IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 14 May 2024, accepted 8 June 2024, date of publication 13 June 2024, date of current version 24 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3413848

RESEARCH ARTICLE

CNN-Based Object Detection via Segmentation Capabilities in Outdoor Natural Scenes

AYSHA NASEER[1], NAIF AL MUDAWI[2], MAHA ABDELHAQ[3], (Member, IEEE), MOHAMMED ALONAZI[4], ABDULWAHAB ALAZEB[2], ASAAD ALGARNI[5], AND AHMAD JALAL[1]

[1]Department of Computer Science, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[3]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[4]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia
[5]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Arar 91911, Saudi Arabia

Corresponding authors: Maha Abdelhaq (msabdelhaq@pnu.edu.sa) and Ahmad Jalal (ahmadjalal@mail.au.edu.pk)

This research was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R97), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Group Funding program grant code (NU/GP/SERC/13/18). This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1445). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number ‘‘NBU-FFR-2024-231-07’’.

ABSTRACT Object recognition along with classification are necessary for many applications, such as surveillance systems, car plate recognition, traffic monitoring, and face detection. Unlike existing approaches, ours incorporates a wide range of important factors to improve recognition precision. The primary phase in the image accumulating process is preprocessing, when semantic segmentation proves its usefulness by accurately defining the physical borders of specific objects inside an image in addition to recognizing them. This paper presents a novel approach to accurate object recognition. Segmentation incorporates previously identified homologous and related groups after employing the K-means clustering technique to group analogous colors and spatial patterns. Convolutional Neural Network (CNN) technology is ultimately used to identify objects in different environmental circumstances. Performance metrics like as F1 Score=0.948, Precision = 0.968, and Recall=0.932 for MSRC and F1 Score=0.921, Precision = 0.951, and Recall=0.891 for Caltech 101 and F1 Score=0.847, Precision = 0.879, and Recall=0.827 over Pascal Voc 2012 demonstrate the efficiency of our strategy. The efficacy of the suggested method is evaluated using multiple benchmark datasets, MSRC-v2, Caltech 101 and Pascal Voc 2012, yielding recognition accuracies of 92.25%, 91.91% and 93.50% respectively, when tested against the Microsoft Research Cambridge (MSRC), California Institute of Technology 101 Object Categories (Caltech 101) and Pascal Voc 2012 datasets.

INDEX TERMS Clustering, machine learning, segmentation, feature fusion, object recognition, convolutional neural network.

I. INTRODUCTION

Object detection and recognition is an emerging and quickly rising topic within the range of image processing and computer vision. An image can be analyzed easily and rapidly by a human. Humans are able to understand images and gather all pertinent information from them with only one glance. A human being is capable of controlling a considerable amount of visual input at once since their brains are incredibly complex processing units. In order to help machines, learn to recognize and comprehend visual information, it centers around the identification and localization of objects inside images or video streams. Object recognition has piqued the interest of scholars over the last decade, who are now delving into and discovering various facets of object detection and recognition problems in an extensive range of fields,

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

including but not limited to robotics [1], surveillance [2], agriculture [3], medicine [4], food industry [5], vehicle detection [6] and facial feature detection [7], [8].

Despite significant contributions to the discipline, there are still disagreements concerning aptly identifying the object of interest. The look, form, and size of the objects are influenced by a variety of circumstances, such as bright occlusion, viewing distances, and backdrop components, making the object identification and recognition task more difficult [9]. The goal of detection is to separate the object from its surroundings. Recognition is concerned with categorizing the object into one of the predetermined categories. It is a method of pinpointing a certain object in a digital image or video [10]. The ultimate objective of segmentation is to make the image's description more understandable by transforming it into something more appropriate and intelligible [11]. Image segmentation is commonly used to recognize boundaries as well as objects in images (such as lines, curves, and so on). The syntheses of object detection [12], recognition [13], and segmentation [14] have been implemented to attain accuracy [15]. In this research study, we discuss a five-step procedure. Firstly, images from the considered dataset undergo image scaling and noise removal during the pre-processing stage. Following that, the K-mean clustering technique [16] is used to group identical colors and regions. Second, Segmentation is carried out through the integration of formerly produced clusters that appear to be similar and related. As we know that a feature that is made up of multiple feature vectors [17] depicts an object, and feature extraction is performed. This feature vector is used to identify and categorize objects. Finally, this study covers the strategy and parameters used for training convolutional neural networks (CNNs) on a variety of real-world objects for accurate and effective object recognition. The implementation is shown on the openly available dataset MSRC-v2. The dataset has a total of 591 images of 213 × 320 and 15 classes of distinct real-world objects, including cow, sheep, duck, car, plane, horse, book, flower and tree, Sign boards, Road, Person, Chair, etc.

In this research, we present combinatorial segmentation technology is embodied in the combination of region-based segmentation with K-means clustering. By merging these two strategies, we take use of their respective advantages: K-means clustering effectively divides the image into groups of comparable pixels, and region-based segmentation offers a structure for integrating previous understanding of the organization and connections within the image. We leverage the power of neural networks by collaboratively bringing together diverse features, allowing the seamless identification of images through an encompassed feature selection. The algorithm at the core of our innovation is thoroughly designed to capture dynamic attributes by identifying essential key points of the objects, facilitating subsequent feature extraction. We rely on the ability of Convolutional Neural Networks (CNN), utilizing its impressive capabilities to accomplish our recognition goals, for the differentiation of objects of particular interest.

The major contributions of our proposed object detection and recognition system are as follows:.

- We used k-mean clustering to form the different clusters based on the different colors.
- To enable clusters more clear, we apply region-based segmentation on the extracted colors.
- We used a combination of different feature extractors to find out the important features to detect the objects.
- To validate our model's capability to recognize objects of variable images, we experimented our proposed methodology on three different datasets.

The article's remaining portion is organized as follows.: Section II describes prior research conducted by numerous researchers using a variety of methodologies; Section III delves into the detailed coverage of the strategy and model architecture of the recommended method; Section IV describes the outcomes of the experiments, information about the used dataset, and analyses of existing approaches; Section V explores the research questions raised by the findings; and Section VI concludes.

II. LITERATURE REVIEW

Conventional approaches have been used by numerous scholars to examine object detection and categorization. These traditional systems compute a variety of characteristics to categorize images and identify objects. A wide range of object detection and recognition techniques have been put to use by numerous researchers.

A. OBJECT SEGMENTATION

Object segmentation, which is the act of applying labels or masks to divide an image into discrete pixel areas that correspond to particular objects, is a crucial aspect of image processing. That being said, despite these developments, object segmentation still faces some inherent difficulties and constraints, especially in situations with complex and detailed backdrops.

Liu et al. [23] provided a framework that incorporates novel features and methodologies to improve the accuracy and robustness of the segmentation process, especially in the setting of complex backdrops, in order to solve the difficulties and constraints in object segmentation. Their framework's use of multiscale contrast, which enables the identification of notable contrast shifts across many spatial scales, is one of its main contributions. This feature allows the framework to recognize important elements in an image even when there are complicated backdrops present. Their framework's use of multiscale contrast, which enables the detection of notable contrast shifts across many spatial scales, is one of its main contributions. This feature allows the framework to recognize important elements in an image even when there are complicated backdrops present. The authors also presented a histogram that shows an object's perimeter as well as its center. The spatial distribution of color information is taken into consideration by this histogram, which helps the framework to recognize and utilize the contextual interactions between pixels. This color-based spatial distribution is used into the segmentation process, which improves its dependability and ability to distinguish objects. Abrar et al. [58] employed Random Forest approach along with the region based segmentation on outdoor datasets to detect the objects and got 86.1% accuracy. Bisma and Ahmad [47] used the Region based segmentation along with the Random forest classifier to recognize the object on distinct dataset UIUC and got 89.45% accuracy. An approach for unsupervised image segmentation that incorporates low-level region merging and local pixel clustering was put out by Kachouri et al. in [27]. The suggested method organizes pixels into clusters based on local similarity. Then uses low-level feature similarity between neighboring clusters to arrange clusters into coherent segments. Evaluations and research on several benchmark datasets determine that the process provides competitive performance with other unsupervised segmentation methods while being computationally efficient. The research gives a thorough analysis of unsupervised image segmentation methods, showing the advantages and disadvantages of various strategies. Lin e al. [29] represented an approach to image segmentation by improving the spanning trees with fractional differential and canny edge detectors.

B. OBJECT DETECTION

Object detection is a computer vision task that involves identifying and localizing objects of interest within an image or a video. The objective is to accurately identify and categorize objects into predetermined categories in addition to detecting their presence.

Object localization and classification of objects are the two steps that most object detection systems use. The method locates the areas of the image that could contain objects during the localization stage. Creating a collection of bounding boxes, also known as regions of interest (ROI), that encircle the objects is a common way to do this. Deep learning techniques [18] may produce incredibly precise and dependable results, they are frequently used in image classification. Tasks that took a lot of time for people to do may now be automated because to these techniques.

Deep learning was used in this work [19] to identify and recognize objects. In recent times, deep convolutional neural networks have proven to outperform humans in tests involving object identification and recognition. A multimodal deep learning feature-based method for RGB-D object recognition was presented by Xu et al. [20]. There are two stages to this method: detect the object at the regional level and evaluating the objects. The datasets SUN RGB-D and NYU Depth v2 were utilized.

Three components make up the technique that Girshick et al. [21] suggested for object recognition and semantic segmentation. Regardless of the object type, region suggestions are produced by the first module. A sizable convolutional neural network is used in the second module to extract feature vectors from every area. A collection of linear SVMs with class definitions are used in the third module. The research yielded noteworthy enhancements in mean Average Precision (mAP), exhibiting a roughly 30% rise in comparison to the preceding cutting-edge results on the PASCAL VOC dataset. A method that integrates real-time object identification with contextual comprehension was presented by Jeonge et al. [22]. Their method efficiently detects and recognizes items by using Deep Neural Networks (DNN) with different parameters.

In order to detect objects more quickly, Girshick et al. [24] added multitasking training and multidimensional training alongside their earlier research [25] on region-relevant pooling. Due to the thorough nature of its operation in each image region, region of interest pooling is computationally expensive. Although the approach processes each image region in detail, there is a significant computational cost associated with it, mainly because of the usage of region of interest pooling. This raises scalability issues and highlights the need for more research to figure out how well the strategy works with large data sets and images of high resolution. a region-proposal network-based approach is proposed in [26], which employs a completely convoluted network for concurrent identification and categorization. Ouadiay et al. [28] present a complete procedure for object detection and posture approximation by drawing bounding boxes that contain the object being pursued and its position. The research's main accomplishment is the generation of bounding boxes on training images and during the testing phase, locating each object in the image. Additionally, each object in the scene has its own set of posture coordinates.

