Electronic Letters on Computer Vision and Image Analysis 24(2):248-272, 2025

Channel Attention based CNN for Improving Unmanned Aerial Vehicle Classification using Micro-Doppler spectrogram Images.

Mainak Bandyopadhyay [1] , Suddhasattwa Das [1] and Subhankar Mishra [2]

[1] *School of Computer Engineering, KIIT Deemed-to-be-University Bhubaneswar, Odisha, India*
[2] *School of Computer Sciences, National Institute of Science Education and Research, Bhubaneswar*

Received 25th of March, 2025; accepted 2nd of December, 2025

Abstract

The increase in technological advancements in Unmanned Aerial Vehicles (UAVs) has led to challenges in detecting drones in flight. The micro-Doppler signatures obtained from radars are used to distinguish and detect different types of drones. Due to relatively similar radar spectrogram image patterns or micro-Doppler signatures, it is sometimes very challenging to classify different types of drones. Previously, Deep Learning methods like transfer learning and residual networks have been proposed to improve the classification accuracy. For further improving the classification efficiency, this paper investigates the integration of channel attention mechanisms like Squeeze and Excitation Net, Efficient Channel Attention, and Gated Channel Transformation in the custom CNN Network (UAVDetect) with three publicly available Micro-Doppler spectrogram UAV datasets. The paper proposes Modified SENet and Modified ECA, which further improves accuracy and convergence. The proposed models with modified channel attention achieves comparable accuracy of 97.13% with comfortable mean inference time of 62 milliseconds on DIAT-μSAT dataset and outperforms in RRN-ATR-Net dateset with 99.6% accuracy and mean inference time of 95 milliseconds.

Keywords: Unmanned Aerial Vehicle, Convolution Neural Network , Channel Attention Mechanism and Micro-Doppler spectrogram Image.

1 Introduction

The rapid development of small cross-sectional drones has provided a tremendous challenge in terms of security and surveillance. The different types of drones nowadays available sometimes camouflage with the environment in such a way that it is undistinguishable by naked eyes and other regular equipments. One of the developments in this area is the use of Doppler radars for the detection of the presence of UAVs. The usage of Doppler radar signature ensures sensing without the constraint of illumination, weather conditions, and sometimes preserves the privacy as the data are spectrographic signatures [1]. Due to these advantages, the Micro-Doppler radar is rapidly used for detection and classification of small UAVs [4]. The radar cross-section signatures provide the

Correspondence to: mainak.bandyopadhyayfcs@kiit.ac.in

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2374
ELCVIA ISSN: 15108-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

characteristic features related to the flying UAVs and objects like birds, drones, quadcopters, etc. However, due to small radar cross-section, it is difficult to classify between different types of UAVs and between UAVs and birds. The Doppler effect of a radar is characterized by the frequency shift or wavelength change generated from the radar signals reflected by the moving target. The micro-movements of parts of the objects contribute to the micro-frequency shifts, also known as the micro-Doppler effect. The micro-Doppler signatures are produced by the moving parts of the objects and therefore depend on the number of rotors, orientation, and speed. The movements are basically the vibrations, rotations, and other types of unique movement patterns specific to the objects. The features extracted from micro-Doppler signatures are used for classifying UAVs. Initial approaches experimented with machine learning techniques like SVM, Naive Bayes classification with extracted features [2][3][4] [5]. Deep Learning is popularly used for extracting and learning features from spectrogram images related to targets for classification [8]. Deep Learning approaches basically custom CNN architectures, transfer learning, and popular residual CNN architectures like ResNET, VGG16, and VGG19 [6], inception-residual neural network (IRNN) [7] are proposed for improving accuracy. From a deep learning perspective, the challenge is the unavailability of diverse samples of correlated UAVs, the small number of classes, and samples taken under various conditions (controlled and uncontrolled environments). Challenges arise in accurately classifying UAVs using micro-Doppler signatures due to their small radar cross-sections and diverse motion patterns. Existing methods suggest the scope for further improvement in the accuracy and training convergence using attention mechanisms. One of the possible improvements can be achieved by employing channel attention blocks in the CNN architectures for extracting distinctive and relevant features.
In the rest of the paper, section 3 and section 4 discuss various deep learning approaches for UAV classification and various popular channel attention mechanisms. Section 5 provides the proposed CNN backbone architecture (UAVDetect), modification of SE and ECA channel attention modules, and integration of channel attention mechanism in the UAVDetect. The experimental results are carried out using three publicly available micro-Doppler spectrogram datasets.

2 Contributions

The following are the key contributions of this paper:-

- The paper identifies that previous works in this domain lack in exploring the use of channel attention mechanism for improving the classification accuracy. Thus, this paper contributes by demonstrating the effectiveness of channel attention mechanisms, i.e., SE-Net, ECA-Net, and GCT, in UAV classification using micro-Doppler spectrogram images.

- The paper proposed modified channel attention mechanisms i.e., Modified SENet and Modified ECA, for improving the accuracy of the UAV spectrogram classification. The integration of these attention mechanisms in the backbone CNN architecture (UAVDetect) demonstrates improvement in accuracy and early training convergence. The accuracy can be further improved by changing the backbone CNN architecture.

3 Related works

3.1 micro-Doppler Radar Signature for UAV

The micro-radar signatures is generated from the motion patterns of the components of the flying object. Out of the various microwave bands L and S band are not generally used for the detection of small UAV. The radar system for the detection of small UAVs can not classify based only on the radar cross-section of the flying object. Various classes of drones and birds have the same radar cross-sections; therefore, for classification, micro radar signatures are required. Mostly, X-band (10GHz), K bands, and W bands are used for collecting

micro-Doppler radar data. X-band data is used in [4], [9],[10], [11],[20],[2],[6] for obtaining spectrogram data for classification and detection of different types of drones and birds. [4] obtained data with X-band (9.5 GHz) for the classification of planes, birds, stationary rotors, quadrocopters, and helicopters. [10] obtained data using X-Band for different maneuvers of drones and seagull birds near the coastal area for the discrimination of drones from seagull. [11] obtained micro-radar data from X-band for different UAV classification (quadcopter, bionic bird-Metafly, and bionic bird-Disco). In [6], X-band (10 GHz) radar is designed for collecting the Micro-Doppler signatures of various types of drones, i.e. a two-blade rotor, a three-short-blade rotor, a three-long-blade rotor, a quad-copter, a bionic bird, and a two-blade rotor and bionic bird. [5] proposed the use of multiple bands, S-band, C-band, and W-band, respectively, for collecting data of two different drones and their different motions.[12] collected K band (24 GHz) and W band (94 GHz) distinctive micro radar signatures for flying objects, drones, and birds. The use of high-frequency bands generates better resolution Doppler signatures, leading to fine features for UAV classifications. [2] proposed using a fusion of micro radar spectrogram data obtained from K-band and X-band for robust classification of drones (Quadcopter, Helicopter, and Hexacopter). [13] proposed the use of Ku-band frequency modulation continuous wave radar signal. The data set merged the Micro-Doppler signature radar (MDS. data with the cadence velocity diagram, which is a frequency domain representation MDS.[14] uses the continuous wave K-band (24Ghz) for the detection and classification of UAVs.[3] effectively demonstrated the use of a single channel Ka band continuous wave system to classify different classes of UAVs like quadcoptor, octocopter, helicopter and fixed wing plane.

3.2 Classification Approaches for UAVs

The approaches for classification of drone micro radar signature spectrogram images vary from machine learning approaches to deep learning approaches. The machine learning approaches for UAV classification are largely dominated by feature extraction methods like PCA, robust PCA, and Singular Value Decomposition, and classifier models like linear SVM, nonlinear SVM models, and decision trees.

For classification of small radar section UAVs, in [4] linear SVM, non-linear SVM, and Naive Bayes models are trained on features extracted from Micro-Doppler radar spectrogram images using PCA. In addition to classifying birds and different types of UAV or drones, [3] [2] proposed feature extraction using PCA and singular value decomposition and classification using a trained SVM classifier. For cost-effective classification [37] demonstrated the use of a low-cost embedded classifier using SVM and robust PCA for outlier rejection. Alternatively, [38] proposed the use of robust PCA and a multi-Gaussian model for classification. For classification using UAV visual images, [41] employed multi-domain visual saliency to extract features, and an SVM classifier was trained for classification. In [39], micro radar signatures of multiple radars are extracted using a singular value decomposition algorithm, i.e., orthogonal matching pursuit algorithm, and classification is performed using SVM and decision tree.

Recently, Deep Learning based classifiers have become a popular choice for UAV detection. The classification problem includes classifying various types of drones and classification of drones from birds. The convolution layers in CNN deep learning architecture extracts a diverse range of features from spectrogram images of micro-Doppler radar data. Initially, [8] demonstrated the possible use of DCNN for both feature extraction and classification of drones based on micro-Doppler radar signature. Multiple parallel convolutions with 3 x 3 , 5 x 5 and 7 x 7 size filters is proposed in CNN-32DC [19] which is trained on RAD-DAR dataset. An improvement over CNN-32DC is proposed by [40] using more layers of multiple parallel blocks. [28] used fully connected convolution for fast detection by proposing to set the effective receptive field for the kernel to the size of sliding window. The proposed using kernels of different sizes i.e. 3 X 5 , 2 X 4 to extract shallow features and 3 x 3 and 1 x 1 to extract more specific features. Also, for real time classification [23] proposed an embedded CNN model for real time recognition with a single convolution layer with 32 filters of size 3 x 3. The proposed model reported better accuracy than the MobileNetV2 and NasNet models. [45] proposes a real-time method for classifying small drones using high-quality micro-Doppler signature (MDS) images from FMCW radars and an ultra-lightweight convolutional neural network (CNN).

The progress for accuracy in classification is lead by transfer learning and residual networks. For better accuracy, pre-trained models are used for classification. In [33] experiments are carried out using pre-trained models like VGG19 , ResNet-18 and ResNet-50 for transfer learning. The work observed that ResNet-18 trains faster than VGG19 with better accuracy. Another exhaustive experimentation by [6] , trained the Micro-Doppler radar drone dataset i.e. DIAT-μSAT dataset using transfer learning. The experiments included pretrained models i.e. VGG16, VGG19, MobileNetV2, InceptionResNetV2, InceptionV3, and DenseNet-201. The experiments reported that VGG16 and VGG19 achieve better accuracy than other pre-trained models with the DIAT-μSAT dataset. In [13] pre-trained GoogLeNet is applied over merged MDS and cadence velocity diagram data. [20] trained a CNN model using the data obtained by the Martin-Mulgrew Model for radar returns from rotating blades. The trained CNN model is for for transfer learning with the real data of drones. Number of works proposed models using Residual Networks involving skip connections. The use of residual connection or skip connection for the detection and classification of different drones with improved accuracy is proposed in [6]. [11] applied residual network ResNet-18 to overcome the vanishing gradient problem. The proposed model, ResNet-SP [11], removed the 5th block in ResNet-18 to classify drone activities and human activities from the radar spectrogram dataset. [7] trained a classification model on RAD-DAR dataset with 2 inception-residual blocks with 16 filters and ELU activation function. [24] trained ResNet-50 architecture on radar signatures for drones and birds collected using K-Band with external noise. A fusion-based approach is proposed in [27] by employing ResNet34 model by fusion at both the data level and feature level.

