Attention-Based Convolutional Neural Network for Anomaly Detection in Multispectral Images of Semi-Natural Ecosystems

Javier López-Fandiño, *Member, IEEE*, Álvaro Ordóñez, *Member, IEEE*, Pablo Quesada-Barriuso, *Member, IEEE*, Alberto S. Garea, *Member, IEEE*, Francisco Argüello, and Dora B. Heras, *Member, IEEE*

***Abstract*—The monitoring of semi-natural ecosystems has become increasingly critical due to the rising impact of ecological disturbances, including natural disasters and unauthorized human-made constructions. Anomaly detection (AD) in multispectral imagery serves as a fundamental tool in this context. Deep-learning (DL)-based techniques are particularly effective at capturing the intricate spectral and spatial patterns of anomalies. This letter proposes a new AD technique called attention-based convolutional neural network (ACNN), designed to enhance AD performance in multispectral images of high spatial resolution for the detection of human-made constructions. The model integrates attention mechanisms to prioritize informative features while suppressing irrelevant background information, thereby improving sensitivity to subtle and rare anomalies. Experimental results on multispectral datasets from semi-natural ecosystems show that the proposed approach outperforms existing DL techniques in terms of detection accuracy. These findings highlight the potential of attention-based models as a robust framework for environmental monitoring and AD in complex remote sensing scenarios.**

***Index Terms*—Anomaly detection (AD), attention mechanism, convolutional neural network (CNN), multispectral image, vegetation.**

I. Introduction

THE rapid development of remote sensing technologies has led to a surge in the availability of high-resolution multispectral imagery, offering unprecedented opportunities for monitoring and analyzing complex ecosystems [1]. Semi-natural ecosystems, landscapes that have been altered by human activity but still retain native species and ecological functions, are highly dynamic environments that play a critical role in maintaining biodiversity [2]. However, these ecosystems are increasingly threatened by anthropogenic activities, such as illegal constructions, resulting in the necessity of advanced methods for timely and accurate detection.

Anomalies in multispectral images of semi-natural ecosystems often correspond to significant ecological disturbances. Spare constructions inside these ecosystems can be considered anomalies, i.e., groups of pixels that are very different from the surroundings. Detecting such anomalies is inherently challenging due to the high dimensionality and data complexity. In particular, due to spectral redundancy and correlation between the spatial and the spectral information. Traditional approaches to AD often rely on hand-crafted features and statistical models, which may struggle to capture the high-dimensional and nonlinear relationships inherent in multispectral data.

Deep learning (DL) methods have demonstrated remarkable success in diverse domains, particularly in the analysis of image data. Convolutional neural networks (CNNs), in particular, have been widely adopted for classification, segmentation, and object detection due to their ability to learn hierarchical features directly from raw data [3]. Nevertheless, the application of CNNs to anomaly detection (AD) presents unique challenges, especially when anomalies exhibit subtle spectral variations or occur in underrepresented contexts.

To address these challenges, attention modules have emerged as a powerful enhancement for DL architectures [4]. By dynamically focusing on the most relevant features or regions in the input data, attention modules enable models to allocate computational resources efficiently and improve performance on complex tasks. The integration of attention modules within CNN architectures has shown promise in applications such as image classification [5], change detection [6], or object tracking [7]. However, their potential in AD for multispectral imagery remains underexplored.

This letter proposes an attention-based CNN (ACNN) tailored for AD in multispectral images, with a specific focus on applications in semi-natural ecosystems. The proposed model leverages attention modules to enhance the network's ability to capture subtle spectral and spatial variations associated with anomalies. By systematically integrating attention modules within the CNN architecture, the model aims to achieve

Received 17 March 2025; revised 2 June 2025; accepted 5 June 2025. Date of publication 9 June 2025; date of current version 19 June 2025. This work was supported in part by MCIN/AEI/10.13039/501100011033 under Grant PID2022–141623NB–I00 and Grant TED2021–130367B–I00; in part by European Union NextGenerationEU/PRTR; in part by the Xunta de Galicia—Consellería de Educación, Ciencia, Universidades e Formación Professional through the Centro de Investigación de Galicia Accreditation under Grant 2024-2027 ED431G-2023/04 and the Reference Competitive Group Accreditation under Grant ED431C-2022/16; and in part by ERDF/EU. *(Corresponding author: Javier López-Fandiño.)*

Javier López-Fandiño, Álvaro Ordóñez, Pablo Quesada-Barriuso, Alberto S. Garea, and Dora B. Heras are with the Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: javier.lopez.fandino@usc.es; alvaro.ordonez@usc.es; pablo.quesada@usc.es; jorge.suarez.garea@usc.es; dora.blanco@usc.es).