Ahmed et al.'s work [33] used a hybrid strategy that included the DBSCAN and k-means algorithms to segment the object. They also applied the Hough transform to precisely determine the location and angle of every item in the surroundings. A genetic algorithm was used to identify the items that were discovered. It is important to note that although this segmentation technique has proven to be resilient in a variety of datasets and scenarios with differing degrees of complexity, there are certain factors to take into account. In particular, the Hough transform may be prone to noise or changes in object morphologies despite being incredibly efficient in object localization and orientation. In [34], Guan et al. developed a rapid RCNN and contextual feature-based region average pooling system for object recognition. An innovative deep learning and traditional features-based object recognition system that is used for machine inspection were introduced by Hussain et al. in [35]. A DNN is utilized in the suggested approach to extract high-level features, and a collection of traditional features, such as texture, color, and form, are used to capture low-level data.

III. THE PROPOSED OBJECT DETECTION AND RECOGNITION SYSTEM

In this article, we proposed an effective object detection and recognition model We elaborate on our object detection system in the following sections of the intended system

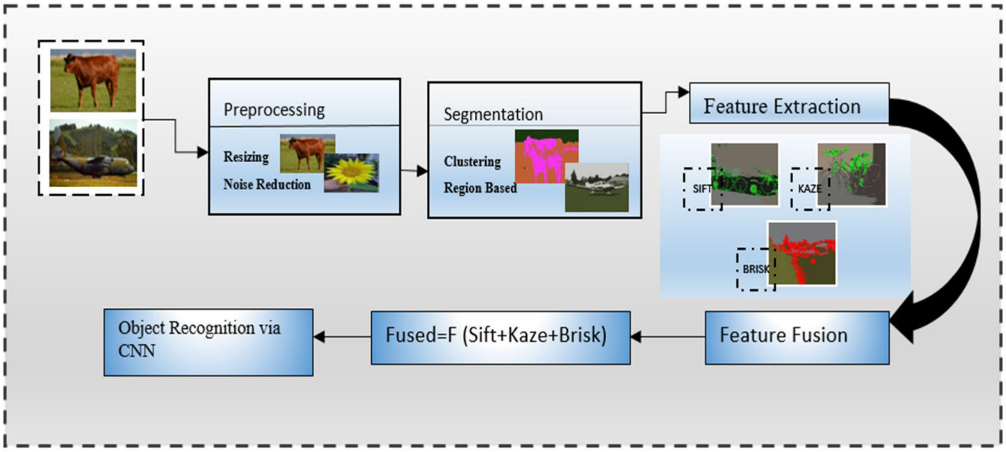


FIGURE 1. Block diagram of the proposed object detection and recognition system.

methodology: (1) pre-processing; First of all, all the images are pre-processed. (2) clustering and segmentation; These pre-processed images are segmented where each pixel was assigned a unique label to extract uniform regions from the images. (3) feature extraction; spatial features are extracted using different descriptors (4) feature fusion; Extracted features are than combined to get more important information and (5) object detection; at last, on the basis of extracted features objects are detected and recognized. All of which are demonstrated in the accompanying visualization. Fig. 1 displays an overview at a glance of the proposed model.

A. PRE-PROCESSING

Image preprocessing is the most basic level of abstraction possible [36]. The process of preprocessing increases the intensity of the image by removing or increasing undesirable elements for further processing [37]. We have applied sharpening filters and contrast enhancement as additional image processing techniques to enhance Figure 2's visual quality. An image must be convolved [38] with a Gaussian kernel in order to be preprocessed using a Gaussian filter. The amount of smoothing applied to the image depends on the filter size, with larger kernel sizes producing more smoothing. Colored images [39] composed of three primary colors (Red, Green, and Blue) pass through the Gaussian filter to remove noise and enhance the image's quality [40]. In this article, a Gaussian filter is used to even the image and eliminate other undesirable features of the image.

$$Gu\left(u,v\right)=\frac{1}{2\pi\sigma_u\sigma_v}e^{\frac{-[(u-\mu_u)^2+(v-\mu_v)^2]}{2\sigma_u\sigma_v}} \tag{1}$$

where u and v are the horizontal and vertical axis distances from the center, respectively while μ = mean and σ = standard deviation. Figure 2 depicts the scaled resulting images after applying the filter.

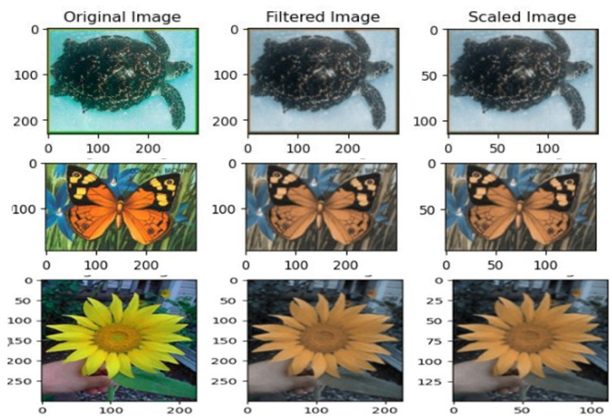


FIGURE 2. Contrast level enhancement images (a) Original (b) filtered by Gaussian (c) Scaled images.

B. SEGMENTATION

To reduce the computational complexity of the model, we applied semantic segmentation to the images before passing it to the CNN algorithm. For this purpose, we applied the combination segmentation techniques.

1) K-MEAN CLUSTERING

After refining the images in a preprocessing step, objects that are similar based on region [41], color [42], and intensity [43] are considered. The K-means technique is employed to group elements of a dataset based on their similarity [44]. The k-mean algorithm is used to cluster homogeneous color regions, and it only requires the number of clusters k at the start, with no other prior knowledge required [44]. If all three values are 255, the color is white; if all three values are muted or zero, the color is black. As a result, the combination of these three will provide us with a certain pixel color shade. Because each integer is an 8-bit number, the values range from 0-255.K-means clustering finds the similarities between objects by using Euclidean distance (See Eq.2).

$$Dis=\sqrt{[(a_2-a_1)^2+(b_2-b_1)^2} \tag{2}$$

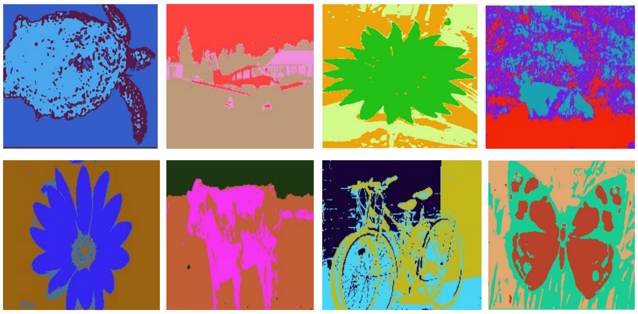


FIGURE 3. Representation of clusters in different images.

where Dis represents the distance between two data points a and b respectively. In K-mean, each cluster has a centroid. Initially, random centroids from each cluster are chosen, and each object's Euclidean distance from the cluster centroid is determined. As a result, the object will join the closest cluster [45]. When an object joins a cluster, a new centroid is computed for this cluster by taking the mean and the process will be repeated until all of the objects in the same cluster remain. K-means clustering has been applied to the mentioned dataset and Figure. 3 presents some examples of the resultant images.

2) REGION-BASED SEGMENTATION

Image segmentation has an extensive spectrum of applications and has been used with many different kinds of images as well as in practically every related area of image processing. Detecting objects and classifying multi-class images are two meticulously performed tasks that can be considerably enhanced by working on them concurrently and feeding knowledge from one to the other. If a region is linked to an object, the class label assigned to that object is limited to the foreground (for example, a "car" object cannot include a "sky" region). The similarities between neighboring pixels [46] are observed using region-based segmentation. Pixels with similar characteristics [47] will form a distinct region. In the paper [48] adjacent pixels in an image are compared to reference intensity values for the region at each pixel. For regions with homogeneous grey levels [49], we use similarity measures such as grey level differences. We employ connectivity to avoid connecting distinct areas of the image. If the difference is less than or equal to the difference threshold (see Eq. 3), the adjacent pixel is selected. Figure. 4 displays segmentation on earlier identified clusters.

$$\left|\mathrm{I}\left[\mathrm{x}\left(\mathrm{i}\right)\right]-\left[\mathrm{x}\left(\mathrm{j}\right)\right]\right|<\mathrm{Thresh} \tag{3}$$

C. FEATURE EXTRACTION

In this section, we extract the distinctive properties from a variety of segmented objects. Different methods for extracting features from deep and machine learning are addressed and expanded. Then, all of these characteristics are combined to successfully identify the objects in the illustrations. Its primary goal is to reduce the complexity by concentrating on the most important details and omitting those that are superfluous or irrelevant to understanding. It uses a feature vector to represent the concentrated part of an image [50]. Consequently, this methodology makes object recognition simpler. In this study, we used diverse feature extractors i.e. SIFT, KAZE, and BRISK to extract the features of the object of interest.

(a) (b) (c) (d)

FIGURE 4. Resultant images (a) Original images (b) Ground Truth (c) Clustered images (d) Segmented images.

1) SCALE INVARIANCE FEATURE TRANSFORM(SIFT)

SIFT to extract the important features of an object of interest. SIFT (Scale Invariance Feature Transform) is an algorithm that detects and describes [51] the local feature of an object. These features consist of curves and lines, corners, borders, points, blobs [52], patterns [53], designs [54], and surfaces [55]. This algorithm is resistant to changes in scale and rotation and more resistant to changes in brightness [56], lighting [57], and viewpoint [58]. Feature vectors indicate the physical dimensions of centroids [59] and the cluster for each object is assigned using Euclidian distance. SIFT generated the set of image features using the following points. The original image is convolved with Gaussian blur to get images over multiple scales and locations using Eq. (4) and Eq. (5).

$$D\left(k,l,\sigma\right)=\left[\left(Gn\left(m,n,p\sigma\right)-Gn\left(m,n,\sigma\right)\right)*H\left(k,l\right)\right] \tag{4}$$

$$G\left(m,n,\sigma\right)=\frac{1}{2\pi\sigma^{2}}e\frac{-(x^{2}+y^{2})}{2\sigma^{2}} \tag{5}$$

where H (k,l) is an input image, m, and n are the distances from points k and l, respectively, and is the scale of the Gaussian. Following the fitting of a model to determine scale and location, key points are chosen based on stability. SIFT [60] regulates a direction for each key point with the intention of defining a feature vector [61] for that key point; a key point has an orientation to hold robustness against rotation variations. Eq. (6) and Eq. (7) showed gradient magnitude mag (k,l) and gradient rotation (k,l) are calculated [62] around collected key points.

$$mag\left(k,l\right)=\sqrt{(H_{k,l}-H_{k+1,l})^{2}+(H_{l,k}-H_{k,l+1})^{2}} \tag{6}$$

$$Rt\left(k,l\right)=atan2[\left(H_{k,l}-H_{k+1,l}\right),\left(H_{k,l+1}-H_{m,n}\right) \tag{7}$$

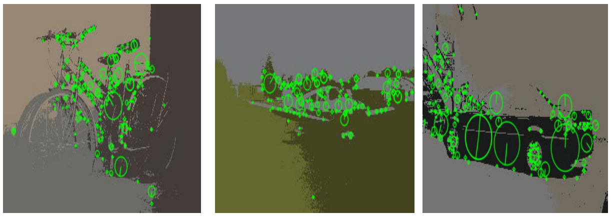


FIGURE 5. Local features extracted through SIFT.

