International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

Performance Analysis of Different Optimization Algorithms for Multi-Class Object Detection

[*1]Jay Laxman Borade, [2]Akkalakshmi Muddana
[*1]Research Scholar, Computer Science and Engineering
GITAM (Deemed to be University), Hyderabad, 502329, India.
[*]Email: 2260716401@gitam.in
[2]Professor, Computer Science and Engineering
GITAM (Deemed to be University), Hyderabad, 502329, India.
Email: amuddana@gitam.edu

Abstract: Object recognition is a significant approach employed for recognizing suitable objects from the image. Various improvements, particularly in computer vision, are probable to diagnose highly difficult tasks with the assistance of local feature detection methodologies. Detecting multi-class objects is quite challenging, and many existing researches have worked to enhance the overall accuracy. But because of certain limitations like higher network loss, degraded training ability, improper consideration of features, less convergent and so on. The proposed research introduced a hybrid convolutional neural network (H-CNN) approach to overcome these drawbacks. The collected input images are pre-processed initially through Gaussian filtering to eradicate the noise and enhance the image quality. Followed by image pre-processing, the objects present in the images are localized using Grid Guided Localization (GGL). The effective features are extracted from the localized objects using the AlexNet model. Different objects are classified by replacing the concluding softmax layer of AlexNet with Support Vector Regression (SVR) model. The losses present in the network model are optimized using the Improved Grey Wolf (IGW) optimization procedure. The performances of the proposed model are analyzed using PYTHON. Various datasets are employed, including MIT-67, PASCAL VOC2010, Microsoft (MS)-COCO and MSRC. The performances are analyzed by varying the loss optimization algorithms like improved Particle Swarm Optimization (IPSO), improved Genetic Algorithm (IGA), and improved dragon fly algorithm (IDFA), improved simulated annealing algorithm (ISAA) and improved bacterial foraging algorithm (IBFA), to choose the best algorithm. The proposed accuracy outcomes are attained as PASCAL VOC2010 (95.04%), MIT-67 dataset (96.02%), MSRC (97.37%), and MS COCO (94.53%), respectively.

Keywords: Object recognition, Hybrid deep learning, improved optimization algorithms, Gaussian filtering, Grey wolf optimization, Multi-class classification

I. INTRODUCTION

The availability of a large number of remote sensing images (RSI) promotes correlation study in comprehending the content of remote sensing images, including scene classification, image retrieval [1], aeroplane recognition [2], detection of vehicles [3], recognition of building [4] etc. In these applications, object-related data is evaluated from very high resolution (VHR) RSI, which requires object detection [5], which is both fundamental and important. Nevertheless, there are man-made objects with strong boundaries that curiously stand out from the background and landscape elements in wide-field images that are analogous to the background. The complexity of item detection is made even more challenging by these varied things. There may be many objects in a distant sensing image, but they are insignificant compared to the intricate background. As a result of shifting viewpoints, lighting conditions and weather, the shape and size of the objects may also alter [6]. In a complex environment, defining and identifying the indescribable little items is difficult.

High spatial resolution remote sensing images present greater obstacles and problems for multi-class object detection, which has garnered much interest and led to the development of numerous object detection techniques. Four categories, namely template-based method, knowledge-based method, machine learning-based method and object-based image analysis method, can be used to categorize these object detection techniques in VHR remote sensing images [7]. By creating diverse rules and knowledge, such as context information [8] and geometry information [9], knowledge-based methods often translate object detection into hypothesis testing problems. Although it is reliable for detecting a single thing, it is ineffective when several objects are present. In order to identify the most effective matches, the template-based object detection approach typically hand-crafts a template for each type of object to be detected before matching the image at every possible point [10]. It is simple to implement but unstable whenever an object changes size or direction. The object-based image analysis approach splits the RSI into numerous objects, representing a generally homogeneous group of pixels.

After image classification, these objects are classified into different categories according to a predetermined set of multi-feature mapping criteria [11]. It provides a framework to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

circumvent the limitations of traditional images based on pixel categorization techniques and the ability to leverage the capabilities of geographic information systems. Since the accuracy of the subsequent image classification depends directly on the delineation quality of the objects, the segmentation method must be extremely precise. Regarding the machine learning approach [12], object detection can be achieved by developing a classifier for a variety of texture features, including SR-based features (sparse representation) [13], scale-invariant feature transformation (SIFT) [14], bag-of-words (BOW) feature [16], Histogram of Oriented Gradients (HOG) [15] and hair-like features [17]. The classifier can typically be trained using a variety of methods, including k-nearest neighbors (kNN), AdaBoost, Support Vector Machines (SVM), Sparse Representation Based Classification (SRC), Conditional Random Fields (CRF) and many others [18]. In some cases, the machine learning-based technique performs better than the alternatives, but choosing hand-built features and training data would significantly impact how well it performed.

Deep learning using Convolutional Neural Networks (CNNs) is an effective hierarchical feature representation system that can automatically learn and extract different levels of robust features through multi-layer perception to increase recognition efficiency and accuracy. The typical CNN-based algorithm faces numerous obstacles when identifying objects in large-scale, complicated remote-sensing imagery [19]. First, the CNN-based approach might have problems and ignore these weak features since the object's feature in the top convolution feature map has been significantly weakened due to the low spatial resolution. Furthermore, the object's feature is lost in the intricately cluttered backgrounds, leaving the object's feature representations lifeless. The backgrounds typically occupy a significant percentage of the remotely sensed image, and during the training phase, there is a difference between the negative and positive data [20]. Because of this, the traditional CNN-based algorithm often overlooks the small object in favor of background noise. The research uses a hybrid convolutional neural network model to reduce these problems. Starting from these existing problems, a novel DL-based methodology with improved results is presented, and the best loss optimizer can be selected in the proposed research work. The main contribution of the proposed research work is stated as follows.

- To introduce an efficient DL based object recognition model called H-CNN to render better recognition outcomes.
- To pre-process the collected images using the Gaussian filtering model to enhance the image quality and eradicate the noise.
- To localize the image objects using GGL and to extract the effective features using the AlexNet model.
- To recognize and classify the suitable objects by replacing the softmax layer of AlexNet with the SVR model to obtain enhanced classification accuracy and lower error rates. The losses in the network model are optimized effectively using the IGW optimization algorithm.
- To compare the performances of the proposed work by varying the loss optimization algorithms like IPSO, IGA, IDFA, ISAA and IBFA to prove the performance superiority of the proposed IGW optimization procedure.

The remaining structure of the paper is specified as follows: Section 2 determines the related works discussing multi-class object detection. Section 3 deliberates the proposed methodology. Section 4 delivers the result and discussion. Finally, section 5 illustrates the conclusion and future scope.

II. RELATED WORKS

Zhu et al. [21] construct shape signature networks (SSN), which serve as a soft constraint to improve the feature capabilities of multi-class discrimination for 3D object identification. Compact and effective as a target and strong over noise and sparsity are two enticing characteristics. The established shape signature served as the foundation for the development of the shape signature networks for object detection from point clouds, which leverage shape information to support multi-class detection through shape-aware heads and shape signature goals.

To simultaneously recognize and track 3D objects from the Lidar point-clouds, Yin et al. [22] suggested a center-based system. In order to generate a heatmap from above and other dense regression outputs, this method employs a typical 3D point-cloud encoder with a few convolutional layers. The simple local peak extraction method for detection is combined with a refinement and closest distance adjustment method for tracking.

A unique VPFNet is proposed by Wang et al. [23] to detect 3D objects, particularly for simultaneously object tracking from many classes and more difficult classifications like pedestrians. The basis of VPFNet is the VPF layer, which merges features according to the geometric relationship and various properties of a voxel-pixel pair. The model is difficult to use and complex to learn and implement.

Chen et al. [24] introduced a neural network architecture called Sample-WeIghted hyPEr Network (SWIPENet) to detect small underwater objects. Also, a sample re-weighting technique called Invert Multi-Class Adaboost (IMA) was introduced to address the noise issue. However, it cannot minimize noise or losses or detect small things.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

To address the issue of ship detection in remotely sensed images, Zhang et al. [25] introduced a few-shot multi-class ship detection technique with an attention feature map and a multi-relation detector (AFMR). The fundamental structure was expanded to include a feature attention map module with Siamese network infrastructure and a multi-relation head. In the feature attention map module, an attention feature map was built by enhancing the object's characteristics in search images using average pooled support features. However, using this method requires a lot of time.

Problem statement: Technological advances are used by people as support to complete everyday actions. The objects possess multiple characteristics, including variability in orientation, size, shape and degree of mobility, which are considered fundamental problems. Recognizing different shapes of objects from images is one of the most demanding applications in computer visualization. Searching for target regions and presenting corresponding categories proves to be complex. Analysis of class probabilities proves difficult due to excessive noise in the images collected by remote sensing. Due to the rapid growth of images in different areas, the computational effort to identify the appropriate objects is high. Over the past few decades, several researchers have developed significant models for detecting different types of objects in different images. Several ML-based algorithms are used in the existing research but are unsuitable due to degraded feature training, ineffective loss minimization, low accuracy, and generalization performance. A novel DL-based methodology with improved results is proposed, and the best loss optimizer can be selected in the proposed research work.

III. PROPOSED METHODOLOGY

Recognition of an object in the image using a hybrid convolutional neural network (H-CNN) is proposed to obtain improved recognition accuracy. AlexNet's integration with SVR is used to design H-CNN, performing the combined feature extraction and object recognition. The attributes are extracted via AlexNet and fed into the SVR to recognize the object. The optimal information learning criteria eliminate the loss associated with the classification. Therefore, an improved Gray Wolf (IGW) optimization is proposed to minimize information loss.