Francisco Argüello is with the Department of Electronics and Computing, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: francisco.arguello@usc.es).

Digital Object Identifier 10.1109/LGRS.2025.3577943

improved sensitivity and precision, particularly for rare or ambiguous anomalies.

The primary contributions of this work are as follows.

1) The development of a novel ACNN model for AD in multispectral imagery, integrating spatial segmentation and channel attention mechanisms to enhance feature discrimination.
2) The design of a model adapted to the detection of anomalies in high-spatial resolution multispectral images. The model is based on a CNN and a simplified attention module that merges the concepts of key and value into a single tensor, enhancing the spectral characteristics of the anomalies present in the image.
3) An experimental evaluation on real-world multispectral datasets of semi-natural ecosystems, demonstrating the high accuracy of the proposed method in the AD task compared with the existing DL approaches.

The remainder of this letter is organized as follows. Section II reviews related work on AD in multispectral imagery and attention modules in DL. Section III details the proposed ACNN architecture and its inclusion in a spectral-spatial model for AD. Section IV presents the experimental results and a comparative analysis of the proposed ACNN with other state-of-the-art alternatives. Finally, Section V concludes this letter and discusses potential directions for future research.

II. Related Work

AD in multispectral images has garnered significant attention due to its importance in applications such as environmental monitoring, precision agriculture, and urban planning. Traditional AD methods rely on statistical models and clustering techniques to identify deviations in spectral signatures. Methods such as principal component analysis [8] and Gaussian mixture models [9] have been widely used to extract features and distinguish normal patterns from anomalies. These approaches often struggle to capture the complex spectral-spatial relationships inherent in multispectral data.

CNNs have been effectively applied to AD in several fields, demonstrating their ability to extract relevant features corresponding to anomalous patterns. For instance, in medical imaging, CNN-based methods have succeeded in identifying irregularities within diagnostic images, improving prognosis capabilities [10]. Similarly, in industrial surface inspection, CNNs have proven effective for detecting defects in manufactured products with high precision and robustness [11].

Attention modules are mechanisms to further enhance the performance of DL models [12]. By enabling networks to focus on the most relevant features in an input, attention modules have proven effective in tasks requiring fine-grained feature discrimination. In hyperspectral and multispectral imaging, attention-based models have been applied to improve classification [13], segmentation [14], image fusion [15], and AD tasks [16]. The use of self-attention and multihead attention modules has shown promise in capturing long-range dependencies and emphasizing critical spectral-spatial interactions, as shown in [17] for road extraction through an attention-assisted U-net. These attention modules can focus on enhancing the spatial or spectral features of the images, or even in both at the same time [18]. Few studies have integrated attention modules into DL for AD in multispectral images. Most focus on classification, ignoring challenges such as data imbalance and undefined anomalies [19], [20]. This gap drives the need for a specialized attention-based CNN architecture.

Incorporating attention modules into a CNN improves both learning efficiency and feature discrimination. By adaptively highlighting the most relevant information, the impact of redundant data is reduced, guiding the model toward meaningful patterns from the outset. This targeted focus accelerates convergence and refines AD by enhancing sensitivity to subtle deviations. Consequently, attention-based CNNs achieve faster training and higher accuracy.

III. Attention CNN

This section introduces the ACNN-based model for AD in multispectral images. As illustrated in Fig. 1, the proposed approach consists of three main stages. First, a preprocessing step applies waterpixel segmentation [21] to enhance spatial features. Then, the ACNN extracts and processes the image patches, where the attention module selectively emphasizes the most relevant spectral features. Finally, the ACNN assigns a label to each patch, which is subsequently propagated to the corresponding region. This approach ensures that segmentation emphasizes the spatial characteristics of the image while the attention module primarily captures spectral features.

