

Deep Wiener Deconvolution Denoising Sparse Autoencoder Model for Pre-processing High-resolution Satellite Images

S. Kiruthika^{1,*}, G. Maria Priscilla², Anna Saro Vijendran³, M. Batumalay⁴, Zhengrui Xu⁵

^{1,2,3}*Sri Ramakrishna College of Arts and Science, Coimbatore and 641006, India*

⁴*Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia*

⁵*Faculty of Liberal Arts, Shinawatra University (SIU) Pathum Thani 12160 Thailand*

(Received: June 22, 2024; Revised: July 12, 2024; Accepted: August 28, 2024; Available online: September 17, 2024)

Abstract

The detection of geospatial objects in surveillance applications faces significant challenges due to the misclassification of object boundaries in noisy and blurry satellite images, which complicates the detection model's computational complexity, uncertainty, and bias. To address these issues and improve object detection accuracy, this paper introduces the Deep Wiener Deconvolution Denoising Sparse Autoencoder (DWDDSAE) model, a novel hybrid approach that integrates deep learning with Wiener deconvolution and Denoising Sparse Autoencoder (DSAE) techniques. The DWDDSAE model enhances image quality by extracting deep features and mitigating adversarial noise, ultimately leading to improved detection outcomes. Evaluations conducted on the NWPU VHR-10 and DOTA datasets demonstrate the effectiveness of the DWDDSAE model, achieving notable performance metrics: 96.32% accuracy, 86.88 edge similarity, 75.47 BRISQUE, 28.05 IQI, 38.08 PSNR (dB), 0.883 SSIM, 98.25 MSE, and 0.099 RMSE. The proposed model outperforms existing methods, offering superior noise and blur removal capabilities and contributing to Sustainable Development Goals (SDGs) such as SDG 9 (Industry, Innovation, and Infrastructure), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action). This research highlights the model's potential for inclusive innovation in object detection applications, showcasing its contributions and novel approach to addressing existing limitations.

Keywords: Geospatial Object Detection, Deblurring, Satellite High-Resolution Images, Deep Wiener Deconvolution, Denoising Sparse Autoencoder Climate

1. Introduction

Detecting geospatial multi-scale objects from high-resolution (HR) and very high-resolution (VHR) aerial and satellite images is a critical and challenging task. Accurate detection is essential for various monitoring applications, where determining both the location and the category of objects at different scales ensures reliability [1]. This task becomes significantly more complex due to the presence of noise and blur in the images, which can arise from various factors, including sensor limitations, atmospheric conditions, and disturbances during data capture. Aerial images often suffer from high noise levels due to these disturbances [2]. Factors such as calibration errors, sensor noise characteristics, random fluctuations in pixel values, as well as motion of the camera or the object, can further introduce imperfections in the data. Additionally, atmospheric conditions like haze, fog, and scattering contribute to image noise and blur, severely affecting the clarity and quality of the captured images [3].

In geospatial object detection, image pre-processing—specifically denoising and deblurring—is a crucial step to restore the clarity of the images and improve detection accuracy [4]. However, this process is highly complex due to the varying scales and types of noise involved. Despite the challenges, numerous methods, including Deep Learning and Machine Learning techniques, have been developed to address the issue of noise and blur in satellite images [5]. While these methods have shown some success, there is still a need for more effective approaches that can handle the diverse nature of geospatial data [6].

The objective of this paper is to propose a novel approach for addressing noise and blur in aerial images using a hybrid model that integrates advanced image processing techniques. The proposed model, named Deep Wiener Deconvolution Denoising Sparse Autoencoder (DWDDSAE), combines the strengths of deep learning, Wiener Deconvolution, and

*Corresponding author: S.Kiruthika (kiruthika@srcas.ac.in)

DOI: <https://doi.org/10.47738/jads.v5i3.357>

This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights

Denoising Sparse Autoencoder methods. In addition, Black Hat Transformation and Gaussian Filtering techniques are employed to further enhance image quality. Black Hat Transformation is used to enhance dark objects in bright backgrounds, while the Gaussian Adaptive Bilateral Filter smoothens the image and preserves edge information. These pre-processing steps significantly improve image clarity, making it easier to detect geospatial objects [7].

The DWDDSAE model applies deep-learned features in the framework of Classical Wiener Deconvolution, refining the features to predict deblurred images and enhance the small-scale clean image features [8]. The Denoising Sparse Autoencoder (DSAE) combines Denoising Autoencoder (AE) and Sparse AE to learn features from both noisy and clean images, effectively reducing noise while improving generalization [9]. This hybrid model is designed to provide a more robust and accurate solution for denoising and deblurring satellite images.

The primary contribution of this paper is the development of the DWDDSAE model, which integrates multiple advanced techniques to address the challenges of noise and blur in geospatial images [10]. The effectiveness of this model is evaluated using two benchmark datasets—NWPU VHR-10 and DOTA—under both real-time and controlled conditions [11], [12]. The structure of this paper is as follows: Section 2 provides a survey of related work, Section 3 explains the proposed DWDDSAE-based image pre-processing model, Section 4 discusses the experimental results, and Section 5 concludes with a summary and suggestions for future research directions [13], [14], [15], [16], [17].

2. Literature Review

Many machine learning (ML) and deep learning (DL) methods have been predominantly employed for denoising and deblurring in the pre-processing stage of geospatial object detection. Asokan and Anitha [1] proposed an optimized bilateral filter using Adaptive Cuckoo Search (ACS), Cuckoo Search (CS), and Particle Swarm Optimization (PSO) for denoising satellite images while preserving edges. The model, tested on Landsat satellite images, demonstrated that the ACS-optimized bilateral filter achieved the best Peak Signal-to-Noise Ratio (PSNR), mean square error (MSE), feature similarity index, and entropy, while reducing computational complexity by 43% in CPU time.

