International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Object Detection in Satellite Images with Canis Hunt Optimized Tetralet Attention enabled Explainable Convolutional Neural Network

Mayur Vijaykumar Tiwari*[1] and Dr. Sanjay Vasant Dudul[2]

Submitted:10/03/2024 **Revised:** 25/04/2024 **Accepted:** 02/05/2024

Abstract: Object detection has always been a research hotspot in computer vision, specifically detection from satellite images remains a challenging research area. Several conventional researches have been developed but failed to work with high-quality images and satellite images. The object detection in the research is proposed with the Canis Hunt Optimized Tetralet Attention enabled Explainable Convolutional Neural Network (CHunt-TetraExNN). The proposed model aims to provide accurate detection from the satellite image inputs. The Tetralet attention module incorporated with the model is composed of triplet attention followed by positional attention, which provides the accurate estimation of the attentional features that are highly helpful in the process of object detection. Further, the research model is supported by the Canis Hunt Optimization, which is the combination of the adaptability and the hunt characteristics of the Lapins and Latrans that belong to the Canis Family. Thus, the model provided accurate estimation outcomes in the research of object detection from the satellite images, which are estimated with an Accuracy of 96.55%, Precision of 94.59%, Recall of 96.56%, and F1 score of 95.6%.

Keywords: *Object detection, Satellite images, Explainable Convolutional Neural Network, Tetralet Attention, and Canis Hunt Optimization*

1. Introduction

The quantity and quality of remote sensing images have increased significantly in recent years due to the rapid development of remote sensing technology. Because it can record representative landmarks like airports, ports, and stations as well as little items like vehicles, ships, and planes, remote sensing image object recognition is becoming more and more common in the military, business, agriculture, and other industries [1]. In computer vision, object detection has long been a popular area of study. Real-world issues such as construction identification, industrial detection, pedestrian detection, and so on often employ it. There are a lot of ground items in these high-resolution remote-sensing image collections, including typical ground objects like cars, trucks, basketball courts, and airplanes. It is crucial to classify, segment, detect, and classify remote sensing images [2] [3] [4]. The method for detecting objects in remote sensing images [5] has been extensively employed in various domains such as land planning, maritime fisheries, military reconnaissance, and deforestation detection [6] [7] [8]. The issue of object identification extends beyond scene-level analysis, which identifies simply the semantic labels of scene images, as it requires simultaneous localization and recognition of several items from a complicated remote sensing scenario. Presently, there has been significant advancement in remote sensing object detectors [1], [9], [10], and [11].

These advancements nearly concentrate on improving feature fusion and regression layer performance, and they have shown encouraging results on a few benchmark datasets. On the other hand, the backbone network is in charge of extracting fundamental semantic features from images, which are essential for locating and identifying objects, and serves as the foundation for the feature fusion network and regression layers [12].Applying deep learning to remote sensing images made sense given the recent advances in deep learning-based object detection. Nonetheless, the adaptation of object detection models meant for natural images to remote-sensing images is not simple [9]. Deep learning (DL) has previously been widely applied in computer vision applications because of its remarkable feature representation capabilities, which significantly increase the efficacy of optical object detection approaches [13]. As such, Salient Object Detection (SOD) is often used as a supplementary technique in real-world applications to highlight relevant regions while reducing background noise. Salient object detection can also be used for other vision tasks such as visual tracking, image retrieval, and object segmentation and recognition [14]. Owing to SOD's significant applicability, it has drawn more attention from sectors like the military, agriculture, and disaster assistance [11]. Multiscale object identification approaches have numerous

[1]*Department of Applied Electronics*
Sant Gadge Baba Amravati University,Camp Area, Near Tapovan Gate, Amravati, Maharashtra, India, 444602
[2]*Department of Applied Electronics*
Sant Gadge Baba Amravati University, Amravati,Maharashtra, India, 444602
Corresponding Author mail id: mayurvtiwari@gmail.com

challenges due to the influence of air particles, light, and other factors on relative sensitivities (RSIs) [15][16]. Because intentionally produced features are not as robust to diversity variations as natural ones are, standard object detection methods are unable to match real-time requirements when subjected to light, object occlusion, and target overlap [10]. In particular, the first detection techniques used in remote sensing images were two-stage detectors, such as region proposals convolutional neural network (RCNN) [12], Fast-RCNN [17], and Faster-RCNN [18].

Despite their great detection accuracy, they struggle to handle large-scale remote sensing images in real time due to their poor detection speed [14]. However, the majority of the algorithms above are two-stage detectors, which perform better in terms of detection accuracy than one-stage detectors. Nonetheless, their inference speed is typically slower. More significantly, the hardware cost of object identification models in real-world applications rises as a result of these algorithms' high hardware environment requirements. On the other hand, the YOLOv5 method in conjunction with a circular smooth label (CSL) [19] not only exhibits superior detection accuracy [20], but it also outperforms other models in terms of inference speed. Owing to the peculiarities of distant sensing imagery, objects are frequently tiny, disorganized, and oriented arbitrarily. Standard convolutional layers in a basic CNN combine the features using filters on regular grids to create deep features, which results in blurry noisy features and inter-class feature coupling. A feature alignment module and a feature decoupling module for de-noising are required for an accurate quick single-stage object detector [9]. A unique single-shot multi-box detector (SSD) network based on feature fusion and dilated convolution was proposed by [16]. Nevertheless, there is no greater improvement in the detection effect of small and overlapping objects. Furthermore,

to improve feature robustness and discriminability, [21] introduced a unique RF block (RFB) module that considered the relationship between RF size and eccentricity [10].

The ultimate intention of the research is to detect the object from the provided satellite images. The research utilizes the CHunt-TetraExNN model to perform the detection of objects, further the segmentation based on Tetralet attention enabled W-Net is utilized in the research. The involved attention mechanisms modify the outcomes of the overall model by achieving effective features.

- **Tetralet Attention mechanism:** the attention mechanism, which is the combination of the triplet and the positional attention mechanisms is named, the Tetralet attention mechanism. The use of the attention mechanism in the research is to encode and obtain the spatial information of the provided input.
- **Tetralet Attention-W-Net based Segmentation:** The Tetralet attention mechanism incorporated with the W-Net added more advantages to the process of extraction such as extraction of the most local information as well as segmenting the most appropriate region of segmentation. Further, the usage of the attention mechanism improved the accuracy of segmentation by concentrating on the most significant features of the image.
- **Canis Socio-Hunt Optimization:** The CHunt optimization is the combination of the hybrid characteristics of Canis Lupus and Latrans. The optimization integrated the characteristics of adaptability as well as hunting, which aids in achieving accurate detection accuracy in the research model of object detection.
- **Canis Hunt Optimized Tetralet Attention enabled Explainable Convolutional Neural Network (CHunt-TetraExNN):** The CHunt-TetraExNN model is utilized in the research that accurately detects the objects from the satellite images. The base explainable CNN model works incorporation with the attention module as well as the optimization.

