**Mersin Photogrammetry Journal**

https://dergipark.org.tr/en/pub/mephoj

**e-ISSN 2687-654X**

# Deep Learning Based Panchromatic2RGB Image Generation from VHR Images

**Ceyda Nigar Ünyılmaz*[1], Bahadır Kulavuz [2], Tolga Bakırman *[3], Bülent Bayram[4]**

[1] *Yildiz Technical University, Geomatics Engineering Department, Türkiye, nigar.unyilmaz@std.yildiz.edu.tr*
[2] *Yildiz Technical University, Geomatics Engineering Department, Türkiye, bahadir.kulavuz@yildiz.edu.tr*
[3] *Yildiz Technical University, Geomatics Engineering Department, Türkiye, bakirman@yildiz.edu.tr*
[4] *Yildiz Technical University, Geomatics Engineering Department, Türkiye, bayram@yildiz.edu.tr*

Cite this study: Ünyılmaz, C.N., & Kulavuz, B. Bakırman, T. Bayram, B. (Year). Deep Learning Based Panchromatic2RGB Image Generation from VHR Images. Mersin Photogrammetry Journal, 6 (2), 79-84.

**Keywords**
Image Colorization
Deep Learning
Remote Sensing
SpaceNet 6

Research Article
DOI:

Received:
Revised:
Accepted:
Published:

**Abstract**

Image colorization is the process of obtaining colored images by assigning RGB color values to a grayscale or panchromatic image. This technique has an important place in the field of computer vision because colored images provide a better visual experience and are widely used in areas such as image recognition and object detection. It also has many practical applications such as coloring historical photographs, adding colors to be used in the analysis of medical images, and improving the analysis of satellite images. Colorization methods are divided into two main categories: brush coloring and sample-based coloring. Both methods have certain limitations. The performance of these methods depends on the selected reference images and may sometimes contain false colors or significant errors. While these methods require operator intervention or pre-defined rules, deep learning based methods are largely automated and uses neural networks to understand the global and local context of an image, leading to more realistic and contextually accurate colorizations.

The presented study uses the Denoising Diffusion Null-Space Model (DDNM) architecture. DDNM is a method that aims to obtain more efficient and high-quality results compared to the coloring approaches available in literature. In the study, the weight data of the DDNM architecture was used to predict colored images from panchromatic images using the SpaceNet 6 open access dataset. The SpaceNet 6 dataset includes a combination of Capella Space 0.5m Synthetic Aperture Radar (SAR) imagery and Maxar's 0.5m electro-optical (EO) imagery. In order to assess the results, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) accuracy metrics are calculated.

## 1.Introduction

Image colorization, a fundamental problem in computer vision, is the process of converting gray-scale or single-band images into color (RGB) images. With this method, the visual interpretation of images is enhanced and their features are more clearly revealed [8]. (Anwar et al., 2025). Conventional colorization methods fall into two main categories: brush-based colorization and sample-based colorization. Both methods have certain limitations. The performance of these methods depends on the reference images selected and can sometimes contain false colors or significant errors. The diversity of features and the extensive area covered by remote sensing images make colorization a particularly challenging task.

Deep learning has emerged as a powerful tool for image colorization, addressing the challenges posed by traditional methods [1]. Deep learning-based image colorization methods demonstrate significant advantages over conventional techniques. These approaches operate with minimal human intervention, leveraging automated processes to enhance efficiency. By learning and incorporating both global and local image context, they generate realistic and contextually accurate color predictions. Unlike traditional methods, they generalize effectively across diverse and complex image datasets, producing vibrant and natural results. Furthermore, deep learning models exhibit scalability, enabling them to process large datasets and adapt to specific tasks with minimal manual adjustments. Their ability to emulate various artistic styles offers added flexibility for creative applications. The use of pretrained models facilitates transfer learning, reducing data requirements and computational effort during training. Additionally, these methods possess semantic understanding, allowing for object recognition and the application of contextually appropriate colors. Their robustness in handling novel or unseen data further underscores their versatility and superiority over traditional methods [2]. In this context, it has been determined that the use of deep learning models, which are a part of our lives, in image colorization studies has increased and different techniques are included in the literature. [3]. Li proposed a generative model with multi-discriminator for colorizing high resolution grayscale satellite images. [4]. Wu aimed to colorize remote sensing data by combining the DCGAN model, which is one of the leading generative models, with the multi-scale convolution Squeeze-and-Excitation Networks (SEnet) structure. [5]. Ji et al. proposed the MR-GAN architecture based on the CycleGan model for SAR image enhancement. The MR-GAN architecture aims to minimize the error by repeating the synthesis stage of SAR and optical satellite images multiple times using 3 CNNs in forward propagation[6]. Wu et al. converted RGB images into YUV images and aimed to produce UV bands using the Y band. In this context, they obtained good results with the DCGAN model with a deep convolution layer. Wang integrated their low-level correlation feature extraction (LCFE) module into the DCGAN model to create a model that preserves salient shallow detail features [7].. Bose et al. (2022) proposed an innovative dual attention fusion of receptive kernels (DARK) framework for coloring panchromatic images from color images using only a few training data[9]. Fu et al. (2024) developed the U-VIT model, a transformer-based generative model that aims to overcome the challenging training conditions in existing generative networks for coloring satellite images and the efficiency problem of Reducing Diffusion Probability Models (DDPM) requiring multiple samples. [10].