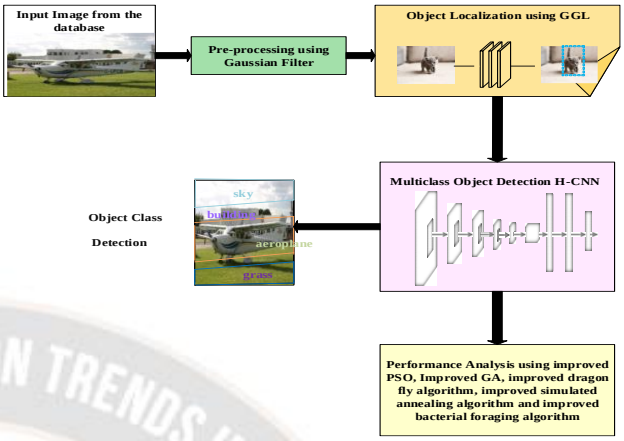


Figure 1: Workflow of the proposed object detection method

The H-CNN object detection workflow is shown in Figure 1. The input image is taken from the dataset for processing the proposed methodology. The artifacts from the image are eliminated using a filtering technique called Gaussian filtering. Then the image localization by the Grid Guided Localization (GGL) technique is used. Finally, multi-class object detection is employed using the H-CNN approach, minimizing information loss through IGW-based optimal learning to improve detection accuracy.

A) Noise Removal using Gaussian Filter

The noise from the input image is initially eliminated using a pre-processing technique where the non-uniform corrections are developed on the high contrast image. Noise removal is used using Gaussian filtering to get a better quality image. Image smoothing and noise reduction are applied to the input image using the Gaussian filter convolution operator. In addition, a better perception result is achieved with the Gaussian blur effect. Let's consider an image $D(C)$ taken from the database and expressed as

$$D(C) = \left\{ x_{pq}, i_{pq}, \ldots i_{lq} \right\}; p = 1,2,3,\ldots l \quad (1)$$

Where, $'x_{pq}'$ refers to the $'x^{th}'$ image in the database from its entire collection intended as l. Then, the expression for estimating the Gaussian filtering is formulated as follows,

$$E(r,s) = \frac{1}{2\pi\alpha} e^{-\left[\frac{r^2+s^2}{2\sigma^2}\right]} \quad (2)$$

Where, $'\alpha'$ indents the standard deviation and (r,s) refers to the location of the pixels. Gaussian filtering is used because of its faster computation capability while performing the core bit estimation. In addition, while removing the noise, it preserves the edge points, enhancing the noise rejection rate.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

The higher value of $'\alpha'$ smooth image is better and is good for removing the salt and pepper noise.

B) Localization of Object

The representation of the objects using the bounding boxes is devised in the object localization phase. Here, the noise removed image using Gaussian filtering is utilized for localizing the object. The Grid Guided Localization (GGL) technique based bounding box representation is utilized in the proposed method for object localization.

1) Grid Guided Localization (GGL)

The object's location is predicted using the GGL using a full convolutional network (FCN), using the grid points to indicate the object's bounding boxes. The grid with the size $R \times R$ is considered for the bounding box, the object with four edges and a centre point. A heat map $R \times R$ is incorporated with the outcome, and the resolution of 56×56 is utilized for predicting the grid points. In this, the heat map observes pixel by pixel sigmoid function to acquire the required map. While estimating the grid point for bounding the object, the secure pixel is chosen during the penetration process for identifying the actual grid point.

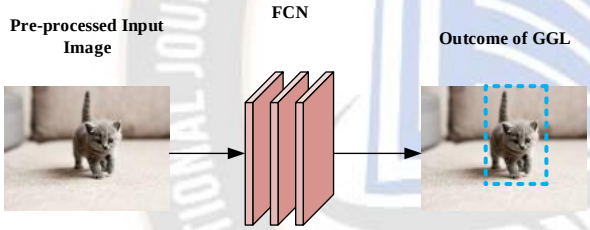


Figure 2: An illustration of GGL

An illustration of GGL is portrayed in Figure 2. Let the original image be indicated (U_r, U_s), and the corresponding heat map is notated as (K_r, K_s). Then, the expression for the original image is formulated as,

$$U_r = V_r + \frac{K_r}{W_z} W_e \tag{3}$$

$$U_s = V_s + \frac{K_r}{g_z} g_e \tag{4}$$

Where, the heat map's width and height are defined as (W_z, g_z) for the output and (W_e, g_e) for the input. Then, the upper left corner of the input image is indicated as (V_r, V_s). The outcome of the object localization is the bounding, which is performed based on the grid points. The bounding box with its four corners is defined as $P = (r_l, s_u, r_c s_b)$, wherein the lower, right, upper and left edges are defined as $r_l, s_u, r_c s_b$. The outcome of the GGL bounding box is depicted in Figure 3.

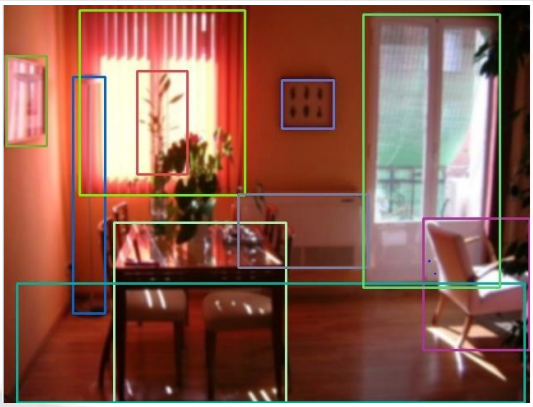


Figure 3: GGL based Bounding Box

Let V_y refers to the projection probability corresponding to the grid E_y with (r_y, s_y) coordinates, where y refers to the grid point. For the edge x, the set of grid points is notated as F_x, which means $y \in F_x$. Then, the formulation for the bounding box is formulated as follows,

$$r_n = \frac{1}{R} \sum_{y \in F_1} r_y V_y, \; s_u = \frac{1}{R} \sum_{y \in F_2} s_y V_y$$
$$y_c = \frac{1}{R} \sum_{y \in F_3} r_y V_y, \; s_b = \frac{1}{R} \sum_{y \in F_4} s_y V_y \tag{5}$$

The deviation is minimized through the bounding box localization of the object within the inner spatial grid points.

C) Multi-Class Object Detection using H-CNN

The most informative attributes are extracted automatically using AlexNet, and SVR employs multiple object detection. Thus, the model designed by hybridizing the SVR and AlexNet constitutes the H-CNN. Here, to hybridize the two architectures, the SVR is integrated into the AlexNet by replacing its softmax layer with SVR to more accurately detect different objects in the image. The significant attribute extraction using the AlexNet enhances the detection accuracy by extracting the cross-class separation and intra-class aggregation attributes. The extracted attributes are used for the classification with the SVR.

2) Attribute Extraction using AlexNet

The AlexNet based feature extraction was developed to preserve the automatic deep features. The architecture of H-CNN, with its layered architecture, is shown in Figure 4.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

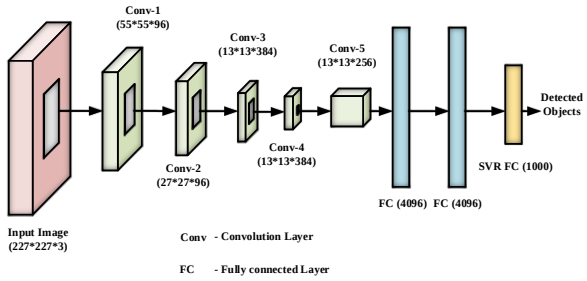


Figure 4: Architecture of H-CNN

The result of the GGL is fed into the input of the proposed H-CNN architecture to recognize the objects in the image. The significant attributes are extracted using the Conv layers, wherein the convolutional kernels C_k are utilized for mapping the features. Finally, the SVR replaces the last fully connected layer of the AlexNet to more accurately detect the objects in the image through the regression criteria. In addition, including the dropout layer to regularize the generated feature maps and the rectified linear unit (ReLU) based nonlinear transformation improves the detection accuracy. The ReLU-based transformation is formulated as follows:

$$f(m) = \max(m,0) \tag{6}$$

Here the role of the ReLu is to solve the overfitting problems of the H-CNN and increase the learning speed for the classifier. The inclusion of the SVR with the AlexNet is to solve the complex issues in classifying the objects more accurately. Let the features extracted by the AlexNet for each sample is indicated as S_X, and the total training data be defined as $\{s_x, m_x\}_{x=1}^{p}$. The total number of classes is represented as R for the total samples p; then, the target class is defined as $m_x \in \{1,2,...R\}$. The regression function for the SVR is formulated as follows,

$$f(s) = W \cdot s + B \tag{7}$$

Where, S refers to the support vector, B refers to bias and W refers to weight. The high dimensional features are mapped using the Radial Basis Function (RBF) kernel function and are formulated as follows,

$$H(s_x \cdot s) = \exp\left(-\frac{\|s_x - s\|^2}{2\alpha^2}\right) \tag{8}$$

Where, a real number with a positive value is indicated as α. The inefficiency in the support vector model is solved using the insensitive factor γ — and is formulated as,

$$|\gamma|_F = \begin{cases} 0, if\ |\gamma| < \beta \\ |\gamma| - \beta, o.w \end{cases} \tag{9}$$

Where, β refers to the error limit and the expression for optimal SVR is formulated as,

$$\min_{B,W,\gamma_X^*,\gamma_X} \frac{1}{2}\|W\|^2 + Q\sum_{x=1}^{p}\left(\gamma_x + \gamma_x^*\right) \tag{10}$$

Where, the minimum error and maximum margin are balanced with the regularization factor Q and γ_X^*, γ_X refers to the slack variables. Here, the IGW algorithm is utilized to optimize the weight vector in the SVR.

3) The Proposed IGW Optimization Algorithm based H-CNN Model

The proposed IGW helps reduce the error in classification and is used to swap the weight values as the worst based on the best solution. The basic idea of this optimization algorithm is to provide hierarchical management and hunting mechanisms for gray wolves in nature [27]. Initially, the wolves are located randomly from the feed. The fitness value depends heavily on the distance calculation between the feed and the GWs. The wolves achieved the best three fitness values emphasized as $\alpha, \beta,$ and δ, respectively. The final position of each wolf is based on the arithmetic mean adjusting the position of three leaders using the following equations:

$$\hat{E} = \left|\hat{D} \times \hat{Y}_p(z) - \hat{Y}(z)\right| \tag{11}$$

$$\hat{Y}(z+1) = \hat{Y}_p(z) - \hat{B} \times \hat{D} \tag{12}$$

$$\hat{D} = 2.\ \hat{b}\ .\ \hat{s}_1 - \hat{b} \tag{13}$$

$$\hat{B} = 2.\ \hat{s}_2 \tag{14}$$

Here $\hat{Y}$ represents the GW vector position, $\hat{Y}_p$ signifies the target location vector, z and $z+1$ manipulates the present and new iterations. Also, $\hat{B}$ and $\hat{D}$ manipulates coefficient vectors, $\hat{s}_1$ and $\hat{s}_2$ signifies the twofold random vector values range between 0 and 1, and $\hat{b}$ manipulates exploitation-exploration vector coefficients, which are defined as follows.

$$\hat{b} = 2 \times \left(1 - \frac{z}{Max_{it}}\right) \tag{15}$$

For obtaining a high-level transition from the time of exploitation to exploration, the new exploration factor $\hat{b}_{new}$ is adopted, which is mathematically represented in the equation

$$\hat{b}_{new} = 2 \times \left(\cos\left(\frac{(\tanh\theta)^2 + (\theta \sin \pi\theta)^l}{(\tanh 1)^2} * \frac{\pi}{2}\right)\right)^2 \tag{16}$$

$$\theta = \frac{z}{Max_{it}} \tag{17}$$

Here z and Max_{it} represents the current and the maximum iterations, respectively, also θ signifies the quotient of z divided

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

by Max_{it}. The factor l adjusts the deterioration curve of the search vector $\hat{b}_{new}$ at the time of the exploration process.