Apart from these aforementioned approaches, other deep learning approaches are also explored. An adversarial model employing GAN for generating a simulated spectrogram to assist in limited sample conditions is suggested in [21]. [26] exploited the high-resolution spectrum sequences to extract the motion patterns of drones and birds. The LSTM model is used for the classification of sequences. [1] proposed a Hybrid Convolutional Neural Network-Memetic (CNN-Memetic) algorithm for classifying flying objects, including drones, helicopters, and birds, using Micro-Doppler Signature (MDS) data from a UAV-mounted radar with promising results in a dynamic environment.

Based on the literature review carried out, the current state of the art in the domain of deep learning for UAV classification using micro-Doppler signature is mostly the classification carried out using either custom CNN models or by fine-tuning pre-trained models. The main objective of the work is to improve the accuracy with a robust model. However, there is still scope to further refine the accuracy of the UAV classification by using attention mechanisms to focus on the specific features of the micro-Doppler signature images.

Another focus area for further development is to improve the inference time and detect multiple UAVs in the same spectrogram radar image.

4 Background

4.1 Channel Attention Mechanism

Channel attention mechanisms endeavor to recalibrate channel weights to enhance the representation abilities of networks. The objective of the channel attention method is to learn the aggregation of global spatial information and adaptively recalibrate the weight of each channel. Given a layer i with c convolution kernels producing feature maps as a 3D tensor with dimensions h x w x c, where c represents the channels in the feature map. The objective is to compute an attention vector that consists of attention weights for each channel such that the representation of more prominent features is weighted significantly than less prominent features.

Popularly, Squeeze and Excitation network, Efficient Channel Attention, and Global Convolution Transformation are used as channel attention mechanisms.

- Squeeze and Excitation (SE Attention module) [29]:- The Attention module is designed for first aggregating global context in each channel (squeeze) and then recalibrating by capturing information interdependence between channels (excitation). The aggregation of contextual information of channels is carried

out using Global Average Pooling on individual channels. The aggregated contextual information is fed into a fully-connected layer to obtain scaling weights for the channels using a sigmoid function. The fully connected layer helps in learning the cross-channel relationships implicitly.

Considering , the output tensor of a convolution transformation as $Y \in \mathbb{R}^{H \times W \times C}$ with , $Y = [y_1, y_2, y_3, \ldots\ldots, y_c]$ and the output of GAP $Z \in \mathbb{R}^{1 \times C}$ as $Z = [z_1, z_2, z_3, \ldots\ldots, z_c]$

The spatial aggregation using global average pooling is given as:-

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} y_c(i, j) \tag{1}$$

The channel-wise attention generated using non-linear fully connected layers is given as

$$s = \sigma(W_2 \delta(W_1 Z)) \tag{2}$$

Recalibrated by scaling the input feature maps with the learned attention weights

$$x_c = s_c \cdot y_c \tag{3}$$

where, $X = [x_1, x_2, \ldots, x_c]$

- Efficient Channel Attention (ECA Attention) [30]:- The attention mechanism focuses on the cross-channel interaction between the neighbors. The ECA attention mechanism replaces the fully connected layers with adaptive kernel size *k* as a function of the number of channels for a 1-D convolution kernel. The experimental results show that ECA outperforms the SE attention mechanism due to the avoidance of dimensional reduction and the incorporation of local cross-channel attention.

  In case of SE attention, the weights calculated by the fully connected layer are given by

$$\omega_i = \sigma \left(\sum_{j=1}^{C} w_i^j z_j \right), z_j \in Z \tag{4}$$

  where , $\omega_i \in W_k$.

  The weight calculation in SE considers the information from all the channels. The ECA restricts the cross-channel interaction locally to *k* neighbours. This is ensured using 1D convolution with *k* neighbouring channels given by

$$s = \sigma \left(conv_1d_k(Z) \right). \tag{5}$$

  where , $conv_1d_k(Z)$ is 1-dimensional convolution kernel of size *k* and $s \in \mathbb{R}^{1 \times C}$

- Gated Channel Transformation (GCT Attention) [31]:- GCT Attention is based on the concept of co-operation and competition. GCT, through cooperation in the shallow layers of the deep learning models, allows learning low-level attributes. Also, competition allows learning high-level discriminative features in deeper layers. The GCT, inspired by SE attention, includes normalization for modeling cross-channel interaction in place of a fully connected layer. The channel normalization layer improves computational complexity by reducing the number of parameters. An adaptive gated mechanism (1+tanh(x) gate activation) is employed to facilitate cooperation and competition. The gating weights, when activated positively, ensure competition by the channel with other channels. The negative gating weights of a channel imply cooperation with other channels.

4.2 Overall System for UAV micro-Doppler signature Classification.

In this section, we present the overall system for the classification of UAVs. The system consists of three components:-

(i) *Sensor and Radar System for capturing the signals.*

The signals are transmitted and then received by the sensor systems. The received signals capture have two components as shown in Eq.6, i.e., (i) translational movement by the UAVs and (ii) rotational movements by the rotors of the object generate the micro-Doppler signals provided in Eq.7.

$$f_D = \frac{2f}{c}[\boldsymbol{v}]_{\text{radial}} + \frac{2f}{c}[\boldsymbol{\omega} \times \boldsymbol{r}]_{\text{radial}} \tag{6}$$

$$f_{microdoppler} = \frac{2f}{c}[\boldsymbol{\omega} \times \boldsymbol{r}]_{\text{radial}} \tag{7}$$

(ii) *Processing system to convert signals into spectrogram images.*

The Micro-Doppler signals captured by the receiver is first processed using a low-pass filter in order to filter out high frequency components, after which it is provided to a short-time Fourier transform module to generate the spectrogram images. Given an input signal as a function of time f(t) , the STFT proceed using a temporary sliding window w (t) and applies a Fourier transform at each time shift (τ) [32].

The continuous form of STFT is provided in Eq.(8)

$$\text{STFT}\{f(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} f(t)w(t-\tau)e^{-i\omega t}dt \tag{8}$$

The discretized form of STFT that is used for practical application is provided in Eq. (9)

$$\text{STFT}(d, k) = \sum_{n=0}^{N-1} f(n)w(n-d)e^{-\frac{j2\pi kn}{N}}. \tag{9}$$

The spectrogram is generated by taking the squared magnitude of the resulting STFT. The time width of the window function determines the instantaneous frequency resolution of the resulting spectrogram

$$\text{spectrogram}\{f(t)\}(\tau, \omega) \equiv |X(\tau, \omega)|^2 \tag{10}$$

(iii) *Classification module of spectrogram images.* Deep learning-based models are trained on various classes of UAV's. CNN models extracts distinctive spectrogram features from the training dataset and the deep layers perform classification. The trained model is used to classify the spectrogram images into different classes of UAV.

5 Channel Attention based CNN for Micro-Droppler UAV detection

5.1 Channel Attention based Deep Learning Module.

For understanding the effectiveness of the attention module, the investigations are carried out on two different publicly available datasets i.e. Real Doppler RAD-DAR (RDRD) dataset and DIAT-μSAT dataset. The Real Doppler RAD-DAR dataset consists of spectrogram images of drones, cars, and people. The DIAT-μSAT dataset consists of micro droppler spectrogram images of 5 different drone categories and images of mixed classes.

Given that the spectrogram images in both datasets have different dimensions and classification objectives. Two separate backbone architectures are used for each dataset. The backbone deep learning architectures are experimented with various channel attention modules. Rather than using transfer learning, the paper is motivated towards designing a model based on the varied designs of CNN architectures provided in the literature

and trained from scratch. Given an input tensor $X \in \mathbb{R}^{H\times W\times C'}$ and a convoluted transformed output $Y \in \mathbb{R}^{H\times W\times C}$. Let $F = [f_1, f_2, f_3 \ldots\ldots\ldots, f_c]$ are the set of kernels or filters at the transformation layer. The output of convoluted transformation $(*)$ is given as $Y = [y_1, y_2, y_3, \ldots\ldots, y_c]$, where

$$y_c = f_c * X = \sum_{i=1}^{c'} f_c^i * x^i \tag{11}$$

The kernel set is defined as

$$f_c = \left[f_c^1, f_c^2, f_c^3, \ldots, f_c^{c'}\right] \tag{12}$$

where each f_c^i is a 2D spatial kernel that operates on the i^{th} channel of the input tensor X, denoted by x^i. As shown in Eq. 11, the c^{th} output channel is obtained by summing the responses across all input channels,

$$y_c = \sum_{i=1}^{c'} f_{c,i} * x^i.$$

Thus, the interdependence of channels is implicitly encoded in the output tensor y.

Therefore, each transformative convolution forwards the interdependence information. Channel attention mechanisms focuses on explicitly selecting the channel dependency so that more informative channels can be transferred to further transformation layers and the lesser one can be suppressed.

5.2 CNN Architectures and Channel Attention based Modification for UAV Detection (UAVDetect)

In this study, to understand the effects of channel attention, two different backbone CNN architectures are considered to accommodate various sizes of spectrogram images. The backbone 1 (UAVDetect1) is considered for input tensors with spectrogram image of small sizes i.e. (11 x 61 , 32 x 32). The backbone 1 is used as up-scaling the small spectrogram image may affect the information content that is provided as input to the CNN model. Rather than a deep network, a shallow neural network is used to extract the features. The backbone 2 (UAVDetect2) is considered for input tensors with spectrogram images of larger size i.e.(224 x 224). The backbone 2 is a deeper network inspired by VGG architecture.

5.2.1 Backbone Model 1 (UAVDetect1)

The CNN network backbone as shown in Fig.1 consist of consecutive layers of convolution kernels with size (3 x 3) with 'Swish' activation function. Global Average Pooling is applied for consolidation of feature extracted in CNN layers. In the dense layer a single layer with 512 neurons are considered with dropout of (0.5). The input layer shape and output neurons depends on the dataset and number of classes. The backbone layer is straightforward and simple to understand the effect of channel attention mechanisms. The total number of trainable parameters are 162,499 with (634.76 KB) memory.

5.2.2 Backbone Model 2 (UAVDetect2)

The backbone architecture, as shown in Fig.2 for feature extraction, consists of 4 blocks of 2D convolution kernels with max-pooling layers and batch normalization. Each block in the backbone architecture consists of two 2D-Convolution Layers of kernel size (3 x 3) with ReLU activation function. For regularization and aggregation, Batch Normalization and MaxPooling layers are incorporated. The fully connected layer consists of 4096 dense neurons followed by a 0.5 dropout layer. The output layers consist of 6 neurons and 4 neurons for the DIAT-μSAT dataset and RRN-ATR-Net dataset, respectively. The total number of parameters is 6,827,334 with a total memory of 26.04 MB. Additionally, a custom VGG19-inspired architecture is employed to introduce the

features extracted from shallow blocks to deeper blocks. In order to investigate the effectiveness of the channel attention mechanisms in improving the accuracy and computational time the attention modules are placed after each block to identify and weight the channels output from each block.

