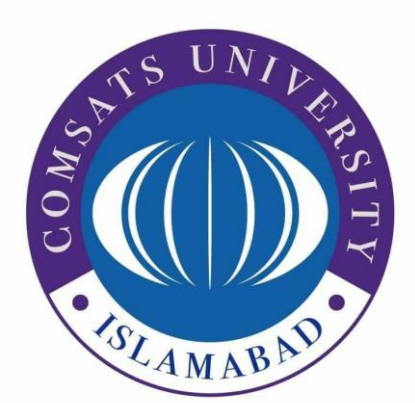
Object Detection and Classification Based on Feature Fusion and Deep Convolutional Neural Network



By

Muhammad Rashid

CIIT/SP17-RCS-009/Wah

MS Thesis

In

Master in Computer Science

COMSATS University Islamabad

Wah Campus - Pakistan

Fall, 2018

COMSATS University Islamabad, Wah Campus

Object Detection and Classification Based on Feature Fusion and Deep Convolutional Neural Network

A Thesis Presented to

COMSATS University Islamabad, Wah Campus

In partial fulfillment

of the requirement for the degree of

MS (CS)

By

Muhammad Rashid

CIIT/SP17-RCS-009/Wah

Fall, 2018

Object Detection and Classification Based on Feature Fusion and Deep Convolutional Neural Network

A Post Graduate Thesis submitted to the Department of Computer Science as partial fulfillment of the requirement for the award of Degree of M.S in Computer Science.

Name	Registration Number
Muhammad Rashid	CIIT/SP17-RCS-009/Wah

Supervisor

Dr. Muhammad Sharif
Associate Professor
Department of Computer Science
COMSATS University Islamabad (CUI)
Wah Campus
January 2018

Final Approval

This thesis titled

Object Detection and Classification Based on Feature Fusion and Deep Convolutional Neural Network

By

Muhammad Rashid

CIIT/SP17-RCS-009/Wah

Has been approved

For the COMSATS University Islamabad, Wah Campus

External Examiner: _____

Dr…….

Supervisor: _____

Department of Computer Science, Wah Campus

Co-Supervisor: _____

Department of Computer Science, Wah Campus

HoD: _____

Department of Computer Science, Wah Campus

Declaration

I **Muhammad Rashid** hereby declare that I have produced the work presented in this thesis, during the scheduled period of study. I also declare that I have not taken any material from any source except referred to wherever due that amount of plagiarism is within an acceptable range. If a violation of HEC rules on research has occurred in this thesis, I shall be liable to punishable action under the plagiarism rules of the HEC.

Date: _____

Signature of the Student

_____

Muhammad Rashid

CIIT/SP17-RCS-009/Wah

Certificate

It is certified that Muhammad Rashid having registration number CIIT/SP17-RCS-009/Wah, has carried out all the work related to this thesis under my supervision at the Department of Computer Science, COMSATS University Islamabad, Wah Campus. The work fulfills the requirement for the award of MS degree.

Date: _____

Supervisor

_____

Dr. Muhammad Sharif

Associate Professor

Department of Computer Science

COMSATS University Islamabad, Wah Campus.

Head of Department:

_____

Dr. Ehsan Ullah Munir

Associate Professor

Department of Computer Science,

COMSATS University Islamabad, Wah Campus.

“In the name of ***Allah*** the most

BENEFICENT the most ***Merciful.*****”**

All praise to **Allah**, Lord of the worlds, who blessed me to Accomplish this task.

It is only because of the blessings of **Allah** that I am able to Complete my Master’s thesis**.**

ACKNOWLEDGEMENT

In the name of **ALLAH**, the Most Gracious and Most Merciful, all praises to **ALLAH** for the strength and his blessing to me for the completion of my master's thesis. I am very thankful to almighty **ALLAH** who showers his blessing always on me. I am grateful to my supervisor **Dr. Muhammad Sharif** who was supportive and helpful to me all the time. He was my source of inspiration and motivation. I am also really thankful to **Dr. Tallha Akram** who help me in all the time. Being truly without his support, guidance, and encouragement this work was not possible. I am thankful for his guidance and cooperation throughout the degree all the way. In this work, I am also thankful to **Dr. Mussarat Yasmin**, **Dr. Jamal Hussain Shah**, and **Dr. Mudassar Raza,** which help me all the time and give me value able comments and suggestions. I am also like to thank **Dr. Ashfaq Ahmed, Dr. Muhammad Awais, Mr. Attique Khan, Mr. Inzamam Mashood Nasir, Mr. Mohsin Khan** and all of my Lab Fellows for their valuable time and fun.