FIGURE 6. Feature points done by utilization of BRISK.

2) BINARY ROBUST INVARIANT SALEABLE KEY POINT(BRISK)

BRISK is a binary robust invariant scalable key point approach designed especially for real-time applications [63]. The BRISK descriptor is a feature extraction approach that, unlike BRIEF or ORB, has a preset sample pattern. Instead of selecting pixels at random, BRISK trials these pixels in a specified way utilizing concentric rings [64]. Each sampling point corresponds to a pixel, and a small patch surrounding that pixel is considered. Prior to running the procedure, the patch is smoothed with Gaussian to reduce noise and improve the robustness of the descriptor [65]. The BRISK algorithm employs the AGAST algorithm to detect corners by constructing a scale-space pyramid of octaves and intra-octaves [66]. In order to reduce redundancy, the FAST score is then calculated for each scale space [67]. By specifying the local gradient for each corner, the BRISK descriptor stores variation [68] and direction invariance [69]. For luminance invariance [70], it evaluates the degree of brightness to obtain results, compares pixel-to-pixel intensity, and generates a string of binary characters [71] of the descriptor. Figure. 6 displays the extracted features using BRISK.

3) KANADE-LUCAS-TOMASI FEATURES(KAZE)

KAZE (Kanade-Lucas-Tomasi Features) is a standard feature extraction approach that is used for image analysis applications like image matching [72], object recognition, and image retrieval [73]. It is a refined version of the well-known Scale-Invariant Feature Transform (SIFT) technique and improves on its predecessor in several ways [74]. KAZE detects and describes key points or points of interest in images. These key points correlate to certain regions of the image that can be identified and matched across images [75]. The method is suitable for a variety of computer vision applications since it is robust to changes in magnitude, rotation, and luminance [32]. The classic nonlinear diffusion formula is shown

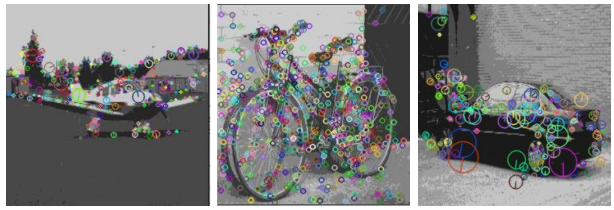


FIGURE 7. Detected and extracted features by KAZE.

in Equation (8).

$$\frac{\partial \mathrm{U}}{\partial t} = \mathrm{dv}(\mathrm{c}\ (\mathrm{a, b, t})\ , \nabla \mathrm{U} \tag{8}$$

where dv is divergence, ∇ is the gradient operator, c is known as the conductivity function [76] and U is the intensity of the image [77]. The parameter "c" is determined by the local image differential structure and can be either a scalar or a tensor. The scale parameter [78] is time t, and bigger values result in simpler visual representations. Fig. 7 shows the results of KAZE features.

D. FEATURE FUSION

In this section, independently computed features i.e. SIFT features $(\boldsymbol{F_{sift}})$, KAZE features $(\boldsymbol{F_{kaze}})$, and BRISK features$(\boldsymbol{F_{brisk}})$ independently are fused in this section. The feature vectors are normalized prior to fusion to ensure the uniformity of the merged feature vector. After normalization [79], a fully fused feature vector is created by fusing together the SIFT, KAZE, and BRISK features as follows Eq. (9).

$$F_{fused} = F_{sift} + F_{kaze} + F_{brisk} \tag{9}$$

For optimal use of these feature extraction [80] methods' contrasting capabilities, SIFT [81], KAZE [82], and BRISK features [83] are directly encompassed. The aim of this fusion strategy [84] is to combine the distinct data that each algorithm captures to provide a more robust and complete image. By adding complementing data, the fusion approach is supposed to increase the feature set's discriminative ability. The fused features may capture a greater variety of visual patterns and variations by integrating the capabilities of many algorithms, which improves their discriminative power to discriminate between various objects or classes.

IV. OBJECT DETECTION AND CLASSIFICATION

A specific type of artificial neural network created especially for the analysis of visual data [85] is the convolutional neural network (CNN). It is frequently employed.in tasks including object identification and image categorization. By dividing the datasets into 70% for training and 30% for testing, a thorough assessment of the CNN models was made possible. The design of CNNs and the significance of the convolution function [86], which enables the extraction of useful features from the image and creates a distinctive representation of each pattern in the image, are the two most crucial factors in how well

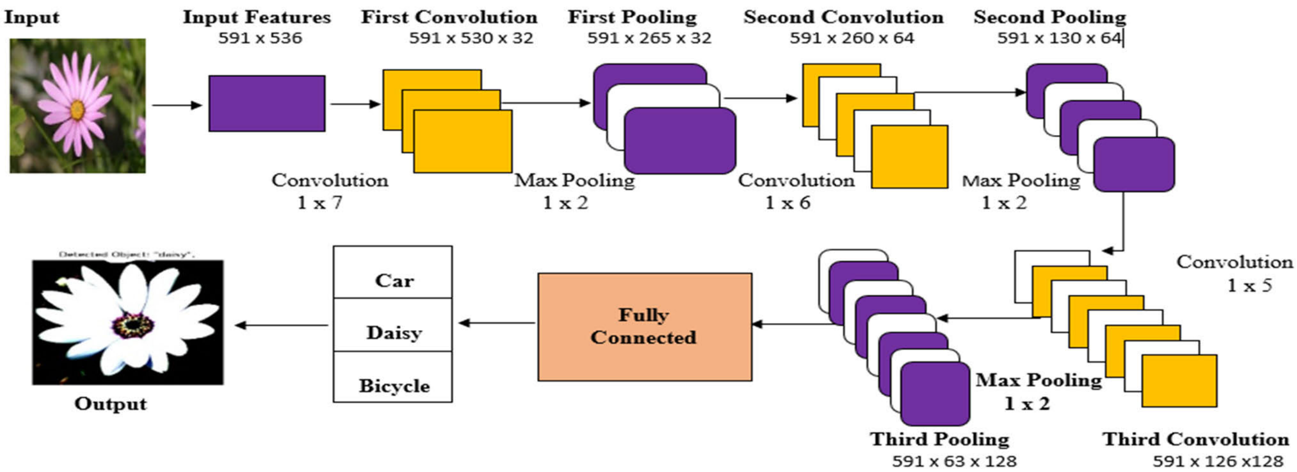


FIGURE 8. The architecture of 1-D CNN to object recognition.

CNNs perform [87]. Additionally, the last layers will make it possible to extract global properties and combine them with extracted local features to produce actual predictions. In part to its ability to gather and understand information from images, CNN offers better classification accuracy [88] than other deep-learning techniques. A limited degree of bias and weights are also used by CNN to attain excellent classification accuracy. In order to effectively categorize the objects, the key features retrieved using the techniques mentioned above are fed into a convolution neural network (CNN) [89]. The MSRC-V2 dataset's acquired attribute set is organized as 591*536 and used as a CNN input in our suggested 1-D CNN model. The number of images in this particular scenario is 591, whereas the feature vectors are represented by 536. The proposed work's representation of a 1-D CNN structure [90] is shown in Fig. 8. Three convoluted layers, three pooling layers, and one fully connected layer make up the proposed CNN model [91]. A fully connected layer used by CNN to predict an accurate class of an object from several classes is the end result. In the first convolution layer, Conv1, the 32 1 × 7-sized kernels convolution with the input matrix. A matrix of 591*536*32 is created as a result. Calculated as (48), the convolution of the matrix on the convolution layer is as follows:

$$Conv_y^{(x+1)}(a,b) = ReLU(c) \tag{10}$$

$$ReLU(c) = \sum_{i=1}^{u} \Omega(a, (b-1+\frac{u+1}{2}))w_y^x(i) + \alpha_y^x \tag{11}$$

where $Conv_y^{(x+1)}(a,b)$ generates the coordinates' convolution results (a, b) of the x + 1 layer with the yth convolution map. Ω is the former level and u is the filter size [92]. The yth convolution filter for the xth layer is designated as w_y^x.

The bias value for the yth layer is represented by α_y^x.

The function of activation ReLU is employed, which is the weights from the previous layer added together and sent to the subsequent layer [93]. The pooling layer Pool 1 is the second layer. By using 1 × 2 max-pooling, the output generated at the first convolution layer Conv1 is sampled at each layer down to a matrix size of 591* 265 * 32. By choosing the greatest value, a 1×2 sliding window is applied to the output of the preceding convolution layer in the pooling layer. As a result, [94] can be used to represent the pooling results of the (x + 1) th layer, y kernel, g row, and h column.

$$Pool_y^{(x+1)}(g,h) = max(Conv_y^{(x)}\left(g, ((h-1)^*(u+v)\right))) \tag{12}$$

where z is the size of the pooling window and 1 ≤ u ≤ v equals v. Using the same procedure for Conv2, a second convolution layer size of 1 × 6, 64 convolution kernels are used. The second and third pooling layers' employ 1 × 2 max-pooling in a similar manner. The output matrix size produced by the third pooling layer is 591 by 63 by 128 [95]. Ultimately, a layer that is completely connected is produced as:

$$FC_m^{(n+1)} = ReLU(\sum_x g_n^x w_{mx}^n + \alpha_m^n) \tag{13}$$

where *FC is fully connected*, w_{mx}^n is the matrix with weight values starting at node an of layer n and going all the way up to node m of layer (n + 1) in the graph. g_n^x denotes the contents of the xth node at layer n [96]. Two convolutional layers with max pooling, a flattening layer, and two fully linked (dense) layers compose the CNN architecture. Softmax activation is implemented in the output layer to perform multiclass classification [97].

our research aimed to determine if the suggested 1-D CNN model architecture is useful for object recognition as well as to assess the performance of CNN models on particular datasets. Although we are aware of the existence of other sophisticated object detection algorithms, the examination and assessment of CNN models is the main focus of our research. The simplicity of the architectural depiction should not be interpreted as a sign that the experiments are not

TABLE 1. Training/ testing details of used datasets.

Datasets	Total images	Training data (70%)	Testing data(30%)
MSRC-v2	591	413	178
Caltech 101	9000	6300	2700
Pascal Voc 2012	11,530	8071	3459

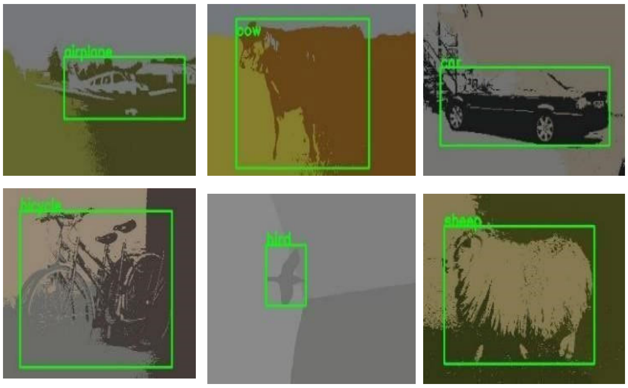


FIGURE 9. Resultant images after object detection and classification.

legitimate or real. Thorough testing and experimentation, including dataset training and testing, performance metrics analysis, and comparison with other methods, are used to assess the efficacy of the suggested design. Better interpretation and comprehension of the model's components and their contributions to overall performance are made possible by the architecture's simplicity.