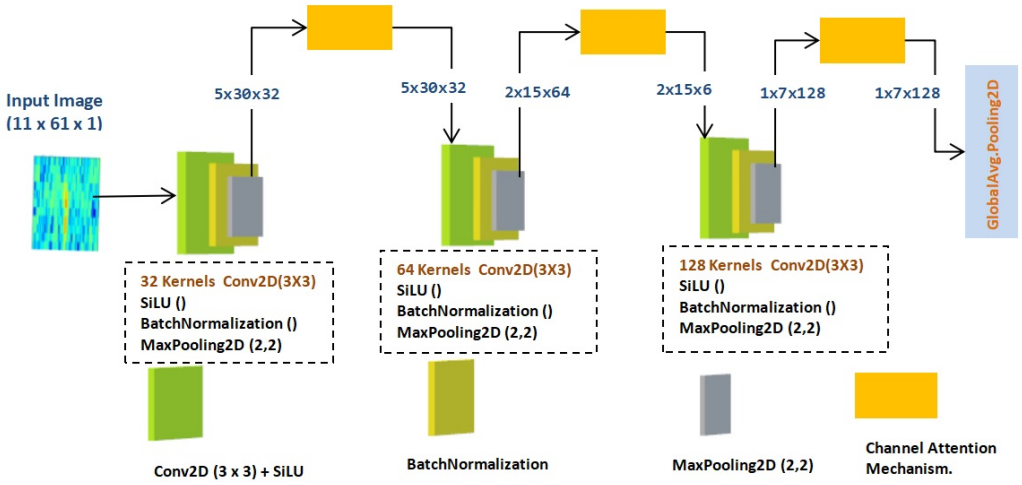


Figure 1: Template Backbone CNN Architecture for Feature Extraction in RAD-DAR dataset.

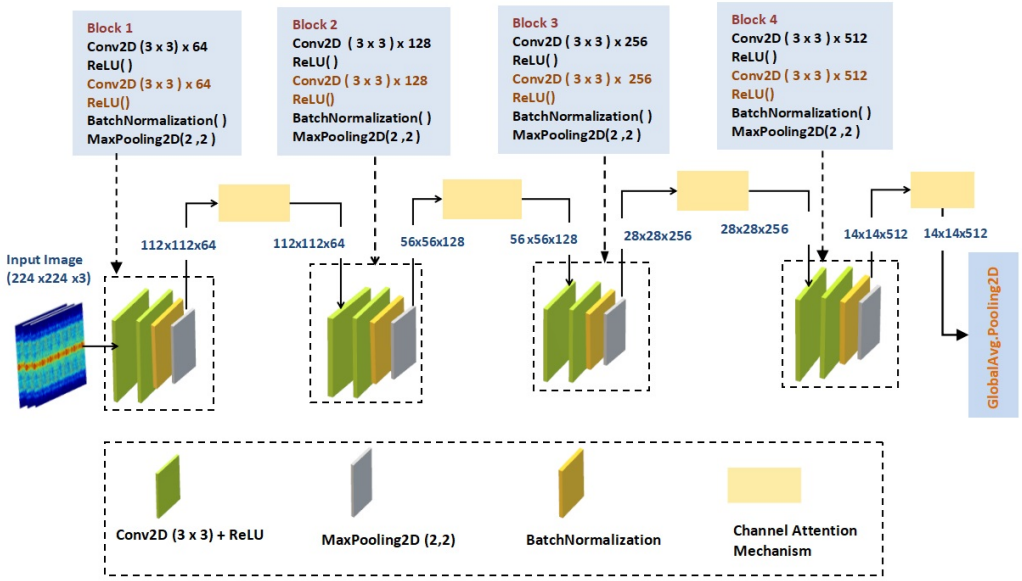


Figure 2: Template Backbone CNN Architecture for Feature Extraction in DIAT-μSAT and RRN-ATR-Net dataset.

Table 1: Details of UAV Detection Architectures

Feature	**UAVDetect1 (RAD-DAR)**	**UAVDetect2 (DIAT-μSAT / RRN-ATR-Net)**
Input Size	(11×61x1)	(224×224x3)
Architecture	Shallow CNN (3 layers)	Deep CNN (4 blocks, Custom VGG19-inspired architecture)
Convolution Layers	3 (32→64→128 filters)	4 blocks (2 conv layers per block, 64→128→256→512 filters)
Kernel Size	(3×3)	(3×3)
Activation	Swish	ReLU
Pooling	GlobalAveragePooling2D	MaxPooling2D (2×2) per block
Batch Normalization	No	After every Conv2D layer
Dense Layers	512 neurons	4096 neurons
Dropout Rate	0.5	0.5
Output Layer	softmax (3 classes)	softmax (6 / 4 classes)
Optimizer	Adam (lr=0.0001)	Adam (lr=0.0001)
Loss Function	SparseCategoricalCrossentropy	CategoricalCrossentropy
Batch Size	16	16
Epochs	50	50
Early Stopping	No	No
Learning Rate Schedule	ReduceLROnPlateau (factor=0.1, patience=3)	ReduceLROnPlateau (factor=0.1, patience=3/4)
Data Augmentation	None	Shear (0.2), Zoom (0.2), HorizontalFlip=True
Attention Mechanisms	Modified SE, Modified ECA, GCT	Modified SE, Modified ECA, GCT
Attention Module Parameters	1627	20482
Total Parameters	162,499 (634.76 KB)	6,827,334 (26.04 MB)
FLOPs	0.0053 G	10.3 G

5.2.3 Integrating Channel Attention on Backbone Model

The parameters and modifications done in SENet, ECANet and GCT are as follows:-

- *Intuition behind using Dual Pooling in Channel Attention*:- Unlike natural images, where pixel intensities often follow relatively smooth spatial correlations, spectrograms are characterized by sparse, heavy-tailed distributions. A large proportion of the pixels correspond to low-energy background or noise, while a small fraction of pixels carry disproportionately high values generated by rotor dynamics. In micro-Doppler spectrograms, the discriminative features that differentiate UAV classes are typically localized, high-energy patterns arising from rotor dynamics, such as harmonic lines generated by propeller rotation, transient peaks from blade-tip motion, or side-band modulations induced by periodic rotor movements.

  Global Average Pooling (GAP), though effective for capturing overall spectral context, dilutes localized high-energy events by averaging them with the dominant low-energy background. In contrast, Global Max Pooling (GMP) preserves the strongest activation within each channel, ensuring that rare but highly

informative spectral peaks linked to UAV rotor mechanics are retained in the channel descriptor. By integrating both, the modified attention modules are better aligned with the statistical structure of spectrogram images, achieving improved robustness and sensitivity to class-specific UAV micro-Doppler signatures.

- The SENet is modified to consider both Global Average Pooling and Global Maximum Pooling as shown in Fig. 3.

  The convoluted transformed output $Y \in \mathbb{R}^{H\times W\times C}$ is downsized with Global Average Pooling and Global Maximum Pooling, to extract channel-wise global descriptors (Z and Z^1) using both average and max pooling, as defined in Eq. (13), which capture the overall channel context and the most salient activations, respectively.

$$Z(c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} Y_c(i,j), \quad Z^1(c) = \max_{i,j} Y_c(i,j), \quad c = 1, \ldots, C \tag{13}$$

  To model cross-channel dependencies efficiently, a bottleneck transformation with reduction ratio r is applied to both descriptors as shown in Eq. 14. The transformed descriptors are fused and passed through a second linear layer followed by a sigmoid activation to generate the channel attention vector as shown in Eq.19. Finally, the input feature map is recalibrated via channel-wise multiplication Eq.16, effectively emphasizing informative channels while suppressing less relevant ones.

$$h = \delta(W_1 Z), \quad h' = \delta(W_1 Z') \tag{14}$$

  where, $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $\delta(\cdot)$ is the ReLU activation.

$$A = \sigma\big(W_2(h + h')\big), \quad A \in \mathbb{R}^{1\times 1\times C}, \tag{15}$$

  where, $W_2 \in \mathbb{R}^{C\times \frac{C}{r}}$.

$$X_c = A_c \cdot Y_c, \quad c = 1, \ldots, C, \tag{16}$$

  For the channel attention in both original and modified SENet a reduction ratio (r) of 8 is used based on an ablation study provided in the Table.2.

Table 2: Accuracy comparison in different Reduction Ratio

Dataset	**r=4**	**r=8**	**r=16 [29]**
DIAT–μ SAT	96.16	96.56	96.28
RRN–ATR–Net	98.08	98.33	98.11
RAD–DAR	96.91	97.08	96.83

- For ECANet, along with the original attention network [30], here a modified version is proposed as shown in Fig. 4. In case of Modified ECA, similar to Modified SENet, the convoluted transformed output $Y \in \mathbb{R}^{H\times W\times C}$ is downsized with Global Average Pooling and Global Maximum Pooling as provided in Eq. (13). The Z captures the overall contextual information of each channel, whereas Z^1 emphasizes the most salient activations. To efficiently model local cross-channel interactions, a 1D convolution with an adaptive kernel size k where C is the channel dimension ($\gamma = 2$, $b = 1$) [30], defined in Eq. (17), is applied to both descriptors producing intermediate responses h and h' as shown in Eq. (18). The

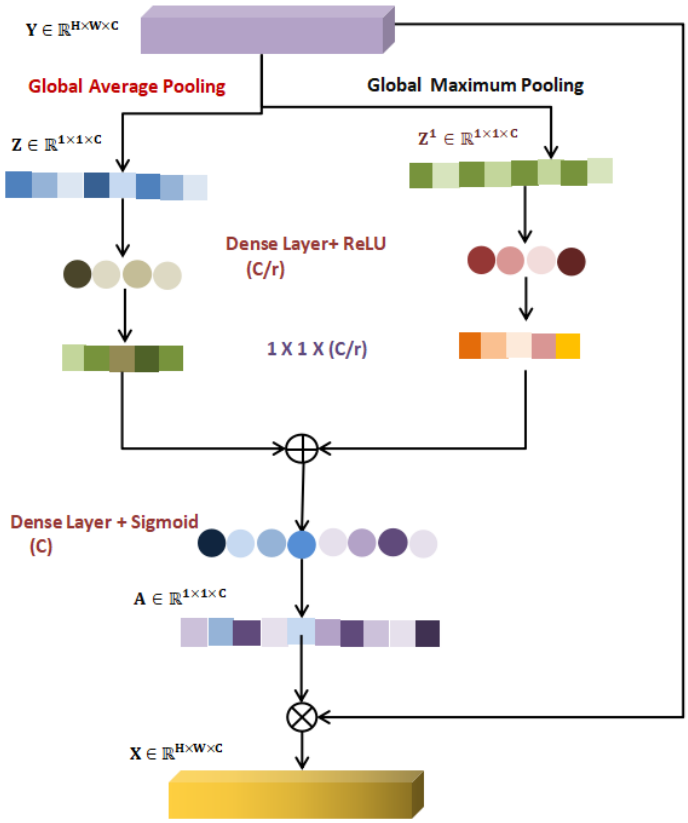


Figure 3: Modified Squeeze and Excitation Network.

two responses are subsequently fused and passed through a sigmoid activation to generate the channel attention vector A, as given in Eq. (19).

$$k = \left| \frac{\log_2(C)}{\gamma} + b \right|_{\text{odd}}, \tag{17}$$

$$h = \text{Conv1D}_k(Z), \quad h' = \text{Conv1D}_k(Z') \tag{18}$$

$$A = \sigma(h + h'), \quad A \in \mathbb{R}^{1\times1\times C}, \tag{19}$$

The input feature map is recalibrated through channel-wise multiplication, as expressed in Eq. (20):

$$X_c = A_c \cdot Y_c, \quad c = 1, \ldots, C. \tag{20}$$

This mechanism effectively emphasizes informative channels while suppressing less relevant ones. By combining the efficiency of local cross-channel modeling, inspired by ECA-Net, with complementary average and max pooling descriptors, it provides a compact and effective attention module.