However, the target position vector is undefined, therefore each wolf in the group contemplates enhancing their location based on $\alpha, \beta,$ and δ as given below:

$$\hat{E}_{\alpha} = \left|\hat{D}_1 \times \hat{Y}_{\alpha}(t) - \hat{Y}(z)\right| \quad (18)$$

$$\hat{E}_{\beta} = \left|\hat{D}_1 \times \hat{Y}_{\beta}(t) - \hat{Y}(z)\right| \quad (19)$$

$$\hat{E}_{\delta} = \left|\hat{D}_3 \times Y_{\delta}(z) - Y(z)\right| \quad (20)$$

$$\hat{Y}_{ad1} = \hat{Y}_{\alpha}(t) - \hat{B}_1 \times \hat{B}_{\alpha} \quad (21)$$

$$\hat{Y}_{ad2} = \hat{Y}_{\beta}(t) - \hat{B}_2 \times \hat{B}_{\beta} \quad (22)$$

$$\hat{Y}_{ad3} = \vec{X}_{\delta}(t) - \vec{A}_3 \times \vec{D}_{\delta} \quad (23)$$

$$\hat{Y}(z+1) = \left(\hat{Y}_{ad1} + \hat{Y}_{ad2} + \hat{Y}_{ad3}\right)/3 \quad (24)$$

Here $\hat{Y}_{\alpha}, \hat{Y}_{\beta}$, and $\hat{Y}_{\delta}$ represents the three wolves position vectors, whereas $\hat{E}_{\alpha}$, $\hat{E}_{\beta}$ and $\hat{E}_{\delta}$ manipulates the distance vectors, and $\hat{Y}_{ad1}$, $\hat{Y}_{ad2}$ and $\hat{Y}_{ad3}$ represents the position adjustments based on three wolves. In addition to this, $\hat{B}_1$, $\hat{B}_2$ and $\hat{B}_3$ contemplates three vector characterizations of $\hat{B}$, and $\hat{D}_1$, $\hat{D}_2$ and $\hat{D}_3$ are the three vector characterizations of $\hat{D}$.

However, detecting the SVR hyperparameters is an optimization problem using a set of real numbers. The objective function can be mathematically represented as,

$$Max_{acc} = SVR(C, \varepsilon, \gamma) \quad here \quad u < u_{\max} \quad (25)$$

Equation (25) manipulates the SVR tuning objective function, D, ε, γ denotes hyperparameter values of l^{th} dimension. The peripheral parameter $u_{\max}$ organizes the number of iterations required for SVR to tune the hyper-parameters. Algorithm 1 represents the pseudo-code of the proposed IGW Optimization Algorithm.

Algorithm 1: Loss optimization using IGW Algorithm

Start

Population initialization of grey wolfs representing the SVR hyperparameters

Assemble training, testing and validation data

For each wolf u, **do**

Decode u into corresponding SVR hyperparameters

Train the model by training data

Evaluate the model based on validation data

Computation of fitness value

End For

Determine the three leading wolves $\alpha, \beta,$ and δ with the best fitness values

While ($_{z < Max_{it}}$)

Update GW population based on the searching process of proposed IGW using equations 18-24.

End While

Recover the appropriate best result Y_{α}

Decode $_{Y_{\alpha}}$ into corresponding SVR hyperparameters

Train the optimized model based on training and validation data combination

Evaluate the optimized SVR model dependent upon test data

Recover the test outcome

End

D) Different optimization algorithms for loss minimization

To evaluate the performance of loss minimization, the proposed H-CNN model has considered different improved optimization algorithms for multi-class object detection. The improved optimization algorithm includes improved particle swarm optimization (IPSO), improved genetic algorithm (IGA), improved dragon fly algorithm (IDFA), improved simulated annealing algorithm (ISAA) and improved bacterial foraging algorithm (IDFA).

4) Improved particle swarm optimization

The PSO algorithm [26] uses a population-based search method that simulates the social behavior of flocks of birds. Particles are individuals flown over the hyper dimensional search area using the PSO method. In order to erase other people's performance, the placements of the particles within the search area are altered according to the people's socio-psychological tendencies. The experience or knowledge impacts how the particles change within the swarm. Simulating this social behavior means the search is processed to return to earlier productive areas of the search space. Specifically, the following equations will change each particle's velocity (C) and position (p):

$$c_{ij}(t+1) = I_w c_{ij}(t) + l_1 a_1 \left(kG_{ij}(t) - p_{ij}(t)\right) + l_2 a_2 \left(hG_{ij}(t) - p_{ij}(t)\right) \quad (26)$$

$$p_{ij}(t+1) = p_{ij}(t) + c_{ij}(t+1) \quad (27)$$

Where $c_{ij}(t+1)$ denotes the particle's velocity i at iteration j and $p_{ij}(t+1)$ denotes its location at the same iteration. I_w Is an inertia weight that will be utilized to reduce the current behaviour of the velocity's influence from its previous behavior? The iteration number is represented by the t, while

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

the cognition learning component is represented as l_1, the social learning factor is l_2 and the remembering ability, a_1 and a_2 are represented by random values in the range [0, 1]. For controlling excessive wandering of the particle outside the search space, the value of each component in C can often be clamped to the range $[_{-C_{\max}}, C_{\max}]$. The PSO method ends when it reaches its maximum generation count or when the best particle location is in the swarm, which cannot be improved after reaching a sufficient number of generations. As a result, the PSO method has proven to be reliable and effective at resolving challenging optimization issues.

Thereby, the learning factors have enhanced linearly based on the change in learning factors. The l_1 and l_2 parameters of the fundamental PSO algorithm are provided beforehand based on experience. However, their values [0, 4] will inhibit the particles' capacity for self-learning. The parameter ranges for l_1 and l_2 are provided in this study. The beginning and final values are $l_1 \in$(2.75, 1.25) and $l_2 \in$ (0.5, 2.25), respectively. The following describes how the learning factor function expresses the linear change:

$$l_1 = l_{1\max} + (l_{1\min} - l_{1\max}) \times t / M \quad (28)$$

$$l_2 = l_{2\max} + (l_{2\min} - l_{2\max}) \times t / M \quad (29)$$

Where $l_{1\max}$ and $l_{2\max}$ represent the starting values of l_1 and l_2 as well as $l_{1\min}$ and $l_{2\min}$ represent their respective final values, M is the total amount of iterations performed and t is the number of iterations currently performed.

5) Improved genetic algorithm

GA is one of the population-based algorithms, where the basic steps are selection, crossover, and mutation. The inspiration for GA [27] comes from the Darwinian theory of evolution while the genes are being replicated. The GA algorithm is specifically used to optimize the loss functions and helps to avoid local optima. Each parameter in GA characterizes a gene, while each solution considered in this algorithm is represented as a chromosome. The fitness of all traits in the population can be examined by looking at the fitness function. Depending on the roulette wheel mechanism, the losses can be optimized with local optima avoidance. The evaluated fitness function is described as follows. The four main stages performed in GA are population initialization formation, best solution selection, crossover, and mutation. Each phase of GA can be explained as follows.

Population Initialization: In GA, a random population are initialized, and for diversity maximization, the population is generated from a Gaussian random distribution. Numerous solutions are encompassed in the initialized population, whereas the generated solutions indicate the individual chromosomes. In the case of the gene or feature stimulation, every chromosome comprises a variable set. The major reason to initialize this step is to spread the individuals uniformly among the search area to enhance population diversity and produce effective solutions.

Selection process: The selection process phase resembles a natural process in which suitable individuals have improved chances of feeding and mating. As a result, the corresponding genes are promoted more extensively to produce the following generation of suitable species. Depending on the knowledge gained, GA uses the roulette wheel mechanism. The probabilities can be assigned to the individuals by a roulette wheel mechanism that selects the genes for the formation of the next generation to the corresponding fitness values. The main benefit of this phase is that it can encourage the natural selection of the fittest individuals in the population.

Crossover process: In selecting the best individuals that will effectively optimize the loss function in the population through a selection process, the crossover phase is performed to bring forth a new generation. Integrating two selected solutions from the dominant phase produces a single solution. Various crossover operations like a uniform crossover, single point, And double point rendering improve the results.

Drawbacks of Genetic algorithm: In the general GA, the main drawback is the generated solutions are stuck with the earlier convergence and local optimum issues. To overcome this issue, IGA is employed, whereas Great Deluge Algorithm (GDA) is blended with GA to enhance the local searching process. By introducing IGA, the local maximum can be effectively avoided to attain the global maximum in the search space through the accomplishment of stochastic child chromosome perturbation.

Great Deluge Algorithm: After the crossover process, GDA is employed for processing the child chromosomes generated in the crossover operation. In this stage, every child chromosome undertakes a series of perturbations until they tend to be weak to be omitted. Otherwise, the chromosome keeps searching for a better solution around the space. Search space exploitation can be effectively enhanced through the combination of GDA.

Mutation process: The final process in GA is a mutation; one or several genes are rehabilitated after creating new solutions. The mutation operator assures population diversity through random differences concerning gene data and aims to preserve the chromosomes from a very high equivalent over numerous generations. The mutation process is based on an independent gene displacement over a similar level. The mutation point is chosen randomly while the gene information is exchanged with the original independent gene. Also, another

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

level of randomness is provided in this process and develops population diversity. Additionally, this phase impedes the solutions from becoming similar and reduces the local solution selection probability as an optimal one.

