

3D Convolutional Autoencoder with Spectral–Spatial Attention (3DCAE-SSA) for Anomaly Detection in Hyperspectral Imagery

Salena Akter¹, Félicia Teferle¹, Abdul Nurunnabi²

¹ Geodesy and Geospatial Engineering, Faculty of Science, Technology and Medicine, University of Luxembourg
6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg. (salena.akter, rebecca.teferle)@uni.lu

² Independent Researcher, Alumni; Geodesy and Geospatial Engineering, University of Luxembourg, Luxembourg
pyal1471@yahoo.com

Keywords: Anomaly Map, Classification, Convolutional Neural Network, Deep Learning, Outlier, Robust Loss, Statistical Distance.

Abstract

Hyperspectral anomaly detection (HAD) refers to the process of identifying pixels or regions that are significantly different from the surrounding background signature of a hyperspectral image (HSI). HAD is highly challenging due to high spectral dimensionality, complex spatial patterns, subtle nature of the real-world environment, and limited knowledge about the signature of interest. This study develops a HAD algorithm combining three-Dimensional (3D) Convolutional Autoencoder (3DCAE), Spectral–Spatial Attention (SSA) and a robust hybrid loss. We use 3DCAE that employs 3D convolution layers (3DCLs) instead of fully connected layers to create a compact latent space from the input image, while the decoder uses deconvolution layers to recover the input image. We attach multi-headed spectral attention and multi-scale spatial attention that help to expand local receptive fields to global context and allow each voxel to attend to the other long-range voxels. In the proposed 3DCAE-SSA, to minimize the image reconstruction error, we introduce a hybrid robust loss to quantify the reconstruction error instead of using mean squared error (MSE) which is sensitive to outliers. To see the performance of the new algorithm, we build an error map based on a robust statistical Mahalanobis distance. For open access, Urban-4 dataset the AUC values show that the new algorithm performs significantly better having AUC value 99.76% than two existing algorithms: Auto-AD (97.47%), and GAED (99.24%).

1. Introduction

Unlike visible (red, green, blue) colour image, hyperspectral image (HSI) provides spectral-spatial information by capturing reflectance measurements across hundreds of contiguous wavelength bands. This high-dimensional representation enables to identify precise objects' material discrimination. Therefore, the use of HSI has become indispensable for a wide range of remote sensing tasks, and hyperspectral anomaly detection (HAD) applications, such as environmental monitoring, precision agriculture, forest health assessment, mineral exploration, and surveillance (Tu et al., 2019; Liu et al., 2024; Zhao et al., 2025). Despite many advantages of HSI, the complex structure and high redundancy of spectral bands present many challenges for HAD.

If pixel-level anomalies are not adequately addressed, they can adversely affect the performance of downstream data analysis tasks, including segmentation, classification, target detection, and change detection. Effective HAD not only facilitates the identification of unusual or potentially hazardous objects but also improves the robustness and accuracy of subsequent analytical processes. Since these tasks depend on consistent and reliable spectral–spatial representations, unmodeled outliers may distort feature distributions, introduce noise, and diminish the generalization capability of machine learning and deep learning models.

Conventional HAD methods, such as Reed–Xiaoli (RX) detectors (Reed and Yu, 1990), its variants (Matteoli et al., 2014), statistical background modeling, and subspace-based approaches (Guo et al., 2024), primarily assume linear relationships or mainly rely on simplified covariance structures. Although computationally efficient, these methods often exhibit limited performance in complex scenes containing small and spectrally

subtle anomalies embedded within highly irregular backgrounds. Since spectroscopic sensors are evolving toward higher spectral and spatial resolutions, these classical approaches are facing increasing difficulty in capturing nonlinear interactions and higher-order correlations inherent in HSI data. This has motivated a raising shift toward deep learning methods, which can model complex spectral–spatial relationships without requiring explicit assumptions about data distributions (Li, et al., 2017).

Autoencoders (AEs) have evolved as a promising deep learning approach for HAD because of their capacity to learn compact representations of background distributions through unsupervised manner (Guo et al., 2024). When trained exclusively or primarily on normal data, AEs are expected to reconstruct background pixels accurately while producing large reconstruction errors for anomalous pixels (Zhao et al., 2025). However, standard two-dimensional (2D) convolutional autoencoders (CAEs) are limited in their capacity to exploit the joint spectral–spatial structure of hyperspectral data, as they typically treat spectral bands as independent channels and miss the model spectral continuity. Similarly, most 1D or 2D spectral models (Hu et al., 2015; Zhao and Du, 2016) capture only limited aspects of HSI structure, lacking a unified treatment of spectral and spatial dependencies to each other. To address these limitations, many recent studies have explored 3D convolutional neural network (3DCNN) that operate directly on HSI data cubes, enabling simultaneous modeling of local spatial relationships and spectral correlations. Although, 3DCAEs provide richer representations than their lower-dimensional counterparts, they still face challenges in emphasizing the most informative spectral wavelengths and important spatial regions within the data. This motivates the integration of attention mechanism. Attention mechanism (AM) has shown remarkable success in natural image processing and sequence modeling (Vaswani et al., 2017).

In HAD, AM can guide the network to focus on salient spectral signatures and spatial patterns that characterize the background structures, thereby improving the robustness and discriminative power of the learned representations.

In this paper, we develop a 3D Convolutional Autoencoder with Spectral-Spatial Attention (3DCAE-SSA) to enhance anomaly detection performance by jointly leveraging local spectral-spatial modeling and global feature weighting. The encoder extracts hierarchical features from hyperspectral patches, while a bottleneck attention module selectively amplifies informative spectral bands and spatial regions. The decoder reconstructs the input data, and anomalies are identified through deviations between original and reconstructed patches. The scientific contributions of the paper are: (i) a new algorithm is developed by using multi-head spectral and multi-scale spatial attention to design an architecture to effectively suppress noise, and capture complex background characteristics of an HSI, (ii) the new algorithm introduces a hybrid metric that uses 3D structural similarity index and a robust loss to quantify the reconstruction error, and guides the model training process, and (iii) produces a precise anomaly map, offering significant improvements over conventional autoencoders and statistical methods.