The research articles are organized as, Section 2 tells the conventional methods with the challenges, Section 3 describes the proposed methodology followed analysis of the outcomes in Section 4, and Section 5 ends in the conclusion.

2. Literature Review

The traditional methods utilized in the research of object detection with satellite images are described in this section as follows,

YunPeng Xu in [14], provided the HOFA-Net model elaborated as, High-Order Feature Association Network. The model could work with the features that provided the accurate detection of dense objects. The convolutional neural networks (CNNs) involved in the classification generated features of the minimal resolution and further SANL is introduced in the research model to address the local feature dependencies. The model failed to detect the pose of the detected objects as well as to include the wide deformation ranges. In [13], Keyan Wang *et al*. presented the MashFormer model that acted as the Multiscale hybrid detector in the process of object detection. The utilized model had CenterNet as the base model, where CNN was added to extract the seamless features to name the object detection through the model. The strong model was infused into the research model to achieve higher relatable outcomes. Though the model achieved high performance, the model ignored the variants of the objects detected in the research.

Tong Zhang *et al*. in [22], described mask image modeling (MIM) with the self-supervised pre-training mechanism (SSP). The described model facilitated capturing the object representation accurately, whereas the masking of the background the provided with the Attention-guided mask generator (AGMG) mechanism that added additional benefits to the model. Though the mechanisms included in the research model provided several benefits, the integration created complexity in the model resulting in poor object detection, and further due to background complexity and wide-scale variance. Linkai Liu *et al*. [8] suggested the distributed inference framework in the research to achieve the most accurate object detection by utilizing the most reasonable resources. Thus, the resulting detection was higher than the repetitive conventional methods. The model needed to be improved in balancing the speed-up ratio and the accuracy, which were relatively proportional. The model ignored the GPU cluster machines due to the unavailability of the resources. Further, the model could be trained efficiently and stably to achieve efficient outcomes.

Yuhui Zhao *et al*. [1] initiated the object detection with the P-SACA improved YOLO-V7 along with the SPD module that carries the spatial information of the input. P-SACA is the combination of the attention mechanisms that provided the most efficient feature extraction mechanism. Further, the research model presented better outcomes with the remote sensing datasets but failed to address the imbalanced data and the complex backgrounds of the input images. Further, the research could be improved to work with minimal parameters and to reduce the computational complexity. Yassin Zakaria *et al*. [9] introduced a Single shot alignment Network (S2A-Net) to detect small-sized objects. The utilized ILD architecture provided complexity during the training process with the fewer data, thus providing less benefit. The utilization of the anchor boxes in the network could improve the efficiency, and in addition, the oriented R-CNN is included in the research to showcase the major efficiency of the research. Further, the model could work with the lightweight attention module, which could be worth analyzing in the research.

Liguo Wang *et al*. [10] suggested the single-shot multi-box detector model to achieve the accurate object detection. The SSD module was implemented in the research to show the efficiency of the process of feature extraction. The model aimed to work with the Multiscale input in the complex input scenes of the remote sensing images. The dataset of the research could be expanded to the optical RSI and further, the bio-inspired optimization algorithms could be added to enhance the efficiency of the research. Longbao Wang *et al*. [11] implemented the CSFFNet, which was the combination of the ResNet, Feature, and pyramidal feature enhancement modules. Further, the model named Lightweight ORSI-SOD was included in the research as the cross-scale fusion. The model fulfilled the criteria of the lightweight modules and extracted the most semantic segmented features. The integration of the pertinent modules enhances the capability of the research to define the objects.

2.1. Challenges

- ➢ The CNN model explored the HOFA network to improve the detection of dense capacity and small objects, but the model was not used to detect tiny objects due to the minimum capacity to accept the maximum size of images [14].
- ➢ The ResNet was utilized in CNN techniques to improve the feature extraction and explain the differences and similarities between the features, but the model was not considered an innovative approach due to an unsatisfied outcome [23].
- ➢ The OCMIM-based SSP process combined with the OCDG was used in downstream object detection tasks with better accuracy but the model was affected due to the task-aware discrepancy [22].
- ➢ The P-SACA model was designed to improve the network performance for object detection, but the model leads to losses of the fine-grained information as well as the model did not remove the unwanted information in the image [1].
- ➢ The ILD model was introduced to enhance the advanced level prediction of bounding boxes, but the model was not fit to use in the real world due to experimental conditions [9].

3. Object detection with CHunt Optimized Tetralet Attention enabled Explainable CNN

The ultimate aim of the research is to detect objects from satellite images or remote sensing images. The images for the research are obtained from the remote sensing datasets such as Airbus Aircraft Detection Dataset [24]. The fetched inputs from the dataset are fed into the Preprocessing stage, where the images are denoised and the Region of Interest (RoI) is extracted accurately. The ROI extracted image undergoes Tetralet Attention-WNET-based segmentation, which can enhance the segmentation accuracy even if the object is varied in shape and size. The Tetralet attention Mechanism can enlighten the important features like edge and shape relevant features within the image. After segmentation, Modified Object Flow based Ternary Pattern (MOFTP), and the Shape and structural informatics features are extracted from different segmented object regions to detect the type of objects. Tetralet attention Resnet 101 and GLCM features are extracted for detecting the objects. To enhance the classification accuracy the CHunt-TetraExNN is used in the research as the classifier, and the hyperparameters of the classifier are tuned by Canis Socio-Hunt optimization with a hybrid error mechanism to achieve high detection accuracy. The workflow of the research is illustrated in Figure 1 as follows,

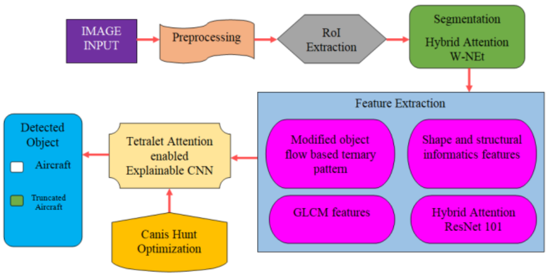


Fig. 1: Block Diagram of the CHunt-TetraExNN model

3.1 Input Image

The input of the research on object detection is engaged with the satellite images provided by the Airbus Aircraft detection dataset [24]. The dataset contains images observed from the commercial satellite constellation as well as the radar constellation. The input obtained from the dataset O is represented as,

$$O = \left\{ O_1, O_2, \ldots. O_b, \ldots., O_d \right\}, \tag{1}$$

where, O_b, is the input image of the dataset containing a total image of d. The input image is with the pixel intensity of $\left(u, v\right)$.

3.2 Preprocessing

Preprocessing of the image is performed to obtain a clear image output, which is achieved through the binary processing of images such as sharpening and smoothening of images. The sharpening mechanism of the research is performed with the “filter2D” mechanism, which converts the values of the pixel intensities in concern with the neighboring pixel intensities of the image. The implementation of the mechanism aids in achieving the sharpened image, represented as, $sharp$.