Zhu et al. [2] presented a blind deblurring method based on a Feature Alignment Module (FAFM) and a Feature Importance Selection Module (FISM) within a Generative Adversarial Network (GAN) framework. Tested on both synthetic remote sensing datasets and real satellite images, this approach demonstrated enhanced texture and detail recovery. Thai et al. [3] proposed a hybrid Riesz-Quincunx UNet variational autoencoder (RQUNet-VAE) to denoise satellite imagery, showing superior noise reduction in image processing tasks.

Li et al. [4] introduced a coupled segmentation and denoising/deblurring model for hyperspectral material identification, which showed significant improvements in denoising and segmentation performance on remote sensing data. Fang et al. [5] developed a multitasking framework for motion deblurring in remote sensing images, which includes an image restoration branch using a U-Net architecture for spatial detail preservation and a multilayer perception network for texture complexity recognition. The model achieved PSNR values of 27.7761 dB in the SydneyBlurred dataset, 27.5526 dB in UCMBlurred, and 27.0272 dB in AIDBlurred, along with strong Structural Similarity Index (SSIM) scores.

Han et al. [6] introduced a non-blind deblurring method to address inaccurately estimated blur kernels. Their model used joint sparsity regularization and an iterative algorithm alternating between latent image estimation and bias correction, effectively managing kernel errors in benchmark and real blurred remote sensing datasets. Song et al. [7] proposed a Wavelet Sub-band Cycle-Consistent Adversarial Network (WavCycleGAN) to remove vertical stripe and wave noise from multispectral satellite images. The model, leveraging wavelet sub-bands for noise direction learning, produced effective noise removal with an average PSNR of 52.32 dB while preserving edges and detailed information.

Zhang et al. [8] presented a joint denoising and deblurring approach targeting blur and stripe noise using residual stripe components and gradient properties. Tested on real blurred and noisy images from orbiting satellite sensors, the method outperformed existing techniques in both quantitative metrics and visual quality. However, the model faced limitations in modeling multiple degradations for real-world imagery across various modalities.

Burdziakowski [10] applied Conditional Generative Adversarial Networks (GANs) to deblur UAV images, improving geometric accuracy and interpretability, particularly in terms of texture and edge sharpness, even for large blur kernels.

However, the model had difficulties with some ground control points and photogrammetric targets that were not properly deblurred.

Jing et al. [12] introduced a Denoising Diffusion Probabilistic Model-Cloud Removal (DDPM-CR) method to remove noise and clouds from optical remote sensing imagery. By utilizing multispectral optical images from Copernicus Sentinel-2 and SAR images from Sentinel-1, their model achieved superior performance, with a mean absolute error (MAE) of 0.0229, root mean square error (RMSE) of 0.0268, PSNR of 31.7712, and SSIM of 0.9033. However, it exhibited some limitations in handling lower resolution images with thick clouds.

Zeng [14] proposed an evolved processing model for high-precision remote sensing images by employing fourth-order differential equations (FD) and Dual-Tree Complex Wavelet Transform (DTCWT) to decompose multi-scale images into low- and high-frequency sub-bands. The FD-DTCWT model achieved a normalized RMS error of 0.02 and a noise variance of 0.03, with structure similarity of 0.74, PSNR of 25.3 dB, and normalized SNR of 0.76. Nevertheless, the model's higher computational complexity posed challenges in shortening denoising time.

Wang et al. [15] proposed a Self-Supervised Hyper Spectral Image Denoising Network (SHDN) that utilized a noise estimator to handle complex real-world noise. Their model outperformed competing methods on airborne, UAV, and ground-based hyperspectral image datasets. Yang et al. [16] developed a super-resolution and deblurring network (SRDN) based on an end-to-end GAN to improve low-quality space target images, which often suffer from poor reliability in manual classification. Experiments on the BUAA-SID-share1.5 satellite dataset showed that SRDN outperformed state-of-the-art deblurring techniques, although it exhibited instability during GAN training.

3. Research Method

The proposed DWDDSAE-based pre-processing model comprises three processes: Black Hat Transformation, Gaussian Adaptive Bilateral filtering and DWDDSAE-based denoising and deblurring. Figure 1 demonstrates the proposed pre-processing model.

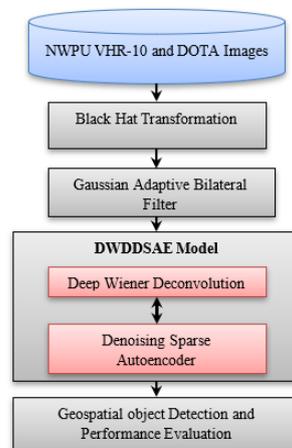


Figure 1. Flow process of Proposed Model

3.1. Black Hat Transformation

Black Hat Transformation is a morphological operation used in image processing to enhance dark objects on a light background [18]. The Black Hat Transformation highlights the features and details those brighter surroundings may overshadow. The input image is subtracted from its closing to attain the Black Hat Transformation. Mathematically, if C is the input image and L is its closing, then the Black Hat transformation H is given by:

$$C = L - H \tag{1}$$

The result shows the highlights of dark structures, objects, or details against a lighter background.