Deep learning architectures for satellite image colorization are often subject-specific and lack sustainability. In this context, the use of the Diffusion Null-Space Model (DDNM) architecture, which is a zero-shot colorization model that does not require training, was tested in the colorization of satellite images. Aim of the study:

- To demonstrate that high-resolution Panchromatic images can be colored with artificial intelligence,
- Evaluating the sustainability of a DDNM architecture that operates without the need for training,
- Identifying potential challenges in colorizing satellite images with diverse objects.

## 2. Material and methods

### 2.1 SpaceNet-6 Dataset

Remote sensing methods vary widely in spatial resolution, sensor type, and intended application. Optical satellites often rely on panchromatic bands to achieve high spatial resolution. Pansharpening, a technique that combines low-resolution multispectral bands with high-resolution panchromatic bands, is commonly used to generate high-resolution color satellite imagery. However, this process can introduce spectral distortions.

This study explores a novel approach to pansharpening by leveraging deep learning techniques to colorize satellite images. The MultiSensor All Weather Mapping (MSAW) dataset [8], comprising high-resolution optical and SAR imagery of Rotterdam port, was selected for this study. The optical images in the data set were obtained from the Maxar Worldview-2 satellite. A cloud-free image strip was collected by satellite on August 31, 2019 at 10:44 AM, with a 236 km$^2$ spatial extent and an 18,4° off-nadir viewing angle. The optical data set consists of 3 different optical data:

- one band panchromatic (0.5m),
- four multi-spectral bands (2.0m): blue, green, red,and near-infrared (NIR)
- four pan-sharpened bands (0.5m): blue, green, red and NIR [8].

This dataset offers diverse geographical features, including dense urban areas, rural landscapes, and coastal environments. For our experiments, we focused on six specific land cover classes: Industry, Forest, Harbor, Road, Building, and Seaside. The dataset, originally composed of 900x900 images, was divided into smaller 256x256 patches for analysis (Figure 1).

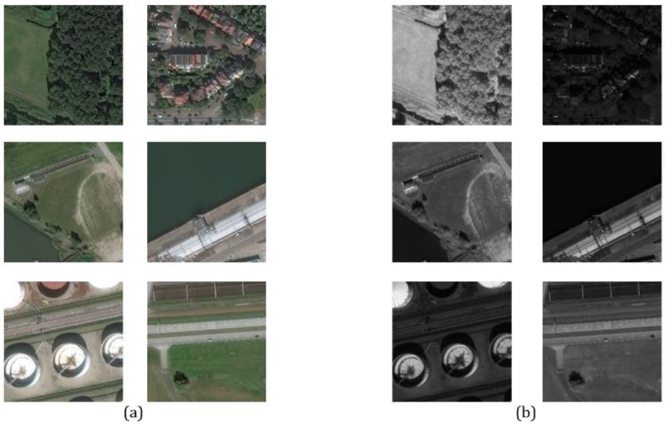


**Figure 1**. RGB(a) and panchromatic(b) imagery within the dataset

### 2.2. Diffusion Null-Space Model

Image Restoration (IR) models are typically tailored to specific tasks, limiting their versatility. The Diffusion Null-Space Model (DDNM) offers a more general approach, addressing various image restoration challenges such as resolution enhancement, colorization, deformation correction, and blur removal without requiring extensive training.

DDNM leverages a pre-trained diffusion model, focusing on refining the null-space [15] (during the reverse diffusion process. This approach ensures realistic and consistent results without the need for optimization, providing a zero-shot solution.

ImageNet and CelebA datasets were used in the training phase of the DDNM architecture.