6) Improved dragon fly algorithm:

The improved dragon fly algorithm (IDFA) is analyzed to determine its performance in reducing classification loss. Traditional dragon fly algorithm (DFA) [28] considers five essential components for updating the position of alive objects from the group of dragon flies. It can be mathematically interpreted based on the movement of the static group and dynamic groups' movements, respectively. The five essential components like separation, symmetry, steadiness, feedstuff, and enemy, can proceed in a step-by-step procedure to address the optimal solution. In addition to this, the step vector and position of DFs are selected randomly based on the upper bounds and lower bounds efficiently. During every iteration, the position of DFs gets altered, and each candidate's neighbours are selected based on the distance between the updated DF's positions. The location gets updated continuously till the stopping condition is obtained.

However, due to premature convergence, the traditional DFA technique effectively reduces the classification loss and fails to achieve a globally optimal solution. To overcome this issue, two alterations must be undertaken to prevent the earlier convergence and local optimal issues. In traditional DFA, two stochastic quantities are the main reason that makes the algorithm for earlier convergence. Hence, the chaos concept is introduced to overcome this issue, and the singer mapping chaos function is utilized. The improved stochastic quantities s_1 and s_2 can be mathematically formulated as,

$$s_1^{i+1} = 1.07\left(7.9s_1^i - 23.3\left(s_1^i\right)^2\right) + 28.7\left(s_1^i\right)^3 - 13.3\left(\left(s_1^i\right)^4\right) \quad (30)$$

$$s_2^{i+1} = 1.07\left(7.9s_1^i - 23.3\left(s_2^i\right)^2\right) + 28.7\left(s_2^i\right)^3 - 13.3\left(\left(s_2^i\right)^4\right) \quad (31)$$

Also, the chaos function uses a quasi-oppositional learning algorithm to generate new solutions, which can solve the early convergence problem. The newly obtained solutions from the second iteration are differentiated with the symmetric values and are effectively selected in this best obtained solution.

Utilizing the t^{th} candidate and location z_t having the constraints $[low\_bound, up\_bound]$ can be mathematically formulated as,

$$\tilde{z}_t = low\_bound_i + up\_bound_i - z_t \qquad i = 1,2,.....,p \quad (32)$$

Here, p represents the dimension. Moreover, the quasi-opponent quantity $\hat{z}_t$ can be mathematically formulated as,

$$\hat{z}_t = rand\left(z_t, 0.5 \times low\_bound_i + up\_bound_i\right) \quad (33)$$

However, the IDFA technique consumes too many iterations and cannot perform well due to varying learning rates.

7) Improved simulated annealing algorithm:

The improved simulated annealing algorithm (ISAA) is analyzed to determine its performance in reducing classification loss. The traditional SAA technique [29] utilizes three functions: the generation of the state function, accept state function, and the updation of the temperature function. In addition, the two principles are also utilized to finalize the inner loop and outer loop effectively. However, the traditional SAA technique easily falls into a local optimum solution and obtaining a globally optimal solution is failed during each iteration. Accepting the probabilities of each state allows the present state to be worse than the middle state during the search process. The ISAA technique is introduced to overcome this issue, which can increase search efficiency and efficiently generate the best optimal location. A detailed explanation of ISAA is presented below:

Phase 1: Assume the primary temperature T_0 that randomly generates the initial solution y and the obtained optimal solution is stored as, $best = y$. Also, when no better solution is obtained, it can be represented as $\dim insh\_tnum$, early convergence as AIM, and Markov chain length as M_l.

Phase 2: $u = u + 1$, if $(u \geq M_l)$ then proceed to phase 10.

Phase 3: Demand all the neighborhood individuals to produce a fresh solution $\hat{y}$.

Phase 4: Estimate the rise in energy E', if $(E' < 0)$ then $y = \hat{y}$, $Num = 0$, $aim = 0$. Then proceed to phase 7.

Phase 5: if $\left(e(-E(y') - E(y))/T_y\right) > rand(float(\,))$, then $y = \hat{y}$, $Num = Num + 1$, then proceed to phase 7.

Phase 6: $Num = Num + 1$, then proceed to phase 7.

Phase 7: if $(aim = AIM)$, then the annealing process is converged earlier and proceeds to phase 10.

Phase 8: if $(Num > \dim insh\_tnum)$ then $aim = aim + 1$, and $Num = 0$.

Phase 9: if $(F(best) \leq F(y))$, then $best = y$, proceed to phase 2 and F is emphasized to evaluate the individual fitness solution.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

Phase 10: Produces the final solution and the iteration process gets completed, and the dropped temperature can be mathematically emphasized as,

$$T_p = \left[\ln\left(\frac{K}{T_0}+1\right)\right]^{-1} \quad (34)$$

Here, T_K represents the temperature at the time p and K represents the total iterations, which is identical to the total dropped temperature. However, the ISAA technique is expensive, and the obtained solution does not effectively enhance the network model performance.

8) Improved bacterial foraging algorithm

The BFO algorithm, a heuristic optimization technique, effectively mimics the essential production processes involved in food acquisition, namely chemotaxis, reproduction, elimination, and dispersal [30]. The traditional BFO has the problem that the efficiency decreases as the search space dimension increases. Searching in high-dimensional space requires a significant amount of time and memory space due to the complex execution structure of BFO. This study uses the Improved Bacterial Foraging Optimization Algorithm to solve this problem.

Each bacteria travels slowly towards its goals by swimming and tumbling due to its ability to quickly escape from hazardous items and get resources. After tumbling, bacteria could not continue swimming in a single direction for search unless their current position worsened or their number of practical movements reached their limit L_n. The new position of the bacterium p in the $q+1^{th}$ chemotaxis, r^{th} reproduction, e^{th} elimination and dispersal is $\phi^p(q+1,r,e)$depicted as the Eq. (1), where $\phi^p(q,r,e)$ denotes the bacteria p previous position, $S(p)$ denotes the step size, and $\Delta(p)$ denotes a random direction vector with all of its elements falling between -1 and 1.

$$\phi^p(q+1,r,e) = \phi^p(q,r,e) + S(p)\times\frac{\Delta(p)}{\sqrt{\Delta^M(p)\Delta(p)}} \quad (35)$$

The Levy flight algorithm has been widely used to mimic the history of foraging behavior. It is used to improve the BFO algorithm. In a way, balancing local exploration and global usage for optimization issues is advantageous when objects move forward during the search process, often using the Lévy flight method with smaller step sizes frequently and bigger durations infrequently. To increase the step size $S(p)$ of the BFO algorithm, the Lévy flight method is employed. The equation is as follows:

$$S(p) = \frac{d}{|u|^{\frac{1}{\alpha}}},\quad \alpha\in[0.3,1.99], d\sim L(0,\sigma_d^2), u\sim L(0,1) \quad (36)$$

$$\sigma_d = \left\{\frac{\Gamma(1+\alpha)\sin\left(\frac{\pi\alpha}{2}\right)}{\Gamma\left(\frac{1+\alpha}{2}\right)\times 2^{\frac{\alpha-1}{2}}\times\alpha}\right\}^{\frac{1}{\alpha}} \quad (37)$$

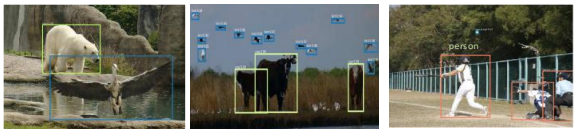
Due to the improvements, the BFO supports enhancing the classifier's performance by minimizing the losses.

IV. RESULTS AND DISCUSSION

This subdivision examines the simulation results of the proposed object recognition study. The results section includes the evaluation of various metrics such as precision, recall, memory, accuracy, mean average precision (mAP), specificity, f1 score, mean absolute error (MAE), root mean squared error (RMSE) and mean squared error (MSE). The proposed object detection system was tested and experimented with four datasets, including MSRC, MIT-67 indoor, PASCAL VOC2010 and MS-COCO datasets. These records are primarily used for scene classification but also provide a variety of object classes. The proposed system for effective object detection is simulated in the standalone computer using the Python platform.

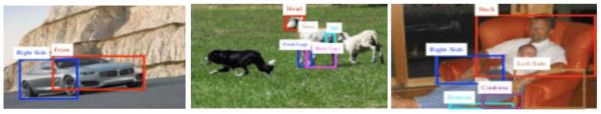
Furthermore, the experimental results of the proposed H-CNN model are compared with existing architectures such as HRNet [31], GCNet [32], RetinaNet [33], YOLOv4 [34], EfficientDet [35] and YOLO-fine architectures [36]. The proposed H-CNN has engaged AlexNet-based CNN architecture to extract deep features. The AlexNet-based CNN architecture consists of five convolutional layers as well as 3 fully connected layers. However, the final fully connected layer has exchanged with the SVR model for detecting objects. Moreover, this architecture takes the input size of 227×227, and the higher level of features have been extracted through the 2nd fully connected layer. Before building the model, the dataset was separated into training, testing and validation in the ratio of 70:20:10. With the support of four datasets, and the SVM was trained and validated after the feature extraction. Furthermore, an IGW optimization algorithm has been used for tuning the SVM model's hyperparameters. The experimental outcomes of the proposed H-CNN model using MSRC, MIT-67 indoor, PASCAL VOC2010 and MS-COCO datasets for multi-class object detection are presented in Figure 5.

(a) MSRC dataset



(b) MS-COCO dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*



(c) PASCAL VOC2010 dataset

(d) MIT-67 indoor dataset

Figure 5: Simulation outcomes of multi-class object detection

E) Dataset description

The datasets used to evaluate the performance of the H-CNN model are described in this subsection. The most widely used data sets computing the H-CNN model are listed below. But each record has different types of objects with different sets.

MSRC V2 Dataset: The MSRC V2 dataset [37] includes 591 images with 23 different classes presented by Microsoft Research. This dataset is typically used for full scene segmentation as well as detection. Among the different classes in the MSRC V2 dataset, only ten object classes, such as Chair, Book, Sign, Flower, Face, Plane, Sky, Tree, Grass and Building, are used to validate the proposed H-CNN model.