3.2. Gaussian Adaptive Bilateral Filtering

Gaussian Adaptive Bilateral Filtering extends the traditional bilateral filter, incorporating adaptive spatial and range weights based on local image statistics. The approach allows the filter to preserve edge and details with intensity variations and varying noise. The bilateral filter balances smoothing and preserving the image considering the spatial proximity of pixels and intensity. The disadvantage of the standard bilateral filter is that it doesn't have fixed spatial and range parameters, which can't be optimal for the images in all the regions. The Gaussian Adaptive Bilateral, locally based on the characteristics of the neighborhood, adjusts the filter of the spatial and range parameters. This adaptation can be achieved by computing the mean and standard deviation to maintain the information and edges along with the filter parameters of the image [19].

3.3. DWDDSAE-based Model

DWDDSAE combines two models: Deep Wiener Deconvolution (DWD) and DSAE. To obtain optimal performance, these two models are integrated. Figure 2 illustrates the proposed DWDDSAE model architecture. DWD is a Neural Network-based model designed for the restoration of image tasks, specifically utilizing principles of Wiener deconvolution to enhance its performance on complex image restoration problems by adding the power of deep learning.

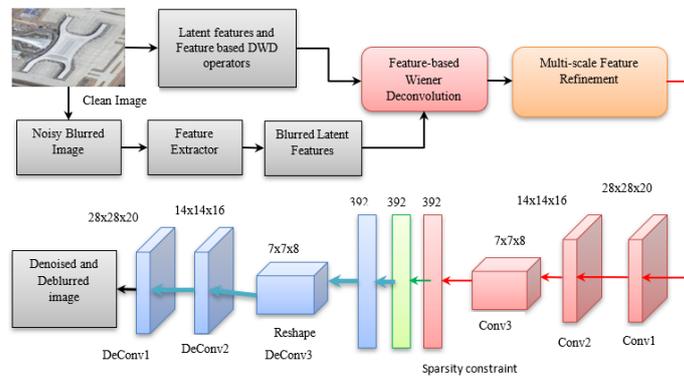


Figure 2. Architecture of Proposed DWDDSAE Model

Deep Wiener deconvolution is a signal processing technique used to recover an original signal from a degraded or blurred version of that signal. The DWD proposed model network consists of multi-scale feature refinement based on a feature deconvolution model [20]. To address the challenge of image deblurring, a DWD network is tailored to incorporate a feature-based Wiener deconvolution (FBWD) module to enhance the deblurring process for blurry input images by leveraging information based on the extracted feature.

$$F_i y = K F_i x + F_i n, \quad \forall i = 1, \dots, M \quad (2)$$

Here, $F_i y, n, k, y, x$ are the forms of vector of $F_i y, n, k, y, x$.

Let $\{f_i\}_{i=1}^M$ is the linear filter set utilized to get the information on the features from the unclear image. The FBWD explicitly unravels the features. $F_i y$ present in the blurry input. This is achieved through the identification of a set of operators tailored for feature-based Wiener deconvolution, allowing us to recover the features as:

$$F_i \hat{x} = G_i F_i y, \quad \forall i = 1, \dots, M \quad (3)$$

Here, \hat{x} denoted as clear image latent. To get the latent features, the mean squared error is minimized.

$$e_i = E(|F_i x - F_i \hat{x}|^2) = E(|F_i x - G_i F_i y|^2) = E(|F_i x - G_i (K F_i x + F_i n)|^2) \quad (4a)$$

$$e_i = (1 - G_i K)(G_i K)^T E(|F_i x|^2) - (1 - G_i K)G_i^T E(F_i x (F_i n)^T) - (1 - G_i K)^T G_i E((F_i x)^T F_i n) + G_i G_i^T E(|F_i n|^2) \quad (4b)$$

Here, Expectation is denoted by E . Assuming that the noise is independent of the clear image and having zero mean, derived as:

$$E(F_i x (F_i n)^T) = E(F_i x) E((F_i n)^T) = 0 \text{ and } E((F_i x)^T F_i n) = E((F_i n)^T) E(F_i x) = 0 \quad (5)$$

Here $E(|F_i x|^2)$ and $E(|F_i n|^2)$ are rewritten as s_i^n and s_i^x in Eq. (4b)

$$e_i = (1 - G_i K)(1 - G_i K)^T s_i^x + G_i G_i^T s_i^n \quad (6)$$

To decrease e_i , estimate the derivation of Eq. (6) with res. to G_i and specify the value to 0:

$$(K^T K s_i^x + s_i^n) G_i - K^T s_i^x = 0 \quad (7)$$

The FBWD operator G_i ,

$$G_i = \frac{K^T}{K^T K + s_i^n / s_i^x} = F^{-1} \left(\frac{\overline{F(K)}}{\overline{F(K)} F(K) + \frac{s_i^n}{s_i^x}} \right) \quad (8)$$

Here, the Fourier transform is implied as F and $\overline{F(K)}$ is the fuse of $F(K)$. The information of the feature is extracted from the unclear images using these functions, and the SD and variance of the unclear feature are assessed to approximately reconstruct the pixels by Wiener Deconvolution. Then, these output images are fed as input to the DSAE module.