The remainder of this paper is organized as follows. Section 2 briefly reviews related works. Section 3 presents the proposed methodology. Section 4 evaluates the proposed algorithm through two experiments using HSI data. Finally, Section 5 concludes the paper.

2. Related Literature

HAD methods have evolved substantially over the past two decades, can be grouped into probability-distribution-based, representation-based, and deep learning-based (Su et al., 2022). Early research predominantly focused on statistical methods, which leveraged the high-dimensional spectral information by modeling background distributions. The RX algorithm is one of the most widely used classical anomaly detection methods, which assumes that background pixels follow a multivariate Gaussian distribution, enabling anomaly identification through Mahalanobis distance type measures. Numerous RX variants, including local RX (Matteoli et al., 2014), weighted RX (Guo et al., 2014) where different weights are given to the background and anomaly pixels, subspace RX (Schaum, 2004), kernel RX (Kwon and Nasrabadi, 2005) that employs a nonlinear mapping of the original data onto a high-dimensional feature space, thereby enabling more effective separation of background and anomaly pixels, and cluster-based RX (Chun-tong et al., 2019), have been developed to enhance performance in heterogeneous environments. These computationally efficient methods remain benchmarks in the field, but their dependence on Gaussian assumptions and linear modelling often limit their pertinence in real-world scenarios characterized by nonlinearity, spectral variability, and the presence of noise.

Deep learning methods can be grouped into supervised and unsupervised learning frameworks. Supervised methods require labeled training data to learn the characteristics of target pixels, whereas unsupervised methods do not require prior knowledge of the target pixels. A well-known unsupervised HAD architecture is AE (Xie et al., 2020; Guo et al., 2024). AE considers that most of the pixels in HSI are background, while the anomalous pixels are only a small part of the image. Thus, reasonably, when an HSI is used in the AE model as inputs, the AE model generally learns the background. AE simultaneously reconstructs both the

anomaly targets and the background, but the lack of prior information and the small number of anomaly pixels can provide false positive results and limit the ability to detect anomalies (Xie et al., 2020).

Beyond Gaussian assumptions-based approaches, subspace and low-rank methods (Sun, et al., 2014; Zhuang et al., 2020) have gained attention for their ability to separate background and anomalies based on structural priors. Sun, et al. (2014) proposed a low-rank and sparse matrix decomposition (LRaSMD)-based anomaly detector that considers background images are low-rank while anomalies are gross errors that are sparsely distributed throughout the hyperspectral image. The method separates the anomalies from the background by solving a convex optimization problem. LRaSMD, and collaborative representation-based hyperspectral anomaly detectors assume that hyperspectral backgrounds lie on low-dimensional manifolds while anomalies exhibit sparse deviations. These improved methods provide better results in cluttered and mixed spectral environments. However, they often struggle with computational scalability when applied to large hyperspectral data. Also, subspace models may fail to capture the rich spectral-spatial interactions essential for distinguishing subtle anomalies from background heterogeneity (Xu et al., 2020; Wang et al., 2021).

To overcome the limitations of purely statistical and subspace-based techniques, researchers have increasingly integrated spatial context into HAD frameworks. Spatial filtering (Ma et al., 2025), superpixel segmentation (Subudhi et al., 2021), morphological profiles (Andikta et al., 2020; Huang et al., 2021), and Markov random field (Altmann et al., 2015)-based methods have been introduced to better utilize local homogeneity and reduce noise-induced false positive detections. While these approaches enhance target localization and suppress isolated noise pixels, they typically rely on hand-crafted features, which may not generalize well across sensors, environments, and data acquisition conditions.

The rise of deep learning architecture has marked a significant shift in anomaly detection methodologies by enabling the extraction of data-driven spectral-spatial representations. Convolutional neural networks (CNNs) have been widely developed for classification, segmentation and objects detection. Their extension to anomaly detection has produced promising results (Hu et al., 2015; Zhao and Du, 2016). 1D-CNNs model only spectral signatures, while 2D-CNNs treat hyperspectral images similarly to multispectral inputs and focus on spatial context, ignoring spectral continuity. These architectures, although informative, cannot fully leverage the spectral-spatial couplings inherent in HSI data.

To address these gaps, 3D-CNNs (Hamida et al., 2018; Zhong et al., 2018) have been introduced, allowing simultaneous processing of both spectral and spatial dimensions. 3D-CNN-based autoencoders (3D-CAEs; Lv et al., 2023) have gained particular importance due to their capability to learn compact representations of background structures in an unsupervised manner. When trained on hyperspectral patches, 3D-CAEs reconstruct background pixels with high fidelity and produce large reconstruction errors for anomalous pixels. Many variants including sparse autoencoders, denoising autoencoders, and variational autoencoders have been proposed to improve reconstruction robustness (Wei et al., 2022). Although effective, standard 3D-CAEs may struggle to emphasize the most informative spectral bands or spatial features, particularly in the presence of high data redundancy, spectral mixing, or fine-grained anomalies.