The obtained sharpened image is further proceeded to the smoothening process, where Fast Non-Local Means (FNLM) Denoising is applied, thus providing an enhanced outcome. The FNLM is utilized to eliminate noise [25] and maintain the high-frequency signal represented as,

$$N_L = \sum w\left(u, v\right) O_b \,, \tag{2}$$

where, $w\left(u, v\right)$, indicates the weight of the pixel intensities $\left(u, v\right)$, which is evaluated as,

$$w\left(u, v\right) = \frac{1}{norm} e^{\frac{-\left\| \vartheta\left(u\right) - \vartheta\left(v\right) \right\|_{2,r}^{2}}{k}} , \tag{3}$$

where, $norm$, indicates normalization constant, $\left\| \vartheta\left(u\right) - \vartheta\left(v\right) \right\|$, represents Euclidean distance, which is enhanced with the Gaussian function that assigns the standard deviation r, applied with the smoothening parameter k that controls the filtering degree.

3.3 RoI extraction

The outcome of the preprocessing is fed into the process of RoI extraction, where the most interesting or informative regions are extracted to achieve accurate processing in the research. In the research of object detection, the objects as airplanes in the research input are considered through RoI. The process of RoI extraction extracts the total number of objects in the single image as,

$$R = \left\{ R_1, R_2, \ldots. R_n \right\}, \tag{4}$$

where, R, is the combined RoI extracted images, and n, is the number of images, which is the same as the total number of objects available in the input image N_L.

3.4 Segmentation of the ROI extracted Image with hybrid attention based W-Net

Segmentation of the RoI extracted images is performed with the Tetralet attention-based W-Net, which provides enhanced outcomes in the research. The provided w-Net works as the unsupervised one, in addition to the Tetralet attention that resulted in the accurate segmentation outcomes [26].

The W-Net is the combination of two convolutional U-Nets, which acts as the complete model of segmentation that remains advantageous as other models stick only to the decoder module for the segmented outcomes. The learning in a complete model that is the encoder and the decoder obtained due to the accurate relation between the input and output. The W-Net model is designed as two sections, where the encoder obtains the feature map and processes the network-based pixel that shows the efficient estimation of the segmented image. Further, the second part named the decoder retrieves the feature information, which is collected by the encoder module. In such processing of recreating the image in the decoder, the encoder provides the segmented output. The basic optimization involved in the W-Net results in reduced Soft N-cut loss at the encoder, and minimized reconstruction loss at the decoder [27]. The encoder and decoder of the W-Net are represented as, E_{nc}, and D_{ec}. The major loss functions declared are, Soft N-cut loss, and reconstruction loss, which are elaborated as follows,

a) Soft N-Cut Loss:

With a certain set of vertices in the determined space, the graph-based normalized cut loss is initiated in W-Net represented as, Soft N-Cut Loss. The graph is made up of points called nodes and edges that are connected through the nodes. By eliminating the weights of the edge that connects the two sets, the similarity could be identified, which is named cut, and should be maintained to be low. N-cut loss is evaluated through the ratio of available nodes to available edges [28]. The maximized N-cut loss resulted in the correlation of the partitions that remained maximum. The incorporation of the Tetralet attention mechanism into the W-Net reduces the soft N-cut loss, and the outcome is represented as, Y_{SNCL} .

b) Reconstruction Loss:

The alignment with the input from the encoder decides the reduction of the reconstruction loss, which is represented as,

$$Y_{RL} = \left\| R - D_{ec}\left(E_{nc}\left(R; Z\left(E_{nc}\right)\right) Z\left(D_{ec}\right)\right)\right\|_2^2 , \quad (5)$$

where, the parameters of encoder and decoder are represented as, $Z\left(E_{nc}\right)$, and $Z\left(D_{ec}\right)$ respectively, and input is represented as, R , which is a collective representation of the RoI extracted images. The minimization Y_{RL} could be performed with the accurate training of the W-Net, and further, the network provides the balance between the accuracy and consistency by involving the Y_{SCNL} , and Y_{RL} parameters.

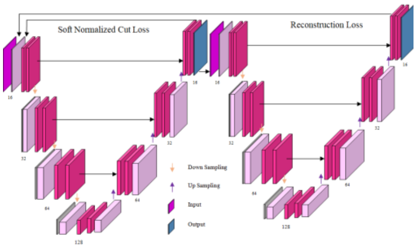


Fig. 2: W-Net in Segmentation of the RoI images.

The entire W-Nets' architecture illustrated in Figure 2 is composed of 14 modules, in which exists 28 convolutional layers. Each module is made up of two convolutional layers followed by batch normalization, ReLU function, and non-linearity function. The encoder block contains seven modules followed by the decoder block having the same seven modules. The encoder block has the input R , which moves through the contracting and expanding path, where the contracting path observes the context, and the expanding path performs localization. The incoming RoI extracted images are processed through convolution in the first module and the obtained features are doubled in further downsampling and further, incorporation with the 2×2 maxpooling layers. The entire process of the contracting path is reversed in the expanding path and the obtained features are reduced by double the time. To overcome the spatial information loss, the input of all contacting path modules is bypassed to the expanding path. The feature vectors obtained at W-Net are mapped with the 1×1 convolution. The Tetralet attention model is incorporated in the first layer of the W-NetThe final segmented outcome is represented as, Seg .

3.5 Feature Extraction

The feature extraction of research is performed to obtain the features that aid the research model in achieving accurate object detection. The feature extraction mechanisms included in the research are Modified object flow-based Ternary Pattern, Grey-level co-occurrence matrix features, shape and structure informatics features, and Hybrid attention-based ResNet-101 features.

3.5.1 Modified Object Flow-based Ternary Pattern

MOFTP is the feature extraction mechanism, which is included in the research to perform the pattern extraction [29]. The MOFTP of the research is estimated through the mathematical expression represented as,

$$COE = \begin{cases} -1, & Pix_{nei} < Pix_{cen} \\ 0, & Pix_{nei} = Pix_{cen} \\ 1, & Pix_{nei} > Pix_{cen} \end{cases} , \quad (6)$$

where, COE , represents the modified coefficient of the neighboring pixel, Pix_{nei} , indicates the neighboring pixels, and Pix_{cen} , shows the center pixel. The obtained coefficients of the MOFTP are divided as left and right side patterns, where left indicates the negative values, whereas right indicates positive values, and the others of both sides are denoted as zero. The negative values identified on the left side are converted into positive values, and further, all the binary values on both sides are converted into decimal values. The converted values are arranged in a counter-clockwise direction, which is represented through the following equation,

$$MOFTP = \sum_{i=0}^{7} COE.2^i , \quad (7)$$

Through the consideration of both side patterns histogram is generated to catch the initial features. Due to the compression provided by the histogram, the detection process speed could be increased, which resulted in the final features $\left(1\times32\times32\right)$.

