical systems to cover extensive areas effectively. However, transmitting the substantial data they generate to ground stations poses challenges due to limited bandwidth. This can result in the data loss in one or more imaging bands. Consequently, the remote sensing community is highly motivated to develop algorithms that can effectively restore missing data from the available bands, recognizing the significance of this endeavor.

Also due to the operational conditions of satellite sensors and the atmospheric environment, remote sensing images frequently exhibit typical image issues, including horizontal stripe artifacts, malfunctioning detectors, absent ports, haze, as well as missing single and multi-band data. Addressing these inherent challenges in spaceborne systems necessitates ground-based solutions to enhance data quality and make it more suitable for users.

While there exist various approaches in the literature for reconstructing and inpainting non-satellite images, there is a notable scarcity of methods tailored for the high-resolution band reconstruction of satellite imagery. Traditional techniques for general images often tackle this problem by employing either parametric or non-parametric patch similarity measures, as seen in (Gupta et al. 2012) and (Ironi et al. 2005). More recently, deep learning-based methods have exhibited remarkable performance improvements in tasks such as inpainting, colorization, and band reconstruction in the context of general images. Notable benchmarks in this realm include pix2pix (Isola et al. 2017) and image colorization (Zhang et al. 2016), along with CycleGAN (J.-Y. Zhu et al. 2017), which leverage Generative Adversarial Networks (GANs) (Goodfellow et al. 2020) to address these challenges.

In the domain of satellite imagery, (Rangekar et al. 2017) harnessed conditional GANs to map coarse-resolution multi-bands to a coarser-resolution hyperspectral domain. Similarly, (Arad and Ben-Shahar, 2016) put forth a range of techniques for extracting hyperspectral bands from undersampled acquisitions within the spectral domain. Meanwhile, (L. Rout 2020) employed a ResNet-based GAN model, augmented with expert regularization, to generate single-band satellite images from multiple bands. Another notable contribution, by (Jiangtao Xu et al. 2021), introduced a Dense U-Net generative adversarial network tailored for near-infrared face images, which demonstrated superior accuracy compared to Pix2pix and CycleGAN models. This method incorporates a Dense U-Net-based generator and a convolutional PatchGAN discriminator.

In our work, our primary objective is to reconstruct a single blue band from one or more available bands, specifically within the context of satellite imagery. Furthermore, our proposed approach is evaluated for cross-satellite validation, aiming to determine whether a model trained on one satellite can be effectively applied to similar satellites without the need for transfer learning, given the absence of Ground Truth (GT) data.

Proposed Method

The problem statement states Artificial Satellite Image Regeneration by synthesizing some bands of images. We are regenerating band because some of the band is faulty or missing,

with high noise, haze or clouds. We are targeting regeneration to have minimal operational cost and complexity of our operations. For our problem we will be using GAN model because they produce promising result with high picture quality and minimum loss in information. In our approach, the GAN model will take a) multispectral Red(R), Green(G), and Near Infrared (NIR) bands as inputs or b) Red, Green, and Near-Infrared images, along with co-registered PAN band images, to generate the missing blue band. This enhancement in the generation process is intended to improve accuracy.

Network Architecture:

For this, we've employed a customized generator based on the Dense U-Net model (Jiangtao Xu et al. 2021), along with a convolutional PatchGAN classifier (Isola et al. 2017) as the discriminator. The role of the discriminator is to penalize deviations in structure at the scale of image patches. The subsequent sections will delve into the intricate details of the network's architecture and the training methodology. To illustrate, we will specifically describe the network's architecture for the synthesis of blue band images using co-registered PAN, Red, Green, and Near-Infrared (NIR) inputs as an example. However, it's important to note that the overall architecture remains consistent when working with different sets of inputs like NIR, R and G bands only. For a visual representation of the architecture, refer to Fig. 1.