Despite its visual simplification, we think the suggested 1-D CNN model offers insightful information and produces encouraging outcomes in object identification tests. Our work aims to investigate CNNs' potential in object identification, and our trials show that the suggested model performs well within the parameters of our investigation.

Algorithm 1 gives complete pseudo code for the proposed model and training/ testing details are tabulated below:

Fig. 9 shows the detected and recognized objects. Bounding boxes show the detected and recognized objects in the images.

Algorithm 1 Pseudo-Code for the Proposed Model

```
Input: RGB Images
Implement /k-means clustering on preprocessing images
# Apply K-means clustering
segmented_image = apply_k_means (image, k)

# Apply region-based segmentation using SLIC
slic_image = apply_slic (image, num_segments)

# Extract features using SIFT, KAZE, and BRISK
sift_features = extract_sift_features(image)
kaze_features = extract_kaze_features(image)
brisk_features = extract_brisk_features(image)

# Fuse the extracted features
def fuse_features (sift_features, kaze_features, brisk_features):
    fused_features = np. concatenate ((sift_features, kaze_features, brisk_features), axis=1)
    return fused_features
# Define the CNN model
model = tf.keras.Sequential([
tf.keras.layers.Conv2D(32, (3, 3), activation='relu', input_shape=(32, 32, 1)),
tf.keras.layers.MaxPooling2D((2, 2)),
tf.keras.layers.Conv2D(64,(3,3),activation='relu'),
tf.keras.layers.MaxPooling2D((2, 2)),
tf.keras.layers.Flatten(),
tf.keras.layers.Dense(64, activation='relu'),
tf.keras.layers.Dense(15, activation='softmax')
])

# Split dataset into training and testing sets
def build_cnn_model (input_shape, num_classes):
    model = Sequential ()
    # Add layers according to your architecture
    model.add(...)
    # Compile the model
    model. Compile (optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
    return model
```

V. EXPERIMENTAL SETUP AND ANALYSIS

For system evaluation and training, Python (version 3.7) was utilized on a machine with an Intel Core i7 CPU running 64-bit Windows 10. The machine is equipped with 16 GB of RAM and a CPU clock speed of 5 GHz. This section highlights the significance of the suggested paradigm by providing a thorough summary of all the experiments carried out in this study and the accompanying results

A. MSRC DATASET

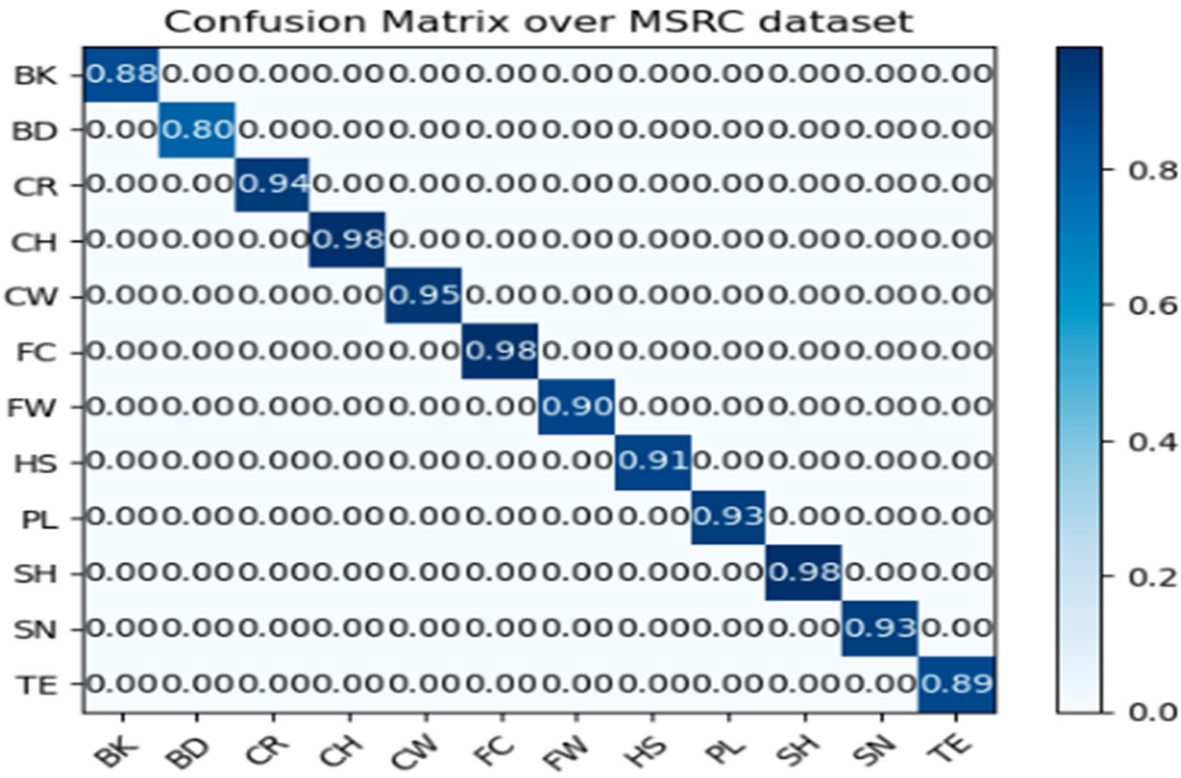
The MSRC-v2 dataset [42], [47], [98] included 591 different kinds of objects in dynamic contexts such as city structures, hilly terrain, traffic signs, and beaches. The dataset consists of 12 distinct classes, such as bike, car, cow, chair, bird, flower, house, plane, signboard, tree, sheep, book, and building. The images in the collection have a 213 × 320 resolution and each image has a complex background.

B. CALTECH 101 DATASET

The Caltech 101 [104], is well-known. It has 101 different object categories, from animals and cars to common objects. The categories "butterfly," "chair," "elephant," "car," and "aero plane," among many others, are noteworthy in this dataset. This dataset stands out for the variety of images with 300 × 255 resolution including various levels of illumination, viewpoints, and backdrops.

C. PASCAL Voc 2012 DATASET

There are twenty different object categories in the PASCAL Visual Object Classes (VOC) 2012 dataset [98], including cars, furniture, pets, and more. Airplanes, bicycles, boats, buses, cars, motorcycles, trains, bottles, chairs, dining tables, potted plants, sofas, TV/monitor, birds, cats, cows, dogs, horses, sheep, and people are some of these categories.



*BK=Bike, BD=Bird, CR=Car, CH=Chair, CW=Cow, FC=Face, FW=Flower, HS= House, PL=Plane, SH=Sheep, SN =Sign, TE= Tree

FIGURE 10. Confusion matrix plot for individual class accuracies over MSRC-v2 dataset using 1D-CNN.

Because each image in the PASCAL VOC 2012 dataset vary in size, there is no predetermined resolution for the collection's images. Every image in the collection may have a different aspect ratio and resolution. This dataset, which is frequently used as a benchmark, is essential for assessing performance on a range of computer vision tasks, including object identification, semantic segmentation, and classification.

D. EXPERIMENT 1: EXPERIMENTAL RESULTS USING PROPOSED APPROACH

We have presented a thorough analysis of our experiments using publically accessible benchmark datasets, such as MSRC-v2 and Caltech 101, in the Experimental Setup and Analysis portion of Chapter 5. We evaluated object recognition accuracy [90], and the tables that follow give a concise summary of our results.

More specifically, the object recognition confusion matrix for the MSRC-v2 [98] [91], Caltech 101 [98], and Pascal VOC 2012 [98] datasets is shown in Tables 2, 3, and 4. We have found via our comparative study that our proposed technique routinely achieves considerable improvements over existing state-of-the-art object recognition algorithms.

In particular, our approach outperforms the state-of-the-art algorithms on the same datasets by 92.25%, 91.91%, and 93.50%. These outcomes demonstrate our proposed approach's efficacy and resilience in object identification tasks, underscoring its potential for practical applications demanding high-performance object detection and classification.

E. EXPERIMENT 2: EXPERIMENTAL RESULTS FOR PRECISON, RECALL AND F1 SCORE

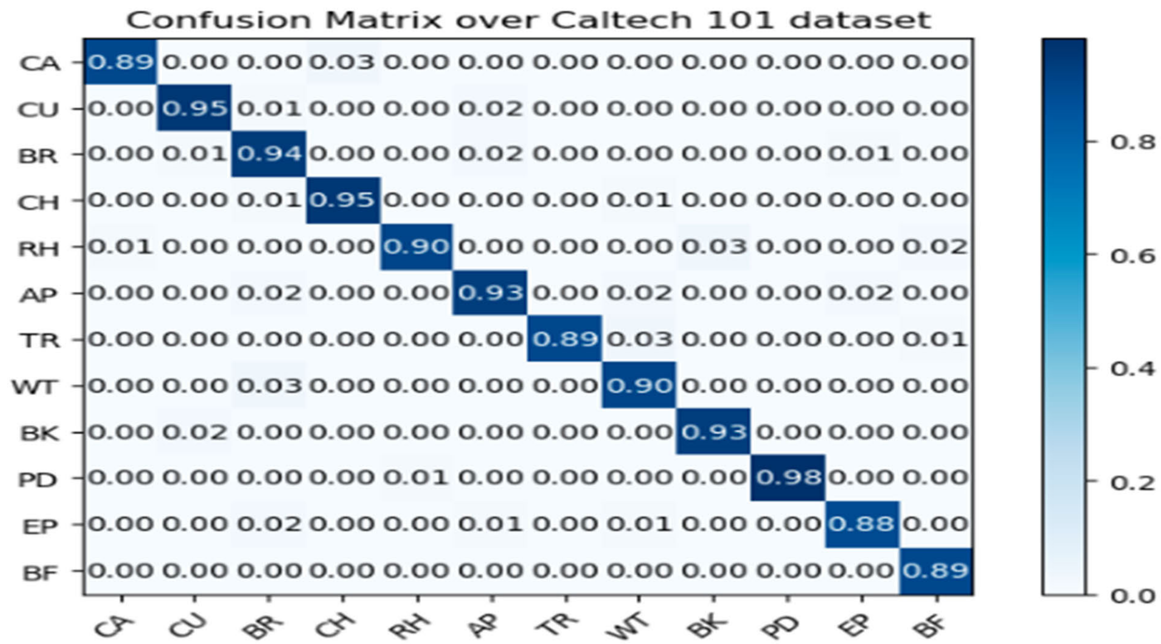
In this section, we provide the precision, recall, and F1 score values for twelve classes from the datasets that were randomly chosen. These outcomes demonstrate the great level of accuracy with which our recognition algorithm can recognize complicated objects. Equations (14), (15), and (16) were used to calculate precision, recall, and F1 scores for each object class in accordance [94]. The F1 score, commonly known as the F measure, is derived from an average weighted of precision and recall. The values range between 0 and 1, with 1 being the most precise.