3.5.2 GLCM features

The GLCM is the textual feature extraction that estimates the co-occurrence matrix [30]. The features of GLCM are as follows, Dissimilarity: The dissimilarity of the input should be high as the extracted local regions have high contrast, which is evaluated as,

$$D = \sum |u - v| N(Seg), \qquad (8)$$

where, N , indicates the normalized GLCM input with the pixel intensities randomly chosen as (u,v).

Energy: The squared elements are summed to obtain the energy, which is otherwise called uniformity and thus, evaluated as,

$$E = \sum N(Seg(u,v))^2 , \qquad (9)$$

Homogeneity: The nearestness of the pixel elements to the diagonal of the GLCM is represented as, homogeneity.

$$H = \sum \frac{N(Seg(u,v))}{1+(u-v)^2}, \qquad (10)$$

Correlation: The dependency of the pixel intensities with the neighboring pixels is estimated through the correlation feature in the GLCM, which is represented as,

$$cor = \sum \frac{(u-\mu_u)(v-\mu_v) N(Seg(u,v))}{\sigma_u \sigma_v}, \qquad (11)$$

where, μ , represents the mean, and σ , indicates the standard deviation of the pixel intensities.

Contrast: The contrast feature is utilized to estimate the distance from the mean diagonal of the GLCM matrix [31], where the highest distance ensures a high weight to the normalized input. The mathematical representation of the contrast is shown as,

$$C = \sum (u-v)^2 N(Seg(u,v)), \qquad (12)$$

The combination of the GLCM features are represented as, $GLCM = [D \,||\, E \,||\, H \,||\, cor \,||\, C]$, with the dimension of (1×6)

3.5.3 Hybrid attention based ResNet-101 features

When the gradient flow is accurately created by residual connections, the Residual Network, or ResNet, is extended. Even while using many convolutional layers in the CNN produced improved results, the key problem still lies in how time-consuming it is to extract the features. In the feature extraction process, ResNet 101 is introduced to carry out with the least amount of time complexity [32]. When the feature map's size is half while maintaining the lowest possible temporal complexity in ResNet 101, the number of filters used to produce the same feature map output is doubled. The ResNet-101 design consists of 33 blocks with 104 convolutional layers each. Of those 33 blocks, 29 are linked to the results of the modules that came before them. These results are approximated using residual connections, which serve as the accumulation operator's first operand. The remaining four of the 33 blocks take into account the input from the block before it and use it in the convolution layer with a filter size of, with batch normalization coming next. The primary goal of avoiding the layers is to eliminate the gradient vanishing problem by applying the attention mechanisms from the previous layer to the subsequent layer, which produces heightened results [33]. The advantage of reutilizing the activation functions is enhanced through incorporating the Tetralet attention mechanisms into ResNet 101, where the attention mechanisms added are elaborated in a further section.

3.5.4 Shape and Structural Informatics features

The shape and the structural informatics features are obtained through edge detection through the canny algorithm and the Local Directional Pattern (LDP). The extraction of the informatics features aids in obtaining the accurate detection of the object.

Canny Edge Detection Algorithm:

As far as image input is concerned, edge detection plays a major role. Hence, to perform the detection of edges, the canny algorithm is chosen as the algorithm as the model never failed to capture the edge or provided the false edge in the image. Further, the model provided the accurate localization of the edges and did not aim to provide the same response in the single edge [34]. To address the provided advantages, the canny algorithm follows six stages as follows,

Initially, the RoI extracted image is converted into the grayscale image, which is mathematically represented as,

$$g(u,v) = 0.299 R_{ed} + 0.587 G_{reen} + 0.114 B_{lue}, \qquad (13)$$

where, $R_{ed}, G_{reen}, B_{lue}$, are the color channels of the input image pixels. Further, to minimize the noise content as well as to smooth the image, the Gaussian function is introduced, and is defined as,

$$G(u,v) = \frac{1}{2\pi\varsigma^2} \exp\left(-\frac{u^2+v^2}{2\pi\varsigma^2}\right) \qquad (14)$$

where, ς , indicates the coefficient representing the space scale in the Gaussian function that works to manage the over-smoothening in the image. The obtained Gaussian function is utilized as the template to estimate the gradient's magnitude and direction [35]. The evaluation of the gradient magnitude is performed with the following mathematical expression,

$$M\left(G\left(u,v\right)\right)=\sqrt{G_u^2+G_v^2}\ , \qquad (15)$$

where, M , shows the magnitude, and additionally the gradient is evaluated as,

$$\theta\left(u,v\right)=\arctan\left(\frac{G_u^2}{G_v^2}\right), \qquad (16)$$

Further, there is required pixel suppression, which is performed through the detection of the non-maximum suppression method that determines the false edge points. The threshold determined is evaluated with the help of the gradient magnitude and direction, where the magnitude is required to be higher than the directional value. The high and the low threshold methods determined are aided in finalizing the edge points to perform the edge detection. The magnitude higher than the high threshold is considered an edge point, while others lower than the low threshold are considered as the non-edge points. Hence, the edge detected through the canny edge algorithm is represented as, K .

Local Directional Pattern:

The LDP of the image acts as the eight-bit binary code, which is assigned to all pixels of the input RoI extracted image. The evaluated image declares the values to encode the input image. Further, the evaluation is carried out with the Kirsch masks that work at eight different orientations [36], where the edge response values are represented as, $L_s, s=0,1,...,7$ .The evaluation of LDP is performed with the mathematical expression as follows,

$$LDP=\sum_{s=0}^{7}a_s\left(L_s-L_f\right).2^s\ , \qquad (17)$$

where, L_f , is the most important f^{th} factor of directional response value, the value of $a_s\left(L_s-L_f\right)$, is determined as,

$$a_s\left(L\right)=\begin{cases}1, & L\geq 0\\ 0, & L<0\end{cases}, \qquad (18)$$

where, $L=\left(L_s-L_f\right)$.

The outcome of the canny edge algorithm and the LDP are concatenated to establish the shape and structural informatics feature, which is represented as,

$$S_F=\left[M\left(G\left(u,v\right)\right)\|\,LDP\right], \qquad (19)$$

Finally, the outcomes of all feature mechanisms such as Modified object flow-based ternary pattern, GLCM features, hybrid ResNet 101 features, and the shape and structural informatics features are concatenated and are represented as,

$$T=\left[MLTP\,\|\,GLCM\,\|\,\mathrm{Re}\,s\,\|\,S_F\right], \qquad (20)$$

3.6 Object detection with Canis Socio-Hunt optimized Tetralet attention enabled Explainable CNN

Object detection from the remote sensing satellite images is quite difficult to the conventional mechanisms in the research sector. The proposed CHunt-TetraExNN provided the most accurate outcome in the research of object detection as the model included mechanisms such as Tetralet attention and Canis Socio-Hunt Optimization, which are hybridized into the base model named Explainable CNN. Due to the errors such as overfitting and error probabilities of traditional CNN, the explainable CNN is chosen as the base model to implement the research of object detection. In explainable CNN, the input image obtained from the feature extraction process is fed, through which Gradient-weighted class activation mapping (Grad CAM++) and the fully graded images are differentiated accurately. In the research, only the Grad CAM ++ is evaluated as it is the localization technique that provides the graphical representation of the CNN network without certain modifications in the structural descriptions [37].

$$Sal_{Map}\left(T\right)=\begin{cases}e_0T+o_0\\ \vdots & \vdots\\ e_nT+o_n\end{cases}, \qquad (21)$$

where, Sal_{Map}, represents the saliency map of GRAD CAM ++, e_0, o_o, represents the initial weight and bias respectively, e_n, o_n, are the final weights and biases respectively.