Muhammad Rashid

CIIT/SP17-RCS-009/Wah

ABSTRACT

Object Detection and Classification Based on Feature Fusion and Deep Convolutional Neural Network

Object detection and classification are challenging task implemented in the domain of pattern recognition and machine learning (ML) due to their emerging applications such as video surveillance and pedestrian detection. In recent times, numerous deep learning-based methods are presented for object classification, but still, several problems/concerns exist which reduce the classification accuracy. Complex background, appearance in congest, and similarity in different objects are real-time challenging issues. To resolve these issues, a new deep learning method is introduced which is based on deep convolutional neural network (CNN) and SIFT point features. First, an improved saliency method is implemented and extract their point features. Then, Deep Convolutional Neural Network (DCNN) features are extracted from two deep CNN models like VGG and AlexNet. The CNN features are extracted by performing activation on fully connected (FC) layer and then perform max pooling to remove the noise factor. Thereafter, Reyni entropy-controlled method is implemented on DCNN pooling and (Scale Invariant features transforms) SIFT point matrix to select the best features. Finally, best features are fused in a matrix by a serial-based method, which is later fed to ensemble classifier for classification. The proposed method is evaluated on five publically available datasets including Caltech-101, Barkley 3D, Pascal 3D+, Birds, and Butterflies datasets and obtained classification accuracy 93.8%, 99%, 88.6%, 100% and 98% shows improved performance as compared to existing methods.

Keywords: Object Detection, Hand Crafted Features, Deep CNN, Features Fusion, Features Reduction, Classification

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
FaNeR	False Negative Rate
FaPoR	False Positive Rate
AuC	Area under Curve
FaNe	False Negative
FaPo	False Positive
NN	Neural Network
HSV	Hue-Saturation-Value
DT	Decision Tree
ROI	Region of Interest
MLP	Multilayer Perceptron
PCA	Principle Component Analysis
RGB	Red-Green-Blue
KNN	K-nearest Neighbor
ROI	Region of Interest
SVM	Support Vector Machine
PPV	Positive Predictive Value
ANN	Artificial Neural Network
TPR	True Positive Rate
FDR	False Discovery Rate
DCNN	Deep Convolutional Neural Network
ObDC	Object Detection and Classification
SIFT	Scale Invariant Feature Transformation
SVM	Support Vector Machine
L-SVM	Linear Support Vector Machine
Q-SVM	Quadratic Support Vector Machine
C-SVM	Cubic Support Vector Machine
WKNN	Weighted K-Nearest Neighbor
DSN	Deep Stack Network
GLCM	Gray Level Co Matrix

HOG	Histogram of Oriented Graphs
SURF	Speeded Up Robust Features

Chapter 1
Introduction

1. Introduction

As the smart cities concept is being implemented rapidly around the globe, recognizing objects automatically has become more common. Objects are recognized and categorized into relevant classes. The Internet is also used commonly to search any object. As there is a lot of information available, classifying this information is required. Classifying information is easy when the target scale is small. Classifying information nowadays by hand is becoming impossible as the classes of objects belong to millions of categories, Classifying the objects into different classes has become necessary.

For people and numerous different creatures, visual observation is a standout amongst the most critical faculties. We vigorously depend on vision at whatever point we communicate with our condition: When we get an object, when we travel through our condition and abstain from knocking into everything in transit or when we perceive our companions by their appearances. For all those errands, object acknowledgment and localization are fundamental. To get a glass, we have to initially figure out which part of our visual impression relates to the glass before we can discover where we need to move our hands to get a handle on it. We never consider these essential preparing steps effectively. In any case, what appears to be so easy for our cerebrum still represents a noteworthy test for counterfeit frameworks like robots that need to process picture content. Existing calculations regularly as it handled a little subset of the distinctive errands important for understanding a picture also, are extremely requesting as far as computational assets and runtime. All together to replicate somewhere around a piece of the human visual discernment capacities, one would need to consolidate a few distinct calculations. Making such a consolidated framework run continuously with the present equipment is a major test. A little advance towards this objective is investigated in this work via preparing a neural system model to realize which parts of a picture are intriguing to human eyewitnesses that scan for an explicit object. This information would then be able to be utilized to accelerate object look in computer vision [1-9].

Object detection and classification are challenging tasks in the domain of computer vision (CV), which are used to classify the objects according to their class labels [10, 11]. This domain got much attention due to their enormous applications including video surveillance, the target recognition, face detection, optical character recognition, video stabilization, image

watermarking, and automated pedestrian detection. Promising results are achieved recently in this area when dealing with simple images of transparent background. But, the problem is not well addressed, when objects containing a complex background, multiple shapes, and appears in congest and scattered background [12].

Researchers work in this domain from last two decades and try to categorize the challenging problem of object detection and classification including complex background, features extraction, best features selection, execution time and accurate classification. They design several features extractions-based methods to detect and classify complex objects using classification methods. Bag of Features (BoF) is one of the most emerging ways for object classification into their specific category and used by several researchers [13]. These features perform well for object classification, but still, they were limiting the ability of image representation. Lazebnik et al. [14] introduced a new technique to deal with this limitation by BoF and made a Spatial Pyramid Matching (SPM) technique, which can split the image into spatial patches, and computes the histogram of each sub-region, which is later used for the creation of a spatial location sensitive vector. Some other features which are used for object classification are SIFT[15], Local Binary Pattern(LBP) [16], color features [17, 18], shape features (HOG) 1-5[16, 19-21], texture features [22] and deep features using Convolutional Neural Network(CNN) [23]. However, features fusion is also an important research area in which combined patterns of multiple descriptors are used to improve the classification accuracy. The prominent fusion methods are a serial-based method and parallel method, which are used in several domains like medical imaging, video surveillance, biometrics, and few more.