Let $z_1 \sim P_{PAN}$, $z_2 \sim P_{NIR,R,G}$ and $x \sim P_{Blue}$where P_{PAN}, $P_{NIR,R,G}$ and P_{Blue}, represent the distribution of PAN, NIR,R,G, and Blue bands respectively. Let z $\sim$ P_z (z_1, z_2), where $z_i \in R^{MXN}$, i = 1, 2 and P_z represents joint source distribution. Let $\hat{x} \sim P_{\hat{y}}$ where $\hat{x} = G(z)$ and $P_{\hat{y}}$ represents generator distribution. The corresponding sample of z in the target distribution is denoted by y $\sim P_y$. The problem of blue band reconstruction is formulated such that the generator learns to sample from the distribution of missing blue band, P_y given the corresponding sample from existing concurrent joint distribution, P_z. The generator model G operates on the source distribution z $\sim$ P_z, with z $\in$ $R^{H \times W \times 4}$ with H and W representing the height and width of image patches, respectively. Let y $\sim$ P_y, be the target distribution related to blue band images with y $\in R^{H \times W \times 1}$. Let $\hat{y} \sim P_{G_\theta}$ where $\hat{y} = G_\theta(z)$ with $\hat{y} \in R^{H \times W \times 1}$, and P_{G_θ}be the generator distribution parameterized by θ. In this setup, a discriminator, D, is designed to classify whether a given ground truth data sample y and a generated data sample $\hat{y}$, conditioned on input z, belong to the real or fake sample space, respectively. D's role is to distinguish between actual data and generated data.

Objective Functions:

The objective function of D is used as that in Wasserstein GANs (WGANs) (Arjovsky et al. 2017) to circumvent the mode collapse and stability issue of conventional GANs. Nonetheless, the WGAN's weight clipping operation restricts itself to a limited set of critic functions, potentially failing to effectively capture complex and challenging-to-learn latent features. To tackle this limitation, we integrate the objective function with Gradient Penalty (GP) (Gulrajani et al. 2017).

In conventional GANs framework, loss function is defined as Eq. 1 and for pixel-to-pixel image translation we used L1 norm regularization for Generator with loss defined as Eq. 2.

$$L_{GAN}(G,D) = E_{x\sim P_{Blue}}\left[\log\left(D(x)\right)\right] + E_{\hat{x}\sim P_{\widehat{Blue}}}\left[\log\left(1-D(\hat{x})\right)\right] \quad (1)$$

$$L_{L1}(G) = E_{x\sim P_{Blue},\ z\sim P_Z(z_1,z_2)}\left[||x - G(z)||_1\right] \quad (2)$$

The objective function for training the generator model G includes two key components: the adversarial loss determined by the discriminator and the mean absolute error (MAE) between the generated sample y and the true sample $\hat{y}$ defined as pixel-to-pixel loss in Eq. 2. Together, these elements define our ultimate objective for the generator is shown in Eq. 3.

$$G^* = arg\min_G \max_D L_{GAN}(G,D) + \lambda * L_{L1}(G) \quad (3)$$

where λ is the regularization parameter, assigned weight of the $L1$ loss in overall generator objective function. The final objective function to train D can be defined as in **Eq. 4**., where λ_{gp} is the weight for GP.

$$D^* = arg\min_D\ E_{\hat{y}\sim P_{G_\theta}\ z\sim P_z}[\mathrm{D}(\hat{y},\mathrm{z})] - E_{y\sim P_y\ z\sim P_z}[\mathrm{D}(\mathrm{y},\mathrm{z})] + \lambda_{gp}E_{\hat{y}\sim P_{\hat{y}}\ z\sim P_z}[(||\nabla_{\hat{y}}\mathrm{D}(\hat{y},\mathrm{z})|| - 1)^2] \quad (4)$$

Datasets and Training Methodology:

The dataset plays a pivotal role in Deep Learning techniques. To effectively train any deep learning model, a sufficient dataset is required, encompassing both the input and corresponding output distributions for supervised learning. In our case, we have opted High-Resolution Cartosat (2S/3) satellite data in order to train a neural network for band reconstruction task. We've acquired Cartosat precision-corrected MX products and bundled products, which contain registered PAN and MX data at their native resolutions. These datasets span various acquisition dates and encompass diverse spectral signatures, such as cityscapes, landscapes, hills, deserts, and water bodies etc. This diverse dataset is essential to ensure that our network learns to accurately reconstruct features of various possible signatures which may come in scenes at inference time.

Each Cartosat-3 scene covers a ground area of 17 km x 17 km, and we've utilized nearly 250 scenes to generate image chips. Our dataset is derived from Cartosat images with an 11-bit 1.1m Ground Sampling Distance (GSD) for MX and 0.28m GSD for PAN images. We've divided the original full-swath images belonging to different spectral classes, into patches of size 256 × 256 pixels to create the datasets. In total, we've generated 4,10,000 such chips, out of which, 3,00,000 chips are used in training, whereas 1,10,000 are used for testing the model.