3.6.1 Tetralet Attention Mechanism

Efficient object recognition is achieved through the integration of positional attention and triplet attention into Tetralet attention. The three parallel branches of the triplet attention [29] are named for the two branches that achieve the cross-dimension interaction, which monitors the input dependencies, and the branch that develops the spatial attention. According to Figure 3, the triplet attention that is shown includes the z-pool that is intended to minimize the zeroth dimension when the attention results are combined [38]. The result of the triplet function's second layer is regarded as weight 1, while the results of the triplet function's first and third layers are regarded as weight 2 in the combined Tetralet attention. Moreover, the sum of the results is averaged to provide the aggregated weight, denoted by wei, which is supplied to the Positional Attention as an input. With the module's self-attention mechanism, the position attention mechanism is connected to the study of obtaining both global knowledge and non-local relationships. The weighted similarity between the places that characterize the attributes is estimated by the process. Generally speaking, the positional attention mechanism's similarity increases with increasing weight. The described positional attention is defined as,

$$Pos = \frac{1}{norm(wei)} \sum_{\forall L} binary(wei)\, unary(wei)$$

(22)

where, the binary function $binary$, along with the unary function $unary$. $norm$, is the normalization factor of the input [39].

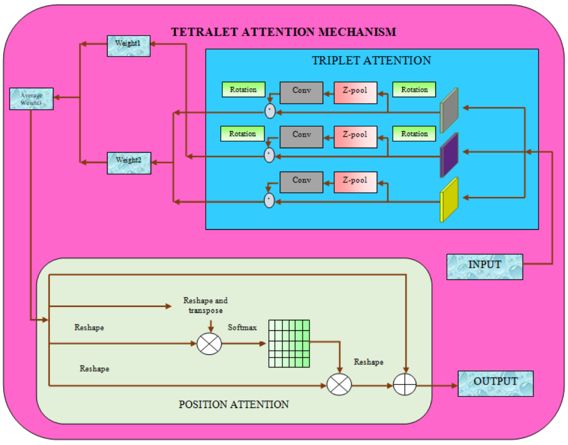


Fig. 3: Tetralet Attention mechanism

3.6.2 Canis Socio-Hunt Optimization

CHunt Optimization is hybridizing Coyote and Grey Wolf with their respective characteristics of adaptability [40] and hunting instincts [41] presenting a compelling approach for enhancing optimization in the domain of object detection from satellite images. The adaptability trait of the Coyote signifies the ability to dynamically adjust strategies and parameters during the optimization process, enabling the algorithm to efficiently navigate complex search spaces. On the other hand, the hunting nature of the Grey Wolf embodies a focused and strategic pursuit towards optimizing objectives, akin to the goal-oriented search for optimal solutions in object detection tasks. The advantages of utilizing this hybrid optimization approach in satellite image-based object detection are multifaceted. Thus, the CHunt optimization achieves highly computationally efficient outcomes.

Inspiration: The adaptable feature of social living is modeled after the Canis Latrans of the family Canidae, which is a band or pack that adapts greatest to its social behavior. The Canis Latrans band is led by a leader named Alpha who helps the group live in cooperative behavior and exchanges members with other members of the group; members typically move from one group to another and adapt to the behavior of that group, which continues to be useful to the group. The Canis Lupus, a member of the dog family, is the source of inspiration for group hunting behavior. It adopts the social behavior of a hierarchical pack, in which the alpha leads the pack and makes all of the group's choices. The beta, which helps with decision-making, resides at the next level of the hierarchical pack, while the omega, who serves as the pack's caregiver, is at the bottom. Although the delta subordinates to the higher level, it rules over the omega and serves as the pack's security manager, protector of its territory, caregiver for the elderly, and aid in food preparation and hunting. The key benefit of the pack is still that the hierarchical pack adheres to group hunting behavior.

Initialization: The initial population of the obtained solutions in the search space is declared as follows,

$$U = [U_1, U_2, \ldots\ldots, U_t], \qquad (23)$$

where, t, indicates the total population, and further, the declared solution is the representation of the weights of the explainable CNN classifier.

Objective Function Declaration: The objective function of the research is represented as,

$$Obj.Fun = \max(Accuracy), \qquad (24)$$

Solution Update: The solution of the CHunt optimization is declared through the availability factor A, as evaluated as follows,

Phase 1: Availability Phase: $|A| \geq 1$

The declared availability factor A shows the availability of the prey within the search space. This declaration aids the solution to follow the strategic moves towards the prey.

Case 1: Time 1 phase: $0 \leq c < \frac{1}{3}(c_{\max})$

The condition represents the solutions' position at the iteration c, and in the range of the lower and upper boundaries as follows,

$$U^c = y + l(z - y), \qquad (25)$$

where, y, indicates the lower boundary, z, is the upper boundary, and l, indicates the multi-iterative factor, which is evaluated as,

$$l = e^{c/c_{\max}} . \left(\frac{c-1}{c}\right), \qquad (26)$$

where, c, ranges from (0) to $\left(\frac{1}{3}\right)$.

Case 2: Time 2 phase $\frac{1}{3}(c_{\max}) \le c < \frac{2}{3}(c_{\max})$

The represented phase is declared as the intermediate phase in the entire availability phase as the time interval lies between the appropriate values of the iteration. The solution enabled at this intermediate phase fastens the hunting through characteristics such as the novel communication between the groups, social organization behavior, and so on. The sharing of information could be extended to the secondary and the low-level solutions as well. The obtained best solutions of the optimization are modeled as,

$$U^{c+1} = \left[U^c + q\delta_1 + h\delta_2\right], \tag{27}$$

where, δ_1, is the relative global best factor, which is evaluated as,

$$\delta_1 = U_\alpha - U^c, \tag{28}$$

where, U_α, is represented as the leading best solution. The relative global mean factor δ_2 is estimated as,

$$\delta_2 = U_{mean} - U_{rand}, \tag{29}$$

where, U_{mean}, is the average mean position of all packs, U_{rand}, is the random position of the solution. q, is the random inertia weight, calculated as,

$$q = 0.5 + \frac{rand()}{2}, \tag{30}$$

where, $rand() \in (0,1)$, and h, is the adaptive inertia factor, which is determined as,

$$h = \frac{\left|Q(U_\alpha) - Q(U^c)\right|}{\left|Q(U_\alpha) + Q(U^c)\right|}, \tag{31}$$

where, $Q(U_\alpha)$, is the fitness of the leading solution.