The ImageNet dataset [11] consists of 14,197,122 images labeled according to the WordNet hierarchy with a size of 30x50. It is widely used in image classification and object detection studies.

The CelebA dataset [12] consists of 202,599 aligned and cropped facial images, each with a resolution of 178x218 pixels, sourced from 10,177 distinct celebrities. Each image is associated with a comprehensive set of 40 binary annotations indicating various facial attributes, including hair color, gender, and age.

The DDNM comprises a forward process that gradually adds noise to the data and a reverse process that aims to reconstruct the original data from the noisy input [13]. A schematic representation of the model is illustrated in Figure 2.

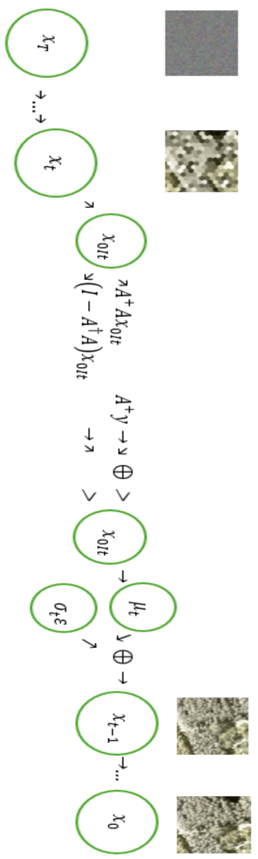

**Figure 2.** Schematic illustration of the model

The hyperparameters of the DDNM architecture are presented in Table 1.

**Table 1.** Hyperparameters

| Hyperparameters | |
|---|---|
| Diffusion steps | 1000 |
| Batch size | 1 |
| Beta start | 0.0001 |
| Beta schedule | Linear |
| Number of channels | 128 |
| Noise schedule | Linear |
| Attention resolution | 16,8 |

### 2.3 Accuracy Metrics

To assess the quality of the colorized images generated through the image colorization process, accuracy metrics are indispensable. This study employs the widely adopted peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metrics, which are prevalent in the literature for evaluating image quality and similarity, to assess the performance of our deep learning model.

#### 2.3.1 Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is an engineering metric employed to quantify the ratio between the maximum potential power of a signal and the disruptive power of noise that compromises its fidelity. Given the extensive dynamic range of many signals, PSNR is typically expressed logarithmically in decibels (dB).

PSNR is a widely-used metric for assessing the reconstruction quality of images and videos subjected to compression-induced loss. It finds application in diverse domains such as image compression, enhancement, and quality assessment. Image compression, the process of reducing image file size, benefits from PSNR evaluation to gauge the performance of compression algorithms. A high PSNR value indicates effective compression that preserves image quality.

Image enhancement, which aims to improve the quality of degraded images, also leverages PSNR to evaluate the efficacy of enhancement algorithms. By effectively removing noise or other artifacts, enhancement techniques can yield images with higher PSNR values.

PSNR offers a reliable and repeatable method for measuring image quality by comparing it against images processed with various algorithms [14].

The mathematical formula for calculating PSNR is presented in Equation (1):

$$PSNR(f,g) = 10\ (\frac{255^2}{MSE(f,g)}) \qquad (1)$$

where

f = reference image,
g = colorized image,
255 = maximum grey value of a pixel,
MSE = mean squared error,
Mean squared error (MSE).

#### 2.3.2 Structural Similarity

Structural Similarity Index Measure (SSIM) is a widely-used technique for assessing the perceptual quality of digital images, including television, cinema, and various other digital media formats. SSIM quantifies the similarity between two images.

As a full-reference metric, SSIM relies on an uncompressed, distortion-free reference image to evaluate the quality of a test image. Unlike other metrics such as Mean Squared Error (MSE) or Peak Signal-to-Noise Ratio (PSNR), which assess absolute error and image details, SSIM focuses on the structural similarity between images [14].
The mathematical formula for calculating SSIM is presented in Equation (2):

$$SSIM(f,g) = \frac{(2\mu_f\mu_g+c_1)(2\sigma_{fg}+c_2)}{(\mu_f^2+\mu_g^2+c_1)(\sigma_f^2+\sigma_g^2+c_2)} \quad (2)$$

where,
$\mu_f$ = average of reference image,
$\mu_g$ = average of colorized image,
$\sigma_f$ = variance of reference image,
$\sigma_g$ = variance of colorized image,
$c1, c2$ = two variables to stabilize the division with weak denominators.

## 3. Results and Discussion

In this study, a workstation equipped with an 11th-generation Intel Core i9-11900 2.50-GHz processor and an NVIDIA Quadro RTX 5000 16-GB graphics card was utilized for the testing of the deep learning architectures.