$$Pr = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{14}$$

$$Rcl = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{15}$$

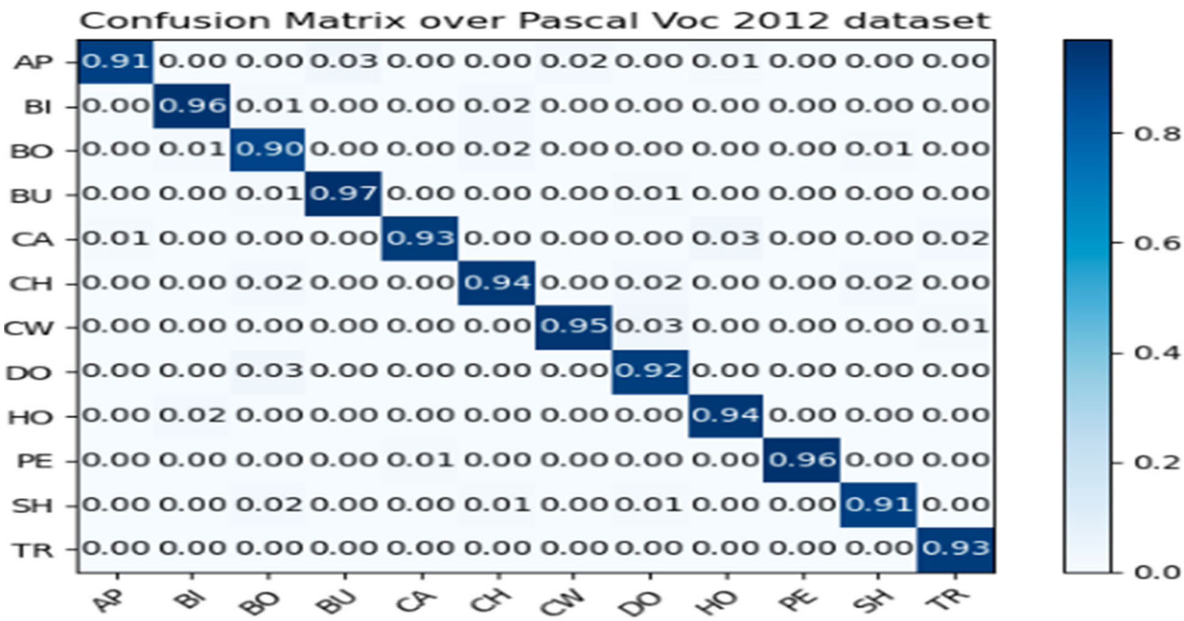
$$F1\ score = \frac{2(Pr * Rcl)}{Pr + Rcl} \tag{16}$$

where Pr= Precision processes the accuracy of positive predictions [93], while Rcl=recall deals with their completeness. Tables 2,3 and 4 present evaluation metrics of Precision [94],



*CA = Camera, CU= Cup, BR = Barrel, CH = Chair, RH = Rhino, AP= Airplane, TR = Tree, WT = Water, BK = Bike, PD = Panda, EP = Elephant, BF = Butterfly.

FIGURE 11. **Confusion matrix plot for individual class accuracies over Caltech 101 dataset using 1D-CNN.**



*AP= Airplane, BI= Bicycle, BO = Boat, BU = Bus, CA = Cat, CH = Chair, CW = Cow, DO= Dog, HO = Horse, PE= people, SH = Sheep, TR = Train,

FIGURE 12. **Confusion matrix plot for individual class accuracies over Pascal Voc 2012 dataset using 1D-CNN.**

Recall [95], and F1 score [96] along the computational time [97] of used datasets.

1) EXPERIMENT 3: COMPUTATIONAL COMPLEXITY OF TIME AND SPACE

The total number of parameters and operations has an immense influence on computational complexity. Although smaller models often use less memory and train more quickly, they may not be able to capture complicated patterns [98]. All three datasets used in this article are middle sized so their Computational complexity of time and Spaces is given below in Table 5.

2) EXPERIMENT 4: INTERSECTION OVER UNION (IoU)

One popular metric for assessing how similar or comparable two sets or areas are to one another is the Intersection over

TABLE 2. Precision, recall, f1 score and computation time over msrc-v2 dataset.

Classes	Precision	Recall	F1 Score	Computation Time
BK	0.855	0.817	**0.836**	101.7
BD	0.770	0.738	**0.754**	115.5
CR	0.799	0.757	**0.778**	96.1
CH	0.763	0.755	**0.759**	121.2
CW	0.843	0.812	**0.826**	130.5
FC	0.828	0.799	**0.813**	112.2
FW	0.810	0.778	**0.794**	108.8
HS	0.785	0.749	**0.767**	100.5
PL	0.807	0.759	**0.782**	98.3
SH	0.816	0.770	**0.792**	105.2
SN	0.843	0.812	**0.827**	99.9
TE	0.763	0.770	**0.765**	90.7
Mean	0.968	0.932	**0.948**	106.71s

TABLE 3. Precision, recall, f1 score and computation time over Caltech 101 dataset.

Classes	Precision	Recall	F1 Score	Computation Time
CA	0.781	0.730	**0.755**	112.0
CU	0.832	0.751	**0.791**	96.5
BR	0.810	0.798	**0.804**	171.0
CH	0.835	0.765	**0.799**	150.2
RH	0.775	0.721	**0.748**	114.1
AP	0.754	0.709	**0.731**	133.2
TR	0.820	0.775	**0.797**	170.9
WT	0.766	0.717	**0.741**	131.2
BK	0.810	0.768	**0.789**	135.8
PD	0.762	0.700	**0.730**	122.2
EP	0.801	0.768	**0.784**	135.0
BF	0.768	0.715	**0.741**	97.5
Mean	0.951	0.891	**0.921**	130.80s

Union (IoU), sometimes referred to as the Jaccard Index. It is frequently used to assess the precision of bounding box or pixel-level segmentation predictions in the context of image segmentation or object recognition. By dividing the area of union between two regions by the area of intersection between them, the IoU is computed. The following formula can be used to determine IoU.

$$\text{IoU} = (\text{Area of Union / Area of Intersection}).$$

F. DISCERNING OUR APPROACH TO CONTEMPORARY SYSTEMS

We compare the performance of our suggested method to current systems in Section E, showing that it performs better on a variety of datasets. A thorough comparison of the recognition accuracy of our suggested model with various cutting-edge techniques using the MSRC-V2, Caltech 101, and Pascal VOC 2012 datasets is given in Tables 9, 10, and 11.

On each data set, our suggested model performs better than the current ones. For example, our model outperforms the best-performing approach by a significant margin, with a mean recognition accuracy of 92.25% and mAP = 0.932 over the MSRC-V2 dataset. Comparatively speaking, our model outperforms all other techniques with mean identification accuracies of 91.91% and mAP = 0.759, 93.50% and mAP = 0.859 respectively, over the Caltech 101 and Pascal VOC 2012 datasets, showing similar trends.

TABLE 4. Precision, recall, f1 score and computation time over Pascal Voc 2012 dataset.

Classes	Precision	Recall	F1 Score	Computation Time
AP	0.871	0.780	**0.822**	131.2
BI	0.852	0.851	**0.851**	114.2
BO	0.910	0.875	**0.892**	188.9
BU	0.835	0.850	**0.842**	170.3
CA	0.875	0.790	**0.830**	105.9
CH	0.784	0.809	**0.796**	156.3
CW	0.920	0.875	**0.896**	199.2
DO	0.866	0.797	**0.830**	157.0
HO	0.900	0.818	**0.857**	162.7
PE	0.812	0.900	**0.853**	113.2
SH	0.901	0.768	**0.829**	114.8
TR	0.930	0.815	**0.868**	138.2
Mean	0.879	0.827	**0.847**	149.82s

TABLE 5. Computational complexities of time and space.

Dataset	Time Complexity	Space Complexity
MSRC-v2	$O(n^2)$	O (1)
Caltech 101	$O(n^2)$	O (1)
Pascal Voc 2012	$O(n^2)$	O(1)

TABLE 6. Intersection over Union over MSRC-v2 dataset.

Objects	IoU	Objects	IoU
BK	0.82	**FW**	0.97
BD	0.92	**HS**	0.89
CR	0.90	**PL**	0.88
CH	0.89	**SH**	0.93
CW	0.92	**SN**	0.97
FC	0.95	**TE**	0.91
Mean IoU = 91.25%			

TABLE 7. Intersection over Union over Caltech 101 dataset.

Objects	IoU	Objects	IoU
CA	0.92	**TR**	0.91
CU	0.85	**WT**	0.89
BR	0.91	**BK**	0.95
CH	0.90	**PD**	0.88
RH	0.89	**EP**	0.87
AP	0.92	**BF**	0.88
Mean IoU = 89.45%			

TABLE 8. Intersection over Union over pascal Voc 2012 dataset.

Objects	IoU	Objects	IoU
AP	0.87	CW	0.96
BI	0.93	DO	0.91
BO	0.90	HO	0.92
BU	0.88	PE	0.86
CA	0.87	SH	0.89
CH	0.92	TR	0.81
Mean IoU = 89. 33%			

TABLE 9. Accuracy recognition comparison between proposed methods and other state of arts methods [99], [100] [102], [103], [104] over MSRC-v2 dataset.

Author/Method	Mean Recognition Accuracy %
A. Rafique et al. [99]	83.10
Z. Ye et al.[100]	77.00
C. Wu et al. [102]	90.30
D. Xie et al. [103]	92.59
A. Ahmed et al. [104]	90.07
Proposed Model	**92.25**

TABLE 10. Accuracy recognition comparison between proposed methods and other state of arts methods [100], [101], [102], [103], [104] over CALTECH 101 dataset.

Author/Method	Mean Recognition Accuracy %
Z. Ye et al.[100]	76.00
Q. Li et al. [101]	78.00
C. Wu et al. [102]	77.93
D. Xie et al. [103]	87.24
A. Ahmed et al. [104]	89.26
Proposed Model	**91.91**

G. ANALYSIS OF RESULTS AND LOSS CURVES

To give a better understanding of the training convergence dynamics and classification accuracy of our suggested model, we examine the results and loss curves in Section F. The results shown in the tables are supported by the accuracy comparison graphs in Figure 13, which demonstrate the improved performance of our model across all datasets.

Furthermore, Figure 14 presents the data loss curves for both training and testing, providing a thorough understanding of the convergence behavior and performance stability of the model.

TABLE 11. Accuracy recognition comparison between proposed methods and other state of arts methods [105], [106], [107], [108], [109] over PASCAL VOC 2012 dataset.

Author/Method	Mean Recognition Accuracy %
M.Yang et. al [105]	74.60
L.C. Chen et. al [106]	75.50
B.Qiang et. al [107]	79.41
Y. Chen et al. [108]	80.90
P.Tang et. al [109]	92.90
Proposed Model	**93.50**

These illustrations provide crucial points of reference for assessing the effectiveness of our suggested methodology and demonstrate its superiority over current techniques with respect to convergence dynamics and accuracy. Through the integration of these comparison studies and visualizations, we improve our paper's analytical depth and offer insightful information about the performance features of our suggested model.