To improve feature discrimination, recent work has increasingly explored the well-known AMs (Vaswani et al., 2017; Hang et al., 2021; Fang et al., 2019; Sun et al., 2020). Inspired by their success in natural image processing and computer vision, they have been applied to HSI classification, segmentation, and object detection, demonstrating strong capability to capture long-range dependencies and highlight salient spectral channels (Fang et al., 2019). However, applications of AMs specifically for HAD remain relatively limited. Existing approaches include spectral attention networks that reweight feature channels, spatial attention modules to highlight informative regions, and hybrid attention that combine both spectral and spatial domains (Sun et al., 2020; Hang et al., 2021). These methods enhance background modeling by guiding networks to focus on relevant spectral-spatial signals, yet many operate within 2D-CNN frameworks or treat spectral and spatial attention separately, limiting the full exploitation of hyperspectral (spectral-spatial) data structure.

More recently, researchers have proposed joint spectral-spatial attention (SSA) architectures that integrate multi-dimensional feature weighting within deep learning models (Sun et al., 2020; Hang et al., 2021; Vaswani et al., 2017). Although they are promising for HSI analysis, such models often require extensive computation or depend on large training datasets, limiting their practical applicability. Moreover, many AM-based detectors lack explicit integration with 3DCNN representations, leading to suboptimal feature fusion. Despite advances in statistical, subspace, spatial-context, and deep learning-based methods, challenges remain in effectively modeling complex spectral-spatial structures. Most of the existing algorithms use single head and single scale attention thus they loss information from different spectral regions and different scale spatial representation. This motivates us to develop a hybrid architecture, name 3DCAE-SSA, that combine 3D convolutional encoding with multi-head and multi-scale attention mechanisms to achieve more discriminative and robust feature representations for HAD.

3. Methodology

This section presents the proposed HAD algorithm, 3DCAE-SSA, whose overall architecture is illustrated in Fig. 1.

Before the implementation of the developed architecture, we preprocess the data using the standard global min-max normalization method to transform all the feature (reflectance)

values to the same scale $[0, 1]$. The standard global min-max equation can be defined as (Noor Asmat et al., 2025):

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (1)$$

where $X \in R^{(H \times W \times B)}$ is a hyperspectral image cube, and H , W , and B are image height, width, and the number of spectral bands respectively. X_{min} and X_{max} are the minimum and the maximum reflectance values respectively of the HSI. Then, we create required patches for each pixel of the normalized HSI, X_N .

3.1 Encoder-Decoder Block

As portrayed in Fig. 1, the new model consists of three encoder blocks (modules) and two decoder blocks with two skip connections. Each of the encoder and decoder blocks consist of five basic tasks: (i) 3D convolution, (ii) batch normalization, (iii) ReLU activation, (iv) multi-head spectral attention (MH-SpecAtt), and (v) multi-scale spatial attention (MS-SpatAtt). Besides the regular tasks. Encoder-1 and encoder-2 use 3D max pooling as the last layer to reduce spatial dimension while decoder blocks use up sampling layer to increase the spatial dimension of feature maps. Encoder-1, encoder-2, and encoder-3 have 16, 32 and 64 output feature dimensions, respectively. The outputs from encoder-3 are used as latent features that are used as the inputs to the decoder-1. Similarly, decoder-1 and decoder-2 have 32 and 16 output feature dimensions. For each encoder and decoder layer, we apply 3D convolution with kernel size (3, 3, 3) to extracts local spectral-spatial features, normalize the extracted feature to stabilize the learning process and then apply the ReLU activation function to learn complex nonlinear spectral-spatial patterns. At the next stage, we use MH-SpecAtt and MS-SpatAtt modules to perform AM.

3.1.1 Multi-head Spectral Attention (MH-SpecAtt)

Module: HSI consists of many spectral bands, many of them are highly correlated and noisy. Typical convolution treats all the spectral bands equally important, whereas spectral attention gives more attention (dynamically reweights) to informative bands and tries to ignore noisy bands. A single head attention captures all relationships among the bands at once, while multi-head attention enables the model to simultaneously learn different types of relationships, attend to multiple positions in parallel, and captures information from diverse subspaces of the input spectra (Vaswani et al., 2017).

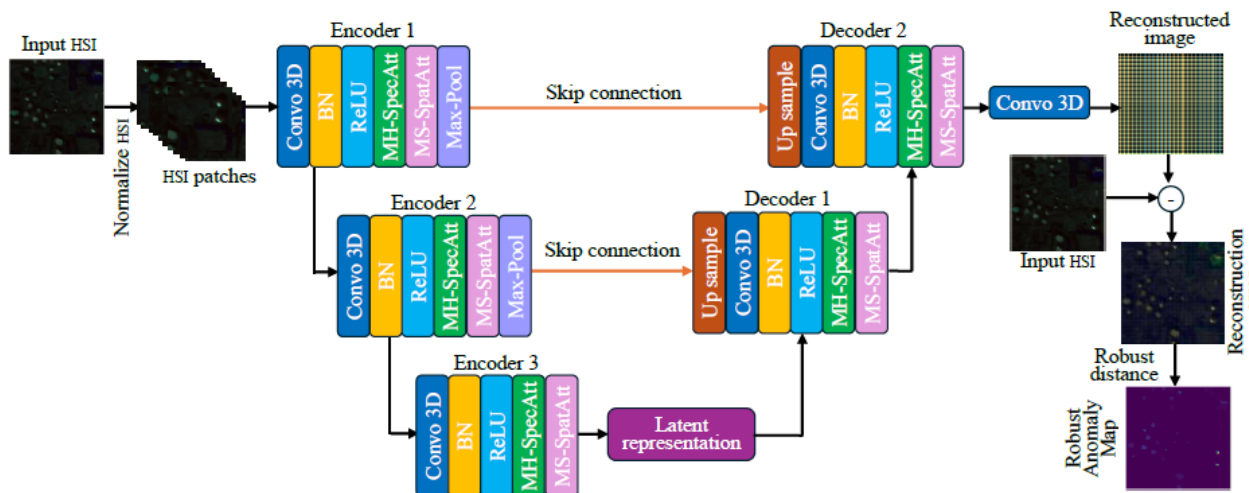


Figure 1. Implementation flow of 3DCAE-SSA and robust anomaly map.