The proposed DSAE is a devised feed-forward unsupervised neural network algorithm used to train and learn a compacted depiction of an input $f(x)$. The standard auto encoders (AE) have encoder and decoder parts. The encoder incorporates decreasing hidden layers, which uses weight and bias to encode the data, and the decoder attempts to reconstruct the actual data with the increasing hidden layers. This type of AE uses multiple layers, requiring more parameters to tune during the training process. The input data of DSAE is taken as $x = [x_1, x_2, \dots, x_m]$ and it transforms the input using the activation function as the hidden feature vector $h = [h_1, h_2, \dots, h_p]$. The output points or the reconstruction vector is represented as $z = [z_1, z_2, \dots, z_m]$. The computation of hidden and output vectors are illustrated below,

$$h = S_g(Wx + b) \quad (9)$$

$$z = S_f(W'y + b') \quad (10)$$

Where, S_g denotes the hidden tier function for activation and S_f is the output layer's activation function. The sigmoid and ReLU are generally used as the activation function. W and W' denotes weights and b' , b are the biases. The cost function of the standard AE is specified as follows,

$$C_1 = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{r}_j} + (1 - r) \log \frac{1 - r}{1 - \hat{r}_j} \right) \quad (11)$$

The proposed DSAE method uses ReLU activation in the hidden layer because of its better performance in the non-stationary data. Next, the non-linear transformation of the output layer is set as \tanh function, and it is expressed as

$$z_i = \tanh \left(\sum_{j=1}^p W_{ij} h_j \right) \quad (12)$$

Further, the weight decay term λ is added to Eq. (11) to limit the overfitting issue, and it is expressed as,

$$C^T = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \frac{\lambda}{2} \sum_{i,k=1}^m \sum_{j=1}^p ((W_{ij})^2 + (W_{jk})^2) + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{r}_j} + (1 - r) \log \frac{1-r}{1-\hat{r}_j} \right) \quad (13)$$

The decay strategy introduces many connecting weights and reduces the sparsity. The quality of reconstruction and sparsity is enhanced by curtailing the negative weights with the assistance of non-negative constraints. The cost function with non-negative constraint is expressed as follows,

$$C^E = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \frac{\delta}{2} \sum_{L=1}^2 \sum_{l=1}^{S_L} \sum_{j=1}^{S_{L+1}} G(W_{jl}^{(L)}) + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{f}_j} + (1-r) \log \frac{1-r}{1-\hat{f}_j} \right) \quad (14)$$

$$\text{Where, } G(W_{jl}^{(L)}) = \begin{cases} (W_{jl}^{(L)})^2 & W_{jl}^{(L)} < 0 \\ 0 & W_{jl}^{(L)} \geq 0 \end{cases}$$

In Eq. (13), the second term represents the non-negative constraint, δ indicates the penalty coefficient, C^E denotes the enhanced cost function, and L th layer node dimension is represented as S_L . To maintain the C^E value as a minimum, the training phases have to adjust the weight $W_{jl}^{(L)}$. The weights are updated utilizing gradient descent with back propagation. It is represented as below,

$$W_{jl}^{(L)} = W_{jl}^{(L)} - \eta \frac{\partial C^M}{\partial W_{jl}^{(L)}} \quad L = 1,2 \quad (15)$$

Here, η denotes the learning rate and $W_{jl}^{(1)} = W_{jk}$, $W_{jl}^{(2)} = W_{ij}$, $\frac{\partial C^M}{\partial W_{jl}^{(L)}} = \frac{\partial C_1}{\partial W_{jl}^{(L)}} + \delta g(W_{jl}^{(L)})$,

$$g(W_{jl}^{(L)}) = \begin{cases} (W_{jl}^{(L)})^2 & W_{jl}^{(L)} < 0 \\ 0 & W_{jl}^{(L)} \geq 0 \end{cases}$$

Now, the DSAE is trained with clean images as input and the corresponding noisy images as target output. During training, the autoencoder learns to encode a sparse representation that captures the essential features of the clean images. After training, the autoencoder can be evaluated on noisy test images. The encoder part of the model is used to obtain a sparse representation of the noisy input, and the decoder reconstructs the denoised image. These images can be fed to the classifier model to detect the geospatial objects more accurately.

3.4. Algorithm: DWDDSAE-based model

The proposed algorithm for denoising and deblurring satellite images begins by taking raw satellite images as input, with the desired output being clean, denoised, and deblurred images. The process initiates by applying a Black Hat transformation to enhance the image contrast, followed by Gaussian Adaptive Bilateral filtering to smoothen the image while preserving edges. For each image, latent features are extracted, and Directional Wavelet Decomposition (DWD) operators are set. The mean squared error (MSE) is then estimated, and noisy, blurred latent features are identified. Subsequently, a Forward Backward Wavelet Decomposition (FBWD) is applied, and the Fourier transform of the FBWD operators is estimated. Using DWD, the pixel values are reconstructed to restore image quality. Figure 3 illustrates the step-by-step process of the proposed DWDDSAE-based model for image denoising and deblurring. This diagram outlines the sequence of operations, including image preprocessing, feature extraction, transformation, and reconstruction, highlighting the key stages involved in the model's workflow.

1	Input: Raw Satellite Images	14	Input the DWD output to the DSAE encoder
2	Output: Denoised and Deblurred Images	15	For each input
3	Initialize the Input image C	16	Transform the input into a hidden feature vector h
4	Applying Black Hat Transformation	17	Compute the hidden and output vectors
5	Applying Gaussian Adaptive Bilateral filtering	18	Estimate the cost function of the DSAE
6	For each image	19	Apply non-linear transformation of the output layer
7	Extract latent features and set DWD operators	20	Perform sparse representation
8	Estimate the mean squared error	21	Reshape the pixels
9	Extract noisy, blurred latent features	22	Apply the denoising process
10	Applying FBWD	23	Estimate cost function with a non-negative constraint
11	Estimating the Fourier transform of FBWD operators	24	End for
12	Reconstruct the pixels by DWD	25	Reconstruct the noisy pixels in the decoder
13	End for		

Figure 3. Algorithm steps of the proposed DWDDSAE-based model.