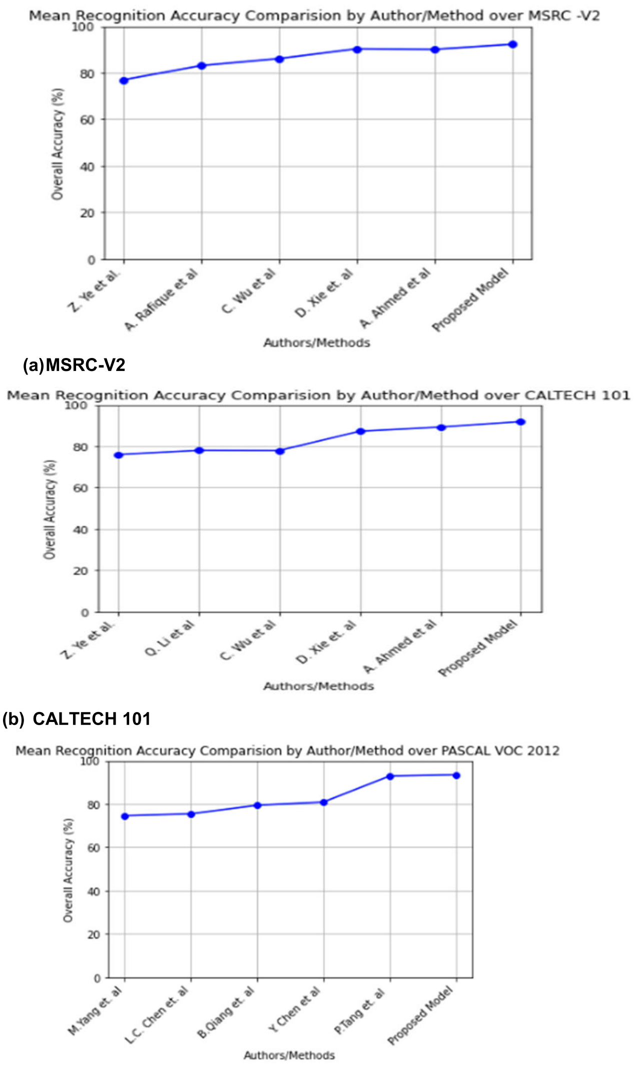
H. ABLATION EXPERIMENTS

Using three datasets—MSRC-v2, Caltech 101, and PASCAL VOC 2012—we used a variety of feature extraction methods, such as BRISK, KAZE, and SIFT, to assess their individual and combined contributions to object classification tasks. Initially, we assessed each feature extraction method's performance independently. BRISK scored 69.75% accuracy on the MSRC-v2 dataset, KAZE scored 72.31%, and SIFT scored 71.11%. The individual accuracy values in the CALTECH 101 dataset were 67.56% for BRISK, 71.63% for KAZE, and 69.79% for SIFT.

However, BRISK, KAZE, and SIFT performed 70.12%, 72.23%, and 71.11%, respectively, on the PASCAL Voc 2012 dataset. The feature fusion techniques were proposed as a way to leverage on the complementing qualities of various feature descriptors.

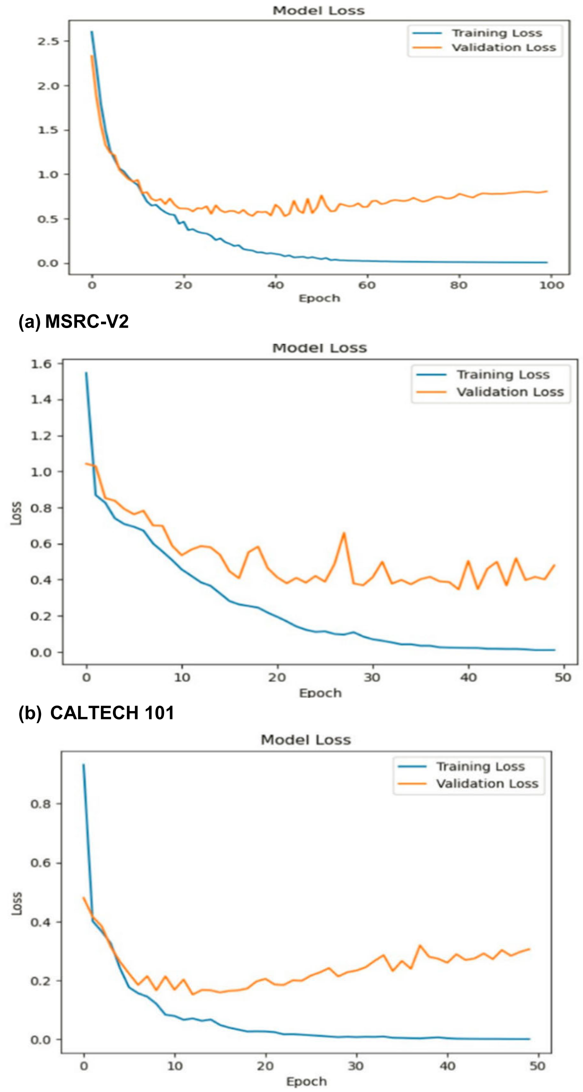
The accuracy rates using a combination of BRISK and KAZE features were 77.67% (MSRC-v2), 75.12% (CALTECH 101), and 78.12% (PASCAL Voc 2012). The accuracy was raised to 85.67% (MSRC-v2), 83.21% (CALTECH 101), and 87.54% (PASCAL Voc 2012) with the integration of BRISK and SIFT features. Comparably, the accuracy of 82.12% (MSRC-v2), 80.36% (CALTECH 101), and 84.14% (PASCAL VOC 2012) was obtained by merging KAZE and BRISK characteristics.

Motivated by the impressive outcomes of feature fusion, we combined three feature descriptors (BRISK, KAZE, and SIFT) into a single feature set. Significant gains were made 101), and 92.71% (VOC 2012) using this comprehensive feature fusion approach.



(c) PASCAL Voc 2012

FIGURE 13. Accuracy comparison of proposed model with SOTA.



(c) PASCAL Voc 2012

FIGURE 14. Data loss curves during training / testing.

The outcomes show how feature fusion is required to accurately describe an object's many complimentary qualities. While individual feature descriptors offer useful information, fusion procedures combine them in a way that best utilizes their strengths to improve performance in object classification tasks across various datasets.

VI. RESEARCH LIMITATIONS AND FUTURE WORK

In our research, we used extensive perspective and imagery issues which resulted in minor variations in our conclusions. When using these datasets, we encountered issues with occlusion and object merging in particular places. Our upcoming studies will concentrate on solving these difficulties using the latest deep-learning techniques and a fresh approach for better results.

VII. CONCLUSION

An approach for object detection across diverse complicated images is presented in this paper. Segmentation is carried out using the suggested system, and numerous features from machine learning approaches, are extracted. After feature fusion, CNN is used to conduct object recognition. The technique of fusing features is essential for raising object recognition rates above those of the benchmark dataset. Numerous real-time applications of the suggested recognition system include robotics, autonomous driving, sports activity recognition, and surveillance systems. When compared to other recognition systems, the method of our proposed system performed superior in terms of recognition accuracy. We're dedicated to expanding our research into more CNN-based semantic segmentation methods, multiple feature extraction, and feature fusion for both general-purpose scene identification and aerial.

ACKNOWLEDGMENT

Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R97), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] M. Dorigo, G. Theraulaz, and V. Trianni, "Reflections on the future of swarm robotics," *Sci. Robot.*, vol. 5, no. 49, pp. 4385–5000, Dec. 2020.

[2] J. Li, L. Han, C. Zhang, Q. Li, and Z. Liu, "Spherical convolution empowered viewport prediction in 360 video multicast with limited FoV feedback," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023, doi: 10.1145/3511603.

[3] J. Li, C. Zhang, Z. Liu, R. Hong, and H. Hu, "Optimal volumetric video streaming with hybrid saliency based tiling," *IEEE Trans. Multimedia*, vol. 25, pp. 2939–2953, 2022, doi: 10.1109/TMM.2022.3153208.

[4] F. Song, Y. Liu, D. Shen, L. Li, and J. Tan, "Learning control for motion coordination in wafer scanners: Toward gain adaptation," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13428–13438, Dec. 2022, doi: 10.1109/TIE.2022.3142428.

[5] M. Mahmood, A. Jalal, and K. Kim, "WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 6919–6950, Mar. 2020.

[6] A. M. Qureshi and A. Jalal, "Vehicle detection and tracking using Kalman filter over aerial images," in *Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS)*, Feb. 2023, pp. 1–6.

[7] Y. Yin, Y. Guo, Q. Su, and Z. Wang, "Task allocation of multiple unmanned aerial vehicles based on deep transfer reinforcement learning," *Drones*, vol. 6, no. 8, p. 215, Aug. 2022, doi: 10.3390/drones6080215.

[8] Y. Di, R. Li, H. Tian, J. Guo, B. Shi, Z. Wang, K. Yan, and Y. Liu, "A maneuvering target tracking based on fastIMM-extended Viterbi algorithm," *Neural Comput. Appl.*, vol. 5, pp. 1–10, Oct. 2023, doi: 10.1007/s00521-023-09039-1.

[9] A. Jalal, S. Kamal, and D. Kim, "Human depth sensors-based activity recognition using spatiotemporal features and hidden Markov model for smart environments," *J. Comput. Netw. Commun.*, vol. 2016, pp. 1–11, Jan. 2016.

[10] N. Khalid, M. Gochoo, A. Jalal, and K. Kim, "Modeling two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system," *Sustainability*, vol. 13, no. 2, p. 970, Jan. 2021.

[11] A. Ahmed, A. Jalal, and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2020, pp. 290–295.

[12] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, Nov. 2019.

[13] A. A. Rafique, A. Jalal, and K. Kim, "Automated sustainable multi-object segmentation and recognition via modified sampling consensus and kernel sliding perceptron," *Symmetry*, vol. 12, no. 11, p. 1928, Nov. 2020.

[14] A. Jalal, N. Khalid, and K. Kim, "Automatic recognition of human interaction via hybrid descriptors and maximum entropy Markov model using depth sensors," *Entropy*, vol. 22, no. 8, p. 817, Jul. 2020.

[15] U. Azmat and A. Jalal, "Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation," in *Proc. Int. Conf. Commun. Technol. (ComTech)*, Sep. 2021, pp. 109–114.

[16] R. Zhang, L. Li, Q. Zhang, J. Zhang, L. Xu, B. Zhang, and B. Wang, "Differential feature awareness network within antagonistic learning for infrared-visible object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, p. 1, Jun. 2023, doi: 10.1109/TCSVT.2023.3289142.

[17] J. Xu, K. Guo, X. Zhang, and P. Z. H. Sun, "Left gaze bias between LHT and RHT: A recommendation strategy to mitigate human errors in left- and right-hand driving," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 10, pp. 4406–4417, Oct. 2023, doi: 10.1109/tiv.2023.3298481.