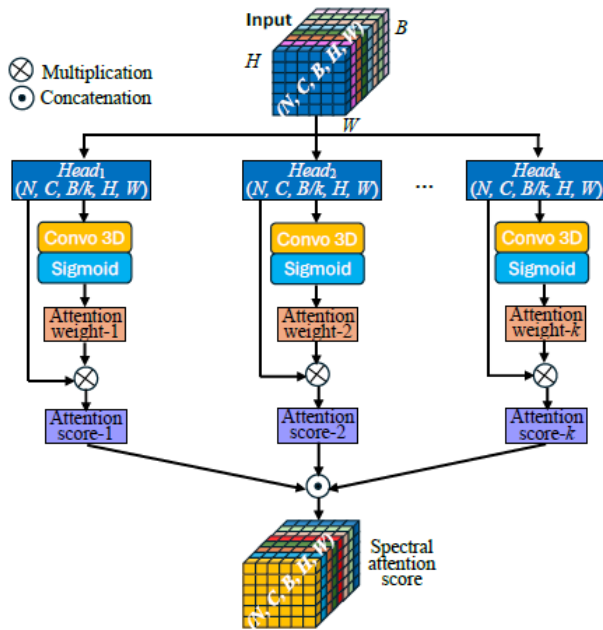


Figure 2. Multi-head spectral attention (MH-SpecAtt) module.

Unlike the standard Transformer, this MH-SpecAtt uses multiple spectral heads to independently learn attention weights for different groups of spectral bands via a lightweight (1×1) convolution and sigmoid gating. This makes it computationally efficient and suitable for HSI to emphasize the informative spectral bands. For MH-SpecAtt, we split the spectral bands into multiple (k) heads ($Head_1, Head_2, \dots, Head_k$) in parallel, where each head learns band-wise attention. In our case, we use $k = 4$. Since different heads learn different patterns, it deals with spectral variability and mixed pixels effects. Fig. 2 shows the construction of the MH-SpecAtt module. We calculate attention scores using Eq. 2.

$$\text{Spectral Attention Score} = \text{concat}(\text{sigmoid}(\text{conv3D}(\text{Head}_1, \text{Head}_2, \dots, \text{Head}_k)) \times (\text{Head}_1, \text{Head}_2, \dots, \text{Head}_k)) \quad (2)$$

3.1.2 Multi-scale Spatial Attention (MS-SpatAtt) Module:

We use the multi-scale spatial attention (MS-SpatAtt; He et al., 2017; Haidarh et al., 2025) module to enhance the new model's ability to capture spatial features at different scales. By employing multiple convolutions of different kernel sizes: (1×1 , 3×3 , 5×5 , and 3×3 with dilation, $d = 2$), the module extracts small, medium, and large-scale spatial information simultaneously. The resulting feature maps are combined and fused by (1×1) convolution and then apply sigmoid activation to transform into a spatial attention map that highlights informative regions while suppressing irrelevant background. Consequently, the network can effectively detect and represent objects of varying sizes and improves feature discrimination. A single convolution kernel cannot capture all these scales effectively. So, this module improves the representation of spectral-spatial features and makes the model more robust to different sizes of objects. Fig. 3 shows the construction of MS-SpatAtt module.

Now we apply 3D max pooling with kernel size (1, 2, 2) to the encoder layer to reduce the spatial resolution by half of the input feature map while preserving spectral dimension. Similarly, decoder layers increase spatial resolution by 3D transpose convolution (deconvolution) with kernel size (1, 2, 2) while preserving spectral information.

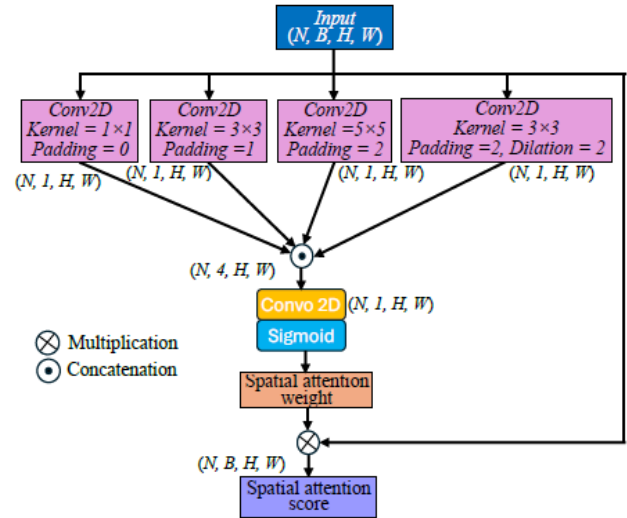


Figure 3. Multi-scale spatial attention (MS-SpatAtt) module.

3.2 Hybrid Loss Function

The assessment of similarity or dissimilarity between two images is a fundamental problem in image processing. Generally, the mean squared error (MSE) has been widely adopted as standard criteria for evaluating image quality and guiding the design and optimization of image-processing algorithms (Wang et al., 2006). Instead of MSE to train the model, we develop a hybrid loss function by combining the 3D Structural SIMilarity (SSIM) index (loss) and robust Welsch loss.