The output from this stage is fed into the Deep Sparse Autoencoder (DSAE) encoder. For each input image, it is transformed into a hidden feature vector h , and hidden and output vectors are computed. The DSAE cost function is estimated, followed by the application of non-linear transformations at the output layer. A sparse representation of the image is performed, reshaping the pixels for further refinement. The denoising process is applied at this stage, where the cost function is optimized with a non-negative constraint to ensure accurate pixel reconstruction. Finally, the decoder reconstructs the noisy pixels into their original form, producing a clean and enhanced satellite image.

4. Results and Discussion

The DWDDSAE-based model is implemented using the PyCharm tool and written in Python. The evaluations are performed using NWPU VHR-10 and DOTA datasets. This dataset has 757 airplanes, 302 steamships, 655 tanks for storage, 390 diamonds, 524 tennis courts, 159 basketball courts, 163 ground field tracks, 224 ports, 124 bridges and 477 transportations. DOTA dataset (v.2.0) has 18 classes, 11,268 images and 1,793,658 objects. The sample input and output images obtained in this study for both datasets are shown in figure 4 and figure 5.



Figure 4. Sample Outputs of NWPU VHR-10 Dataset



Figure 5. Sample Outputs of DOTA Dataset

The proposed DWDDSAE-based model is evaluated in terms of accuracy, edge similarity, Blind/Reference less Image Spatial Quality Evaluator (BRISQUE), IQI, PSNR, Structural Similarity Index (SSIM), MSE, RMSE and Execution time per image.

Accuracy is computed between the true denoised pixels and the noisy pixels. It is also called pixel accuracy.

$$\text{Accuracy} = \frac{\text{Number of true pixels}}{\text{Number of all pixels}} \quad (16)$$

Edge similarity is often used to compare edges detected in images based on the Euclidean distance between edge features. If $E1$ and $E2$ are two edges with their respective feature vectors as $F1$ and $F2$, the feature vectors might include properties such as edge strength, orientation, and location.

$$\text{Edge Similarity} = \sqrt{\sum_{i=1}^n (F1_i - F2_i)^2} \quad (17)$$

BRISQUE is a no-reference image quality assessment process, which means it can evaluate the quality of an image without requiring a reference image for comparison. It operates by extracting a set of spatial domain features from the input image, which capture various statistics and properties of the image and is given by

$$\text{BRISQUE} = \sum_{i=1}^N w_i \times f_i \quad (18)$$

Here, N denotes the number of features, w_i denotes the weight of the i-th feature and f_i denotes the value of the i-th feature extracted from the image.

IQI is a metric used to assess the quality of an image, typically by comparing it to a reference image or by analyzing its characteristics. It is computed by

$$\text{IQI} = \frac{\text{Quality of Processed Image}}{\text{Quality of Original Image}} \times 100\% \quad (19)$$

PSNR measures the ratio between the maximum possible power of an image and the power of corrupting noise that affects the fidelity of its representation.

$$\text{PSNR} = 10 \times \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (20)$$

Here, MAX is the maximum possible pixel value of the image, and MSE is the mean square error.

SSIM is a metric used to quantify the similarity between two images, typically an original image and a processed or reconstructed image. It combines local image structure, luminance, and contrast into a single local quality score. In this metric, structures are patterns of pixel intensities, especially among neighboring pixels, after normalizing for luminance and contrast. For two images, x and y, SSIM is given by

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (21)$$

Here, μ_x, μ_y denotes the mean values of x and y, respectively, σ_x^2, σ_y^2 denotes the variances of x and y, respectively, σ_{xy} denotes the covariance of x and y, and c_1, c_2 denotes the stabilizing constants.

MSE is used to measure the difference between the pixel values of two images, such as an original image and a reconstructed image. It is given by

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (I_i - K_i)^2 \quad (22)$$

Here, N denotes the number of image pixels, I_i denotes the pixel value of the original image at position i, and K_i denotes the pixel value of the reconstructed image at position i.

RMSE is used to quantify the accuracy of a predictive model or the difference between two sets of images. RMSE is computed as the square root value of the MSE.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_i - K_i)^2} \quad (23)$$

Execution time per image is the time taken by the proposed model to estimate and remove the noise and blur in an image. The results are obtained by varying the noise and blur levels consistently. Table 1 presents the performance results of the proposed DWDDSAE-based model on two datasets: NWPU VHR-10 and DOTA. The table summarizes various evaluation metrics across different levels of noise and blur.

For the NWPU VHR-10 dataset, the accuracy, edge similarity, BRISQUE, IQI, PSNR, SSIM, MSE, and RMSE are reported for noise/blur levels of 0%, 10%, 20%, and 30%. The results show that as the noise/blur level increases, the accuracy, edge similarity, and PSNR tend to decrease, while MSE and RMSE generally increase. Despite these changes, the model maintains relatively high performance, with accuracy ranging from 92.49% to 94.32% and PSNR values from 30.57 dB to 36.55 dB. For the DOTA dataset, similar metrics are evaluated. The model performs exceptionally well, with accuracy values ranging from 98.32% to 98.53% across the different noise/blur levels. Edge similarity and PSNR values are high, particularly at 0% and 10% noise/blur levels. However, as the noise/blur level increases to 30%, there is a notable decline in edge similarity and PSNR, with corresponding increases in MSE and RMSE.