Case 3: Time 3 phase: $\frac{2}{3}(c_{\max}) \le c < c_{\max}$

The phase works on considering the final interval of the entire exploitation or the prey availability phase. In this time interval, the solution with the best fitness communicates with the entire population and guides them to make the attack successful. To perform a successful attack, the division of the group into the least popularities takes place, which establishes the hierarchical communication and organization. Through the described characteristics, a successful hunt takes place, which is evaluated as,

$$U^{c+1} = \frac{U_1 + U_2 + U_3 + U_4}{4}, \tag{32}$$

where, U_1, U_2, U_3, U_4, are the low-level hierarchical structure, which is estimated as,

$$U_1 = U^c + q\delta_1 + h\delta_2, \tag{33}$$

$$U_2 = U_\beta - B_1\left[\left|P_1.U_\beta - U^c\right|\right], \tag{34}$$

$$U_3 = U_\gamma - B_2\left[\left|P_2.U_\gamma - U^c\right|\right], \tag{35}$$

$$U_4 = U_\eta - B_3\left[\left|P_3.U_\eta - U^c\right|\right], \tag{36}$$

where, $U_\beta, U_\gamma, U_\eta$, are the second, third, and least preferable best solutions, and B_1, B_2, B_3, are the amplification factors that are evaluated generally as,

$$\left[B_1, B_2, B_3 = 2px - p\right], \tag{37}$$

where, p, is the linear decrement factor that ranges from $(2,0)$, and the random factor x ranges from $(0,1)$. Further, P_1, P_2, P_3, are the terms that show the mutable ability of the solutions, evaluated as,

$$\left[P_1, P_2, P_3 = 2j\right], \tag{38}$$

where, j, shows the communication index between the solutions.

Phase 2: Unavailability Phase: $|A| < 1$

The provided condition describes that the availability factor is less than one, and thus the group of solutions diverges to search for the prey. Thus the individual hunt phase begins or else, after an appropriate food search, the solutions converge to attack the prey.

Reevaluate the objective function: The objective function of the updated solution is re-evaluated, and further, the updated solution is declared as the best solution.

Termination condition: Once the best solution is declared, and the termination condition called $c > c_{\max}$, is achieved, the loop ends. The overall workflow of the CHunt optimization is represented in Figure 4

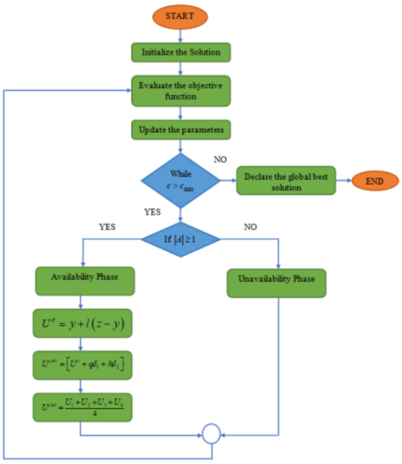


Fig. 4: Flowchart of the CHunt optimization

4. Results

The outcomes in the research of object detection with the CHunt-TetraExNN are analyzed in the section.

4.1 Experimental Setup

The present research is designed to evaluate the efficiency of the proposed model in object detection. The design is implemented in the Python software in the system having configuration of Windows 11, and 16 GB RAM storage.

4.2 Dataset Description

The utilized aircraft airbus dataset [24] contains the satellite images of the airport that are used in the research of object detection. Typically, airports are home to aircraft. Regularly, Earth observation satellites such as Airbus's Pleiades twin satellites take images of airports worldwide. It is possible to apply DL to automatically determine the quantity, kind, and size of aircraft on the property. Consequently, this can furnish data regarding the operations of any airport.

4.3 Experimental Analysis

The image results of the object detection research are shown in Figure 5, which is described in this section.

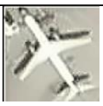
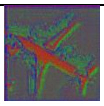
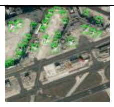
PROCESS	SAMPLE INPUT
Preprocessing	
Feature extraction	
Segmentation	
Mapped Image	

Fig. 5: Image analysis of CHunt-TetraExNN in object detection

4.4 Evaluation Metrics

The performance of the object detection with CHunt-TetraExNN is analyzed with metrics such as Accuracy, Precision, Recall, and F1-score. Accuracy measures the proportion of correctly detected instances among all instances. Precision measures the proportion of correctly detected positive instances among all detected positive instances. Precision is particularly useful when dealing with class imbalance, where certain classes are rarer but more critical to detect accurately. Recall measures the proportion of correctly detected positive instances among all actual positive instances. Recall is essential for tasks where it's crucial to capture all instances of a particular class, ensuring comprehensive coverage. The F1 score is the harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives, and further, F1 score is particularly valuable in scenarios with imbalanced classes, as it combines precision and recall to gauge the overall model performance.

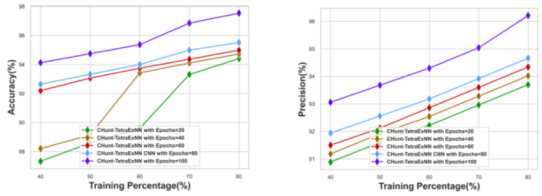
4.5 Comparative Methods

The proposed CHunt-TetraExNN is compared with the conventional methods such as Long Short Term Memory (LSTM) [42], Bidirectional Long Short Term Memory (BiLSTM) [43], Residual Network (ResNet-CNN) [23], AAFMCNN [44], TetraExNN, COA-TetraExNN, and GWO-TetraExNN. The outcomes are evaluated in terms of the training percentage (TP) and K-Fold.

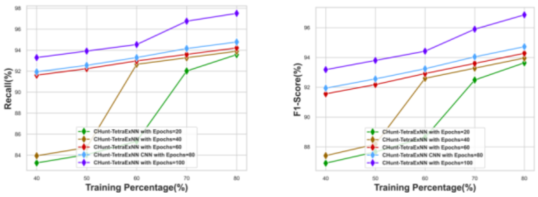
4.6 Result Analysis of CHunt-TetraExNN model

4.6.1 Performance Evaluation

The research of object detection with CHunt-TetraExNN at TP 80% and epoch 100% achieved a maximum accuracy of 97.53%. The achievement is due to the processing model that integrated Explainable CNN, Tetralet attention mechanism, and CHunt optimization. The proposed model achieved greater results similarly in F1-score at 96.85%., precision at 96.21%, and recall at 97.51%. The outcomes obtained in the research of object detection with CHunt-TetraExNN are depicted in Figure 6.