- As suggested in [31], the performance of GCT attention is better when placed before the convolution layers. The choice of l-norm i.e., l1-norm and l2-norm for the global context embedding performs equally with consistent non-negative weights. Here, in experiments, it is also observed that the L1 and L2 norms do not significantly alter the accuracy. Here, we have considered L2-norm for global context embedding, with the value of epsilon being 1e-5.

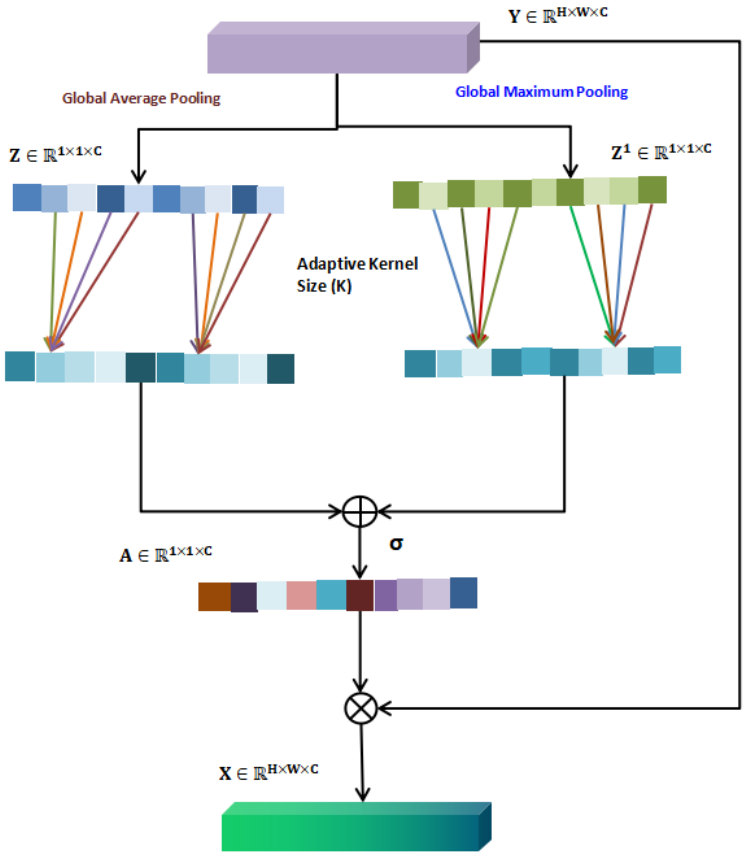


Figure 4: Modified Efficient Channel Attention.

1. **Attention Mechanism in UAVDetect1:-** The number of classes in the RAD-DAR dataset is only 3. The dataset contains spectrograms of different objects, i.e. car, people and drones altogether. The input image size is 11 x 61. The channels in the feature extraction layers are 32, 64 and 128. The attention blocks are placed after each layer of the backbone model 1 for highlighting the significant channels.

2. **Attention Mechanism in UAVDetect2:-** The channel attention modules guide the network towards selecting channels suitable for classification. In the Backbone Model 2, the block 1, block 2, block 3 and block 4 outputs 64, 128, 256, and 512 channels respectively. The spectrogram signature images in DIAT-μSAT dataset has a recurring pattern within the images. Thus, across different channels of the output activation, it is very difficult to observe diversity, unlike other object classification problems. As an initial attempt, the attention blocks are placed after each convolution block to enforce the significant channels across each block. In this modification, the attention module is not applied to the *stem* of the model. The initial layers of the model, also termed as stem, is involved in extracting more local features [42]. In case of large images the receptive field of 7 x 7 with stride of 2 is used to extract local features which in the later layers (with receptive field 3 x 3) can be associated with global features. In case of RRN-ATR-Net dataset, there are very less visibly distinctive features present in different classes of UAVs. The attention modules are applied after each block and also after the last two blocks. However, in this case also by applying the attention modules after blocks with sufficiently large number of channels the accuracy is satisfactory.

6 Experimental Setup and Dataset Description

The training simulations are carried out on NVIDIA GeForce RTX 3060 GPU with 16 GB GDDR6 memory and with 3584 CUDA cores. The model architecture is implemented using Python(3.9.18), Tensorflow(2.10.0) and Keras(2.10.0) libraries.The additional libraries for the setup includes NumPy(1.26.3), Pandas(2.1.4), scikit-learn(1.4.1.post1) and Augmentor(0.2.12). For optimization of weights, Adam optimizer is used with an initial learning rate of 0.0001. Learning rate is reduced by a factor of 0.1 if a lower validation accuracy is not found in three consecutive epochs. The training is carried out end-to-end using the entire dataset with a batch size of 16 and 50 epochs. For cross dateset validation 5-fold validation is used.

6.1 Spectrogram Image Dataset description

Basically the investigations on the effectiveness is carried out using three publically available micro-Doppler dataset on UAV classification.

- The RAD-DAR dataset provided by the Microwave and Radar Group from the UPM consist of 17000 micro-doppler spectrogram samples of drones, cars and people. As provided in Fig.5 the samples have a dimension of 11 x 61. The data is collected using a frequency-modulated continuous wave (FMCW) on a frequency band centered at 8.75 GHz with a maximum bandwidth of 500 MHz. The objective is to classify between people, car, and drones.

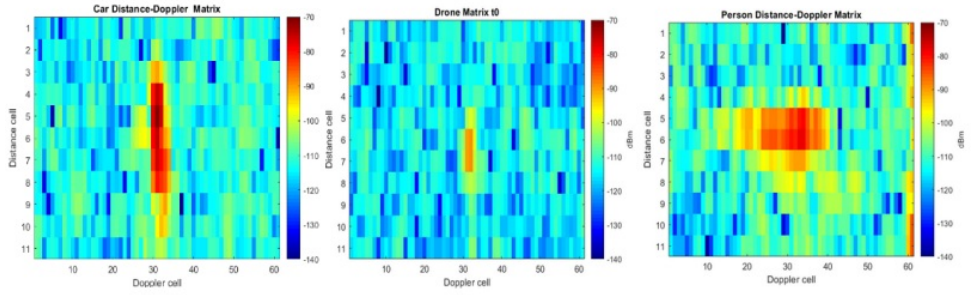


Figure 5: Micro-Doppler Samples from RAD-DAR dataset.

- DIAT-μSAT dataset collected using continuous wave (CW) (X-band: 10 GHz) radar comprises 4849 samples of micro-Doppler images of six different classes of small UAVs. The various small UAV targets include RC Plane, three-short-blade rotor, three-long-blade rotor, quadcopter, bionic bird, and mini-helicopter + bionic bird as shown in Fig.6. The objective is to classify small UAVs into different classes. To increase the number of samples, data augmentation is applied.
- RRN_ATR_Net dataset consists of micro-Doppler signatures of four different classes of UAVs, i.e., quadrotor (Phantom3s), fixed-wing (Cessna182), helicopter (T-REX450), and bionic bird (Gogo Bird1020) as shown in Fig.7. The dataset consists of 1200 sample signatures acquired using Texas Instruments' AWR1642BOOST radar sensor and the DCA1000EVM high-speed data acquisition card.

7 Result Analysis

The comparative results provided in the Table.3, Table.4, and Table.5 for DIAT-μSAT, RRN-ATR-Net, and RAD-DAR datasets indicate that Modified SE and Modified ECA attention mechanism performs better than the original channel attention mechanisms. Also, for the backbone architectures, the channel attention effectively increases the accuracy. However, the results are very contrasting in the case of DIAT-μSAT dataset. There is an increase in the accuracy with Modified ECA and Modified SE by 0.7% and 1.0 % respectively, in comparison

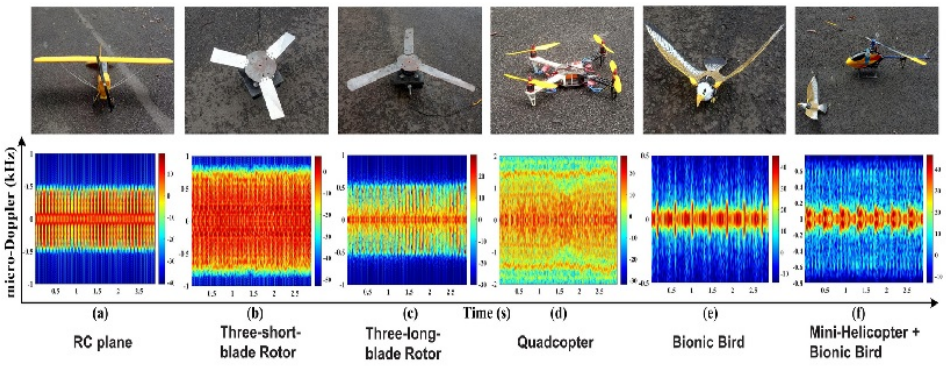


Figure 6: Micro-Doppler Samples from DIAT-μSAT dataset.

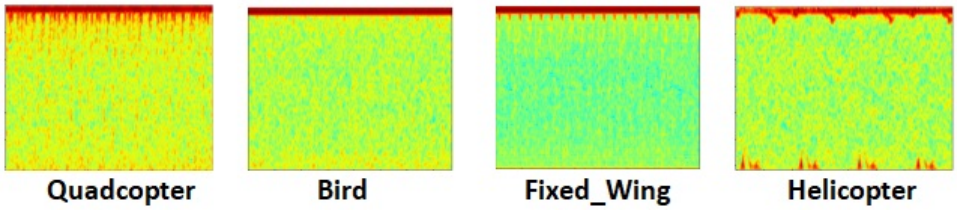


Figure 7: Micro-Doppler Samples from RRN_ATR_Net dataset.

to original ECA and SE channel attention. However, the performance of GCT attention is not better than SE attention in the case of DIAT-μSAT and RAD-DAR datasets. As provided in Fig.8 and Fig.10, the convergence of training is relatively early in the case of modified ECA in comparison to other channel attention mechanisms and without an attention mechanism. The detailed analysis for the different datasets is as follows.