MS (Microsoft)-COCO: The Microsoft common objects in context (MS COCO) [38] is reflected as an extensive data set for object recognition with 328k images, a total of 2.5M labeled objects and 91 object classes. This dataset has been established through a group of personnel via a distinctive user interface mainly for object segmentation, object spotting and class detection. By employing this dataset, the ten classes, such as clock, cake, cup, zebra, bird, bench, truck, bus, car and person, are employed for performance validation.

PASCAL VOC2010: Another prominent dataset used in the proposed H-CNN model for object detection is PASCAL VOC (Visual Object Classes) [39]. This dataset encompasses 21738 labelled images using 20 dissimilar object classes. Indoor, animal, vehicle and person are the primary categories of the PASCAL VOC dataset. The H-CNN model is validated in ten different classes, including Sofa, Bottle, Train, Boat, Bike, Sheep, Horse, Dog, Cow, and Cat.

MIT-67 Indoor Dataset: The MIT-67 indoor data set offers a variety of object classes and is often used to classify indoor scenes [40]. With 15620 annotated images, the MIT-67 indoor dataset has 67 indoor object classes. Several images are available under different classes, including 100 interior images. The proposed H-CNN model used ten object classes, such as a wall, cup, monitor, floor, chair, table, person, picture, flower and food, for object recognition.

F) Evaluation metrics

To assess the performance of the H-CNN model, various assessment metrics such as precision, recall, memory, accuracy, mAP, specificity, f1 score, MAE, MSE, and RMSE are described in this section. The mathematical formulation of these assessment measures is specified below as follows:

$$R_{CLL} = \frac{X\_{tp}}{X\_{tp} + Y\_{fn}} \tag{38}$$

$$P_{CSNN} = \frac{X\_{tp}}{X\_{tp} + Y\_{fp}} \tag{39}$$

$$F1_{scree} = \frac{2 * P_{CSNN} * R_{CLL}}{P_{CSNN} + R_{CLL}} \tag{40}$$

$$Sp_{FYY} = \frac{X\_{tn}}{X\_{tn} + Y\_{fp}} \tag{41}$$

$$Acc_{CYY} = \frac{X\_{tp} + X\_{tn}}{X\_{tp} + X\_{tn} + Y\_{fp} + Y\_{fn}} \tag{42}$$

$$mAP = \sum_{p=1}^{P} \frac{AP(p)}{P} \tag{43}$$

$$\Phi_{mae} = \frac{1}{P}\sum_{1}^{P} |y(p) - \ast(p)| \tag{44}$$

$$\Phi_{rmse} = \sqrt{\frac{1}{P}\sum_{p=0}^{P-1} \left(y(p) - \mathcal{Y}(p)\right)^2} \tag{45}$$

$$\Phi_{mse} = \frac{1}{P}\sum_{p=0}^{P-1} \left(y(p) - \mathcal{Y}(p)\right)^2 \tag{46}$$

where, R_{CLL} shows the recall, P_{CSNN} indicates the precision, $F1_{scree}$ establishes the f1-score, Sp_{FYY} signifies specificity, Acc_{CYY} indicates the accuracy, X_{tp} symbolizes true positive, X_{tn} characterizes true negative, Y_{fp} signifies true positive, Y_{fn} characterizes false positive, P signifies the number of classes and $AP(p)$ states the average precision of class p. Further, Φ_{mae} resembles mean absolute error, Φ_{rmse} describes RMSE, Φ_{mse} designates MSE, $y(p)$ and $\mathcal{Y}(p)$ indicates the original and noise less images.

G) Performance analysis

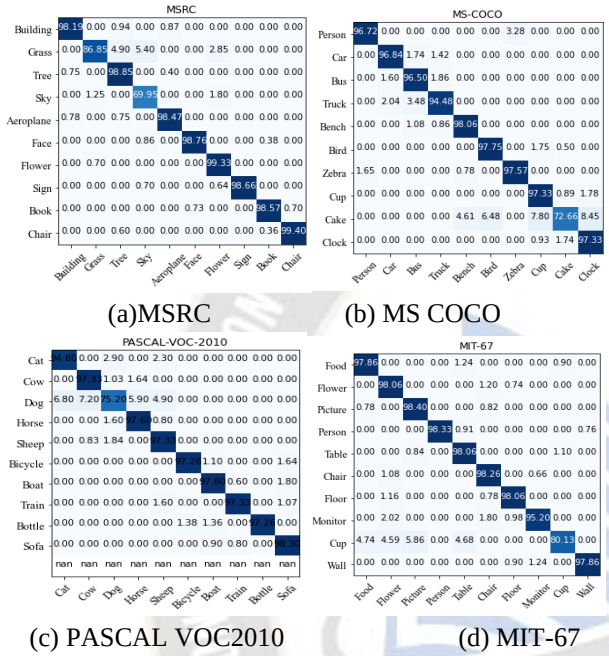
In the sub-section, the simulation outcomes of the proposed H-CNN model for object detection have been evaluated. Here, the outcomes of the H-CNN model are compared with prevailing architectures. Each classification model is evaluated with MSRC V2 Dataset, MS COCO, MIT-67 Indoor and PASCAL VOC2010 datasets for undertaking a fair assessment. The H-CNN model is also correlated for validating the progressions in several evaluation metrics. In the following subsections, a brief discussion of the evaluation for effective object detection can be delivered.

9) Analysis in terms of different evaluation metrics

In this sub-division, the result of the H-CNN model for object detection is compared with existing architectures.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

Results of four datasets, including MSRC, PASCAL VOC2010, MS COCO, and MIT-67, are shown as confusion matrices in Figure 6 (a)-(d). Ten object classes from each dataset have been selected for performance validation. Based on accuracy metrics, the confusion matrix results are calculated by comparing actual and predicted values. The graphical representation shows that each object class has better accuracy outcomes with a low loss percentage for four datasets. Because of this, the H-CNN model for multi-object identification can accurately identify several things that are present in the images.



(a)MSRC (b) MS COCO

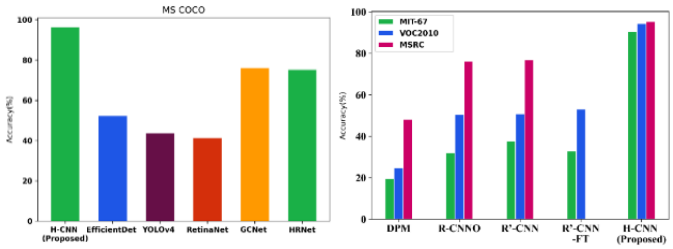
(c) PASCAL VOC2010 (d) MIT-67

Figure 6: Confusion matrix of the proposed H-CNN model

The overall results of the proposed H-CNN model for multi-class object detection are exposed in Table 1. MSRC, PASCAL, MS COCO, and MIT-67 are the four data sets employed to evaluate the proposed object identification framework. Some parameters, including F1-score, precision, recall, mAP, and accuracy, are provided in Table 1. The maximum accuracy of 97.37% has obtained through the H-CNN model while experimenting with the MSRC dataset. The MIT-67 dataset yields the second-highest accuracy of 96.02%, followed by PASCAL VOC2010 at 95.04% and MS COCO at 94.53%, respectively.

TABLE 1: Analysis of proposed H-CNN in terms of recall, f1-score, precision, mAP, and accuracy

Datasets	F1-score (%)	Recall (%)	Precision (%)	mAP or AP (%)	Accuracy (%)
MSRC Dataset	0.90	0.91	0.87	89.19	97.37
MS COCO	0.95	0.92	0.93	90.03	94.53
PASCAL VOC2010	0.88	0.95	0.90	87.11	95.04
MIT-67 indoor dataset	0.81	0.87	0.84	84.31	96.02

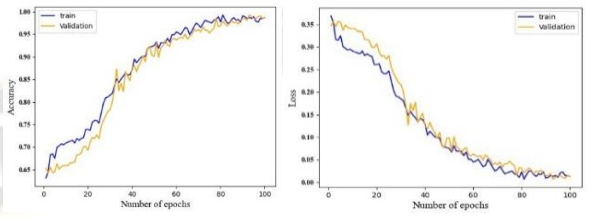


(a) MS COCO dataset (b) MSRC, VOC2010, MIT 67 datasets

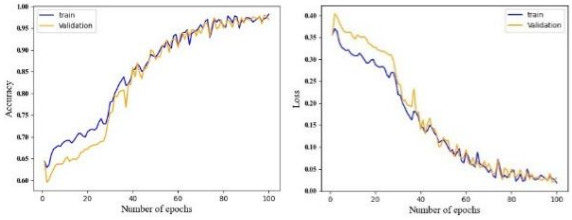
Figure 7: Comparison of accuracy with proposed H-CNN and existing architectures

The comparative analysis of the H-CNN model on dissimilar datasets for multi-class object detection is portrayed in Figure 7. In Figure 7(a), the comparison of accuracy with proposed H-CNN and DL architectures is provided using the MS COCO dataset. On the other hand, Figure 7(b) provides the accuracy comparison using MSRC, PASCAL and MIT-67. The graphical representation shows that the H-CNN model accurately detects multi-class objects. But, the existing classifiers have obtained the least accuracy owing to lower convergence complexity and degraded detection rates with over-fitting problems. Thus, the proposed H-CNN model is suitable for multi-class object detection.

Figure 8 demonstrates the accuracy and loss of training and validation data using four datasets: MSRC, PASCAL VOC2010, MS COCO and MIT-67. The training and validation process has been accomplished based on 70% and 10% of the data for each dataset. While the number of epochs has maximized for each dataset, the level of accuracy maximizes progressively. On the other hand, the loss level for both training and validation data is also diminished while maximizing the number of epochs.