During the prediction phase, an investigation was conducted to identify the optimal data format for the satellite image colorization model. Initially, it was aimed to use the raw images at their original size (900x900 pixels) to assess the impact of image resolution on colorization accuracy. However, analysis of the results revealed that the model was resizing the input images during processing, leading to a loss of detail in the satellite imagery. Furthermore, it was determined that the model encountered integration issues due to the pixel depth exceeding 8 bits. Based on these observations, the images were pre-processed by performing normalization based on their maximum pixel values and conversion to 8-bit depth. Subsequently, the images were resized into smaller tiles of 256x256 pixels for model prediction. Table 2 presents the PSNR and accuracy values achieved for the colorized satellite images.

**Table 2.** Accuracy Results

| Class | PSNR | SSIM |
|---|---|---|
| Industry | 9.0685 | 0.4797 |
| Forest | 7.6632 | 0.4839 |
| Harbor | 10.6253 | 0.2373 |
| Road | 20.8840 | 0.9249 |
| Building | 11.5581 | 0.1486 |
| Seaside | 14.8631 | 0.7171 |

Visual inspection of the generated colorized images (Figure 3) revealed inconsistencies in water region coloring. While objects on water surfaces exhibited homogeneous colorization, the model struggled to differentiate objects within dense forest areas, successfully coloring only the tree canopy. Conversely, the model demonstrated an ability to distinguish objects in mixed-class regions, such as images containing residential areas, roads, and green spaces. These objects were assigned distinct colors, suggesting class-specific colorization attempts.

It was determined that SSIM and PSNR values varied for different classes, but the road class achieved the highest coloring accuracy. Analysis of the prediction images and accuracy measurements further confirmed this observation.

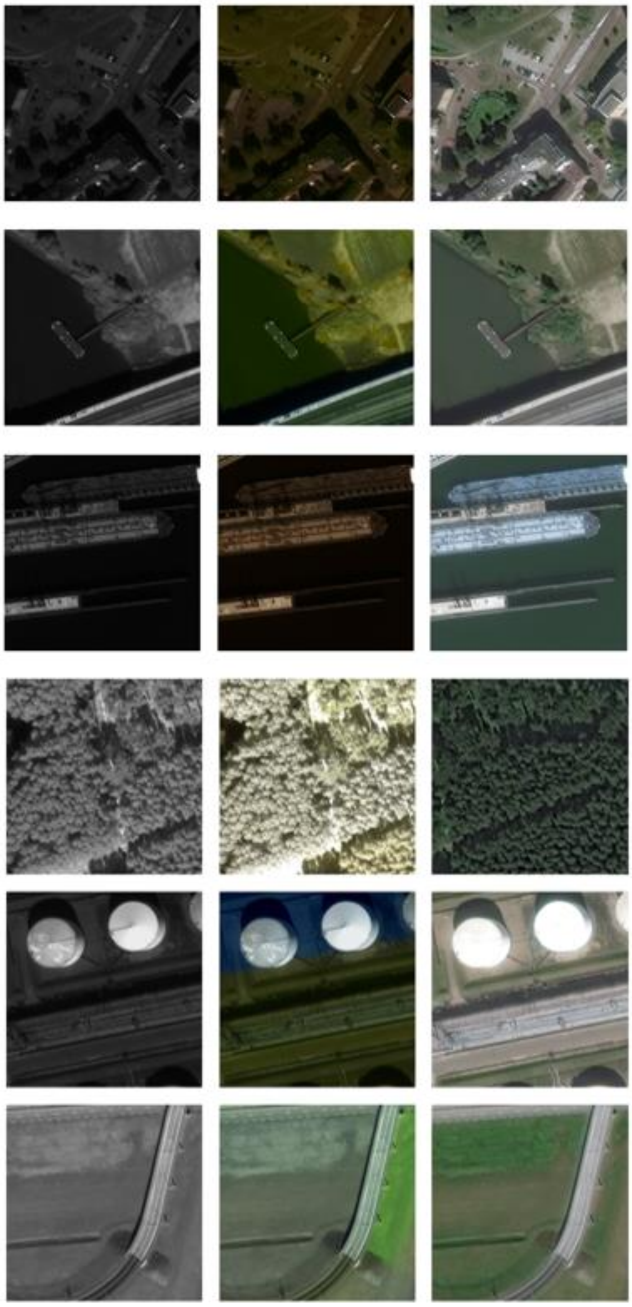

**Figure 3**. Comparison of prediction images (Left to Right Panchromatic image, Prediction image, RGB image)

## 4. Conclusion

The achieved accuracy metrics indicate the model's potential for real-world application due to its class-agnostic nature, meaning it can process unseen classes. However, the current performance falls short of expectations. While the model exhibits the capability to colorize satellite imagery based on object density, its performance deteriorates in regions with dense forests and water bodies. This limitation is attributed to the potential lack of satellite images containing such features within the training data used for the pre-trained weights employed during inference.