- *Analysis of Classification Results for DIAT-μSAT Dataset*:- The DIAT-μSAT dataset demonstrates improved classification accuracy with the incorporation of channel attention mechanisms, particularly the Modified ECA Net, which effectively enhances the model's ability to select relevant features and focus on distinctive structural traits for each class. Classes with unique visual features, such as *Bird* and *Drone_1*, consistently achieve 100% accuracy across all models, highlighting their inherent separability. Rotor-based classes, like *3_long_blade_rotor* and *3_short_blade_rotor_1*, show notable performance gains; for instance, *3_long_blade_rotor* improves from 91.82% in the baseline model to 93.08% with Modified ECA Net, reflecting the efficacy of attention mechanisms in differentiating visually similar classes. Despite these advances, some structurally similar classes, such as *Bird+mini-helicopter_1* and *RC_plane_1*, still exhibit minor mis-classification trends, with overlapping aerodynamic and rotor structures contributing to confusion. The Modified ECA Net, while reducing errors compared to baseline and GCT models, achieves 93.87% accuracy for *Bird+mini-helicopter_1*, with most misclassifications occurring with *RC_plane_1* (3.86%). Overall, attention mechanisms enhance the differentiation of challenging classes but face limitations with structurally similar features, as reflected in confusion matrix trends.

- *Analysis of Classification Results for RRN_ATR_Net:*-The classification results for the RRN_ATR_Net model demonstrate that incorporating channel attention mechanisms significantly improves the model's ability to differentiate between structurally similar classes, particularly the *Quadrotor* class. Without attention mechanisms, the model achieves perfect classification accuracy (100%) for the *Bird*, *Fixed_Wing*, and *Helicopter* classes, while the *Quadrotor* class is classified with 91.67% accuracy, with 8.33% of its samples misclassified as *Fixed_Wing*. This mis-classification highlights the challenge posed by overlapping aerodynamic and structural characteristics between *Quadrotor* and *Fixed_Wing* classes. The introduction of SE_Net improves the *Quadrotor* classification accuracy to 93.33%, reducing the mis-

classification rate to 6.67% while maintaining 100% accuracy for the other three classes. ECA_Net delivers the highest improvement, achieving a *Quadrotor* classification accuracy of 98.33% and reducing the mis-classification rate to just 1.67%, indicating its superior capacity to focus on class-specific features. GCT also enhances the *Quadrotor* classification accuracy to 96.67%, with a 3.33% mis-classification rate, positioning it between SE_Net and ECA_Net in terms of effectiveness. Across all configurations, the *Bird*, *Fixed_Wing*, and *Helicopter* classes consistently achieve 100% accuracy, suggesting that these classes possess distinct and readily separable features that do not heavily rely on attention mechanisms for differentiation. Overall, the results reveal that attention mechanisms systematically improve the classification of *Quadrotor*, the most challenging class, by effectively amplifying distinguishing features and reducing confusion with similar classes. Among the mechanisms tested, ECA_Net proves the most effective, followed by GCT and SE_Net, emphasizing its ability to refine feature representation in aerial classification tasks.

- *Classification Results Analysis for RAD-DAR Dataset*:- The RAD-DAR dataset demonstrates notable classification performance improvements with the application of channel attention mechanisms. Without attention, *People* achieve the highest accuracy (98.73%), followed by *Drones* (96.54%) and *Cars* (94.32%), with mis-classifications primarily between *Cars* and *Drones*. Modified SE_Net improves performance across all classes, with *People* and *Drones* reaching 98.88% and 97.14% accuracy, respectively, and *Cars* at 94.93%. Modified ECA_Net further boosts *People* to 99.33%, though *Cars* drop slightly to 93.79%, reflecting ongoing mis-classification with *Drones*. GCT attention achieves balanced performance, with *People* at 98.88%, *Drones* at 96.35%, and *Cars* at 94.84%. Overall, Modified SE_Net and GCT effectively reduce confusion between *Cars* and *Drones*, while Modified ECA_Net excels for the *People* class, underscoring the role of channel attention mechanisms in addressing overlapping features among classes.

Table 3: F1-Score for different classes and channel attention mechanisms on various datasets.

Dataset	Classes	Without Attention	Original-ECA	Modified-ECA	Original-SE	Modified-SENET	GCT
DIAT-μSAT	3 long blade rotor	0.93	0.93	0.94	0.93	0.94	0.94
	3 short blade rotor	0.93	0.92	0.95	0.91	0.92	0.91
	Bionic-bird	1.00	1.00	1.00	1.00	1.00	1.00
	Bird+Mini-Helicopter	0.93	0.95	0.97	0.96	0.95	0.96
	RC Plane	0.95	0.94	0.97	0.94	0.95	0.93
	Quadcopter	1.00	1.00	1.00	1.00	1.00	1.00
	Accuracy	**95.56**	**96.35**	**97.14**	**96.18**	**96.56**	**96.00**
RRN-ATR-Net	Bird	0.98	1.00	1.00	1.00	1.00	1.00
	Fixed Wing	0.97	0.98	0.99	0.98	0.97	0.98
	Helicopter	0.98	0.98	1.00	0.99	1.00	1.00
	Quadcopter	0.97	1.00	0.99	0.97	0.97	0.98
	Accuracy	**97.50**	**99.17**	**99.60**	**98.33**	**98.33**	**99.17**
RAD-DAR	Car	0.95	0.96	0.95	0.96	0.96	0.96
	Drones	0.94	0.95	0.95	0.95	0.95	0.95
	People	0.99	0.99	0.99	0.99	0.99	0.99
	Accuracy	**96.33**	**96.88**	**96.63**	**96.94**	**97.08**	**96.83**

Table 4: Recall for different classes and channel attention mechanisms on various datasets.

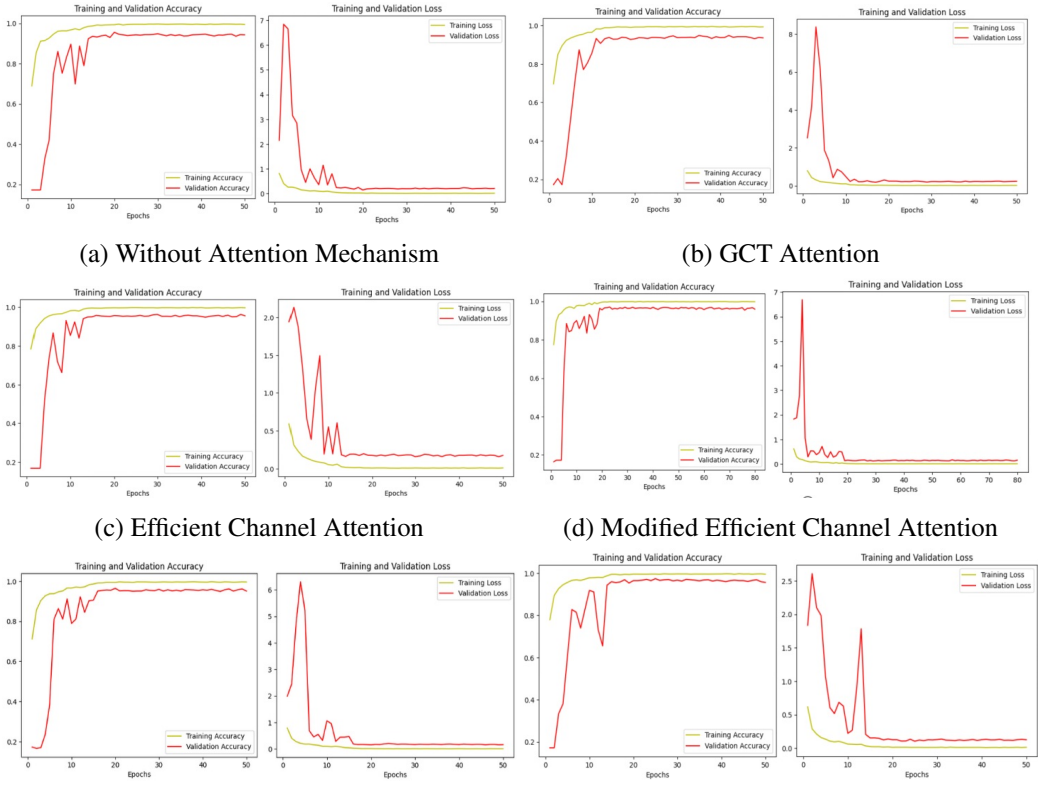
Dataset	Classes	Without Attention	Original-ECA	Modified-ECA	Original-SE	Modified-SENET	GCT
DIAT-μSAT	3 long blade rotor	0.92	0.91	0.93	0.92	0.92	0.92
	3 short blade rotor	0.96	0.97	0.97	0.97	0.97	0.96
	Bionic-bird	1.00	1.00	1.00	1.00	1.00	1.00
	Bird+Mini Helicopter	0.88	0.90	0.94	0.91	0.91	0.93
	RC Plane	0.99	0.95	0.98	0.97	0.97	0.93
	Quadcopter	1.00	1.00	1.00	1.00	1.00	1.00
RRN-ATR-Net	Bird	1.00	1.00	1.00	1.00	1.00	1.00
	Fixed Wing	1.00	0.97	1.00	1.00	1.00	1.00
	Helicopter	0.97	1.00	1.00	0.98	1.00	1.00
	Quadcopter	0.93	1.00	0.98	0.95	0.93	0.97
RAD-DAR	Car	0.94	0.95	0.94	0.95	0.95	0.95
	Drones	0.95	0.97	0.97	0.97	0.97	0.96
	People	0.99	0.99	0.99	0.99	0.99	0.99

Table 5: Precision for different classes and channel attention mechanisms on various datasets.

Dataset	Classes	Without Attention	Original-ECA	Modified-ECA	Original-SE	Modified-SENET	GCT
DIAT-μSAT	3 long blade rotor	0.94	0.95	0.95	0.94	0.97	0.95
	3 short blade rotor	0.90	0.87	0.92	0.87	0.87	0.86
	Bionic-bird	1.00	1.00	1.00	1.00	1.00	1.00
	Bird+Mini Helicopter	1.00	1.00	1.00	1.00	1.00	1.00
	RC Plane	0.91	0.94	0.96	0.94	0.94	0.94
	Quadcopter	1.00	1.00	1.00	1.00	1.00	1.00
RRN-ATR-Net	Bird	0.97	1.00	1.00	1.00	1.00	1.00
	Fixed Wing	0.94	1.00	0.98	0.95	0.94	0.97
	Helicopter	1.00	0.97	1.00	1.00	1.00	1.00
	Quadcopter	1.00	1.00	1.00	0.98	1.00	1.00
RAD-DAR	Car	0.95	0.97	0.97	0.97	0.97	0.97
	Drones	0.93	0.94	0.93	0.94	0.94	0.94
	People	0.99	0.99	0.99	0.99	0.99	0.99

Table 6: Performance comparison with previous works.

Sl.No	Dataset	Baseline Accuracy	Modified SENET	Modified ECA
1.	DIAT-μSAT	[95 (VGG16), 97 (VGG19) [6] **(2021)**] , 95.4 [**?**] **(2022)**	96.56	97.14
2.	RRN-ATR-Net	96.55 [46] **(2025)**	98.33	99.60
3.	RAD-DAR	96.86 [19] **(2022)** , 95.43 (ResNet18)	97.08	96.63

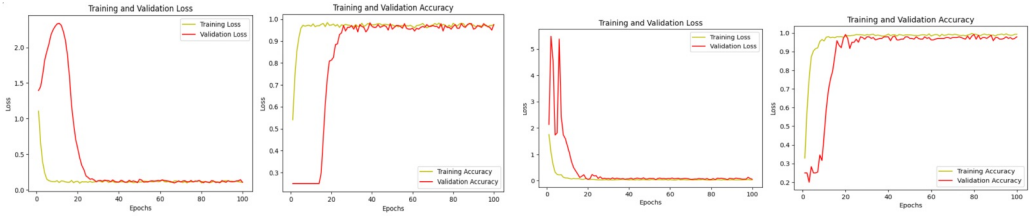


(a) Without Attention Mechanism (b) GCT Attention

(c) Efficient Channel Attention (d) Modified Efficient Channel Attention

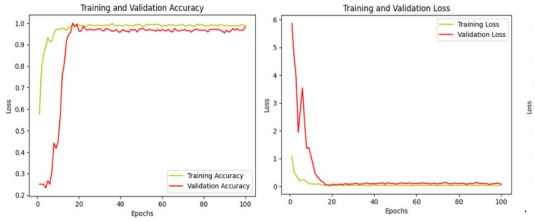
(e) Squeeze Excitation Attention (f) Modified Squeeze Excitation Attention

Figure 8: The Accuracy and Loss graphs for DIAT-μSAT dataset.