3.2.1 Structural SIMilarity Index (SSIM): SSIM index measures the structural similarity between two images that compare local pattern of pixel intensities that have been normalized for luminance (brightness), contrast, and structure comparisons (Wang et al., 2004). The SSIM between two image X , and Y is defined as:

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (3)$$

where $l(x, y)$, $c(x, y)$, and $s(x, y)$ are the luminance, contrast, and the structural comparisons, respectively. The corresponding functions are the typical average, standard deviation and normalized functions of the signals, respectively, and $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are the parameters used to adjust the relative importance of the three components. SSIM loss is defined as:

$$\text{SSIM}_{\text{Loss}}(x, y) = \frac{1 - \text{SSIM}(x, y)}{2} \quad (4)$$

This SSIM loss is calculated in local window-like patch. We use 3D SSIM loss since 3D patch is used as input to the model. The reader is referred to Wang et al. (2004) for the detail about the SSIM index.

3.2.2 Welsch Loss: To reduce the influence of outliers during optimization, we use the Welsch loss, which is a robust loss function. The Welsch loss gradually limits the contribution of large residuals, makes itself particularly useful when the data contain noise, mismatches, or outlying observations. It can keep away AEs from reconstructing anomalies. The Welsch loss is defined as:

$$\text{Welsch}_{\text{loss}}(x, y) = 1 - \exp\left(-\frac{1}{2}(x - y)^2\right). \quad (5)$$

Finally, the hybrid loss (HL) is defined as:

$$HL = \alpha \times \text{WelschLoss}(x, y) + \beta \times \text{SSIM}_{\text{Loss}}(x, y), \quad (6)$$

where α and β are constants, determined by empirical study.

3.3 Robust Anomaly Map Generation

An anomaly map is a spatial map in which each pixel is assigned a value that quantifies its degree of abnormality relative to the rest of the image. That means, anomalous pixels have higher anomaly scores than background pixels. Typically, the mean squared error (MSE) is used as an anomaly score. However, we avoid using MSE because its squared-residual formulation makes it highly sensitive to extreme observations, which are often treated as outliers. Consequently, large residuals receive disproportionately high weights, potentially reducing the robustness of anomaly detection (Nurunnabi et al., 2018). We use multivariate robust distance (RD) to produce robust anomaly map that can sharply visualize the identified anomalies in HSI.

To generate a robust anomaly map, we estimate the robust mean and covariance matrix from the reconstructed residual vectors and subsequently compute a Mahalanobis distance-type robust distance (RD) for each pixel, resulting in an alternative anomaly map. RD can be defined as:

$$RD = (X - \bar{X}_{\text{FMCD}})^T \Sigma_{\text{FMCD}}^{-1} (X - \bar{X}_{\text{FMCD}}), \quad (7)$$

where $\bar{X}_{\text{FMCD}}$ and Σ_{FMCD} are the fast minimum covariance determinant (FMCD) based robust mean and covariance matrix (Rousseeuw and Driessen, 1999; Nurunnabi et al., 2012).

4. Experiments, Evaluation and Discussion

We experimented on several benchmark HSI datasets. Two of them [Airport-4 (Gulfport) and Urban-4 (Los Angeles)] are used here to show the performance of the newly developed algorithm. Both datasets were acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor that captures spectral range of 400–2500 nm, contain up to 224 bands (Tu et al., 2019, Wang et al., 2023). To measure the performance of the new

model, we use Area Under the ROC (receiver operating characteristic) Curve (AUC) which is one of the most widely used metrics to evaluate the performance of HAD algorithms. The AUC is defined as (Liu et al., 2024):

$$\text{AUC} = \int_{-\infty}^{+\infty} \text{DP}(H) \text{FAR}(H) dH, \quad (8)$$

where $\text{DP}(H)$ denotes how many positive samples occurs among all samples, when threshold is H . $\text{FAR}(H)$ denotes how many false positive samples occur among all negative samples.

The proposed 3DCAE-SSA algorithm was implemented using the PyTorch framework on a Windows 11 Pro operating system and executed on an NVIDIA GeForce RTX 4090 GPU with an Intel Core i7 processor and 64 GB of RAM. For each dataset, the network is trained using the Adam optimizer with a learning rate of 0.001 for 200 epochs. The number of attention heads in MH-SpecAtt is set to 5. The remaining hyperparameters are specified separately for each experiment. The performance of the proposed algorithm (3DCAE-SSA) is evaluated against two recently introduced AE-based algorithms, Auto-AD (Wang et al., 2022) and GAED (Xiang et al., 2022).

4.1 Experiment 1: Airport-4 (Gulfport) Dataset

For the first experiment, we consider the Airport-4 dataset, which was acquired in Gulfport Airport, Mississippi, USA. In this dataset, spectral bands substantially affected by moisture and noise that were excluded from further analysis (Tu et al., 2019, Wang et al., 2023). After removing noisy bands, the dataset retains 191 spectral bands. The size of the image is 100*100 pixels with spatial resolution 3.4m. This dataset captures greenery, runway, tarmac, and three aircrafts. Fig. 4 displays: (a) false-color image, (b) ground truth image, and (c) mean spectral profile of background and anomalies. Sixty pixels corresponding to three aircraft are considered anomalies, accounting for 0.6% of all pixels in the dataset. This dataset has several challenges: very few anomalies (0.6%), background variability (presence of runway, tarmac, vegetation and other materials), low spatial resolution (an aircraft occupies only a part of pixels), and some of aircraft pixels have spectral characteristics that closely resemble those of the surrounding background materials.