The training process was conducted on a server equipped with 4 TESLA V100 GPUs, employing a batch size of 8 chips and images of size 256x256 pixels. Training across all four GPUs took approximately 20 minutes for each epoch, and we completed 500 epochs. We trained the model using ADAM (Kingma and J. Ba 2015), optimizer with a learning rate of 1e − 5. The weights for gradient penalty and L1 losses are λ_{gp} = 10 and λ = 100, respectively.

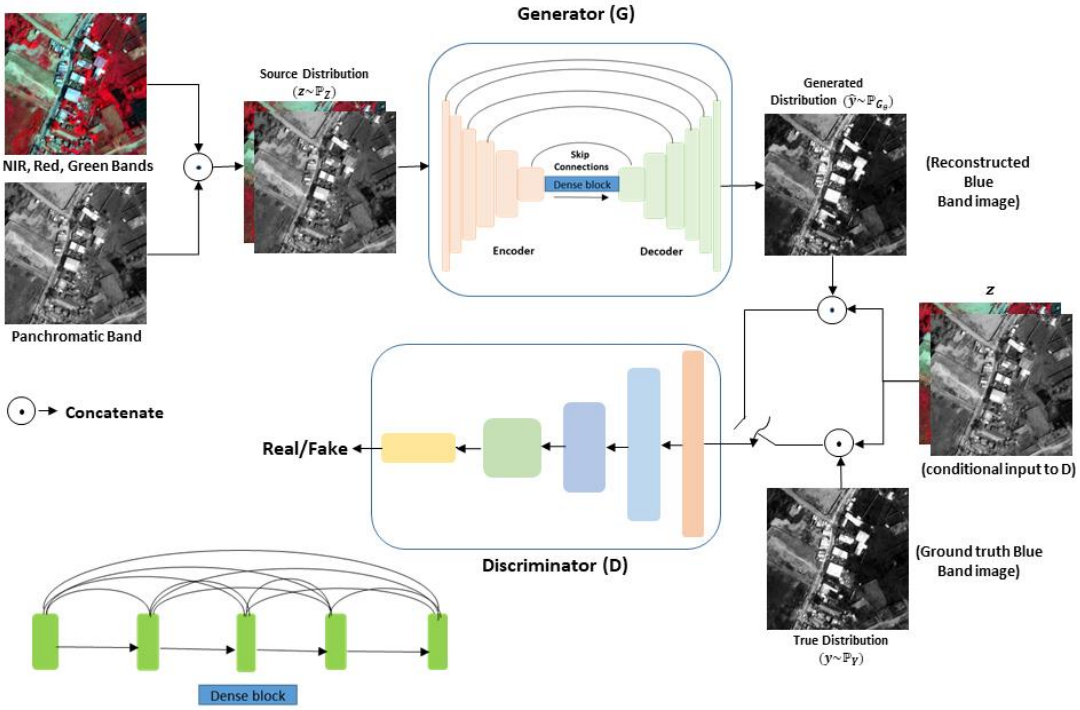

Fig. 17 Overview of our proposed Method and Model Architecture.

Results and Discussion

A GAN model's evaluation can be approached in two distinct ways. Firstly, the Qualitative Approach involves the comparison of images generated by the Generator with actual images. However, this method has its limitations, as the Generator can become so adept at producing realistic images that distinguishing them from genuine ones becomes visually challenging. Additionally, one can analyze the loss values of both the generator and discriminator. On the other hand, the Quantitative Approach entails the use of predefined metrics to assess the model's performance, providing a mathematical perspective on its accuracy. In this study, two metrics, PSNR and FID, are employed. After running the model for over 500 epochs on the primary dataset, the results indicated that the model converges as training progresses. Notably, key model parameters, such as generator and discriminator losses, exhibit minimal changes with ongoing training. To gain a comprehensive understanding of the model's performance, a comparative analysis using both Qualitative and Quantitative Approaches is conducted. Despite training our model with both approaches, utilizing NIR, R, G, and PAN+NIR, R, G bands for input distribution in the generation of blue bands, we have chosen to exclusively present results for the PAN+ (NIR, R, G) model owing to its better performance.