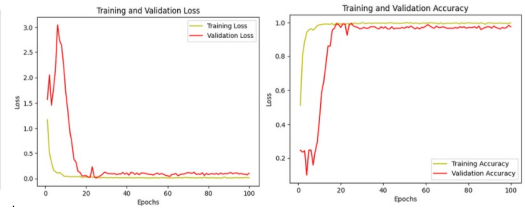


(a) Without Attention Mechanism

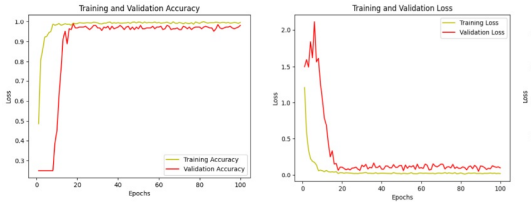
(b) GCT Attention



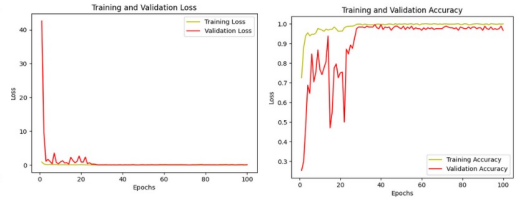
(c) Efficient Channel Attention



(d) Modified Efficient Channel Attention

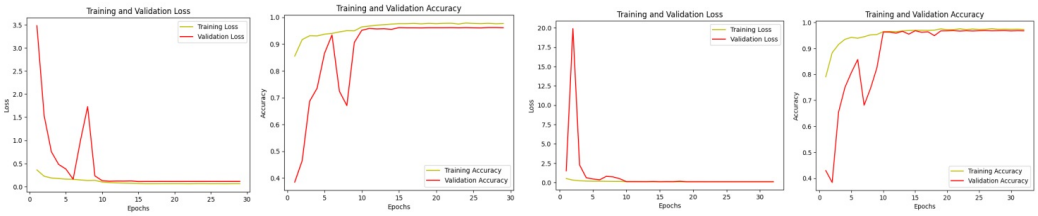


(e) Squeeze Excitation Attention



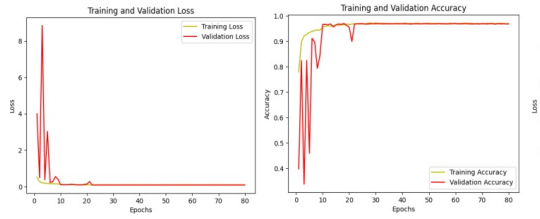
(f) Modified Squeeze Excitation Attention

Figure 9: The Accuracy and Loss graphs for RRN-ATR-Net dataset.

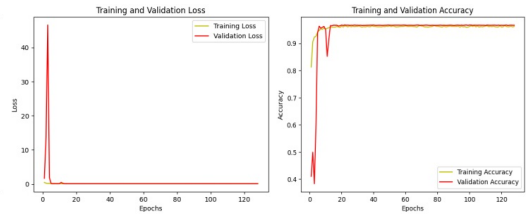


(a) Without Attention Mechanism

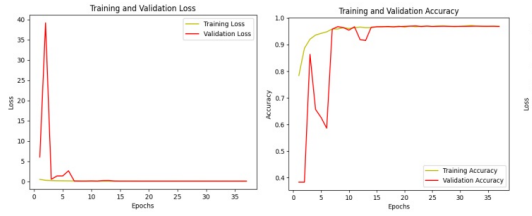
(b) GCT Attention



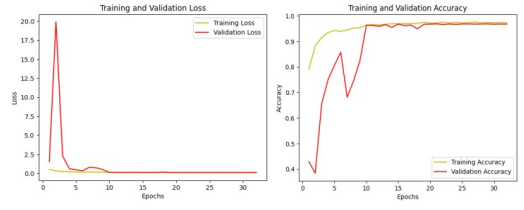
(c) Efficient Channel Attention



(d) Modified Efficient Channel Attention

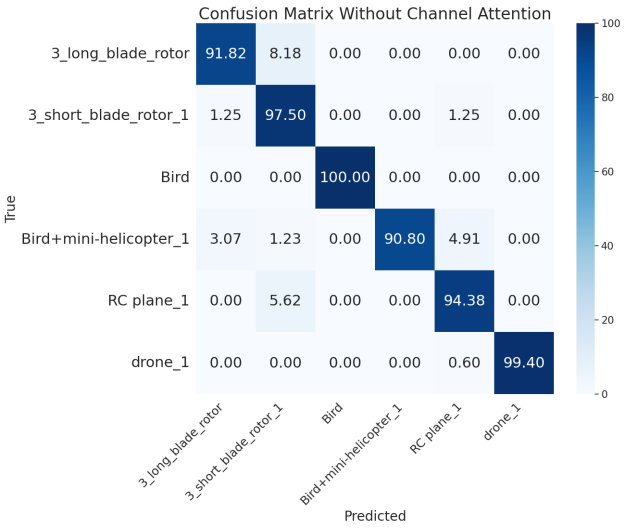


(e) Squeeze Excitation Attention

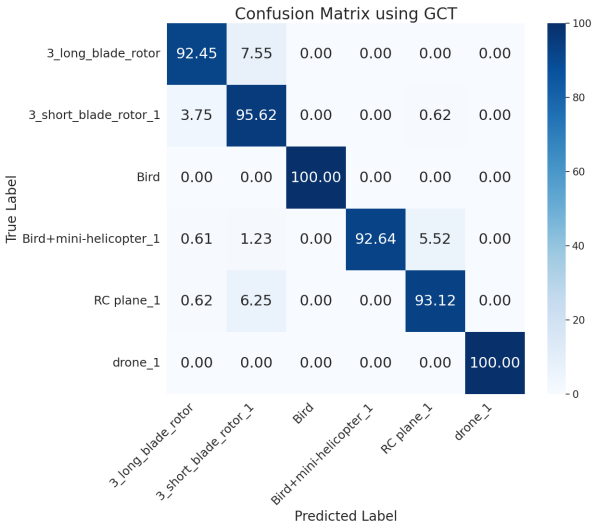


(f) Modified Squeeze Excitation Attention

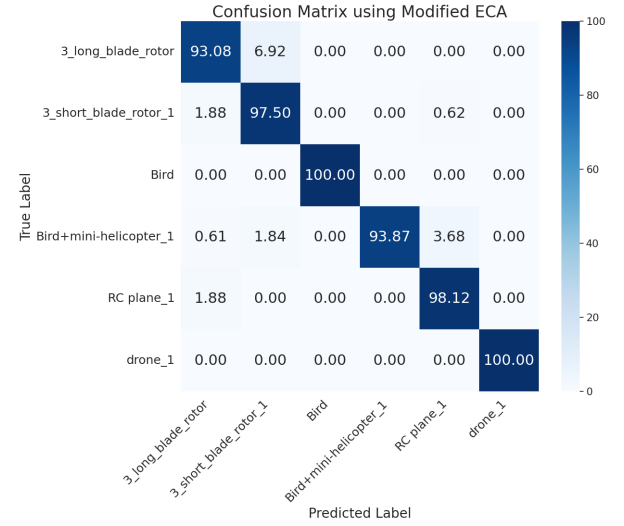
Figure 10: The Accuracy and Loss graphs for RAD-DAR dataset.