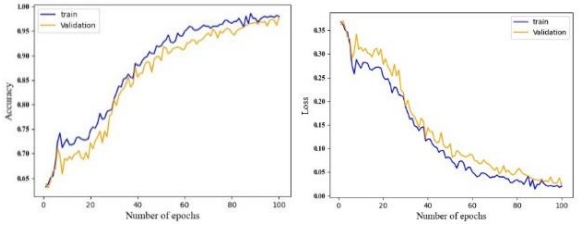


(a) MSRC

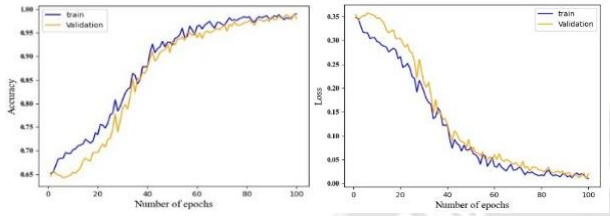


(b) MS COCO

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*



(c) PASCAL VOC2010



(d) MIT-67

Figure 8: Analysis of training and validation accuracy and Loss

Furthermore, the H-CNN model's performance has been analyzed with other recently existing architectures to evaluate the proposed study's efficacy. The existing architectures, such as HRNet, MREEP-Net, DCNN, YOLOv4 and EfficientDet architectures, are considered for comparison. The comparative analysis with state-of-art models is presented in Table 2. Here, the effectiveness of H-CNN has related to state-of-art models in terms of recall, precision and accuracy. On testing with the MS COCO dataset, the proposed H-CNN model's accuracy is better owing to object localization and multi-class object recognition based on GGL and H-CNN.

TABLE 2: Assessment against state-of-art architectures based on accuracy, precision and recall

Technique	**Dataset**	**Recall (%)**	**Precision (%)**	**Accuracy (%)**
HRNet	MS COCO	81.2	76.2	75.10
MREEP-Net	MS COCO	-	52	63.9
DCNN	PASCAL VOC	-	55.3	50.9
YOLOv4	MS COCO	-	44.4	43.5
EfficientDet	MS COCO	-	-	52.2
YOLO-fine	VEDAI	74	80	75.17
	MUNICH	100	96	99.69
	XVIEW	72	87	84.34
Proposed H-CNN model	MSRC	91	87	97.37
	PASCAL VOC2010	95	90	95.04
	MS COCO	92	93	**94.53**
	MIT-67	87	84	96.02

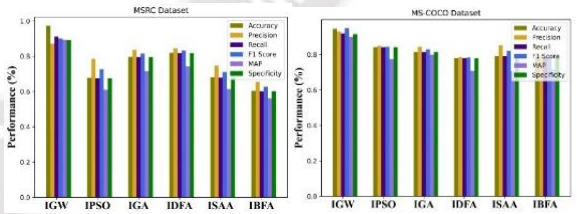
10) Analysis of loss minimization

In this sub-section, the outcomes of the proposed H-CNN model has evaluated by varying the loss function. Initially, the H-CNN model is analyzed the loss minimization using the IGW algorithm. Then, the proposed H-CNN employed different optimization algorithms, namely IPSO, IGA, IDFA, ISAA and IBFA, for loss minimization during classification. Further, in order to analyze the efficiency of IGW in loss minimization, the results of the IGW optimization algorithm for loss minimization are compared with other existing algorithms.

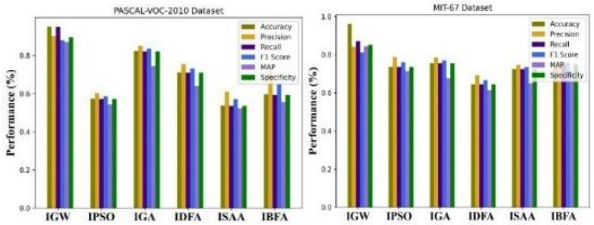
i) Evaluation based on performance measures (accuracy, recall, specificity, precision, f1-score and mAP):

Here, the evaluation metrics, including specificity, recall, precision, accuracy, f1-score and mAP, are used to analyze the performance of loss minimization. Figure 9 provides the performance evaluation of the proposed H-CNN using IGW for loss minimization and other optimization algorithms.

Typically, accuracy has been reflected as the primary metric for loss minimization during multi-class object classification. Figure 9 presents the performance analysis using various optimization algorithms for loss minimization regarding the accuracy, recall, specificity, precision, f1-score and mAP. In the graphical representation, the proposed H-CNN using IGW for loss minimization has obtained better accuracy for four datasets. In contrast, the proposed H-CNN using another optimization algorithm such as IPSO, IGA, IDFA, ISAA and IBFA for loss minimization has obtained lower accuracy than the IGW optimization algorithm. Another primary evaluation measure is precision, which signifies the rate of acceptable classified objects. The above graphical representation reflects the precision value of IGW and existing optimization. The H-CNN model using IGW retains precision of 84% for MIT-67, 93% for MS-COCO, 90% for PASCAL-VOC-2010, and 87% for the MSRC dataset and the existing implemented optimization algorithms accomplish minimum precision score. The recall metric is defined as the fraction of retrieved relevant instances. While comparing, the H-CNN model based on IGW retains a high recall score for all datasets.



(a) MSRC dataset (b) MS COCO dataset



(c)PASCAL VOC 2010 (d) MIT 67

Figure 9: Performance analysis using various optimization algorithms for loss minimization

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

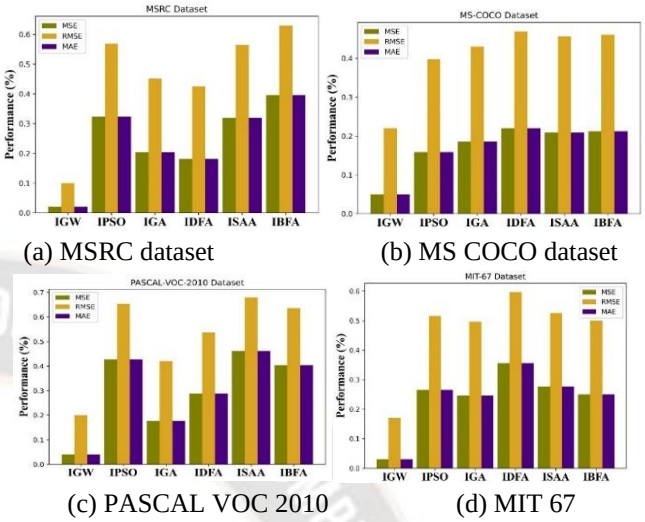
Moreover, specificity is quantified as the extent of the true negative, and the proposed H-CNN accomplishes a greater specificity score. The maximum specificity and accuracy of H-CNN are owing to the wide usage of weight as well as effective and robust optimization algorithms for loss minimization. Subsequently, better f1-score and mAP are also witnessed in the figure for the proposed H-CNN model based on IGW compared to improved optimization algorithms for loss minimization. Table 3 demonstrates the evaluation of performance measures under the H-CNN model using IGW and H-CNN model with existing IPSO, IGA, IDFA, ISAA and IBFA.

TABLE 3: Evaluation of performance measures under the H-CNN model using IGW and existing optimization algorithms

MSRC Dataset						
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**mAP (%)**	**Specificity (%)**
IPSO	67.69	78.53	67.45	72.57	60.97	67.45
IGA	79.62	83.70	79.48	81.54	71.49	79.48
IDFA	81.92	84.46	81.82	83.12	74.16	81.82
ISAA	68.08	74.61	67.88	71.08	61.40	67.88
IBFA	60.38	65.44	60.16	62.69	56.11	60.16
IGW	**97.37**	**87.00**	**91.00**	**90.00**	**89.19**	**89.15**
MS-COCO Dataset						
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**mAP (%)**	**Specificity (%)**
IPSO	84.17	84.78	84.14	84.46	77.50	84.14
IGA	81.47	84.41	81.52	82.94	79.89	81.52
IDFA	77.99	78.61	77.96	78.29	70.94	77.96
ISAA	79.15	85.33	79.07	82.08	70.65	79.07
IBFA	78.76	79.55	78.73	79.14	71.61	78.73
IGW	**94.53**	**93.00**	**92.00**	**95.00**	**90.03**	**91.55**
PASCAL VOC 2010 Dataset						
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**mAP (%)**	**Specificity (%)**
IPSO	57.31	60.23	57.10	58.62	54.25	57.10
IGA	82.31	85.04	82.20	83.59	74.49	82.20
IDFA	71.15	75.37	71.00	73.12	63.97	71.00
ISAA	53.85	60.94	53.53	57.00	52.22	53.53
IBFA	59.62	71.60	59.33	64.89	55.54	59.33
IGW	**95.04**	**90.00**	**95.00**	**88.00**	**87.11**	**89.55**
MIT-67 Dataset						
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**mAP (%)**	**Specificity (%)**
IPSO	73.48	78.63	73.48	75.97	71.32	73.48
IGA	75.37	78.39	75.37	76.85	67.54	75.37
IDFA	64.39	69.19	64.39	66.70	61.34	64.39
ISAA	72.34	74.48	72.34	73.40	65.02	72.34
IBFA	75.00	76.36	75.00	75.67	70.58	75.00
IGW	**96.02**	**84.00**	**87.00**	**81.00**	**84.31**	**85.00**

ii) Evaluation based on error measures (MAE, MSE and RMSE):

Here, the significant error metrics, including MAE, MSE and RMSE, are considered to analyze the performance of loss minimization. Figure 10 demonstrates the performance analysis of the error metric using various optimization algorithms like IPSO, IGA, IDFA, ISAA and IBFA for loss minimization.



(a) MSRC dataset (b) MS COCO dataset

(c) PASCAL VOC 2010 (d) MIT 67

Figure 10: Performance analysis of error metric using various optimization algorithms for loss minimization

Considering the error metric scenario, the H-CNN based IGW is compared to determine better balance as well as accuracy in the process of optimization. The meta-heuristic optimization algorithms are more influential in several applications for maximizing efficiency and avoiding complexities. The IPSO algorithm has reached its efficacy and robustness in resolving complex optimization issues; however, at the training phase, the percentage of training error gets maximized. IGA is very efficient in minimizing loss, but the computational complexity is high, so the error measure gets raised. In IDFA, the parameter and computation efficiency decrease because of excessive steps, thereby increasing the error. For ISAA, the tradeoff between the time taken and result quality is minimum. The IBFA exhibits weak convergence accuracy, and it is complex to balance between exploration and exploitation. Owing to these shortcomings, the error metrics of the existing optimization algorithms provided high outcomes in terms of MAE MSE as well as RMSE scores, respectively.