A. Waterpixel Segmentation and Patch Extraction

To improve efficiency and optimize AD, the process operates at the superpixel level instead of analyzing each pixel. This is achieved by the waterpixel segmentation stage [21] that groups pixels into homogeneous regions, reducing noise, preserving structural boundaries, and ensuring spatially coherent inputs. A representative patch is then extracted for each superpixel by determining its minimal bounding quadrilateral. The central pixel of the superpixel serves as a reference, defining an $(N \times N)$ patch. The superpixel is assigned the class of its central pixel, and this label is propagated throughout the region. This structured preprocessing improves both computational efficiency and AD accuracy.

B. Attention Module

A channelwise attention module is designed to enhance important features by dynamically reweighting feature maps. This approach is aligned with the methodology proposed in [18]. The proposed attention merges the concepts of key and value into a single tensor, simplifying the attention module. Instead of computing similarities between separate queries and keys and then using those similarities to weight distinct values, this method directly generates a weight map that is applied to the input.

This module consists of two pointwise convolutional layers (using 1×1 kernels) and nonlinear activations, as shown in Fig. 2. Given an input tensor of shape (C, H, W), where C represents the number of channels, the first convolution reduces the number of channels to $C/8$, effectively creating a

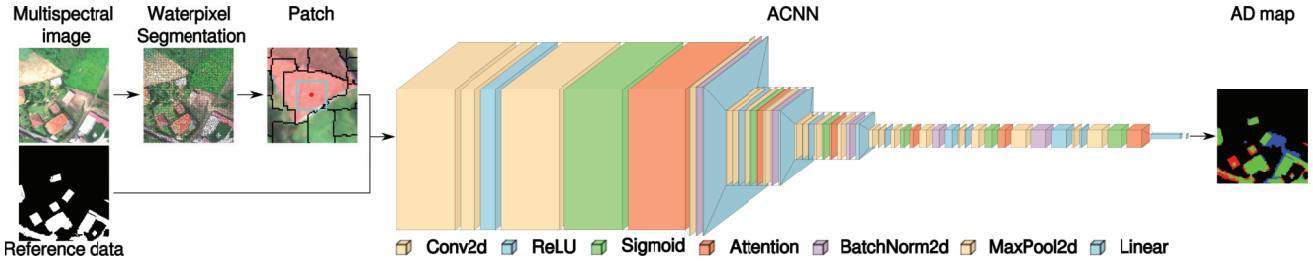

Fig. 1. Proposed ACNN-based AD model.

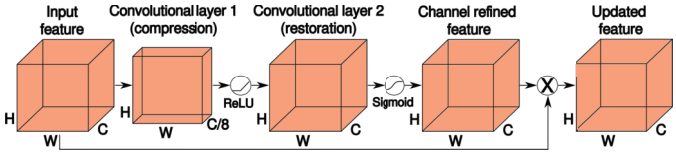

Fig. 2. Detail of a channel attention module of the ACNN.

TABLE I
ACNN ARCHITECTURE DESCRIPTION

Layer	Operation	Output Size
Input	-	$W \times H \times C$
Conv1	$3 \times 3, s = 1, p = 1, C \rightarrow 2048$	$W \times H \times 2048$
Attention1	-	
Conv2	$3 \times 3, s = 1, p = 2, 2048 \rightarrow 16$	
BatchNorm + ReLU	-	$W \times H \times 16$
MaxPool	$2 \times 2, s = 2$	
Attention2	-	
Conv3	$5 \times 5, s = 1, p = 2, 16 \rightarrow 32$	
BatchNorm + ReLU	-	$\frac{W}{2} \times \frac{H}{2} \times 32$
MaxPool	$2 \times 2, s = 2$	
Attention3	-	
Conv4	$3 \times 3, s = 1, p = 1, 32 \rightarrow 64$	
BatchNorm + ReLU	-	$\frac{W}{4} \times \frac{H}{4} \times 64$
MaxPool	$2 \times 2, s = 2$	
Attention4	-	
Conv5	$3 \times 3, s = 1, p = 1, 64 \rightarrow 128$	
BatchNorm + ReLU	-	$\frac{W}{8} \times \frac{H}{8} \times 128$
Attention5	-	
Conv6	$3 \times 3, s = 1, p = 1, 128 \rightarrow 256$	
BatchNorm + ReLU	-	$\frac{W}{8} \times \frac{H}{8} \times 256$
Attention6	-	
Flatten	-	$\frac{W}{8} \times \frac{H}{8} \times 256$
FC1	$Linear(\cdot \rightarrow 512)$	512
FC2	$Linear(\cdot \rightarrow n_{\text{classes}})$	n_{classes}