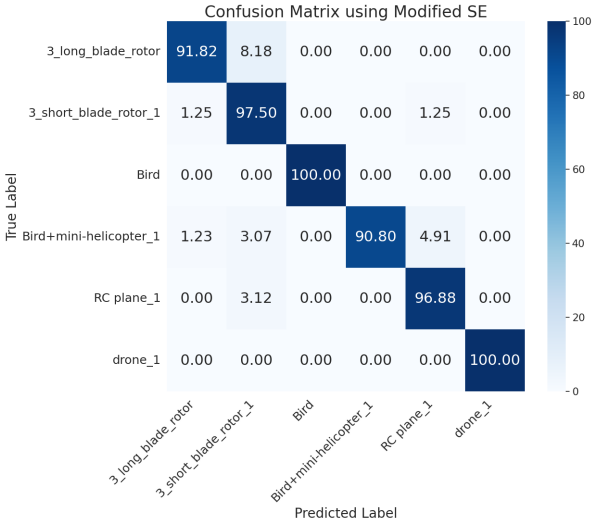
(a) Without Attention Mechanism



(b) GCT

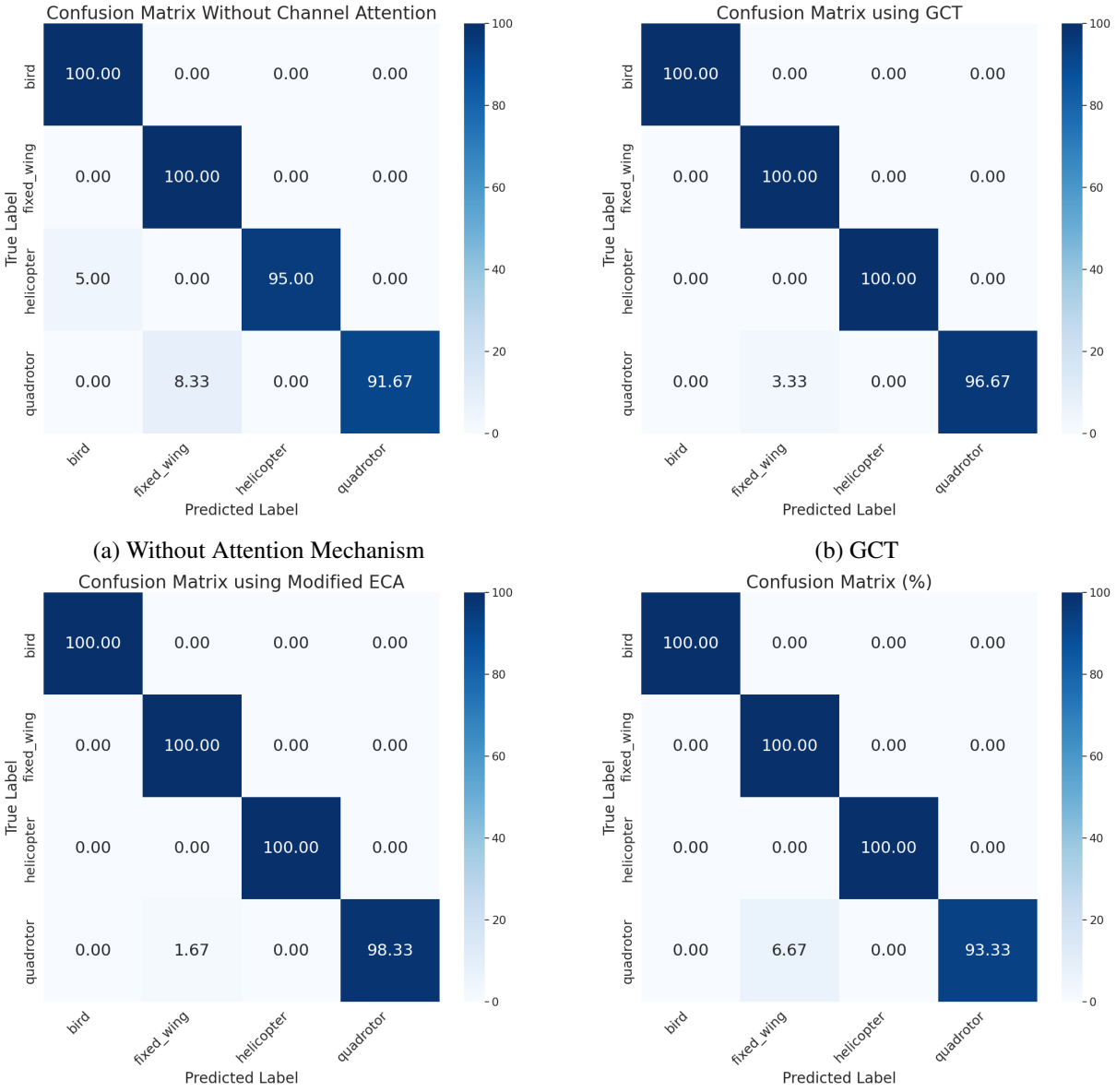


(c) Modified ECA



(d) Modified SE

Figure 11: Confusion Matrix showing classification on DIAT-μSAT dataset.



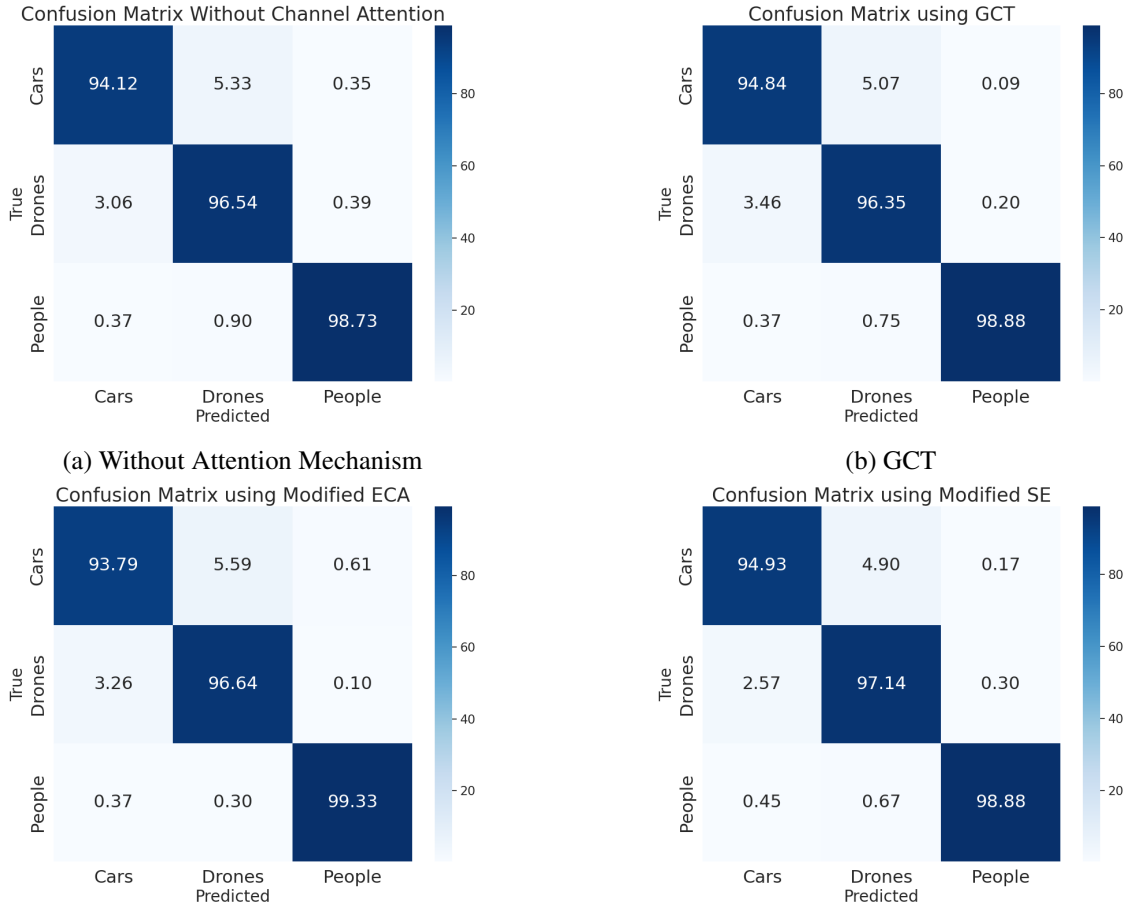
(a) Without Attention Mechanism

(b) GCT

(c) Modified ECA

(d) Modified SE

Figure 12: Confusion Matrix showing classification on RRN-ATR-Net dataset.



(a) Without Attention Mechanism

(b) GCT

(c) Modified ECA

(d) Modified SE

Figure 13: Confusion Matrix showing classification on RAD-DAR dataset.

7.1 Performance Efficiency and Statistical Analysis

The efficiency of the proposed model with integrated channel attention is compared using MACs , total parameters and disk size for a batch size of 10 samples as provided in Table.7. For statistically validating the run time efficiency , inference time is reported after 5 simulation runs. For comparing inference efficiency for critical real time situations, the experiments are carried out on 1GB GPU system. For the DIAT dateset, the model proposed by [25] considers depth-separable convolution networks to reduce the number of parameters at the same time increasing the layers to 50-70. Thus, the mean inference time is better but the accuracy (95.4%) suffers due to reduction in parameters . The proposed UAVDetect2 with channel attention maintains a decent mean inference time of 80 ms with a satisfactory accuracy of (97.14%). In case of RRN-ATR-Net datasets, proposed UAVDetect2 although having slight increase in mean inference time than baseline, the accuracy of classification is significantly high (99.60%). For the RAD-DAR dateset the proposed UAVDetect1 is having comparable inference time but significant increase in accuracy when compared to baseline models.

8 Limitation and Future work

The study has certain limitations that need to be addressed for real-world applicability. First, the models were trained and tested on datasets collected under controlled conditions, and their performance in noisy, cluttered environments, such as urban areas with heavy RF interference or adverse weather, remains unverified. Addi-

Table 7: Comparison of Real Time Inference Performance across UAV datasets. (Batch size = 10 samples.)

Dataset	Model	MACs (G)	Batch latency (ms)	Parameters (M)	Disk size
DIAT-SAT	VGG16 [6]	15.4703	300 ± 50	138	528 MB
	VGG19 [6]	19.6321	330 ± 50	143	549 MB
	[25]	0.59	25 ± 40	0.49	2.2 MB
	UAVDetect2	10.3	62 ± 30	6.8	26 MB
RRN-ATR-Net	[46]	8.3	90 ± 20	6	22 MB
	UAVDetect2	10.3	95 ± 30	6.8	26 MB
RAD-DAR	[19]	0.0034	4 ± 1.5	0.14	620 KB
	ResNET18	0.006	4 ± 2	0.18	3 MB
	UAVDetect1	0.0053	4 ± 1.5	0.16	634 KB

tionally, the system's ability to generalize to unseen UAV classes can be improved, as the training data may not encompass the full diversity of micro-Doppler signatures from emerging or modified drones. The dependency on specific radar frequency bands (e.g., X-band) further restricts adaptability to different sensor configurations.

Future work should explore robustness enhancements through adversarial training and advanced data augmentation. Modern architectures like vision transformers (ViTs) could be investigated to improve feature extraction, while model optimization techniques like quantization and pruning could enable real-time deployment. Extending the framework for multi-target detection and integrating cross-domain adaptation methods would further enhance generalizability. Combining radar data with other sensors (e.g., LiDAR or RF signals) could also improve classification accuracy in complex environments. These improvements would make the system more practical for real-world UAV detection and surveillance applications.

9 Conclusion

The study underscores the significance of employing channel attention mechanisms to enhance UAV classification using micro-Doppler radar signatures. The results reveal that attention mechanisms, particularly the Modified ECA Net, substantially improve classification accuracy by enabling models to focus on class-specific features and differentiate between structurally similar classes. Across datasets such as DIAT-μSAT, RAD-DAR, and RRN_ATR_Net, the Modified ECA and Modified SENet consistently outperform baseline models, demonstrating their ability to refine feature selection and mitigate mis-classifications, particularly in challenging scenarios involving overlapping aerodynamic characteristics. The findings highlight that classes with distinct visual traits, such as birds and drones, achieve near-perfect classification accuracy, underscoring their inherent separability. However, structurally similar classes, such as quadrotors and fixed-wing drones or birds and mini-helicopters, pose persistent challenges. Channel attention mechanisms, especially the Modified ECANet, reduce misclassification rates significantly but still face limitations in scenarios with substantial feature overlap. This work also emphasizes the robustness and scalability of attention-based approaches, as demonstrated by their performance across multiple datasets and architectures. The integration of attention mechanisms leads to faster convergence during training and more reliable differentiation of complex patterns. The results obtained shows that the mean inference time and the classification accuracy is comfortably balance as accuracy is increased and mean inference time decreases. While the current methodologies achieve notable accuracy improvements, the study identifies areas for further exploration, including enhancing robustness in noisy conditions, improving inference times, and enabling the detection and classification of multiple UAVs in the same spectrogram. Overall, this research contributes to advancing the state of the art in UAV classification, offer-

ing insights into leveraging attention mechanisms for more accurate, efficient, and robust micro-Doppler radar signature analysis. These findings pave the way for future work in real-time aerial target classification and the development of more sophisticated models for challenging operational scenarios.