Table 1. Performance results for nwpu vhr-10 and dota datasets

Dataset	Noise/ Blur level (%)	Accuracy (%)	Edge Similarity	BRISQUE	IQI	PSNR (dB)	SSIM	MSE	RMSE	Execution Time (s)
NWPU VHR-10	0	94.32	60.63	75.46	28.06	36.55	0.882	96.57	0.0982	17.09
	10	92.49	60.69	70.13	27.97	34.87	0.84	98.84	0.0994	16.9998
	20	93.006	54.98	71.59	27.99	31.80	0.754	99.41	0.0997	17.0241
	30	93.21	54.64	74.49	27.96	30.57	0.713	99.84	0.0999	16.9707
DOTA	0	98.32	99.91	75.48	28.11	39.61	0.884	99.93	0.0999	17.0718
	10	98.47	99.91	72.81	28.09	32.97	0.642	100.4	0.1001	17.0389
	20	98.53	98.64	71.32	28.06	31.18	0.535	101.5	0.1007	17.2583
	30	98.45	85.20	66.64	28.01	29.91	0.430	103.4	0.1016	17.0938

Table 2 presents a comparative analysis of the average performance results of various image denoising and deblurring methods, including the proposed DWDDSAE-based model, evaluated on the NWPU VHR-10 and DOTA datasets. The table shows that the proposed DWDDSAE-based model outperforms its counterparts in several key metrics. Specifically, it achieves the highest accuracy of 96.32%, reflecting its superior ability to correctly identify and process image details.

Table 2. Average comparison results for nwpu vhr-10 and dota datasets

Methods	Accuracy (%)	Edge Similarity	BRISQUE	IQI	PSNR (dB)	SSIM	MSE	RMSE	Execution Time (s)
ACS bilateral filter [1]	91.67	56.54	69.59	26.40	31.31	0.7865	102.6	0.179	18.23
FAFM-FISM-GAN [2]	90.87	78.80	72.87	27.11	35.33	0.828	101.8	0.189	19.99
RQUNet-VAE [3]	94.44	68.32	59.82	26.90	33.45	0.814	100.6	0.101	23.48
U-Net [5]	94.47	71.23	61.23	25.50	31.18	0.871	99.99	0.322	18.33
WavCycleGAN [7]	93.5	75.45	67.75	25.54	30.70	0.865	103.4	0.345	20.87
Joint Model [8]	93.35	80.65	66.67	24.93	32.54	0.855	103.9	0.114	21.40
Conditional GAN [10]	92.98	82.42	66.66	22.91	33.57	0.836	99.45	0.234	19.67
DDPM-CR [12]	91.19	79.91	68.84	24.22	36.67	0.876	101.1	0.201	20.03
FD-DTCWT [14]	93.56	80.32	71.71	25.98	35.04	0.844	98.89	0.115	19.41
SHDN [15]	90.06	77.43	72.65	23.34	33.28	0.835	99.76	0.278	17.94

SRDN [16]	93.23	75.54	65.43	27.67	29.89	0.837	100.3	0.154	18.66
Proposed DWDDSAE	96.32	86.88	75.47	28.05	38.08	0.883	98.25	0.099	17.12

Additionally, the DWDDSAE model excels in edge similarity, with a score of 86.88, indicating its effective preservation of image edges. The model also demonstrates the highest BRISQUE score of 75.47, underscoring its capability to maintain high image quality. In terms of IQI, the DWDDSAE model records a notable score of 28.05, further suggesting improved image integrity. Its PSNR of 38.08 dB is the highest among the methods tested, signifying exceptional performance in denoising and deblurring. The model's SSIM score of 0.883 reflects superior structural similarity between the processed and original images. Moreover, it exhibits lower MSE (0.099) and RMSE (17.12), indicating reduced reconstruction errors. Despite these strong results, the DWDDSAE model maintains a competitive execution time of 17.12 seconds, making it both effective and efficient. Overall, the proposed DWDDSAE-based model demonstrates notable advancements in performance across most evaluation metrics, proving its effectiveness in enhancing image quality compared to other existing methods.

The proposed denoising and deblurring method of DWDDSAE demonstrated superior performance across a comprehensive set of evaluation metrics for the NWPU VHR-10 and DOTA datasets. The results of the DWDDSAE model are also equated against models from the literature for a broader comparison. Table II demonstrates the comparison results of the DWDDSAE against the literature methods at a 0% noise level.

The results in figure 6 show that the proposed DWDDSAE has a higher accuracy of 96.32%, edge similarity of 86.88, BRISQUE of 75.47 and reduced MSE of 98.25. The enhanced data pre-processing and effective handling of the contextual information have resulted in these improvements. Effective pre-processing with denoising and filtering techniques likely reduced noise and artifacts, leading to cleaner images and lower MSE values.

The results in figure 7 show that the proposed DWDDSAE has a higher IQI of 28.05 and PSNR of 38.08. The improvement in IQI values suggests that the DWDDSAE model has been optimized to better preserve image quality during operations such as enhancement, or restoration. The increase in PSNR values indicates a reduction in noise and distortion in the processed images compared to their original image. The improvement in IQI and PSNR values reflects a higher degree of fidelity and similarity between the processed and original images.