[18] M. M. Afsar, S. Saqib, M. Aladfaj, M. H. Alatiyyah, K. Alnowaiser, H. Aljuaid, A. Jalal, and J. Park, "Body-worn sensors for recognizing physical sports activities in exergaming via deep learning model," *IEEE Access*, vol. 11, pp. 12460–12473, 2023.

[19] A. Nadeem, A. Jalal, and K. Kim, "Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness," *Symmetry*, vol. 12, no. 11, p. 1766, Oct. 2020.

[20] X. Xu, Y. Li, G. Wu, and J. Luo, "Multi-modal deep feature learning for RGB-D object detection," *Pattern Recognit.*, vol. 72, pp. 300–313, Dec. 2017.

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[22] H. Jeong, S. Choi, S. Jang, and Y. Ha, "Driving scene understanding using hybrid deep neural network," *J. Photogramm. Remote Sens.*, vol. 8, no. 2, pp. 1–4, Feb. 2019.

[23] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[24] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[25] E. Ohn-Bar and M. M. Trivedi, "Multi-scale volumes for deep object detection and localization," *Pattern Recognit.*, vol. 61, pp. 557–572, Jan. 2017.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[27] R. Kachouri, M. Soua, and M. Akil, "Unsupervised image segmentation based on local pixel clustering and low-level region merging," in *Proc. 2nd Int. Conf. Adv. Technol. for Signal Image Process. (ATSIP)*, Monastir, Tunisia, Mar. 2016, pp. 177–182.

[28] F. Z. Ouadiay, H. Bouftaih, E. H. Bouyakhf, and M. M. Himmi, "Simultaneous object detection and localization using convolutional neural networks," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, Fez, Morocco, Apr. 2018, pp. 1–8.

[29] J. Lin, T. Guo, and Q. Yan, "Image segmentation by improved spanning tree using Canny edge detector," *J. Algorithm Comput. Technol.*, vol. 13, pp. 1–13, Jan. 2019.

[30] Y. Zhao and A. Cai, "A novel relative orientation feature for shape-based object recognition," in *Proc. IEEE Int. Conf. Netw. Infrastruct. Digit. Content*, Nov. 2009, pp. 686–689.

[31] M. Songhui, S. Mingming, and H. Chufeng, "Objects detection and location based on mask RCNN and stereo vision," in *Proc. 14th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Changsha, China, Nov. 2019, pp. 369–373.

[32] W. Zheng, S. Lu, Y. Yang, Z. Yin, and L. Yin, "Lightweight transformer image feature extraction network," *PeerJ Comput. Sci.*, vol. 10, p. e1755, Jan. 2024, doi: 10.7717/peerj-cs.1755.

[33] A. Ahmed, A. Jalal, and A. A. Rafique, "Salient segmentation based object detection and recognition using hybrid genetic transform," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Taxila, Pakistan, Aug. 2019, pp. 203–208.

[34] C. Guan, K. K. F. Yuen, and Q. Chen, "Towards a hybrid approach of K-means and density-based spatial clustering of applications with noise for image segmentation," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jun. 2017, pp. 396–399.

[35] N. Hussain, M. A. Khan, M. Sharif, S. A. Khan, A. A. Albesher, T. Saba, and A. Armaghan, "A deep neural network and classical features based scheme for objects recognition: An application for machine inspection," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14935–14957, Apr. 2020.

[36] A. Jalal, M. Z. Sarwar, and K. Kim, "RGB-D images for objects recognition using 3D point clouds and RANSAC plane fitting," in *Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2021, pp. 518–523.

[37] Z. Xiao, H. Li, H. Jiang, Y. Li, M. Alazab, Y. Zhu, and S. Dustdar, "Predicting urban region heat via learning arrive-stay-leave behaviors of private cars," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10843–10856, Jun. 2023, doi: 10.1109/TITS.2023.3276704.

[38] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and $\mathfrak{R}$ transformation for smart homes," *Indoor Built Environ.*, vol. 21, no. 1, pp. 184–190, Feb. 2012.

[39] R. Sun, Y. Dai, and Q. Cheng, "An adaptive weighting strategy for multisensor integrated navigation in urban areas," *IEEE Internet Things J.*, vol. 10, no. 14, pp. 12777–12786, Jul. 2023, doi: 10.1109/jiot.2023.3256008.

[40] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using Hue component of HSV colour space," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 56–68, Mar. 2022.

[41] Y. Wang, R. Sun, Q. Cheng, and W. Y. Ochieng, "Measurement quality control aided multisensor system for improved vehicle navigation in urban areas," *IEEE Trans. Ind. Electron.*, vol. 71, no. 6, pp. 6407–6417, Jun. 2024, doi: 10.1109/TIE.2023.3288188.

[42] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection," in *Proc. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2009, pp. 895–902.

[43] Y. Ren, Z. Lan, L. Liu, and H. Yu, "EMSIN: Enhanced multi-stream interaction network for vehicle trajectory prediction," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 1, pp. 1–15, Feb. 2024, doi: 10.1109/TFUZZ.2024.3360946.

[44] Z. Cai, X. Zhu, P. Gergondet, X. Chen, and Z. Yu, "A friction-driven strategy for agile steering wheel manipulation by humanoid robots," *Cyborg Bionic Syst.*, vol. 4, p. 64, Jan. 2023, doi: 10.34133/cbsystems.0064.

[45] H. Jiang, S. Chen, Z. Xiao, J. Hu, J. Liu, and S. Dustdar, "Pa-Count: Passenger counting in vehicles using Wi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, Mar. 2023, doi: 10.1109/TMC.2023.3263229.

[46] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 457–473, Feb. 2023, doi: 10.1109/JSAC.2022.3227027.

[47] B. R. Chughtai and A. Jalal, "Object detection and segmentation for scene understanding via random forest," in *Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS)*, Feb. 2023, pp. 6–13.

[48] Z. Xiao, H. Fang, H. Jiang, J. Bai, V. Havyarimana, H. Chen, and L. Jiao, "Understanding private car aggregation effect via spatio-temporal analysis of trajectory data," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2346–2357, Apr. 2023, doi: 10.1109/TCYB.2021.3117705.

[49] G. Sun, Y. Zhang, H. Yu, X. Du, and M. Guizani, "Intersection fog-based distributed routing for V2V communication in urban vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2409–2426, Jun. 2020, doi: 10.1109/TITS.2019.2918255.

[50] G. Sun, L. Song, H. Yu, V. Chang, X. Du, and M. Guizani, "V2V routing in a VANET based on the autoregressive integrated moving average model," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 908–922, Jan. 2019, doi: 10.1109/TVT.2018.2884525.

[51] X. Zhao, Y. Fang, H. Min, X. Wu, W. Wang, and R. Teixeira, "Potential sources of sensor data anomalies for autonomous vehicles: An overview from road vehicle safety perspective," *Expert Syst. Appl.*, vol. 236, Feb. 2024, Art. no. 121358, doi: 10.1016/j.eswa.2023.121358.

[52] Z. Qu, X. Liu, and M. Zheng, "Temporal–spatial quantum graph convolutional neural network based on Schrödinger approach for traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 99, no. 1, pp. 1–9, Jan. 2022, doi: 10.1109/TITS.2022.3203791.

[53] A. A. Rafique, A. Jalal, and K. Kim, "Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2020, pp. 271–276.

[54] Z. Xiao, J. Shu, H. Jiang, G. Min, J. Liang, and A. Iyengar, "Toward collaborative occlusion-free perception in connected autonomous vehicles," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4918–4929, May 2024, doi: 10.1109/TMC.2023.3298643.

[55] G. Sun, Y. Zhang, D. Liao, H. Yu, X. Du, and M. Guizani, "Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7550–7563, Aug. 2018, doi: 10.1109/TVT.2018.2828651.

[56] J. Luo, G. Wang, G. Li, and G. Pesce, "Transport infrastructure connectivity and conflict resolution: A machine learning analysis," *Neural Comput. Appl.*, vol. 34, no. 9, pp. 6585–6601, May 2022, doi: 10.1007/s00521-021-06015-5.

[57] J. Yu, L. Lu, Y. Chen, Y. Zhu, and L. Kong, "An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 337–351, Feb. 2021, doi: 10.1109/TMC.2019.2947468.

[58] A. Ahmed, A. Jalal, and K. Kim, "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Dec. 2019, pp. 1–12.

[59] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and P. Zhao, "An incentive mechanism of incorporating supervision game for federated learning in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14800–14812, Dec. 2023, doi: 10.1109/TITS.2023.3297996.

[60] G. Sun, L. Sheng, L. Luo, and H. Yu, "Game theoretic approach for multipriority data transmission in 5G vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24672–24685, Dec. 2022, doi: 10.1109/TITS.2022.3198046.

[61] C. Ding, C. Li, Z. Xiong, Z. Li, and Q. Liang, "Intelligent identification of moving trajectory of autonomous vehicle based on friction nano-generator," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 3090–3097, Mar. 2024, doi: 10.1109/TITS.2023.3303267.

[62] J. Mou, K. Gao, P. Duan, J. Li, A. Garg, and R. Sharma, "A machine learning approach for energy-efficient intelligent transportation scheduling problem in a real-world dynamic circumstances," *IEEE Trans. Intell. Transp. Syst.*, 2022, doi: 10.1109/TITS.2022.3183215.

[63] W. Gao, M. Wei, and S. Huang, "Optimization of aerodynamic drag reduction for vehicles with non-smooth surfaces and research on aerodynamic characteristics under crosswind," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 2023, May 2023, Art. no. 095440702311734, doi: 10.1177/09544070231173471.

[64] Z. W. Deng, Y. Q. Zhao, B. H. Wang, W. Gao, and X. Kong, "A preview driver model based on sliding-mode and fuzzy control for articulated heavy vehicle," *Meccanica*, vol. 57, no. 8, pp. 1853–1878, 2022.

[65] L. Yin, W.-T. Pan, J. Kuang, and M. Zhuang, "Application of bootstrap-DEA with fuzzy computing in performance evaluation of forklift leasing supplier," *IEEE Access*, vol. 8, pp. 66095–66104, 2020.

[66] C. Lu, J. Zheng, L. Yin, and R. Wang, "An improved iterated greedy algorithm for the distributed hybrid flowshop scheduling problem," *Eng. Optim.*, vol. 56, no. 5, pp. 792–810, May 2024, doi: 10.1080/0305215x.2023.2198768.

[67] R. Luo, Z. Peng, J. Hu, and B. K. Ghosh, "Adaptive optimal control of affine nonlinear systems via identifier–critic neural network approximation with relaxed PE conditions," *Neural Netw.*, vol. 167, pp. 588–600, Oct. 2023, doi: 10.1016/j.neunet.2023.08.044.

[68] J. Zhao, D. Song, B. Zhu, Z. Sun, J. Han, and Y. Sun, "A human-like trajectory planning method on a curve based on the driver preview mechanism," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11682–11698, May 2023, doi: 10.1109/TITS.2023.3285430.