Note: s denotes stride, p denotes padding.

compact representation of the feature maps. This is followed by an ReLU activation, introducing nonlinearity and improving feature discrimination. The second convolution restores the original channel dimension C, and a sigmoid activation is applied to generate an attention map that determines the relative importance of each channel. During the forward pass, the computed attention is elementwise multiplied with the original input tensor, effectively scaling the activation of each channel according to its learned importance. This approach helps the network focus on more relevant features, improving AD performance.

C. ACNN Architecture

The proposed ACNN is a CNN with integrated attention modules consisting of multiple convolutional layers, batch normalization, ReLU activations, and max-pooling operations, with attention modules enhancing feature representations at several stages. The ACNN is described in detail in Table I.

TABLE II
DATASET DESCRIPTION

Name	z1	z2	e1	e2
Width	3807	2081	3629	1094
Height	2141	957	961	707
Spectral bands	5	5	5	5
Anomalies (px)	321710	249167	164488	25539
Non-anomalies (px)	7829077	1742350	3322981	747919
Anomalies (%)	3.95	12.51	4.72	3.30
Non-anomalies (%)	96.05	87.49	95.28	96.70
Size (MB)	163.0	39.8	69.7	15.5

The network begins with an input layer that accepts inputs of size (W, H, C), where W and H are the spatial dimensions of a patch and C represents the number of input channels, i.e., the number of spectral bands. An initial convolutional layer expands the feature dimension to 2048 channels, followed by an attention module to refine the extracted features. Subsequent layers progressively refine feature maps while reducing spatial dimensions through max-pooling operations. After the final convolutional layer, the feature maps are flattened to a 1-D vector. Finally, the fully connected layers, consisting of two consecutive linear transformations, map the output to one of the two possible classes for the AD task.

IV. EXPERIMENTS

A. Dataset

The experiments were carried out using data acquired with the MicaSense RedEdge multispectral sensor mounted on a specialized uncrewed aerial vehicle (UAV). This sensor captures imagery across five specific spectral bands: blue (475 nm), green (560 nm), red (668 nm), red edge (717 nm), and near-infrared (NIR) (840 nm). Aerial images of fluvial ecosystems were collected during the summer of 2018 in the Galicia region of Spain from an altitude of 120 m, achieving a high spatial resolution of 8.2 cm/pixel [22].

The dataset, publicly available in [23], contains four different images representing semi-natural ecosystems in densely vegetated areas. Human-made structures, such as buildings and roads, are identified as anomalies. It is essential to note that detecting all the anomalies is critical for this application, as failing to identify them could result in ecological harm.

Table II provides an overview of the key features of the four images considered. Anomalies account for between 3% and 12% of the total pixels in each image. Each image size is calculated based on pixel data stored in a 4-byte format.