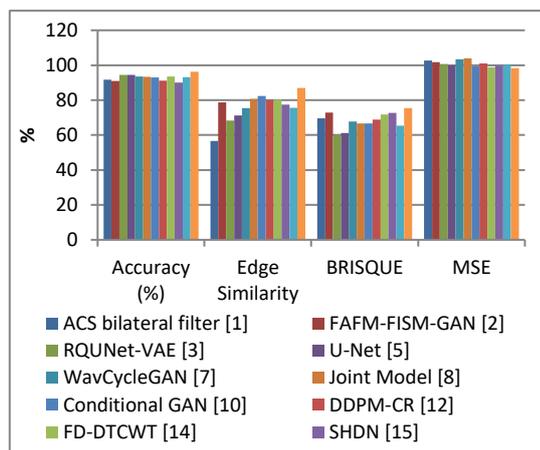


Figure 6. Accuracy, Edge Similarity, BRISQUE and MSE

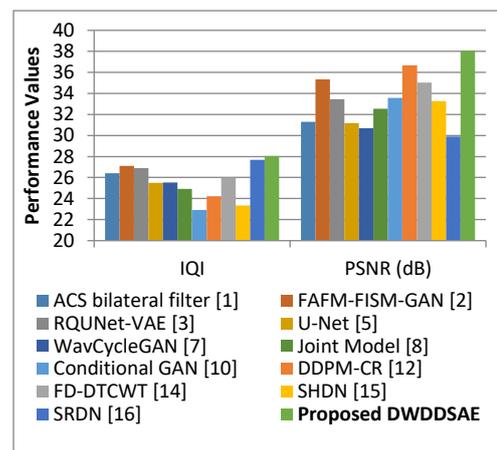


Figure 7. IQI and PSNR

The results in figure 8 show that the proposed DWDDSAE has a higher SSIM of 0.883 and a reduced RMSE of 0.099. The increase in SSIM and decrease in RMSE indicates that the processed or reconstructed image is closer in pixel values to the original image. Effective noise reduction techniques applied during image processing can help minimize pixel-wise discrepancies and improve the overall accuracy of reconstructed images. Figure 9 show the execution time per image is less than the existing methods. Efficient preprocessing of input data can streamline subsequent processing

steps and reduce computation time. This involves implementing more efficient processing and reducing computational complexity.

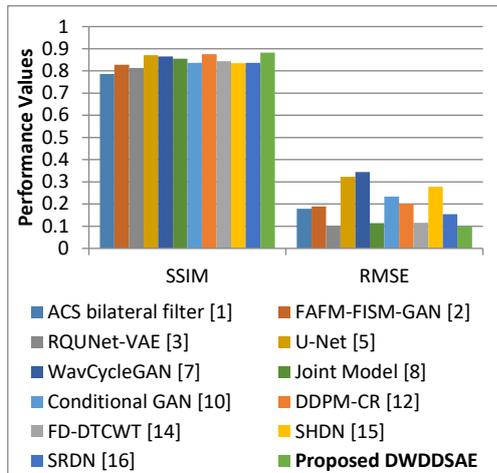


Figure 8. SSIM and RMSE

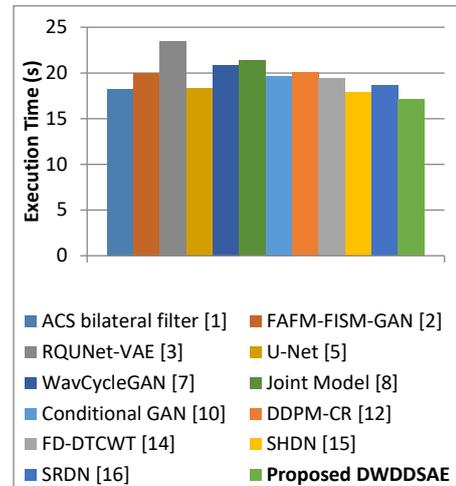


Figure 9. Execution time

Thus, the proposed DWDDSAE model showed a superior performance on the set of evaluation metrics compared to the other existing models. For preserving the images' critical details and sharp edges, the DWDDSAE model exhibited remarkable accuracy in denoising and deblurring tasks. The method effectively mitigated noise and blur artifacts, improving visual fidelity. The Deep Wiener Deconvolution Denoising Sparse Autoencoder demonstrated efficient execution times per image despite its advanced capabilities. This ensures that improvements are observed not only in pixel-wise fidelity but also in terms of perceived image quality.

5. Conclusions

This paper addresses the challenges associated with detecting geospatial objects in surveillance applications, particularly focusing on issues like misclassification of object boundaries due to noise and blur in satellite images. Such complexities can significantly hinder detection accuracy, leading to increased uncertainty and bias in the detection model. To overcome these challenges, the DWDDSAE model has been proposed, which integrates deep learning techniques, specifically Wiener Deconvolution and Denoising Sparse Autoencoder (DSAE) methods.

The performance of the DWDDSAE model was evaluated using the NWPU VHR-10 and DOTA datasets. The results demonstrate that the DWDDSAE model outperforms existing methods in noise and blur removal, as evidenced by its high accuracy of 96.32% and edge similarity score of 86.88% on the average comparison results. The model achieved notable improvements in PSNR (38.08 dB) and SSIM (0.883), while maintaining a low MSE of 0.099 and RMSE of 17.12 seconds, compared to other state-of-the-art techniques.

Despite these promising results, further research is needed to fine-tune the model for diverse datasets and those with unique characteristics to enhance its generalizability. Additionally, environmental conditions such as weather and seasonal changes can impact satellite imagery quality, which may affect model performance. Future work should focus on improving the robustness of the DWDDSAE model to ensure reliable performance across a broader range of operational scenarios.