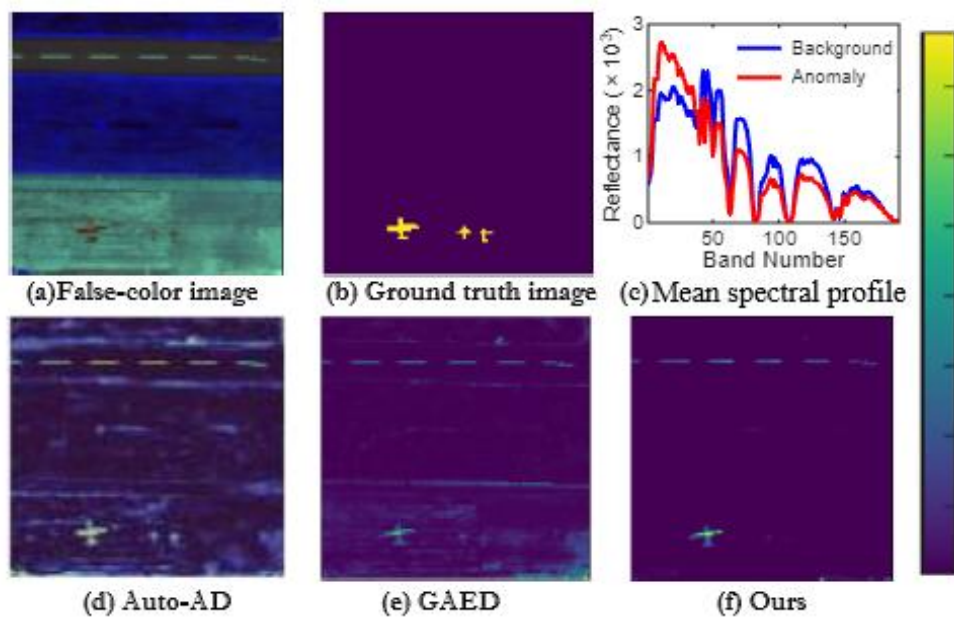


Figure 4. Airport-4 dataset; (a) False-color image, (b) Ground truth image, (c) Mean spectral plot of background and anomaly pixels; anomaly map generated by: (d) Auto-AD, (e) GAED, and (f) 3DCAE-SSA (ours).

Dataset	Algorithm		
	Auto-AD	GAED	Ours (3DCAE-SSA)
Airport-4	98.38	96.60	98.98
Urban-4	97.47	99.24	99.76

Table 1. Quantitative performance evaluation; AUC values (%) for Auto-AD [Wang et al., 2022], GAED [Xiang et al., 2022], and 3DCAE-SSA (ours).

To implement MH-SpectAtt, the number of spectral bands must be divisible by the number of attention heads. Therefore, we use the first 190 spectral bands in our experiments, as 190 is divisible by five. For this dataset, the remaining hyperparameters are set to a patch size of (8, 8), a stride of 1, a batch size of 4, $\alpha = 0.7$, and $\beta = 0.1$.

For both Auto-AD and GAED, we use the default parameter settings recommended in the respective papers. As shown in Table 1, the proposed algorithm achieves higher AUC of 98.98%, compared to 98.38% for Auto-AD and 96.60% for GAED. Although Auto-AD outperforms GAED, it remains inferior to 3DCAE-SSA. The corresponding error maps are presented in Fig. 4(d–f), further demonstrating the effectiveness of the proposed 3DCAE-SSA and supporting its advantages over the competent methods.

4.2 Experiment 2: Urban-4 (Los Angeles) Dataset

In the 2nd experiment, we perform all three HAD algorithms via the well-known Urban-4 dataset. This dataset was collected in Los Angeles, USA. The image size of this dataset occupies 100*100 pixels with 7.1 m spatial resolution. Excluding the noisy bands, the dataset contains 205 bands (Tu et al., 2019). Two hundreds and seventy-two buildings' pixels are considered as anomalies which is 2.72% of all pixels in this dataset (Küçük, 2026). This dataset is selected for the experiment because its anomalous objects exhibit larger spatial structures, lower spatial resolution, and greater spectral differences than the aircrafts in the Airport-4 dataset.

Therefore, it can test whether an anomaly detector can exploit both spectral characteristics and spatial context. The hyper parameters for this dataset are set to a patch size (12, 12), a stride of 4, a batch size of 4, $\alpha = 0.4$ and $\beta = 0.3$. Fig. 5 displays (a) the false-color image, (b) ground truth image, and (c) mean spectral profile of background and anomalies, and (d–f) the anomaly maps generated by the three different methods.

The performance of this dataset also measured by the AUC values. Despite having lower spatial resolution, our methods perform better than the existing two methods. The AUC values in Table 1 demonstrate that the proposed algorithm achieves superior performance (99.76%) compared with Auto-AD (97.47%) and GAED (99.24%). The error maps generated by Auto-AD, GAED, and the proposed method are shown in Fig. 5(d–f), respectively. For this dataset, GAED outperforms Auto-AD; however, the new algorithm achieves superior performance on both the datasets. This indicates that the proposed algorithm is more generalizable than the other two existing methods.

4.3 Discussions

The two datasets present several challenging characteristics, including a very low proportion of anomalous pixels (0.6%), substantial background variability (e.g., runway, tarmac, and vegetation), low spatial resolution, and high spectral similarity between some anomalous pixels and the surrounding background.