Additionally, it is hypothesized that preserving the original image depth during pre-processing could lead to improved results in future experiments. The model's high computational cost necessitates further investigation into model architecture modifications and hyperparameter optimization. Despite these shortcomings, the model offers significant advantages over traditional methods: faster processing times and the ability to directly generate data without extensive training, making it suitable for preliminary information gathering. This study demonstrates the potential of DDNM for colorization tasks. Furthermore, it suggests that colorization approaches utilizing smaller, regionally-specific datasets for training, as opposed to pre-trained models, could potentially yield more accurate results. This shows the scalability potential of the DDNM model for applicability to other satellite datasets and tasks, enabling broader adoption in diverse remote sensing applications. These findings highlight the potential for sustainable and efficient use of deep learning models in real-world applications, particularly by reducing reliance on extensive training datasets and enabling faster, region-specific colorization processes.

## Acknowledgments

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

## References

1. Zhang, R., Isola, P., Efros, A.A. (2016). Colorful Image Colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9907. Springer, Cham. https://doi.org/10.1007/978-3-319-46487-9_40
2. Huang, S., Jin, X., Jiang, Q., & Liu, L. (2022). Deep learning for image colorization: Current and future prospects. Engineering Applications of Artificial Intelligence, 114, 105006.
3. Li, F., Ma, L., & Cai, J. (2018). Multi-discriminator generative adversarial network for high resolution gray-scale satellite image colorization. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 3489-3492). IEEE.
4. Wu, M., Jin, X., Jiang, Q., Lee, S. J., Guo, L., Di, Y., ... & Huang, J. (2019). Remote sensing image colorization based on multiscale SEnet GAN. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-6). IEEE
5. Ji, G., Wang, Z., Zhou, L., Xia, Y., Zhong, S., & Gong, S. (2020). SAR image colorization using multidomain cycle-consistency generative adversarial network. IEEE Geoscience and Remote Sensing Letters, 18(2), 296-300
6. Wu, M., Jin, X., Jiang, Q., Lee, S. J., Liang, W., Lin, G., & Yao, S. (2021). Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space. The Visual Computer, 37, 1707-1729.
7. Wang, Y., Yu, J., & Zhang, J. (2022). Zero-shot image restoration using denoising diffusion null-space model. arXiv preprint arXiv:2212.00490
8. Anwar, Saeed, Muhammad Tahir, Chongyi Li, Ajmal Mian, Fahad Shahbaz Khan, and Abdul Wahab Muzaffar. "Image colorization: A survey and dataset." Information Fusion 114 (2025): 102720.
9. Bose, R., Banerjee, A. S. B., & Chaudhuri, S. (2022). DARK: Few-shot remote sensing colorization using label conditioned color injection. IEEE Geoscience and Remote Sensing Lett
10. Fu, Q., Xia, S., Kang, Y., Sun, M., & Tan, K. (2024). Satellite Remote Sensing Grayscale Image Colorization Based on Denoising Generative Adversarial Network. Remote Sensing, 16(19), 3644
11. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255)
12. Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision* (pp. 3730-3738)
13. Wang, J., Nie, J., Chen, H., Xie, H., Zheng, C., Ye, M., & Wei, Z. (2022). Remote sensing image colorization based on joint stream deep convolutional generative adversarial networks. In *Proceedings of the 4th ACM International Conference on Multimedia in Asia* (pp. 1-8)
14. Ertan, Z., Korkut, B., Gördük, G., Kulavuz, B., Bakırman, T., & Bayram, B. (2024). Enhancement of underwater images with artificial intelligence. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 149-156
15. Schwab, J., Antholzer, S., & Haltmeier, M. (2019). Deep null space learning for inverse problems: convergence analysis and rates. Inverse Problems, 35(2),
16. Shermeyer, J., Hogan, D., Brown, J., Van Etten, A., Weir, N., Pacifici, F., ... & Lewis, R. (2020). SpaceNet 6: Multi-sensor all weather mapping dataset. In Proceedings of

the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 196-197)