MSE is generally defined as altering the original and the detected object's quality. The lower MSE value indicates that the H-CNN based IGW is effective and accomplishes better outcomes for multi-class object classification. In the graphical representation, it is understood that the proposed object detection model acquires a minimum MSE, and the IGW is more perfect against existing optimization algorithms. Besides, regarding multi-object detection, MAE defines the amount of average error magnitude, and the graphical representation is provided above. Subsequently, the root mean square of the MSE, i.e. RMSE metric, is employed to determine the variation among samples. From the graphical representation, it is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

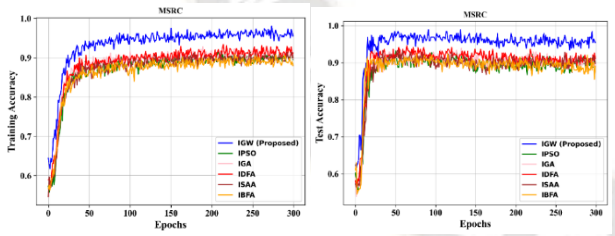
observed that IGW obtained improved outcomes than existing optimization algorithms. Table 4 evaluates error metrics under the H-CNN model using the IGW and H-CNN model with existing IPSO, IGA, IDFA, ISAA and IBFA.

TABLE 4: Evaluation of error measures under the H-CNN model using IGW and existing optimization algorithms

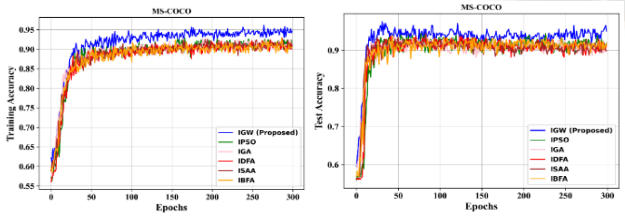
MSRC Dataset						
	IPSO	**IGA**	**IDFA**	**ISAA**	**IBFA**	**IGW**
MSE	0.32	0.2	0.18	0.32	0.4	**0.02**
MAE	0.3	0.2	0.1	0.3	0.5	**0.02**
RMSE	0.57	0.45	0.43	0.57	0.63	**0.10**
MS-COCO Dataset						
	IPSO	**IGA**	**IDFA**	**ISAA**	**IBFA**	**IGW**
MSE	0.16	0.19	0.22	0.21	0.21	**0.05**
MAE	0.16	0.19	0.22	0.21	0.21	**0.05**
RMSE	0.40	0.43	0.47	0.46	0.46	**0.22**
PASCAL VOC 2010 Dataset						
	IPSO	**IGA**	**IDFA**	**ISAA**	**IBFA**	**IGW**
MSE	0.43	0.18	0.29	0.46	0.4	**0.04**
MAE	0.4	0.17	0.29	0.47	0.4	**0.04**
RMSE	0.65	0.42	0.54	0.29	0.64	**0.2**
MIT-67 Dataset						
	IPSO	**IGA**	**IDFA**	**ISAA**	**IBFA**	**IGW**
MSE	0.26	0.24	0.35	0.27	0.25	**0.03**
MAE	0.27	0.25	0.36	0.28	0.25	**0.03**
RMSE	0.514	0.496	0.596	0.52	0.5	**0.17**

iii) Analysis in terms of accuracy-loss:

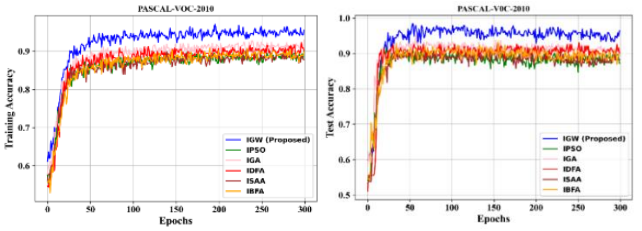
The accuracy and loss of the proposed H-CNN model based on IGW and existing algorithms are analyzed with training and testing data using the MIT-67, PASCAL-VOC-2010, MS-COCO and MSRC datasets. In the proposed work of multi-class object detection, 70% of data is engaged for training and 20% for testing.



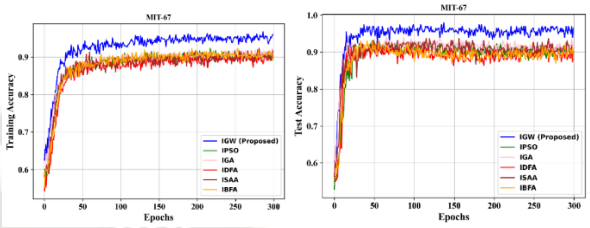
(a) MSRC dataset



(b) MS COCO dataset



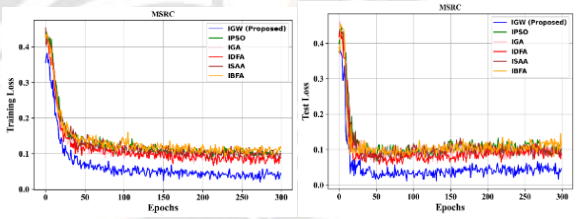
(c) PASCAL VOC 2010



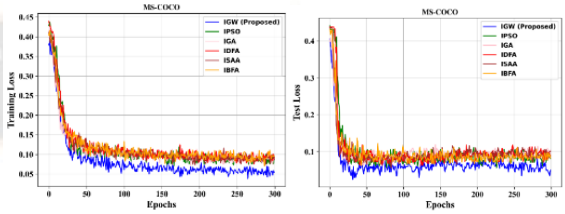
(d) MIT 67

Figure 11: Training and testing accuracy for the H-CNN model using IGW and existing optimization algorithms for loss minimization

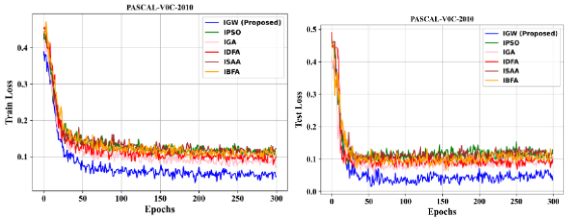
Figure 11 (a)-(d) portrays the performance of training and testing accuracy for the H-CNN model using the IGW and H-CNN model using IPSO, IGA, IDFA, ISAA and IBFA for loss minimization. By modifying the epoch size from 0 to 300, the accuracy performance of the H-CNN model using IGW is assessed, and it is provided in a graphical representation against the existing optimization algorithms. The accuracy computed for testing, and training is almost comparable. Also, accuracy increases at maximum epoch size, and loss decreases at minimum epoch size.



(a) MSRC dataset

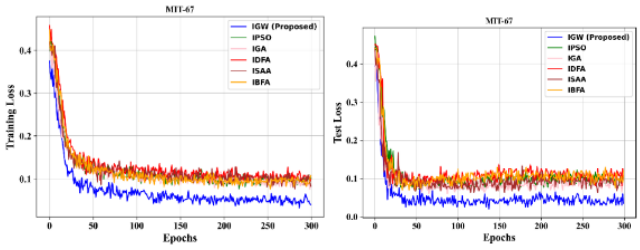


(b) MS COCO dataset



(c) PASCAL VOC 2010

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*



(d) MIT 67

Figure 12: Training and testing loss for the H-CNN model using IGW and existing optimization algorithms for loss minimization

Similarly, the training and testing loss is inspected for the H-CNN model using IGW and existing IPSO, IGA, IDFA, ISAA and IBFA and trained for 300 epochs. When the size of the epoch is extended, the H-CNN model using IGW acquires a minimum loss against the existing one. Figure 12 (a)-(d) depicts the loss performance accomplished through diverse phases of the proposed H-CNN model using IGW and existing optimization algorithms. Due to the effective data training process through the H-CNN model using IGW, the proposed study acquired minimal losses.

H) Analysis of time complexity

The time complexity study for the proposed H-CNN model and existing architectures is shown in Table 5. The tabular representation shows that the proposed H-CNN model required less training and inference time than the existing architectures such as CNN, Inception_v3 and ResNet. This minimized the time complexity of finding different classes for multi-class object detection. The training time achieved by the H-CNN model is 15.12 seconds for the MSRC dataset, 21.14 seconds for PASCAL VOC 2010, and 15.76 seconds for the MS COCO dataset and 16.78 seconds for the MIT 67 dataset. On the other hand, the inference time achieved by the proposed H-CNN model is 0.24 seconds for the MSRC dataset, 0.42 seconds for PASCAL VOC 2010, and 2.16 seconds for the MS COCO dataset and 2.43 seconds for the MIT 67 -Record. Furthermore, the proposed H-CNN model for training and validating the instances is obvious because the training efficiency improves the testing efficiency. Therefore, it is proven from the table that the H-CNN model has minimized the time complexity compared to the existing architectures.

TABLE 5: Time evaluation for proposed H-CNN against existing methods

Dataset	**Techniques**	**Inference time (sec)**	**Training time (sec)**
MSRC	CNN	0.74	34.21
	Inception_v3	0.44	46.73
	ResNet	0.57	25.74
	Proposed H-CNN	0.24	15.12
MS COCO	CNN	6.23	21.62
	Inception_v3	7.14	38.17
	ResNet	7.62	43.72
	Proposed H-CNN	2.16	15.76
PASCAL VOC2010	CNN	0.72	37.23
	Inception_v3	0.66	40.63
	ResNet	0.64	41.78
	Proposed H-CNN	0.42	21.14
MIT-67 indoor dataset	CNN	5.42	20.97
	Inception_v3	3.64	40.56
	ResNet	3.97	50.74
	Proposed H-CNN	2.43	16.78

I) Ablation study

Ablation research was adopted in the context of DL to elucidate a process in which specific neural network components are eliminated to better realize the network's behaviour. This part conducts thorough ablation research on two different practices to validate the effectiveness of the proposed design. For the proposed H-CNN model, the merging of object localization and detection using the GGL-AlexNet and the IGW model has suggested itself. The proposed H-CNN model has been divided into modules, Module I and Module II, using MSRC, PASCAL VOC, MS COCO and MIT 67 datasets. The overall accuracy provided by these two modules is determined independently to find the efficiency of each phase of the H-CNN model. Module I specifies the proposed H-CNN without any pre-processing phase, and Module II specifies the proposed H-CNN model without performing IGW optimization for loss minimization.