Accuracy **Precision**

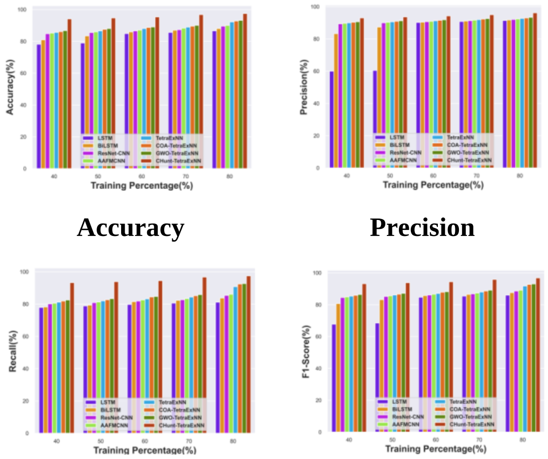


Recall F1-score

Fig. 6: Performance evaluation concerning TP

4.6.2 Comparative Evaluation

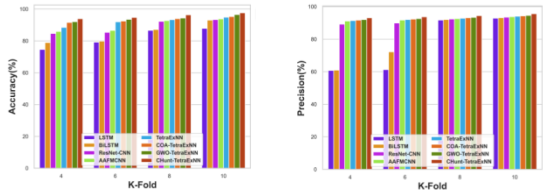
The comparison of the CHunt-TetraExNN with the existing methods showed an improvement of 7.64%, which is average among the improvements shown from the outcomes of LSTM, BiLSTM, ResNet-CNN, TetraExNN, COA-TetraExNN, and GWO-TetraExNN. The average improvement addressed at F1 score, Recall, and precision are 7.96%, 11.75%., and 3.93% respectively. As a result, the detection of objects with CHunt-TetraExNN produced superior results than the previous approaches. The better result is displayed with the K-fold 10 as being comparable to the comparison of the TP outcomes. Figures 7 and 8 display an examination of the results using the K-fold and TP, respectively.



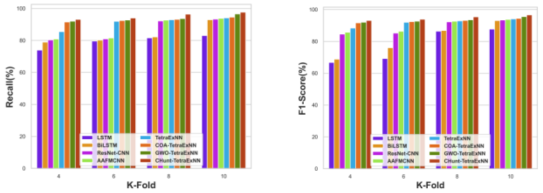
Accuracy Precision

Recall F1-score

Fig. 7: The comparison of outcomes with TP.



Accuracy Precision



Recall F1-score

Fig. 8: The comparison of outcomes with K-fold.

4.7 Comparative Discussion

The proposed CHunt-TetraExNN model achieved high accuracy results when compared with the other conventional methods. In terms of both LSTM and BiLSTM, there are sequencing issues as Satellite images are typically 2D grids, and applying the method directly can be challenging due to the sequential nature of these models designed for time-series or sequential data. Further capturing long-range dependencies in satellite images using BiLSTM might be inefficient, as they can struggle with remembering information over long sequences. Deploying very deep networks like ResNet for object detection can be computationally intensive, requiring significant resources, and ResNet-CNNs were originally designed for natural images, and directly applying them to high-resolution satellite images might not capture fine-grained details crucial for object detection. Implementing attention mechanisms within Explainable CNNs for object detection in satellite imagery requires careful design and tuning. Attention mechanisms can introduce additional computational overhead, and in addition, understanding and interpreting attention weights in the context of satellite image analysis can be challenging due to the unique characteristics of satellite data. Tuning parameters for optimized Explainable CNNs can be non-trivial and time-consuming. Due to the described drawbacks of the conventional methods, the proposed CHunt-TetraExNN is introduced in the research that achieved high efficiency. The comparison between the methods is tabulated in Table 1.

Table 1: Comparison results of CHunt-TetraExNN

Analysis/Methods			LSTM	BiLSTM	ResNet-CNN	TetraExNN	COA-TetraExNN	GWO-TetraExNN	CHunt-TetraExNN
Airbus Aircraft	TP	Accuracy (%)	86.65	88.11	89.58	90.08	92.26	92.84	93.22
		Precision (%)	91.46	91.78	92.10	92.42	92.74	93.06	93.38
		Recall (%)	81.18	83.78	85.42	86.10	90.81	92.46	92.72

		F1 score (%)	86.01	87.60	88.63	89.15	91.77	92.76	93.05
	K-Fold	**Accuracy (%)**	87.88	93.07	93.41	93.83	94.89	95.29	96.55
		Precision (%)	92.91	93.20	93.50	93.80	94.09	94.33	94.59
		Recall (%)	83.03	92.94	93.31	93.80	94.09	94.58	96.56
		F1 score (%)	87.69	93.07	93.40	93.80	94.09	94.45	95.56

5. Conclusion

The efficient object detection with the satellite images is performed with the CHunt-TetraExNN model that achieves high-efficiency outcomes. The CHunt-TetraExNN is comprised of the Explainable CNN base model in which the Tetralet attention mechanism and the CHunt optimization. The triplet attention and positional attention make up the Tetralet attention module that is included in the model. The module accurately estimates attentional features, which is very useful during the object detection phase. The segmentation in the research is carried out with the Canis Hunt Optimization, which combines the dominance and hunt traits of Latrans and Lapins, members of the Canis Family, further supporting the research concept. As a result, the model produced precise estimation results with an accuracy of 96.55%, precision of 94.59%, recall of 96.56%, and F1 score of 95.56% in the study of object detection from satellite photos. To improve the research model's efficiency, future research could be expanded to include working with learning algorithms.

References

[1] Zhao, Y., Yang, R., Guo, C. and Chen, X., “Parallel Space and Channel Attention for Stronger Remote Sensing Object Detection.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2023.

[2] Makantasis, K., Voulodimos, A., Doulamis, A., Doulamis, N. and Georgoulas, I., “Hyperspectral image classification with tensor-based rank-R learning models.” In 2019 IEEE International Conference on Image Processing (ICIP), 3148-3125,2019.

[3] Makantasis, K., Doulamis, A.D., Doulamis, N.D. and Nikitakis, A., “Tensor-based classification models for hyperspectral data analysis.” IEEE Transactions on Geoscience and Remote Sensing, 56(12), pp.6884-6898, 2018.

[4] Makantasis, K., Karantzalos, K., Doulamis, A. and Loupos, K., “Deep learning-based man-made object detection from hyperspectral data.” In Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I, 11, pp. 717-727, 2015.

[5] Ding, J., Xue, N., Xia, G.S., Bai, X., Yang, W., Yang, M.Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M. and Zhang, L., “Object detection in aerial images: A large-scale benchmark and challenges.” IEEE transactions on pattern analysis and machine intelligence, 44(11), pp.7778-7796, 2021.

[6] Kaselimi, M., Voulodimos, A., Daskalopoulos, I., Doulamis, N. and Doulamis, A., “A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring.” IEEE Transactions on Neural Networks and Learning Systems. 2022.

[7] Yi, D., Su, J. and Chen, W.H., “Probabilistic faster R-CNN with stochastic region proposing: Towards object detection and recognition in remote sensing imagery.” Neurocomputing,459, pp.290-301, 2021.

[8] Liu, L., Liu, Y., Yan, J., Liu, H., Li, M., Wang, J. and Zhou, K., “Object detection in large-scale remote sensing images with a distributed deep learning framework.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, pp.8142-8154, 2022.