6. Declarations

6.1. Author Contributions

Conceptualization: S.K., G.M.P., A.S.V., M.B., and Z.X.; Methodology: G.M.P. and M.B.; Software: S.K. and Z.X.; Validation: S.K., G.M.P., and A.S.V.; Formal Analysis: S.K., G.M.P., and Z.X.; Investigation: A.S.V. and S.K.; Resources: G.M.P. and M.B.; Data Curation: Z.X. and G.M.P.; Writing Original Draft Preparation: S.K., G.M.P., and A.S.V.; Writing Review and Editing: M.B., A.S.V., and G.M.P.; Visualization: S.K. and Z.X.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] A. Asokan and J. Anitha, "Adaptive Cuckoo Search based optimal bilateral filtering for denoising of satellite images," *ISA Trans.*, vol. 97, no. 1, pp. 237–246, Nov. 2019, doi: 10.1016/j.isatra.2019.11.008
- [2] B. Zhu, Q. Lv, Y. Yang, X. Su, Y. Zhang, Y. Tang and Z. Tan, "Blind Deblurring of Remote-Sensing Single Images Based on Feature Alignment," *Sensors*, vol. 22, no. 20, pp. 7894, Oct. 2022, doi: 10.3390/s22207894
- [3] D. H. Thai, X. Fei, M. T. Le, A. Züfle and K. Wessels, "Riesz-Quincunx-UNet Variational Autoencoder for Unsupervised Satellite Image Denoising," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 1, pp. 1-19, 2023, Art no. 5404519, doi: 10.1109/TGRS.2023.3291309. doi: 10.1109/TGRS.2023.3291309
- [4] F. Li, M. K. Ng, and R. J. Plemmons, "Coupled segmentation and denoising/deblurring models for hyperspectral material identification," *Numer. Linear Algebra Appl.*, vol. 19, no. 1, pp. 153–173, Sep. 2010, doi: 10.1002/nla.750
- [5] J. Fang, X. Cao, D. Wang, and S. Xu, "Multitask Learning Mechanism for Remote Sensing Image Motion Deblurring," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, no. 1, pp. 2184–2193, Jan. 2021, doi: 10.1109/JSTARS.2020.3047636
- [6] J. Han, S. Zhang, and Z. Ye, "A Nonblind Deconvolution Method by Bias Correction for Inaccurate Blur Kernel Estimation in Image Deblurring," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 1, pp. 1–14, 2022, doi: 10.1109/TGRS.2022.3165609
- [7] J. Song, J. H. Jeong, D.-S. Park, H. Kim, D. Seo, and J. C. Ye, "Unsupervised Denoising for Satellite Imagery Using Wavelet Directional CycleGAN," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6823–6839, Aug. 2021, doi: 10.1109/TGRS.2020.3025601
- [8] J. Zhang, X. Zhou, L. Li, T. Hu, and C. Fansheng, "A Combined Stripe Noise Removal and Deblurring Recovering Method for Thermal Infrared Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, pp. 1–14, Jan. 2022, doi: 10.1109/TGRS.2022.3196050
- [9] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Vis. Comput. Ind. Biomed. Art*, vol. 2, no. 1, p. 7, Jan. 2019, doi: 10.1186/s42492-019-0015-9
- [10] P. Burdziakowski, "A Novel Method for the Deblurring of Photogrammetric Images Using Conditional Generative Adversarial Networks," *Remote Sens.*, vol. 12, no. 16, p. 2586, Aug. 2020, doi: 10.3390/rs12162586
- [11] P. Liu, Y. Zhao, W. Yang, S. Wang, and S. Chen, "Blind Deblurring Using Space Target Features," *IEEE Access*, vol. 7, no. 1, pp. 131818–131829, Jan. 2019, doi: 10.1109/ACCESS.2019.2940817
- [12] R. Jing, F. Duan, F. Lu, M. Zhang, and W. Zhao, "Denoising Diffusion Probabilistic Feature-Based Network for Cloud Removal in Sentinel-2 Imagery," *Remote Sens.*, vol. 15, no. 9, pp. 2217-2225, Apr. 2023, doi: 10.3390/rs15092217
- [13] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, no. 1, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016
- [14] W. Zeng, "Application of DTCWT Decomposition and Partial Differential Equation Denoising Methods in Remote Sensing Image Big Data Denoising and Reconstruction," *J. Appl. Math.*, vol. 2022, no. 1, pp. 1–13, Oct. 2022, doi: 10.1155/2022/8553330

-
- [15] X. Wang, Z. Luo, W. Li, X. Hu, and L. Zhang, "A Self-Supervised Denoising Network for Satellite-Airborne-Ground Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, pp. 1–16, Jan. 2022, doi: 10.1109/TGRS.2021.3064429
- [16] X. Yang, X. Wang, N. Wang, and X. Gao, "SRDN: A Unified Super-Resolution and Motion Deblurring Network for Space Image Restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, pp. 1–11, Jan. 2022, doi: 10.1109/TGRS.2021.3131264
- [17] T. K. Hock, G. T. Chala, and H. H. Cheng, "An innovative hybrid steam-microwave sterilization of palm oil fruits at atmospheric pressure," *Innov. Food Sci. Emerg. Technol.*, vol. 60, no. 1, pp. 102289-102298, Mar. 2020, doi: 10.1016/j.ifset.2020.102289
- [18] H. T. Sukmana and J. I. Kim, "Exploring the Impact of Virtual Reality Experiences on Tourist Behavior and Perceptions," *Int. J. Res. Metav.*, vol. 1, no. 2, pp. 94-112, 2024
- [19] B. H. Hayadi and I. M. M. El Emary, "Enhancing Security and Efficiency in Decentralized Smart Applications through Blockchain Machine Learning Integration", *J. Curr. Res. Blockchain.*, vol. 1, no. 2, pp. 139–154, Sep. 2024.
- [20] A.R.Yadulla, G.S.Nadella, M.H.Maturi, H.Gonaygunta, "Evaluating Behavioral Intention and Financial Stability in Cryptocurrency Exchange App: Analyzing System Quality, Perceived Trust, and Digital Currency in Indonesia," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 2, pp. 103-124, 2024.