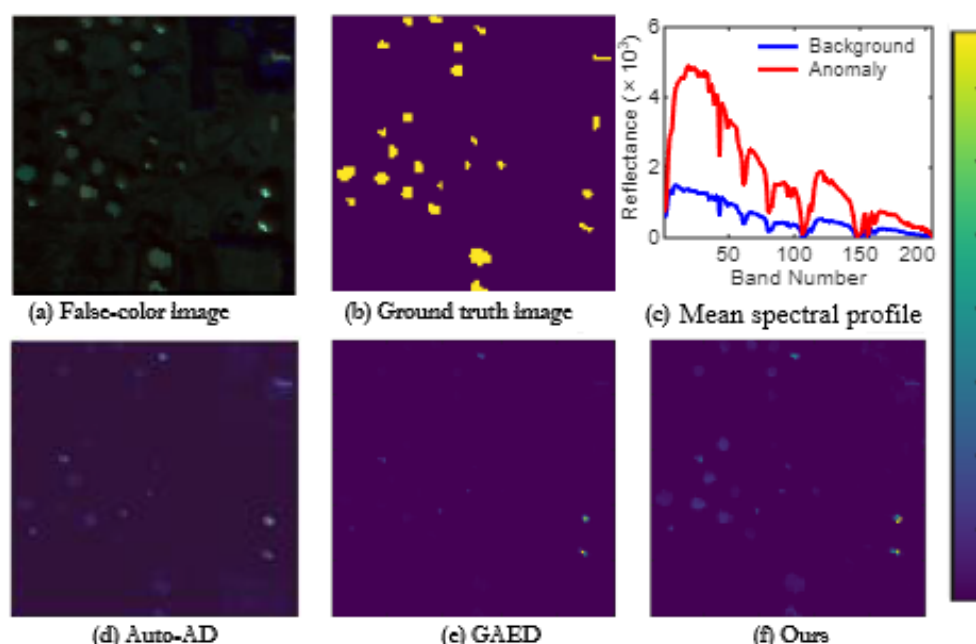


Figure 5. Urban-4 dataset. (a) False-color image, (b) Ground truth image, (c) Mean spectral plot of background and anomaly pixels; anomaly map generated by: (d) Auto-AD, (e) GAED, and (f) 3DCAE-SSA (ours).

Importantly, the nature and severity of these challenges differ between the two datasets, providing a more comprehensive assessment of the robustness and generalizability of the proposed method.

The Airport-4 dataset contains a small number of anomalies, some of which exhibit spectral characteristics like those of the background, making anomaly detection exceptionally challenging. In contrast, the Urban-4 dataset is characterized by lower spatial resolution and relatively large anomalies. To address these challenges, the proposed multi-head spectral attention effectively models spectral similarities and discriminates subtle spectral differences between anomalous and background pixels, while the multi-scale spatial attention captures anomaly features at different spatial scales, enabling the detection of both small and large anomalies. As a result, the proposed method can successfully detect anomalies in both datasets despite variations in spatial resolution, anomaly size, and spectral contrast between anomalies and background pixels. Experimental results demonstrate that the proposed approach is robust across diverse scenarios, whereas existing methods often perform well on one dataset but fail to generalize effectively to others.

5. Conclusions

The proposed HAD algorithm effectively suppresses noise, highlights semantically meaningful features and isolates anomalies. In the new algorithm, multi-head spectral attention (MH-SpecAtt) and multi-scale spatial attention (MS-SpatAtt), effectively capture diverse spectral relationships and multi-scale spatial features, respectively, leading to improved spectral-spatial representation and enhanced model performance. The new hybrid loss combining structural similarity (SSIM) index and Welsch loss quantifies the image reconstruction error successfully. Furthermore, a robust distance metric was utilized to construct the anomaly map, enabling more accurate anomaly characterization and sharper visualization of anomalous targets. Results in two experiments justify the necessity of the 3DCAE-SSA. The new algorithm is more generalizable and robust than the two existing methods.

Acknowledgements

Ms Salena Akter is funded via the University of Luxembourg. Project PRECISION is funded by the Luxembourg National Research Fund (FNR) and the Ministry of Agriculture, Food and Viticulture (MAAV) under the Agriculture BRIDGES Call 2022, reference 17949865.

References

Altmann, Y., McLaughlin, S., Hero, A., 2015. Robust linear spectral unmixing using anomaly detection. *IEEE Trans. Comput. Imaging*, 1(2), 74–85.

Andika, F., Rizkinia, M., Okuda, M., 2020. A hyperspectral anomaly detection algorithm based on morphological profile and attribute filter with band selection and automatic determination of maximum area. *Remote Sensing*, 12(20), 3387.

Chun-tong, L., Shi-xin, M., Hao, W., Yang, W., Hong-cai, L., 2019. A density-based cluster kernel RX algorithm for hyperspectral anomaly detection. *Spectrosc. Spectr. Anal.*, 39(6), 1878–1884.

Fang, B., Li, Y., Zhang, H., Chan, J.C.W., 2019. Hyperspectral images classification based on dense convolutional networks with spectral-wise attention mechanism. *Remote Sens.* 11(2), 159.

Guo, Q., Zhang, B., Ran, Q., Gao, L., Li, J., Plaza, A., 2014. Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 7(6), 2351–2366.

Guo, Q., Yi, C., Zhang, L., Zhang, Y., Hu, S., Liu, X., 2024. Robust multi-stage progressive autoencoder for hyperspectral anomaly detection. *Int. J. Appl. Earth Obs. Geoinf.*, 134.

Haidarh, M., Mu, C., Liu, Y., 2025. Hybrid multiscale spectral-spatial swin transformer for hyperspectral image classification. *Opt. Laser Technol.*, 192, 113923.

Hamida, A.B., Benoit, A., Lambert, P., Amar, C.B., 2018. 3-D deep learning approach for remote sensing image classification. *IEEE Trans Geosci Remote Sens.*, 56(8), 4420–4434.

Hang, R., Li, Z., Liu, Q., Ghamisi, P., Bhattacharyya, S.S., 2021. Hyperspectral image classification with attention-aided CNNs. *IEEE Trans Geosci Remote Sens.*, 59(3), 2281–2293.

He, M., Li, B., Chen, H., 2017, September. Multi-scale 3D deep convolutional neural network for hyperspectral image classification. *Proc. Int. Conf. Image Process. ICIP*, 3904–3908.

Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.*, (1), 258619.

Huang, J., Liu, K., Xu, M., Perc, M., Li, X., 2021. Background purification framework with extended morphological attribute profile for hyperspectral anomaly detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, 14, 8113–8124.