[9] Zakaria, Y., Mokhtar, S.A., Baraka, H. and Hadhoud, M., “Improving Small and Cluttered Object Detection by Incorporating Instance Level Denoising Into Single-Shot Alignment Network for Remote Sensing Imagery.” IEEE Access, 10, pp.51176-51190, 2022.

[10] Wang, L., Shoulin, Y., Alyami, H., Laghari, A.A., Rashid, M., Almotiri, J., Alyamani, H.J. and Alturise, F., “A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images.” 2022.

[11] Wang, L., Long, C., Li, X., Tang, X., Bai, Z. and Gao, H., “CSFFNet: Lightweight cross-scale feature fusion network for salient object detection in remote sensing images.” IET Image Processing, 18(3), pp.602-614, 2024.

[12] Girshick, R., Donahue, J., Darrell, T. and Malik, J., “Rich feature hierarchies for accurate object detection and semantic segmentation.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

[13] Wang, K., Bai, F., Li, J., Liu, Y. and Li, Y., “MashFormer: A novel multiscale aware hybrid detector for remote sensing object detection.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, pp.2753-2763, 2023.

[14] Xu, Y., Wu, X., Wang, L., Xu, L., Shao, Z. and Fei, A., “HOFA-Net: A High-Order Feature Association Network for Dense Object Detection in Remote Sensing.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2023.

[15] Yu, Y., Gao, J., Liu, C., Guan, H., Li, D., Yu, C., Jin, S., Li, F. and Li, J., “OA-CapsNet: A one-stage anchor-free capsule network for geospatial object detection from remote sensing imagery.” Canadian Journal of Remote Sensing, 47(3), pp.485-498, 2021.

[16] Qu, J., Su, C., Zhang, Z. and Razi, A., “Dilated convolution and feature fusion SSD network for small object detection in remote sensing images.” IEEE Access, 8, pp.82832-82843, 2020.

[17] Girshick, R., “Fast r-cnn.” In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448, 2015.

[18] Ren, S., He, K., Girshick, R. and Sun, J., “Faster r-cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems, 28, 2015.

[19] Chen, C., Gong, W., Chen, Y. and Li, W., “Object detection in remote sensing images based on a scene-contextual feature pyramid network.” Remote Sensing, 11(3), p.339, 2019.

[20] Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M. and Guo, Z., “Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks.” Remote sensing, 10(1), p.132, 2018.

[21] Liu, S. and Huang, D., “Receptive field block net for accurate and fast object detection.” In Proceedings of the European conference on computer vision (ECCV), pp. 385-400, 2018.

[22] Zhang, T., Zhuang, Y., Chen, H., Chen, L., Wang, G., Gao, P. and Dong, H., “Object-Centric Masked Image Modeling Based Self-Supervised Pretraining for Remote Sensing Object Detection.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2023.

[23] Haryono, A., Jati, G. and Jatmiko, W., “Oriented object detection in satellite images using convolutional neural network based on ResNeXt.” ETRI Journal. 2023.

[24] Airbus Aircraft Detection Dataset, https://www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset accessed April, 2024.

[25] Kang, S.H. and Kim, J.Y., “Application of fast non-local means algorithm for noise reduction using separable color channels in light microscopy images.” International Journal of Environmental Research and Public Health, 18(6), p.2903, 2021.

[26] Xia, X. and Kulis, B., “W-net: A deep model for fully unsupervised image segmentation.” arXiv preprint arXiv:1711.08506, 2017.

[27] Saleh, E.S., Haridas, T.M. and Supriya, M.H., “March. Unsupervised image segmentation model based on w net architecture and conditional random field for underwater images.” In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1, pp. 34-39, 2021.

[28] Shi, J. and Malik, J., “Normalized cuts and image segmentation.” IEEE Transactions on pattern analysis and machine intelligence, 22(8), pp.888-905, 2000.

[29] Rangsee, P., Raja, K.B. and Venugopal, K.R., “modified local ternary pattern based face recognition using SVM.” In 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), (Vol. 3, pp. 343-350), 2018.

[30] Jafarpour, S., Sedghi, Z. and Amirani, M.C., “A robust brain MRI classification with GLCM features.” International Journal of Computer Applications, 37(12), pp.1-5, 2012.

[31] Zulpe, N. and Pawar, V., “GLCM textural features for brain tumor classification.” International Journal of Computer Science Issues (IJCSI), 9(3), p.354, 2012.

[32] Demir, A., Yilmaz, F. and Kose, O., “Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3.” In 2019 medical technologies congress (TIPTEKNO), (pp. 1-4), 2019.

[33] Ghosal, P., Nandanwar, L., Kanchan, S., Bhadra, A., Chakraborty, J. and Nandi, D., “Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network.” In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), (pp. 1-6), 2019.

[34] Yan, X. and Li, Y., “A method of lane edge detection based on Canny algorithm.” In 2017 Chinese Automation Congress (CAC) (pp. 2120-2124), 2017.

[35] Song, R., Zhang, Z. and Liu, H., “Edge connection based Canny edge detection algorithm.” Pattern

Recognition and Image Analysis, 27, pp.740-747, 2017.

[36] Kar, A., Bhattacharjee, D., Basu, D.K., Nasipuri, M. and Kundu, M., “An adaptive block based integrated LDP, GLCM, and Morphological features for Face Recognition.” arXiv preprint arXiv:1312.1512, 2013.

[37] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., “Grad-cam: Visual explanations from deep networks via gradient-based localization.” In Proceedings of the IEEE international conference on computer vision, (pp. 618-626), 2017.

[38] Misra, D., Nalamada, T., Arasanipalai, A.U. and Hou, Q., “Rotate to attend: Convolutional triplet attention module.” In Proceedings of the IEEE/CVF winter conference on applications of computer vision, (pp. 3139-3148), 2021.

[39] Zhuang, H., Qin, Z., Wang, X., Bendersky, M., Qian, X., Hu, P. and Chen, D.C., “Cross-positional attention for debiasing clicks.” In Proceedings of the Web Conference 2021, (pp. 788-797), 2021.

[40] Pierezan, J. and Coelho, L.D.S., “Coyote optimization algorithm: a new metaheuristic for global optimization problems.” In 2018 IEEE congress on evolutionary computation (CEC), (pp. 1-8), 2018.

[41] Mirjalili, S., Mirjalili, S.M. and Lewis, A., “Grey wolf optimizer.” Advances in engineering software, 69, pp.46-61, 2014.

[42] Teng, Z., Duan, Y., Liu, Y., Zhang, B. and Fan, J., “Global to local: Clip-LSTM-based object detection from remote sensing images.” IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-13, 2021.

[43] Peng, Y., Jia, S., Xie, L. and Shang, J., “An Attention-BiLSTM Model for Satellite Operation Prediction with Correlation Telemetry.” 2023.

[44] Wu, Z.Z., Wang, X.F., Zou, L., Xu, L.X., Li, X.L. and Weise, T., “Hierarchical object detection for very high-resolution satellite images.” Applied Soft Computing, 113, p.107885, 2021.