TABLE III
ACNN AD RESULTS AND COMPARISON TO OTHER METHODS

Image	Algorithm	Parameters	TP	TN	FP	FN	Recall	AUC	Youden's J	MCC
z1	ACNN	3962432	271 514.50 ± 7240.16	7 749 899.50 ± 10 048.93	71 397.50 ± 10 048.93	49 871.50 ± 7240.16	0.8440 ± 0.0200	0.9170 ± 0.0100	0.8357	0.8102
	CNN	17762	250 524.70 ± 5205.87	7 778 719.70 ± 5510.53	42 577.30 ± 5510.53	70 861.30 ± 5205.87	0.7787 ± 0.0200	0.8863 ± 0.0100	0.7741	0.8091
	ResNet	195506	263 877.90 ± 7874.79	7 773 559.50 ± 13 703.31	47 737.50 ± 13 703.31	57 508.10 ± 7874.79	0.8202 ± 0.0200	0.9067 ± 0.0100	0.8150	0.8271
	ViT	6350594	232 907.50 ± 31 703.22	7 778 140.70 ± 20 330.04	35 374.30 ± 20 330.04	88 153.50 ± 31 703.22	0.7240 ± 0.1000	0.8590 ± 0.0500	0.7209	0.7860
	BAGAN	1429036	171 779.50 ± 80 218.94	7 788 129.20 ± 56 626.91	37 057.80 ± 56 626.91	149 768.50 ± 80 218.94	0.5340 ± 0.2500	0.7645 ± 0.1200	0.5295	0.6523
	SwinT	27477518	225 745.70 ± 51 636.71	7 767 250.20 ± 31 561.10	46 264.80 ± 31 561.10	95 315.30 ± 51 636.71	0.7017 ± 0.1600	0.8472 ± 0.0800	0.6972	0.7551
z2	ACNN	3962432	204 019.70 ± 19 359.69	1 701 822.90 ± 16 443.86	38 789.10 ± 16 443.86	44 901.30 ± 19 359.69	0.8188 ± 0.0800	0.8979 ± 0.0300	0.7973	0.8059
	CNN	17762	141 999.90 ± 56 230.93	1 722 283.00 ± 12 459.91	18 329.00 ± 12 459.91	106 921.10 ± 56 230.93	0.5699 ± 0.2300	0.7794 ± 0.1100	0.5599	0.6806
	ResNet	195506	95 644.60 ± 41 641.04	1 737 261.90 ± 6399.91	3350.10 ± 6399.91	153 276.40 ± 41 641.04	0.3839 ± 0.1700	0.6908 ± 0.0800	0.3823	0.5817
	ViT	6350594	195 520.50 ± 19 417.48	1 713 525.70 ± 28 194.28	25 348.30 ± 28 194.28	53 153.50 ± 19 417.48	0.7847 ± 0.0800	0.8843 ± 0.0300	0.7717	0.8123
	BAGAN	1429036	155 549.70 ± 41 609.43	1 689 747.20 ± 48 370.53	51 733.80 ± 48 370.53	93 494.30 ± 41 609.43	0.6243 ± 0.1700	0.7971 ± 0.0700	0.5949	0.6444
	SwinT	27477518	201 977.30 ± 35 183.20	1 697 782.50 ± 46 822.78	41 091.50 ± 46 822.78	46 696.70 ± 35 183.20	0.8106 ± 0.1400	0.8927 ± 0.0600	0.7886	0.7963
e1	ACNN	3962432	147 528.60 ± 3581.81	3 298 053.80 ± 4944.41	18 327.20 ± 4944.41	16 627.40 ± 3581.81	0.8969 ± 0.0200	0.9448 ± 0.0100	0.8932	0.8888
	CNN	17762	145 903.50 ± 4193.55	3 300 003.30 ± 4133.55	16 377.70 ± 4133.55	18 252.50 ± 4193.55	0.8870 ± 0.0300	0.9402 ± 0.0100	0.8839	0.8887
	ResNet	195506	147 749.50 ± 5774.05	3 303 637.80 ± 4592.55	12 743.20 ± 4592.55	16 406.50 ± 5774.05	0.8982 ± 0.0400	0.9463 ± 0.0200	0.8962	0.9059
	ViT	6350594	132 248.50 ± 18 539.37	3 300 401.10 ± 13 925.44	15 979.90 ± 13 925.44	31 907.50 ± 18 539.37	0.8040 ± 0.1100	0.8988 ± 0.0500	0.8008	0.8407
	BAGAN	1429036	122 922.20 ± 21 585.82	3 298 824.30 ± 17 885.91	19 206.70 ± 17 885.91	41 316.80 ± 21 585.82	0.7473 ± 0.1300	0.8702 ± 0.0600	0.7426	0.7957
	SwinT	27477518	143 206.60 ± 9198.51	3 299 334.30 ± 11 241.51	17 046.70 ± 11 241.51	20 949.40 ± 9198.51	0.8706 ± 0.0600	0.9319 ± 0.0300	0.8672	0.8772
e2	ACNN	3962432	18 290.50 ± 4327.49	739 176.90 ± 589.26	2644.10 ± 589.26	7023.50 ± 4327.49	0.7162 ± 0.1700	0.8531 ± 0.0800	0.7190	0.7883
	CNN	17762	17 921.50 ± 865.90	738 814.60 ± 614.73	3006.40 ± 614.73	7392.50 ± 865.90	0.7017 ± 0.0300	0.8457 ± 0.0200	0.7039	0.7719
	ResNet	195506	17 174.40 ± 1511.11	738 324.80 ± 1005.86	3496.20 ± 1005.86	8139.60 ± 1511.11	0.6725 ± 0.0600	0.8309 ± 0.0300	0.6737	0.7433
	ViT	6350594	9963.20 ± 1115.89	740 974.30 ± 812.66	846.70 ± 812.66	15 350.80 ± 1115.89	0.3901 ± 0.0400	0.6926 ± 0.0200	0.3924	0.5948
	BAGAN	1429036	8278.30 ± 3985.75	740 501.60 ± 2803.78	2843.40 ± 2803.78	17 091.70 ± 3985.75	0.3241 ± 0.1600	0.6589 ± 0.0800	0.3225	0.4824
	SwinT	27477518	12 754.30 ± 2641.65	739 007.60 ± 3230.66	2813.40 ± 3230.66	12 559.70 ± 2641.65	0.4994 ± 0.1000	0.7455 ± 0.0500	0.5001	0.6335