Qualitative Results: Images generated by Generator from our proposed approach can be visualized as in Fig. 2 when applied to a variety of test samples. These samples have been thoughtfully chosen to exemplify the method's ability to reproduce diverse spectral characteristics. Outputs are generated from PAN, NIR, Red, and green bands, which serve as inputs for our GAN model. The first and third rows of the figure display the generated images, while the second and fourth rows display the corresponding ground truth images. In

the generated image, we have replaced the generated blue band for visual comparison with the ground truth RGB images, including the target blue band. Blue band generation was performed for a portion of the complete Cartosat-3 satellite scene over Ahmedabad region shown in Fig. 3. The newly generated band is substituted in RGB visualizations. The left image represents the generated result, while the right image displays the original RGB input images. The proposed technique can successfully reconstruct complete blue band images with high visual quality, similar to the input images.

Quantitative Results: In this context, we evaluate the model's accuracy through mathematical formulas and predefined functions. We have calculated metrics such as Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and the coefficient of determination ($\boldsymbol{R^2}$) for the generated blue band. The computed values across all test samples are as follows: 18.06 for MAE, 43.86 dB for PSNR, 99.26 for SSIM, and 0.98 for $\boldsymbol{R^2}$. The low MAE value suggests a favorable generation capability, as it signifies a minimal average error. Furthermore, the high PSNR value indicates superior visual image quality. The exceptionally high SSIM value suggests that the structural similarity remains well-preserved when compared to the ground truth images.

Training convergence analysis: For training convergence analysis, we used FID and PSNR values generated for test samples. The FID value of the model was observed from epoch 70 to epoch 500, as shown in the left image of Fig. 5. The FID value should be as low as possible. A lower FID value indicates that the images generated by the Generator are of higher quality as the number of epochs increases. The PSNR value of the generated image varies at different epochs, as depicted in the right image of Fig. 5. It is evident that the overall PSNR value increases from 70 epochs to 500 epochs. We concluded our model training at 500 epochs based on an analysis of model loss, PSNR, and FID values.

Fig. 4 illustrates comparison between original and reconstructed pixel values for blue band. The R^2 value 0.98 indicate that there is a strong correlation between pixel values for the original and generated images during the reconstruction process.

Cross Satellites validation and Analysis:

For blue band generation training was done for Cartosat-3 and Cartosat-2E (GSD 0.6 m PAN and 1.6m MX) separately. One of the applications of this method is cross satellite single band image generation, where one of the bands (like Blue Band) does not exist during satellite system design in order to reduce the system complexity. Proposed method is also tested for cross satellite validation. We have trained with Cartosat-3 (C3) satellite data and applied to Cartosat-2E (C2E) satellite and vice versa. In Fig. 6(a), the plotted curve represents the fitting results obtained using the Cartosat-3 blue band model and applied to Cartosat-3 data, yielding an impressive R-square value of approximately 0.98. Moving on to Fig. 6(b), we can observe the fitting curve derived from the Cartosat-3 model and its application to Cartosat-2E data, resulting in an R-square value of 0.86. In Fig. 6(c), we explore the application of the C2E model to Cartosat-3 data, which yields an R-square value of 0.75. Finally, in Fig. 6(d), the C2E model is employed with C2E data, showcasing an impressive R-square value of 0.975.

This analysis reveals that our blue band generation network demonstrates commendable generation accuracy, both within the same sensor and in cross-sensor applications. Additionally, it is worth noting that this model can be effectively utilized with data from other high-resolution satellites, to achieve generation accuracy of approximately 86%. Table 1 provides further insights by presenting an analysis of the difference images between generated blue band data and ground truth, emphasizing lower standard deviation and mean counts as indicators of better model accuracy.

Fig. 2 Images generated by Generator from our proposed approach with PAN, NIR, Red, Green bands as inputs, can be visualized to a variety of test samples.

Fig. 3 Blue band generation for part of Ahmedabad scene of Cartosat-3 satellite. Left is Generated blue band replaced RGB image and right is input Ground Truth RGB images.