References

[1] P. Mandal, L. P. Roy, and S. K. Das, "Flying objects classification based on micro–Doppler signature data from UAV borne radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024. https://doi.org/10.1109/LGRS.2024.3351980

[2] P. Zhang, L. Yang, G. Chen, and G. Li, "Classification of drones based on micro-Doppler signatures with dual-band radar sensors," in *Proc. 2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)*, pp. 638–643, IEEE, 2017. https://doi.org/10.1109/PIERS-FALL.2017.8293215

[3] L. Fuhrmann, O. Biallawons, J. Klare, R. Panhuber, R. Klenke, and J. Ender, "Micro-Doppler analysis and classification of UAVs at Ka band," in *Proc. 18th Int. Radar Symp. (IRS)*, pp. 1–9, IEEE, 2017. https://doi.org/10.23919/IRS.2017.8008112

[4] P. Molchanov, R. I. A. Harmanny, J. J. M. de Wit, K. Egiazarian, and J. Astola,"Classification of small UAVs and birds by micro-Doppler signatures," *Int. J. Microw. Wireless Technol.*, vol. 6, no. 3–4, pp. 435–444, 2014. https://doi.org/10.1017/S1759078714000282

[5] D. Dhulashia, N. Peters, C. Horne, P. Beasley, and M. Ritchie, "Multi-frequency radar micro-doppler based classification of micro-drone payload weight," *Frontiers in Signal Processing*, vol. 1, p. 781777, 2021. https://doi.org/10.3389/frsip.2021.686341

[6] H. C. Kumawat, M. Chakraborty, A. A. B. Raj, and S. V. Dhavale, "DIAT-μSAT: Small aerial targets' micro-Doppler signatures and their classification using CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021. https://doi.org/10.1109/LGRS.2021.3061559.

[7] H. Le, V.-S. Doan, H.-H. Nguyen, T. Huynh-The, K. Le-Ha, V.-P. Hoang, *et al.*, "Micro-Doppler-radar-based UAV detection using inception-residual neural network," in *Proc. Int. Conf. Adv. Technol. Commun. (ATC)*, pp. 177–181, IEEE, 2020. https://doi.org/10.1109/ATC49230.2020.9247926

[8] B. Choi and D. Oh, "Classification of drone type using deep convolutional neural networks based on micro-Doppler simulation," in *Proc. Int. Symp. Antennas Propag. (ISAP)*, pp. 1–2, IEEE, 2018. https://doi.org/10.23919/ISAP.2018.8639483

[9] Á. D. Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. López, "Drone detection with X-band ubiquitous radar," in *Proc. 19th Int. Radar Symp. (IRS)*, pp. 1–10, IEEE, 2018. https://doi.org/10.23919/IRS.2018.8447954

[10] K. K. Gümen, S. Paker, and K. Savci, "Examination of Micro-Doppler Signatures of Drone and Seagulls with X-Band Noise Radar," in *Proc. 32nd Signal Process. Commun. Appl. Conf. (SIU)*, pp. 1–4, IEEE, 2024. https://doi.org/10.1109/SIU62312.2024.10543798

[11] D. Park, S. Lee, S. Park, and N. Kwak, "Radar-spectrogram-based UAV classification using convolutional neural networks," *Sensors*, vol. 21, no. 1, p. 210, 2020. https://doi.org/10.3390/s21010210

[12] S. Rahman and D. A. Robertson, "Radar micro-Doppler signatures of drones and birds at K-band and W-band," *Sci. Rep.*, vol. 8, no. 1, p. 17396, 2018. https://doi.org/10.1038/s41598-018-35880-9

[13] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged Doppler images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, 2016. https://doi.org/10.1109/LGRS.2016.2624820

[14] A. Kyritsis, R. Makri, and N. Uzunoglu, "Detection of Low RCS Unmanned Air Systems Using K-Band Continuous Wave Doppler Radar," *J. Electr. Comput. Eng.*, vol. 2023, no. 1, p. 5683661, 2023. https://doi.org/10.1155/2023/5683661

[15] J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. J. Soraghan, and S. Marshall, "Robust PCA micro-Doppler classification using SVM on embedded systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 2304–2310, 2014. https://doi.org/10.1109/TAES.2014.120716

[16] J. Ren and X. Jiang, "A three-step classification framework to handle complex data distribution for radar UAV detection," *Pattern Recognit.*, vol. 111, p. 107709, 2021. https://doi.org/10.1016/j.patcog.2020.107709

[17] G. Lu and Y. Bu, "Mini-UAV Movement Classification Based on Sparse Decomposition of Micro-Doppler Signature," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024. https://doi.org/10.1109/LGRS.2024.3364966

[18] A. Aouto, T. Jun, J.-M. Lee, and D.-S. Kim, "DopeNet: Range–Doppler Radar-based UAV Detection Using Convolutional Neural Network," in *Proc. 14th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, pp. 889–893, IEEE, 2023. https://doi.org/10.1109/ICUFN57995.2023.10199464

[19] A. J. Garcia, A. Aouto, J.-M. Lee, and D.-S. Kim, "CNN-32DC: An improved radar-based drone recognition system based on Convolutional Neural Network," *ICT Express*, vol. 8, no. 4, pp. 606–610, 2022. https://doi.org/10.1016/j.icte.2022.03.012

[20] D. Raval, E. Hunter, I. Lam, S. Rajan, A. Damini, and B. Balaji, "Detecting drones with radars and convolutional networks based on micro-Doppler signatures," in *Proc. IEEE Radar Conf. (RadarConf22)*, pp. 1–6, IEEE, 2022. https://doi.org/10.1109/RadarConf2248738.2022.9764267

[21] A. Huizing, M. Heiligers, B. Dekker, J. de Wit, L. Cifola, and R. Harmanny, "Deep learning for classification of mini-UAVs using micro-Doppler spectrograms in cognitive radar," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 11, pp. 46–56, 2019. https://doi.org/10.1109/MAES.2019.2925573

[22] K. Jung, J.-I. Lee, N. Kim, S. Oh, and D.-W. Seo, "Classification of space objects by using deep learning with micro-Doppler signature images," *Sensors*, vol. 21, no. 13, p. 4365, 2021. https://doi.org/10.3390/s21134365

[23] I. Roldan, C. R. del-Blanco, Á. Duque de Quevedo, F. I. Urzaiz, J. G. Menoyo, A. A. López, *et al.*, "DopplerNet: A convolutional neural network for recognising targets in real scenarios using a persistent range–Doppler radar," *IET Radar Sonar Navig.*, vol. 14, no. 4, pp. 593–600, 2020. https://doi.org/10.1049/iet-rsn.2019.0440

[24] S. E. Abdelsamad, M. A. Abdelteef, O. Y. Elsheikh, *et al.*, "Vision-Based Support for the Detection and Recognition of Drones with Small Radar Cross Sections," *Electronics*, vol. 12, no. 10, p. 2235, 2023. https://doi.org/10.3390/electronics12102235

[25] H. C. Kumawat, M. Chakraborty, and A. A. B. Raj, "DIAT-RadSATNet—A novel lightweight DCNN architecture for micro-Doppler-based small unmanned aerial vehicle (SUAV) targets' detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 71, p. 8504011, 2022. https://doi.org/10.1109/TIM.2022.3175053

[26] J. Duan, L. Zhang, Y. Wu, Y. Zhang, Z. Zhao, and X. Guo, "Classification of birds and drones by exploiting periodical motions in Doppler spectrum series," *J. Syst. Eng. Electron.*, vol. 34, no. 1, pp. 19–27, 2023. https://doi.org/10.23919/JSEE.2023.000003

[27] X. Chen, C. Ma, C. Zhao, and Y. Luo, "UAV Classification Based on Deep Learning Fusion of Multi-Dimensional UAV Micro-Doppler Image Features," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024. https://doi.org/10.1109/LGRS.2024.3353846

[28] J. Tian, C. Wang, J. Cao, and X. Wang, "Fully Convolutional Network-Based Fast UAV Detection in Pulse Doppler Radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–12, 2024. https://doi.org/10.1109/TGRS.2023.3340576

[29] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 7132–7141, 2018. https://doi.org/10.1109/CVPR.2018.00745

[30] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 11534–11542, 2020. https://doi.org/10.1109/CVPR42600.2020.01155

[31] Z. Yang, L. Zhu, Y. Wu, and Y. Yang, "Gated channel transformation for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 11794–11803, 2020. https://doi.org/10.1109/CVPR42600.2020.01181

[32] D. B. Herr and D. Tahmoush, "Data-driven STFT for UAV micro-Doppler signature analysis," in *Proc. IEEE Int. Radar Conf. (RADAR)*, pp. 1023–1028, 2020. https://doi.org/10.1109/RADAR42522.2020.9114755

[33] K. Jung, J.-I. Lee, N. Kim, S. Oh, and D.-W. Seo, "Classification of space objects by using deep learning with micro-Doppler signature images," *Sensors*, vol. 21, no. 13, p. 4365, 2021. https://doi.org/10.3390/s21134365

[34] D. Raval, E. Hunter, I. Lam, S. Rajan, A. Damini, and B. Balaji, "Multifrequency classification of drones using Doppler radar micro-Doppler signatures," in *Proc. IEEE Radar Conf. (RadarConf21)*, pp. 1–6, 2021. https://doi.org/10.1109/RadarConf2147009.2021.9455333

[35] X. Ding, J. Wang, C. Liu, C. Li, H. Zhang, and Z. Zhang, "Drone detection using convolutional neural networks and multi-sensor fusion for micro-Doppler radar and infrared thermal sensor," *Sensors*, vol. 20, no. 15, p. 4255, 2020. https://doi.org/10.3390/s20154255

[36] W. Kim, K. Kim, and J.-S. Nam, "Deep neural network classification of UAVs using micro-Doppler radar," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, 2020. https://doi.org/10.1109/ICC40277.2020.9148677

[37] J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. J. Soraghan, and S. Marshall, "Robust PCA micro-Doppler classification using SVM on embedded systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 2304–2310, 2014. https://doi.org/10.1109/TAES.2014.120716

[38] J. Ren and X. Jiang, "A three-step classification framework to handle complex data distribution for radar UAV detection," *Pattern Recognit.*, vol. 111, p. 107709, 2021. https://doi.org/10.1016/j.patcog.2020.107709

[39] G. Lu and Y. Bu, "Mini-UAV Movement Classification Based on Sparse Decomposition of Micro-Doppler Signature," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024. https://doi.org/10.1109/LGRS.2024.3364966

[40] A. Aouto, T. Jun, J.-M. Lee, and D.-S. Kim, "DopeNet: Range–Doppler Radar-based UAV Detection Using Convolutional Neural Network," in *Proc. 14th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, pp. 889–893, IEEE, 2023. https://doi.org/10.1109/ICUFN57995.2023.10199464

[41] N. Yao, J. Qin, Y. Zhao, and G. Wu, "UAV image target recognition algorithm based on machine learning," in *J. Phys.: Conf. Ser.*, vol. 2558, no. 1, p. 012009, 2023. https://doi.org/10.1088/1742-6596/2558/1/012009

[42] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," in *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019. https://doi.org/10.48550/arXiv.1906.05909

[43] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1135–1144, 2016. https://doi.org/10.1145/2939672.2939778

[44] C. Agarwal and A. Nguyen, "Explaining image classifiers by removing input features using generative models," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2020. https://doi.org/10.1007/978-3-030-69544-6_7

[45] J. Park and J.-S. Park, "Classification of small drones using low-uncertainty micro-Doppler signature images and ultra-lightweight convolutional neural network," *IEEE Trans. Image Process.*, vol. 33, pp. 2979–2994, 2024. https://doi.org/10.1109/TIP.2024.3389041

[46] Z. Wu and Z. Liu, "RRN-ATR-Net: A Robust Relation Network for Aerial Targets Recognition Based on Two-Stream Micro-Doppler Signatures Fusion," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025. https://doi.org/10.1109/TIM.2024.3512351