*Colors indicate First, second and third best results for each metric and image.

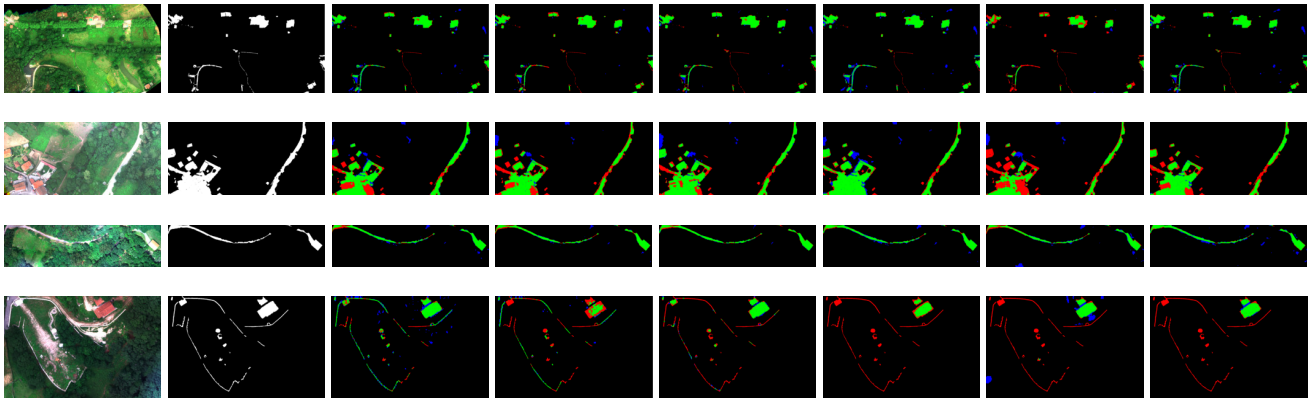

Fig. 3. AD maps for $z1$, $z2$, $e1$, and $e2$ images. The first two columns show RGB compositions and reference data of anomalies (anomalies in white). The remaining columns display AD maps from ACNN, CNN, ResNet, ViT, BAGAN, and SwinT. In AD maps, green = TP, blue = FP, red = FN, and black = TN.

B. Results

The effectiveness of the ACNN is compared with five state-of-the-art deep neural networks. These include a traditional CNN, a ResNet [24], which allows the construction of very deep networks that better capture the complex patterns of anomalies by avoiding the problem of vanishing gradients, a vision transformer (ViT) [25], an architecture that captures long-range dependencies across an image through self-attention mechanisms, a balancing GAN (BAGAN) [26], which adds to the generative nature of a GAN a balancing augmentation technique improving the learning of minority classes, and a SwinT Transformer [27], which uses shifted window-based self-attention to model long-range dependencies while maintaining computational efficiency for image recognition tasks.