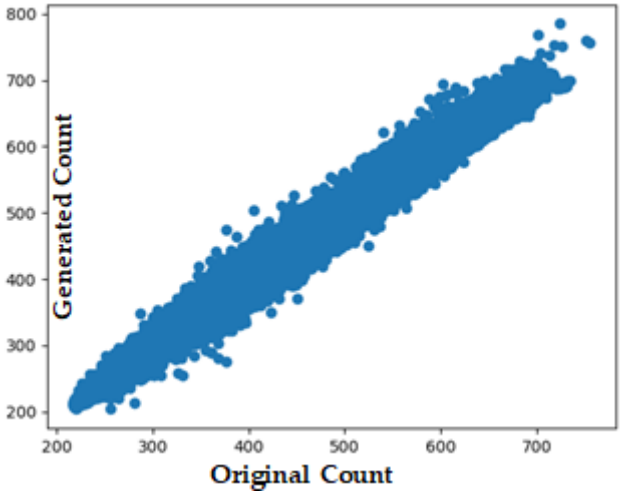

Fig. 4 Original and generated counts plot for blue band.

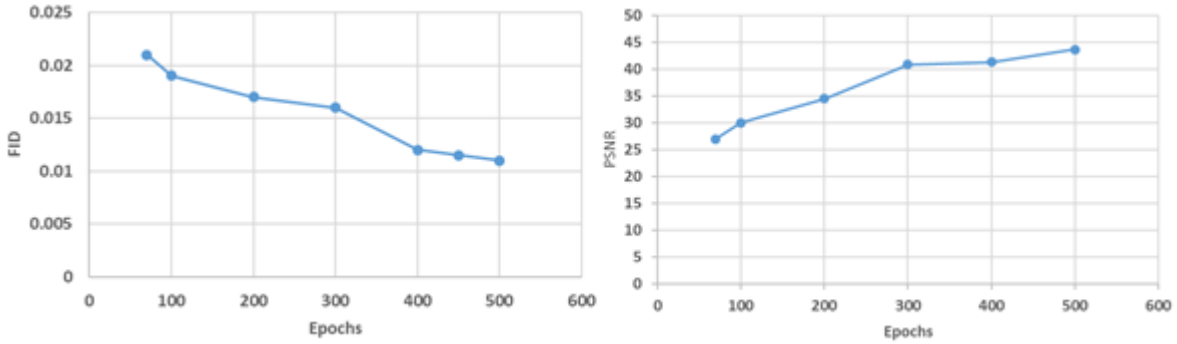

Fig. 5 FID (left) and PSNR (right) graphs with respect to training epochs.

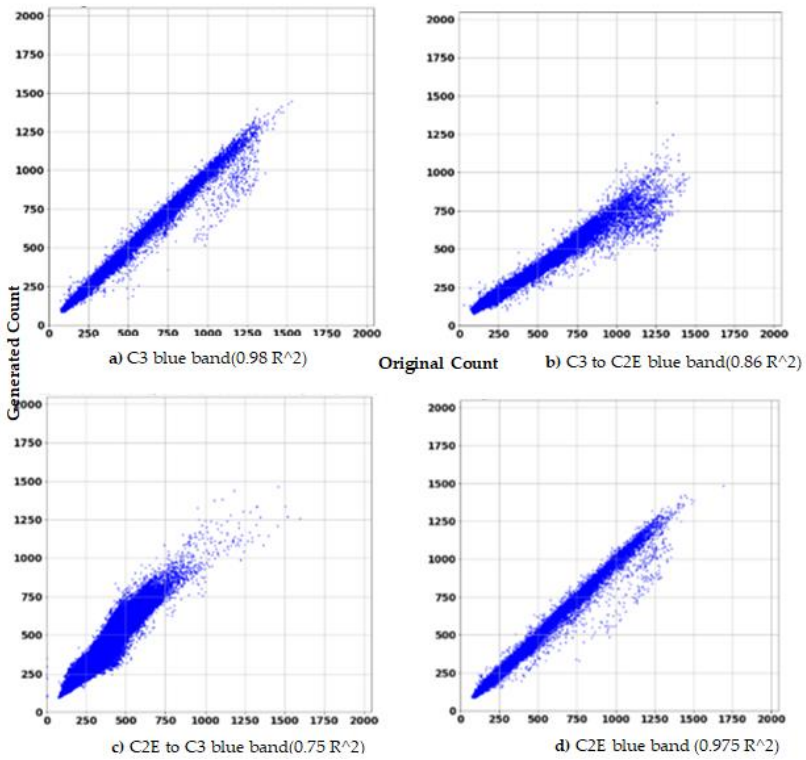

Fig. 6 Cross satellite validation results for Generated Vs Original counts with R-square values

Table 8 Difference image statistics between generated and ground truth blue band.