TABLE 6: Ablation study for proposed H-CNN model

Module	**Accuracy (%)**			
	Dataset			
	MSRC	MS COCO	PASCAL VOC	MIT 67
I	67.58	64.84	73.21	70.62
II	91.23	91.29	92.67	92.06
Proposed HCNN	**97.37**	**94.53**	**95.04**	**96.02**

Table 6 shows the performance of the proposed H-CNN model acquired in the ablation study for multi-class object detection. The ablation study was performed for each dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

Compared to each module, module I achieves a minimum performance by excluding pre-processing. Here, the inputs from MSRC, PASCAL VOC, MS COCO and MIT 67 datasets are passed directly to the object localization stage, and the further process occurs. The proposed H-CNN model could not perform better without pre-processing the input image from each dataset due to higher noise and lower data quality. On the other hand, no IGW is used to minimize losses in Module II during the classification phase. This reduces the efficiency of the proposed multi-class object detection model.

V. CONCLUSION

In this research work, a multi-class recognition of object model called H-CNN is proposed to enhance the classification accuracy by suppressing the loss functions. Initially, the collected images are pre-processed, so the noises present are extensively reduced, and the image quality is highly enhanced. Followed by this, the objects present in the images are localized due to the assistance of bounding box generation. The feature extraction process for extracting the relevant features and classification are integrated to improve the multi-class object detection performance. Moreover, while analyzing the outcomes, the integrated model has promoted enhanced accuracy outcomes with effective optimization of losses. The accuracy outcomes of the MIT-67 dataset are 96.02%, MSRC at 97.37%, PASCAL VOC2010 at 95.04%, and MS COCO at 94.53%, respectively. Also, the performances of the proposed model are analyzed by varying the loss optimization algorithms, including IPSO, IGA, IDFA, ISAA and IBFA. In comparison, effective outcomes are obtained using the IGW optimization algorithm, chosen as the best model. In future, the presented research work can be further extended to improve the scale variants capability and complex object detection based on deep level feature learning models with lesser time complexity.

ACKNOWLEDGMENT

None

REFERENCES

[1] U.A. Khan, A. Javed, R. Ashraf, “An effective hybrid framework for content based image retrieval (CBIR).” Multimedia Tools and Applications pp. 26911-26937, vol. 80, 2021,

[2] R. Gao, S. Zhang, H. Wang, J. Zhang, H. Li, Z. Zhang, “The Aeroplane and Undercarriage Detection Based on Attention Mechanism and Multi-Scale Features Processing.” Mobile Information Systems pp. 2022, 2022,

[3] M. Mandal, M. Shah, P. Meena, S. Devi, S. K. Vipparthi, “AVDNet: A small-sized vehicle detection network for aerial visual data.” IEEE Geoscience and Remote Sensing Letters pp. 494-498, vol. 17, no. 3, 2019,

[4] K. Liu, Z. Jiang, M. Xu, M. Perc, X. Li, “Tilt correction toward building detection of remote sensing images.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing pp. 5854-5866, vol. 14, 2021,

[5] Z. Cai, N. Vasconcelos, “Cascade R-CNN: high quality object detection and instance segmentation.” IEEE transactions on pattern analysis and machine intelligence pp. 1483-1498, vol. 43, no. 5, 2019,

[6] X. Gao, G. Xing, S. Roy, H. Liu, “Ramp-cnn: A novel neural network for enhanced automotive radar object recognition.” IEEE Sensors Journal pp. 5119-5132, vol. 21, no. 4, 2020,

[7] Y. Gong, Z. Xiao, X. Tan, H. Sui, C. Xu, H. Duan, D. Li, “Context-aware convolutional neural network for object detection in VHR remote sensing imagery.” IEEE Transactions on Geoscience and Remote Sensing pp. 34-44, vol. 58, no. 1, 2019,

[8] Q. Zhang, R. Cong, C. Li, M.M. Cheng, Y. Fang, X. Cao, Y. Zhao, S. Kwong, “Dense attention fluid network for salient object detection in optical remote sensing images.” IEEE Transactions on Image Processing pp. 1305-1317, vol. 30, 2020,

[9] R. Souza, S. Azevedo, G. Cardim, E. Antonio, "Semiautomatic Method for Reconstruction of Road Network Detected from Satellites Image."

[10] M. Rudorfer, “Towards Robust Object Detection and Pose Estimation as a Service for Manufacturing Industries.”

[11] L. Barsanti, L. Birindelli, P. Gualtieri, “Water monitoring by means of digital microscopy identification and classification of microalgae.” Environmental Science: Processes & Impacts pp. 1443-1457, vol. 23, no. 10, 2021,

[12] A. Dasgupta, M. Manuel, R. S. Mansur, N. Nowak, D. Gračanin, “Towards real time object recognition for context awareness in mixed reality: a machine learning approach.” In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 262-268, IEEE, 2020,

[13] Y. Sun, Z. Zhang, W. Jiang, Z. Zhang, L. Zhang, S. Yan, M. Wang, “Discriminative local sparse representation by robust adaptive dictionary pair learning.” IEEE Transactions on Neural Networks and Learning Systems pp. 4303-4317, vol. 31, no. 10, 2020,

[14] S. Chen, S. Zhong, B. X. X. Li, Liaoying Zhao, C.I. Chang, “Iterative scale-invariant feature transform for remote sensing image registration.” IEEE Transactions on Geoscience and Remote Sensing pp. 3244-3265, vol. 59, no. 4, 2020,

[15] C.R. Rahmad, A. Asmara, D. R. H. Putra, I. Dharma, H. Darmono, I. Muhiqqin, “Comparison of Viola-Jones Haar Cascade classifier and histogram of oriented gradients (HOG) for face detection.” In IOP conference series: materials science and engineering, IOP Publishing, p. 012038, vol. 732, no. 1, 2020,

[16] W. A. Qader, M. M. Ameen, B. I. Ahmed, “An overview of bag of words; importance, implementation, applications, and challenges.” In 2019 international engineering conference (IEC), pp. 200-204. IEEE, 2019,

[17] J. Ai, R. Tian, Q. Luo, J. Jin, B. Tang, “Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery.” IEEE Transactions on Geoscience and Remote Sensing pp. 10070-10087, vol. 57, no. 12, 2019,

[18] B. Cheng, Z. Li, Q. Wu, B. Li, H. Yang, L. Qing, B. Qi, “Multi-class objects detection method in remote sensing image based on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6400
***Article Received**: 04 February 2023 **Revised**: 07 March 2023 **Accepted**: 15 March 2023*

direct feedback control for convolutional neural network." IEEE Access 144691-144709, vol. 7, 2019,

[19] J. Pang, C. Li, J. Shi, Z. Xu, H. Feng, "$\mathcal {R}^ 2$-CNN: fast Tiny object detection in large-scale remote sensing images." IEEE Transactions on Geoscience and Remote Sensing pp. 5512-5524, vol. 57, no. 8, 2019,

[20] Z. Shao, P. Tang, Z. Wang, N. Saleem, S. Yam, C. Sommai, "BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images." Remote Sensing pp. 1050, 12, no. 6, 2020,

[21] X. Zhu, Y. Ma, T. Wang, Y. Xu, J. Shi, D. Lin, "Ssn: Shape signature networks for multi-class object detection from point clouds." In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16, pp. 581-597, Springer International Publishing, 2020,

[22] T. Yin, X. Zhou, P. Krahenbuhl, "Center-based 3d object detection and tracking." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11784-11793, 2021,

[23] C.H. Wang, H.W. Chen, L.C. Fu, "Vpfnet: Voxel-pixel fusion network for multi-class 3d object detection." arXiv preprint arXiv: 2111.00966 2021,

[24] L. Chen, Z. Liu, L.Tong, Z. Jiang, S. Wang, J. Dong, H. Zhou, "Underwater object detection using Invert Multi-Class Adaboost with deep learning." In 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1-8, 2020,

[25] H. Zhang, X. Zhang, G. Meng, C. Guo, Z. Jiang, "Few-Shot Multi-Class Ship Detection in Remote Sensing Images Using Attention Feature Map and Multi-Relation Detector." Remote Sensing pp. 2790, vol. 14, no. 12, 2022,

[26] A. G. Gad, "Particle swarm optimization algorithm and its applications: a systematic review." Archives of computational methods in engineering pp. 2531-2561, vol. 29, no. 5, 2022,

[27] P. Kumar, R. Anil Kumar, A. Mandal, B. Vaferi, "Genetic algorithm optimization of deep structured classifier-predictor models for pressure transient analysis." Journal of Energy Resources Technology pp. 023003, vol. 145, no. 2, 2023,

[28] C. M. Rahman, T. A. Rashid, "Dragonfly algorithm and its applications in applied science survey." Computational Intelligence and Neuroscience vol. 2019, 2019,

[29] F. He, Q. Ye, "A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm." Sensors pp. 1410, vol. 22, no. 4, 2022,

[30] N. Hakimuddin, I. Nasiruddin, T. S. Bhatti, "Generation-based automatic generation control with multisources power system using bacterial foraging algorithm." Engineering Reports e12191, vol. 2, no. 8, 2020,

[31] W. Jingdong, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, "Deep high-resolution representation learning for visual recognition," IEEE transactions on pattern analysis and machine intelligence. 2020,

[32] L. Aziz, M.S. FC, S. Ayub, "Multi-level refinement enriched feature pyramid network for object detection," Image and Vision Computing. pp. 104287, vol. 115, 2021,

[33] D. Cao, Z. Chen, L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," Human-centric Computing and Information Sciences. pp. 1-22, vol. 10, 1, 2020,

[34] B. Alexey, C.Y. Wang, Mark H.Y. Liao, Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020,

[35] T. Mingxing, R. Pang, Q.V. Le, "Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10781-10790, 2020,

[36] P. Minh-Tan, L. Courtrai, C. Friguet, S. Lefèvre, A. Baussard, "One-stage detector of small objects under various backgrounds in remote sensing images," Remote Sensing, pp. 2501, vol. 12, no. 15, 2020,

[37] M. Tomasz, A. Alexei, A Improving spatial support for objects via multiple segmentation. 2007,

[38] L. Tsung-Yi, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, "Common objects in context," In European conference on computer vision, Springer, Cham. pp. 740-755, 2014,

[39] E. Mark, L.V. Gool, C.K.I. Williams, J. Winn, A. Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision. pp. 303-338, vol. 88, no. 2, 2010,

[40] Q. Ariadna, A. Torralba, Recognizing indoor scenes. "In 2009 IEEE Conference on Computer Vision and Pattern Recognition," IEEE. pp. 413-420, 2009,