[69] B. Zhu, Y. Sun, J. Zhao, J. Han, P. Zhang, and T. Fan, "A critical scenario search method for intelligent vehicle testing based on the social cognitive optimization algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 7974–7986, May 2023, doi: 10.1109/TITS.2023.3268324.

[70] H. S. Dadi and G. K. Mohan Pillutla, "Improved face recognition rate using HOG features and SVM classifier," *IOSR J. Electron. Commun. Eng.*, vol. 11, no. 4, pp. 34–44, Apr. 2016.

[71] F. Ahmad, "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 200–218, Jun. 2022, doi: 10.1049/cit2.12083.

[72] Y. Jiang, Y. Yang, Y. Xu, and E. Wang, "Spatial–temporal interval aware individual future trajectory prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 1, pp. 1–6, Nov. 2024, doi: 10.1109/TKDE.2023.3332929.

[73] Y. Zhang, S. Li, S. Wang, X. Wang, and H. Duan, "Distributed bearing-based formation maneuver control of fixed-wing UAVs by finite-time orientation estimation," *Aerosp. Sci. Technol.*, vol. 136, May 2023, Art. no. 108241, doi: 10.1016/j.ast.2023.108241.

[74] C. Zhang, L. Zhou, and Y. Li, "Pareto optimal reconfiguration planning and distributed parallel motion control of mobile modular robots," *IEEE Trans. Ind. Electron.*, vol. 71, no. 8, pp. 9255–9264, Jun. 2024, doi: 10.1109/TIE.2023.3321997.

[75] M. Mahmood, A. Jalal, and M. A. Sidduqi, "Robust spatio-temporal features for human interaction recognition via artificial neural network," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Islamabad, Pakistan, Dec. 2018, pp. 218–233.

[76] J. Chen, Q. Wang, H. H. Cheng, W. Peng, and W. Xu, "A review of vision-based traffic semantic understanding in ITSs," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 19954–19979, Nov. 2022, doi: 10.1109/TITS.2022.3182410.

[77] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[78] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, Jan. 2020.

[79] Q. She, R. Hu, J. Xu, M. Liu, K. Xu, and H. Huang, "Learning high-DOF reaching-and-grasping via dynamic representation of gripper-object interaction," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1–14, Jul. 2022.

[80] Q. Li and X. Wang, "Image classification based on SIFT and SVM," in *Proc. IEEE/ACIS 17th Int. Conf. Comput. Inf. Sci. (ICIS)*, Sydney, NSW, Australia, Jun. 2018, pp. 762–765.

[81] J. Chen, M. Xu, W. Xu, D. Li, W. Peng, and H. Xu, "A flow feedback traffic prediction based on visual quantified features," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 10067–10075, Jul. 2023, doi: 10.1109/TITS.2023.3269794.

[82] I. Abaspur Kazerouni, G. Dooly, and D. Toal, "Underwater image enhancement and mosaicking system based on A-KAZE feature matching," *J. Mar. Sci. Eng.*, vol. 8, no. 6, p. 449, Jun. 2020.

[83] J. Chen, Q. Wang, W. Peng, H. Xu, X. Li, and W. Xu, "Disparity-based multiscale fusion network for transportation detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18855–18863, Oct. 2022, doi: 10.1109/TITS.2022.3161977.

[84] S. Li, J. Chen, W. Peng, X. Shi, and W. Bu, "A vehicle detection method based on disparity segmentation," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 19643–19655, May 2023, doi: 10.1007/s11042-023-14360-x.

[85] S. A. Rizwan, A. Jalal, M. Gochoo, and K. Kim, "Robust active shape model via hierarchical feature extraction with SFS-optimized convolution neural network for invariant human age classification," *Electronics*, vol. 10, no. 4, p. 465, Feb. 2021.

[86] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Mar. 2018, pp. 1–10.

[87] Y. Gu, Z. Hu, Y. Zhao, J. Liao, and W. Zhang, "MFGTN: A multi-modal fast gated transformer for identifying single trawl marine fishing vessel," *Ocean Eng.*, vol. 303, Jul. 2024, Art. no. 117711.

[88] Y. Chen, N. Li, D. Zhu, C. C. Zhou, Z. Hu, Y. Bai, and J. Yan, "BEVSOC: Self-supervised contrastive learning for calibration-free BEV 3-D object detection," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22167–22182, Jun. 2024, doi: 10.1109/JIOT.2024.3379471.

[89] Md. Z. Uddin, W. Khaksar, and J. Torresen, "Facial expression recognition using salient features and convolutional neural network," *IEEE Access*, vol. 5, pp. 26146–26161, 2017.

[90] D. Li, X. Dai, J. Wang, Q. Xu, Y. Wang, T. Fu, A. Hafez, and J. Grant, "Evaluation of college students' classroom learning effect based on the neural network algorithm," *Mobile Inf. Syst.*, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/7772620.

[91] A. Naseer and J. Ahmad, "Pixels to precision: Features fusion and random forests over labelled-based segmentation," in *Proc. IBCAST*, 2023, pp. 1–7.

[92] D. Cai, R. Li, Z. Hu, J. Lu, S. Li, and Y. Zhao, "A comprehensive overview of core modules in visual SLAM framework," *Neurocomputing*, vol. 590, Jul. 2024, Art. no. 127760, doi: 10.1016/j.neucom.2024.127760.

[93] Z. Li, Y. Wang, R. Zhang, F. Ding, C. Wei, and J.-G. Lu, "A LiDAR-OpenStreetMap matching method for vehicle global position initialization based on boundary directional feature extraction," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 1, Apr. 2024, Art. no. 127760, doi: 10.1109/TIV.2024.3393229.

[94] H. Zhu, D. Xu, Y. Huang, Z. Jin, W. Ding, J. Tong, and G. Chong, "Graph structure enhanced pre-training language model for knowledge graph completion," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 1, Apr. 2024, Art. no. 3393229, doi: 10.1109/TETCI.2024.3372442.

[95] D. Li and M. Zakarya, "Machine learning based preschool education quality assessment system," *Mobile Inf. Syst.*, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/2862518.

[96] D. Li, R. Hu, and Z. Lin, "Vocational education platform based on block chain and IoT technology," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, Aug. 2022, doi: 10.1155/2022/5856229.

[97] X. Xu and Z. Wei, "Dynamic pickup and delivery problem with transshipments and LIFO constraints," *Comput. Ind. Eng.*, vol. 175, May 2022, Art. no. 108835.

[98] Z. Al-Huda, B. Peng, Y. Yang, and R. N. A. Algburi, "Object scale selection of hierarchical image segmentation with deep seeds," *IET Image Process.*, vol. 15, no. 1, pp. 191–205, Jan. 2021.

[99] A. A. Rafique, A. Jalal, and A. Ahmed, "Scene understanding and recognition: Statistical segmented model using geometrical features and Gaussian Naïve Bayes," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, London, U.K., Aug. 2019, pp. 225–230.

[100] Z. Ye, P. Liu, W. Zhao, and X. Tang, "Hierarchical abstract semantic model for image classification," *J. Electron. Imag.*, vol. 24, no. 5, Oct. 2015, Art. no. 053022.

[101] Q. Li, Q. Peng, J. Chen, and C. Yan, "Improving image classification accuracy with ELM and CSIFT," *Comput. Sci. Eng.*, vol. 21, no. 5, pp. 26–34, Sep. 2019.

[102] C. Wu, Y. Li, Z. Zhao, and B. Liu, "Image classification method rationally utilizing spatial information of the image," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 19181–19199, Jul. 2019.

[103] D. Xie, Q. Li, W. Xia, S. Pang, H. He, and Q. Gao, "Multi-view classification via adaptive discriminant analysis," *IEEE Access*, vol. 7, pp. 36702–36709, 2019.

[104] A. Jalal, A. Ahmed, A. A. Rafique, and K. Kim, "Scene semantic recognition based on modified fuzzy C-mean and maximum entropy using object-to-object relations," *IEEE Access*, vol. 9, pp. 27758–27772, 2021.

[105] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.

[106] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[107] B. Qiang, R. Chen, M. Zhou, Y. Pang, Y. Zhai, and M. Yang, "Convolutional neural networks-based object detection algorithm by jointing semantic segmentation for images," *Sensors*, vol. 20, no. 18, p. 5080, Sep. 2020.

[108] Y. Chen, J. Tao, L. Liu, J. Xiong, R. Xia, and K. Yang, "Research of improving semantic image segmentation based on a feature fusion model," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 11, pp. 1–13, 2022.

[109] P. Tang, X. Wang, B. Shi, X. Bai, W. Liu, and Z. Tu, "Deep FisherNet for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2244–2250, Jul. 2019, doi: 10.1109/TNNLS.2018.2874657.

AYSHA NASEER received the M.S. degree in computer engineering from the Center for Advanced Studies in Engineering (CASE), Islamabad. She is currently pursuing the Ph.D. degree in computer science with Air University, Islamabad, Pakistan. Her research interests include artificial intelligence, computer vision, machine learning algorithms, deep learning, image and video processing, and intelligent systems.

NAIF AL MUDAWI received the master's degree in computer science from La Trobe University, Australia, in 2011, and the Ph.D. degree from the Collage of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. He has many published research and scientific articles in many prestigious journals in various disciplines of computer science. During the master's degree, he was a member of Australian Computer Science Committee.

MAHA ABDELHAQ (Member, IEEE) received the B.Sc. degree in computer science and the M.Sc. degree in securing wireless communications from The University of Jordan, Jordan, in 2006 and 2009, respectively, and the Ph.D. degree from the Faculty of Information Science and Technology, National University of Malaysia, Malaysia, in 2014. She is currently an Associate Professor with the College of Computer and Information Sciences, Princess Nourah Bint Abdul Rahman University, Saudi Arabia. Her research interests include vehicular networks, MANET routing protocols, artificial immune systems, network security, and intelligent computational. She is a member of ACM and the International Association of Engineers (IAENG).

MOHAMMED ALONAZI received the B.Sc. degree in computer science from King Saud University, in 2008, the M.Sc. degree in computer science from Florida Institute of Technology, Melbourne, FL, USA, in 2015, and the Ph.D. degree in informatics from the University of Sussex, U.K., in 2019. He is currently an Assistant Professor with the Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. His research interests include human–computer interaction, UX/UI, digital transformation, cyber security, and machine learning.

ABDULWAHAB ALAZEB received the B.S. degree in computer science from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree in computer science from the Department of Computer Science, University of Colorado Denver, USA, in 2014, and the Ph.D. degree in cybersecurity from the University of Arkansas, USA, in 2021. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. His research interests include cybersecurity, cloud and edge computing security, machine learning, and the Internet of Things.

ASAAD ALGARNI received the Ph.D. degree in software engineering from North Dakota State University, USA. He is currently an Assistant Professor with the Department of Computer Sciences, College of Computing and Information Technology, Northern Borders University, Saudi Arabia. His research interests include software engineering, computer vision applications, and machine learning.

AHMAD JALAL received the Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. He is a Postdoctoral Research Fellow with POSTECH. His research interests include multimedia contents, artificial intelligence, and machine learning.