To demonstrate how the inclusion of the attention module enhances the AD capabilities of the network, experiments are carried out with only 5% of samples for training in the images with a higher number of anomalies (i.e., $z1$ and $z2$) and 15% for the remaining images ($e1$, $e2$). 5% of samples is used for validation in all the cases. In addition, the number of training epochs is limited to five across all the networks and images, allowing us to assess whether the attention module accelerates the learning process. The number of epochs was empirically determined to achieve optimal performance. A batch size of 100 and a patch size of 32×32 are consistently maintained in all the experiments. ACNN, CNN, and ResNet were trained using the Adam optimizer, a `CosineAnnealingLR` scheduler, and `CrossEntropyLoss`. ViT and SwinT were trained with AdamW and a `LambdaLR` scheduler implementing a linear warm-up followed by cosine decay, both using `CrossEntropyLoss`. BAGAN used Adam with a learning rate schedule that halved the rate at each step and used `MSELoss` as the loss function. A single superpixel segmentation is computed for each image and applied across all the experiments to ensure that the comparison remains focused on differences in model architectures rather than segmentation variations. All the reported results represent the average performance over ten independent runs.

Table III summarizes the AD results for the evaluated state-of-the-art techniques. The number of trainable parameters for each algorithm is included for contextualization. Metrics include true and false positives and negatives (TP, TN, FP, FN), recall (the percentage of correctly detected anomalies), and area under the receiver operating characteristic curve (AUC),

which measures the discrimination ability. Youden's J statistic is reported to assess the tradeoff between sensitivity and specificity, and the Matthews correlation coefficient (MCC) provides a balanced evaluation under class imbalance.

As shown in Table III, ACNN outperforms both the traditional CNN and the other architectures included for comparison. More in detail, ACNN detects up to 2% more anomalies and obtains the highest AUC and Youden's J for three of the four images. ACNN also ranks first or second in MCC for all the images. ACNN obtains the best TP and FN rates, which is key to the use case presented in this work, where missed alarms could pose an ecological risk. For image $e1$, ACNN and ResNet achieve similar AUC. This image differs from the others, containing only two large anomaly structures. ResNet captures their edges, detecting a number of anomaly pixels comparable to ACNN.

Fig. 3 shows the AD maps obtained by ACNN, together with those obtained for the networks used for comparison, for the dataset introduced in Section IV-A. It can be seen that ACNN detects most of the anomaly structures in the reference data while maintaining a small number of false alarms, outperforming all the other alternatives. Waterpixel segmentation helps refine the shape of detected anomalies, ensuring that the alarms more accurately align with the actual object boundaries.

V. Conclusion

This letter introduced an ACNN-based model for AD in multispectral images, specifically targeting semi-natural ecosystems. The proposed model is based on a CNN that incorporates attention modules to enhance sensitivity to subtle and rare anomalies by effectively focusing on the most relevant spectral and spatial features.

The experimental evaluation was conducted on four multispectral images of semi-natural ecosystems of Galicia. ACNN achieves the best accuracy compared with the conventional CNN and other state-of-the-art approaches from the literature, achieving up to 2% points higher accuracy in AD.

The findings highlight the potential of ACNNs for environmental monitoring, offering a robust framework for detecting anomalies in complex multispectral data. The main limitations of this work are the limited number of public datasets and the computational overhead from attention modules. Further research work would be necessary to explore the implemented attention mechanism under different experimental conditions, such as different AD problems. The computational cost is also a concern that requires a more detailed analysis, especially for larger datasets.

References

[1] X. Hu et al., "Hyperspectral anomaly detection using deep learning: A review," *Remote Sens.*, vol. 14, no. 9, p. 1973, Apr. 2022.

[2] D. S. Rhee, Y. D. Kim, B. Kang, and D. Kim, "Applications of unmanned aerial vehicles in fluvial remote sensing: An overview of recent achievements," *KSCE J. Civil Eng.*, vol. 22, no. 2, pp. 588–602, Feb. 2018.