Kwon, H., Nasrabadi, N. M., 2005. Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 43(2), 388–397.

Küçük, F., 2026. A wavelet-integrated framework for feature extraction and background refinement in hyperspectral anomaly detection. *Scientific Reports*, 16(1), 8862.

Li, W., Wu, G., Du, Q., 2017. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, 14(5), 597–601.

Liu, S., Li, Z., Wang, G., Qiu, X., Liu, T., Cao, J., Zhang, D., 2024. Spectral-spatial feature fusion for hyperspectral anomaly detection. *Sensors*, 24(5), 1652.

Lv, S., Zhao, S., Li, D., Pang, B., Lian, X., Liu, Y., 2023. Spatial-spectral joint hyperspectral anomaly detection based on a two-branch 3D convolutional autoencoder and spatial filtering. *Remote Sensing*, 15(10), 2542.

Ma, Z., Zhang, Y., Lian, Y., Jiang, X., Liu, X., Cai, Z., 2025. Multiscale memory autoencoder and spatial filtering for hyperspectral anomaly detection. *IEEE Geosci. Remote Sens. Lett.*, 22, 1–5.

Matteoli, S., Veracini, T., Diani, M., Corsini, G., 2014. A locally adaptive background density estimator: An evolution for RX-based anomaly detectors. *IEEE Geosci. Remote Sens. Lett.*, 11(1), 323–327.

- Noor Asmat, B., Syed Muhammad Bilal, H., Uddin, M.I., Khalid Karim, F., Mostafa, S.M., Varela-Aldás, J., 2025. Enhancing feature learning of hyperspectral imaging using shallow autoencoder by adding parallel paths encoding. *Scientific Reports*, 15(1), 17363.
- Nurunnabi, A., Belton, D., West, G., 2012. Diagnostic-robust statistical analysis for local surface fitting in 3D point cloud data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, I-3, 269–274.
- Nurunnabi, A., Sadahiro, Y., Laefer, D. F., 2018. Robust statistical approaches for circle fitting in laser scanning three-dimensional point cloud data. *Pattern Recognition*, 81, 417–431.
- Reed, I.S., Yu, X., 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoust. Speech Signal Process.*, 38(10), 1760–1770.
- Rousseeuw, P.J., Driessen, K.V., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212–223.
- Schaum, A., 2004. Joint subspace detection of hyperspectral targets. *IEEE Aerosp. Conf. Proc.* (IEEE cat. no. 04TH8720) (vol. 3).
- Su, H., Wu, Z., Zhang, H., Du, Q., 2022. Hyperspectral anomaly detection: A survey. *IEEE Geosci. Remote Sens. Mag.*, 10 (1), 64–90.
- Subudhi, S., Patro, R.N., Biswal, P.K., Dell’Acqua, F., 2021. A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 14, 5015–5035.
- Sun, W., Liu, C., Li, J., Lai, Y.M., Li, W., 2014. Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery. *J. Appl. Remote Sens.*, 8(1), 083641.
- Sun, H., Zheng, X., Lu, X., Wu, S., 2020. Spectral–spatial attention network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 58(5), 3232–3245.
- Tu, B., Li, N., Liao, Z., Ou, X., Zhang, G., 2019. Hyperspectral anomaly detection via spatial density background purification. *Remote Sensing*, 11(22), 2618.
- Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.*, 30.
- Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process*, 13(4), 600–612.
- Wang, Z., Bovik, A.C., 2006. *Modern Image Quality Assessment*. Springer.
- Wang, R., Lu, Y., Zhang, Q., Nie, F., Wang, Z., Li, X., 2021. Ensemble and random collaborative representation-based anomaly detector for hyperspectral imagery. *arXiv:2101.01976*.
- Wang, S., Wang, X., Zhang, L., Zhong, Y., 2022. Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–14.
- Wang, H., Yang, M., Zhang, T., Tian, D., Wang, H., Yao, D., Meng, L., Shen, H., 2023. Hyperspectral anomaly detection with differential attribute profiles and genetic algorithms. *Remote Sensing*, 15(4), 1050.
- Wei, J., Zhang, J., Xu, Y., Xu, L., Wu, Z., Wei, Z., 2022. Hyperspectral anomaly detection based on graph regularized variational autoencoder. *IEEE Geosci. Remote Sens. Lett.*, 19, 1–5.
- Xie, W., Liu, B., Li, Y., Lei, J., Du, Q., 2020. Autoencoder and adversarial learning-based semi supervised background estimation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.*, 58(8), 5416–5427.
- Xiang, P., Ali, S., Jung, S.K., Zhou, H., 2022. Hyperspectral anomaly detection with guided autoencoder. *IEEE Trans. Geosci. Remote Sens.* 60, 1–18.
- Xu, Y., Du, B., Zhang, L., Chang, S., 2020. A low-rank and sparse matrix decomposition-based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection. *IEEE Geosci. Remote Sens. Lett.*, 17(7), 1248–1252.
- Zhao, W., Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Trans Geosci Remote Sens.*, 54(8), 4544–4554.
- Zhao, Z., Song, J., You, M., Xiang, P., Zhao, D., Zhang, J., Zhou, H., Wang, D., Wang, X., 2025. Hyperspectral anomaly detection via cascaded convolutional autoencoders with adaptive pixel-level attention. *Expert Syst. Appl.*, 279, 127366.
- Zhong, Z., Li, J., Luo, Z., Chapman, M., 2018. Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans Geosci Remote Sens.*, 56(2), 847–858.
- Zhuang, L., Gao, L., Zhang, B., Fu, X., Bioucas-Dias, J.M., 2020. Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations. *IEEE Trans Geosci Remote Sens.*, 60, 1–17.