Trained Model	Input Bands	Output Band	Abs. Mean count	Std.(count)	R^2
C3	C3-PAN, NIR, R, G	C3-Blue	6	13	0.98
C2E	C2E-PAN, NIR, R, G	C2E-Blue	2.5	9.5	0.975
C3	C2E-PAN, NIR, R, G	C2E-Blue	14	26	0.86
C2E	C3-PAN, NIR, R, G	C3-Blue	25	31	0.75

Conclusion and Future Scope

The primary objective of this work is to create a deep learning model specifically designed to generate artificial high-resolution blue band images for satellites. The project involved an exploration of various GAN models tailored to our specific needs, with the ultimate choice being the Dense U-Net model. The training process involved the fine-tuning of parameters to minimize loss and mitigate any risk of model failure. The model achieved an impressive overall accuracy of 0.98, as assessed by the R2 score. Key metrics such as PSNR and FID are utilized to evaluate the model's performance during different training stages, with the final accuracy measured using the R^2 score.

For cross-satellite validation, the model trained on C3 data and tested on C2E data achieved a fitting accuracy of approximately 86%, while the model trained on C2E data and applied to C3 data achieved a fitting accuracy of around 75%. This cross-satellite application is beneficial for generating blue band images for satellites that lack this band, facilitating the visualization of natural color composites for various products.

Despite the model's achievements, there is room for improvement in several areas. Future work will focus on enhancing the accuracy of the cross-satellite validation model by

adjusting parameters and updating the model architecture. Additionally, incorporating more advanced and contemporary deep learning methods, such as transformer networks, will contribute to overall model accuracy enhancement.

Acknowledgements

We express their sincere gratitude to Shri. Nilesh M. Desai, Director, Space Applications Centre (SAC) Ahmedabad for encouragement to carry out this work. We also wish to express our gratitude to the National Remote Sensing Centre (NRSC) and our colleagues at SAC for their valuable support.

References

R. K. Gupta, A. YS. Chia, D. Rajan, E. S. Ng, and H. Zhiyong (2012), "Image colorization using similar images," in Proceedings of 20th ACM, pp. 369–378.

R. Ironi, D. Cohen-Or, and D. Lischinski (2005), "Colorization by example.," Rendering techniques, vol. 29, pp. 201–210.

P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros (2017), "Image-to-image translation with conditional adversarial networks,"in Proceedings of CVPR, pp. 1125–1134.

R. Zhang, P. Isola, and A. A. Efros (2016), "Colorful image colorization," in ECCV 2016: Proceedings. Springer, pp. 649–666.

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision.

I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D.W. Farley, S. Ozair, A. Courville, and Y. Bengio (2020), "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144.

A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, and M. Hoffman (2017), "Aerial spectral super-resolution using conditional adversarial networks," arXiv preprint arXiv:1712.08690.

B. Arad and O. Ben-Shahar (2016), "Sparse recovery of hyperspectral signal from natural rgb images," in ECCV 2016, Proceedings. Springer, pp. 19–34.

L. Rout (2020), "Alert: Adversarial learning with expert regularization using tikhonov operator for missing band reconstruction," IEEE TGRS, pp. 4395–4405.

Jiangtao Xu, Kaige Lu, Xingping Shi, Shuzhen Qin, Han Wang∗, Jianguo Ma (2021), A Dense Unet Generative Adversarial Network for Near-Infrared Face Image Colorization, Signal Processing 183(11):108007 DOI: 10.1016/j.sigpro.2021.108007.

M. Arjovsky, S. Chintala, and L. Bottou (2017), "Wasserstein generative adversarial networks," 06–11, vol. 70, pp. 214–223, PMLR.

I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville (2017), "Improved training of Wasserstein gans," Advances in NIPS, vol. 30.

D. P. Kingma, J. Ba (2015), "Adam: A Method for Stochastic Optimization", International Conference for Learning Representation, San Diego.

Citation

Hossain, Md. A., Singh, S.K., Singh, A.K., Naidu, S.D (2024). Band Reconstruction for High Resolution Satellite Images using Deep Learning. In: Dandabathula, G., Bera, A.K., Rao, S.S., Srivastav, S.K. (Eds.), Proceedings of the 43rd INCA International Conference, Jodhpur, 06–08 November 2023, pp. 547–556, ISBN 978-93-341-2277-0.