[3] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2021.

[4] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.

[5] X. Tang, Q. Ma, X. Zhang, F. Liu, J. Ma, and L. Jiao, "Attention consistent network for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2030–2045, 2021.

[6] J. Qu, S. Hou, W. Dong, Y. Li, and W. Xie, "A multilevel encoder–decoder attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518113.

[7] X. Lu, J. Ji, Z. Xing, and Q. Miao, "Attention and feature fusion SSD for remote sensing object detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[8] N. Merrill and C. C. Olson, "Unsupervised ensemble-kernel principal component analysis for hyperspectral anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 507–515.

[9] J. Qu, Q. Du, Y. Li, L. Tian, and H. Xia, "Anomaly detection in hyperspectral imagery based on Gaussian mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9504–9517, Nov. 2021.

[10] R. Siddalingappa and S. Kanagaraj, "Anomaly detection on medical images using autoencoder and convolutional neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 148–156, 2021.

[11] B. Staar, M. Lütjen, and M. Freitag, "Anomaly detection with convolutional neural networks for industrial surface inspection," *Proc. CIRP*, vol. 79, pp. 484–489, Jan. 2019.

[12] Y. Chen, H. Zhang, Y. Wang, Y. Yang, X. Zhou, and Q. M. J. Wu, "MAMA net: Multi-scale attention memory autoencoder network for anomaly detection," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 1032–1041, Mar. 2021.

[13] Z. Xie, J. Hu, X. Kang, P. Duan, and S. Li, "Multilayer global Spectral–Spatial attention network for wetland hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518913.

[14] Y. Gui, W. Li, X.-G. Xia, R. Tao, and A. Yue, "Infrared attention network for woodland segmentation using multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5627214.

[15] Q. Yang, Y. Xu, Z. Wu, and Z. Wei, "Hyperspectral and multispectral image fusion based on deep attention network," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–5.

[16] S.-S. Young, C.-H. Lin, and Z.-C. Leng, "Unsupervised abundance matrix reconstruction transformer-guided fractional attention mechanism for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 5, pp. 9150–9164, May 2025.

[17] A. Akhtarmanesh, D. Abbasi-Moghadam, A. Sharifi, M. H. Yadkouri, A. Tariq, and L. Lu, "Road extraction from satellite images using attention-assisted UNet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1126–1136, 2024.

[18] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[19] W. Zhang, H. Guo, S. Liu, and S. Wu, "Attention-aware spectral difference representation for hyperspectral anomaly detection," *Remote Sens.*, vol. 15, no. 10, p. 2652, May 2023.

[20] J. Wang, T. Ouyang, Y. Duan, and L. Cui, "SAOCNN: Self-attention and one-class neural networks for hyperspectral anomaly detection," *Remote Sens.*, vol. 14, no. 21, p. 5555, Nov. 2022.

[21] V. Machairas, M. Faessel, D. Cárdenas-Pena, T. Chabardes, T. Walter, and E. Decenciére, "Waterpixels," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3707–3716, Nov. 2015.

[22] F. Argüello, D. B. Heras, A. S. Garea, and P. Quesada-Barriuso, "Watershed monitoring in Galicia from UAV multispectral imagery using advanced texture methods," *Remote Sens.*, vol. 13, no. 14, p. 2687, Jul. 2021.

[23] J. López-Fandino, Á. Ordóñez, P. Quesada-Barriuso, A. S. Garea, F. Argüello, and D. B. Heras, "Galician rivers multispectral anomaly detection dataset," 2025. Accessed: Feb. 12, 2025, doi: 10.5281/zenodo.14852117.

[24] J. Liang, "Image classification based on RESNET," *J. Phys., Conf. Ser.*, vol. 1634, no. 1, Sep. 2020, Art. no. 012110.

[25] R. Rad, "Vision transformer for multispectral satellite imagery: Advancing landcover classification," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jun. 2024, pp. 8176–8183.

[26] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, and C. Malossi, "BAGAN: Data augmentation with balancing GAN," 2018, *arXiv:1803.09655*.

[27] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.