Stochastic discriminant analysis (SDA) [24], fusion of low level and mid-level features [25], and transfer based features fusion techniques are famous algorithms in this domain. Therefore, the major aim of feature reduction techniques is to solve the issue of dimensionality and also remove the redundant information. The most existing feature reduction and classification methods are Principal Component Analysis (PCA) [26] , Entropy-based feature selection [27], Genetic Algorithm(GA) [28], Particle Swarm Optimization (PSO) [29, 30], and Canonical Correlation Analysis(CCA) [31]. The reduced features are finally classified by classification algorithms like Linear Support Vector Machine (L-SVM) [32, 33], Cubic-SVM (C-SVM),

Quadratic-SVM (Q-SVM), Fine K-Nearest Neighbor (F-KNN), Cubic-KNN (C-KNN), Ensemble Subspace-KNN (EKNN), and few more. These methods are also known as supervised learning methods. To overcome the drawback of conventional algorithms of features extraction like handcrafted approaches, deep learning is a suitable candidate for this domain due to their natural ability. In deep learning, CNN is a subtype of deep architecture [34], have shown improved performance for classification and recognition with the successful applications such as machine learning and pattern recognition [35]. Several pre-trained DCNN models are introduced by several researchers such as VGG [36], AlexNet [37], ResNet [38], YOLO model, and GoogleNet [39]. These model are used for several purposes such as image classification [40], action recognition [41], medical imaging [42], agricultural plants, and few more. Dmytro et al. [32] introduce a new CNN based approach for object classification. The introduced method is utilized Convolutional layers and extracts 4D descriptors, which are work as object classification. However, these methods are not performed well when we have large of images and most of the methods are validated on a maximum of 30 images per class.

As before, no one fused two pre-trained DCNN models because each model has a different number of inputs, which make the problem for fusion. Moreover, patterns fusion of two DCNN models provides better classification performance as compared to the individual model. To inspire this approach, we introduced a novel DCNN method for object classification from static images. The proposed method employed in two parallel steps. The first step has an improved saliency-based method and SIFT point features are extracted from a mapped RGB image. Then, VGG and AlexNet pre-trained DCNN models are used and extract deep CNN features by employing activation on the FC layer. Thereafter, Reyni entropy-controlled method is proposed, which is employed on DCNN and SIFT Point feature matrices to select the best features. But the size of inputs for each model is different, which takes a problem in the fusion process. To resolve this problem, we perform the augmentation and make the equal size of both matrices. Then fused both feature matrices using a serial-based approach and stored features in a new matrix, which are passed to ensemble classifier for classification.

1.1 Object Detection in Computer Vision

The capacity to identify the objects resides in a picture or scene is a standout amongst the necessities with regards to associating with one's condition. While it appears to be easy with

people and in truth, most creatures, attempting to instruct PCs to see - and furthermore "comprehend" what they are seeing - has demonstrated to a great degree troublesome.

The way to understanding visual scenes are three firmly related sub-issues. The simplest one will be called classification in the accompanying. For classification, the one overwhelming object in a given picture ought to be resolved and marked. The following additional requesting errand is object localization: notwithstanding marking the predominant object, it additionally should be restricted in the picture, generally by deciding a bounding box around the picture locale that is involved by the object. The trouble of this undertaking again increments if one, as well as all objects in a picture, should be marked, what's more, various objects of a similar classification can show up in one picture. This errand is called object detection. Various varieties of this errands exist, and the wording utilized in this proposal won't generally be acclimated with different sources. What's more, object detection is a major and extremely dynamic field of research, so the accompanying passage will give an exceptionally concise diagram about late advances

1.2 Challenges

Tackling and getting the most out of a dataset with multiple classes is one of the key challenges of the presented thesis. Imbalance data in sample classes implies a significant difference in the classification results. If the dataset is imbalanced, the decision boundary for a total error of majority class is biased by the masking of minority class [43]. Thus the classes having a higher number of samples are tend to be adopted over the classes with smaller samples [44]. To overcome the issue of imbalanced datasets, a solution is implemented by class equalization. Minimum images in each class were set to as the selected images from each class.

1.3 Research Motivation

In the past, object classification has been used for transmission and storage of information. Dealing with the image in an efficient and integrated way is the main objective of many companies. Recognizing the appropriate class of objects and then assigning that image to its class is the main objective of this thesis.

Object classification can be used and needed in many ways [45]. It is considered an important step in the grouping, classifying and analyzing the images semantically. For several applications, classification is the main step.

Detection and classification [46] are utilized in various therapeutic imaging applications. Distinguishing destructive cells, abnormalities, sicknesses, uncovering inward structures covered up by the skin also, bones, and in addition diagnosing and treating infections are the principal objectives. The information can be gotten from various imaging procedures including X-beam radiography, infinitesimal pictures, medicinal ultrasonography or ultrasound, and attractive reverberation imaging. Video reconnaissance is valuable for the accompanying things: expanding security, observing and recording exercises for different purposes, forestalling the loss of products, offering office insurance, upgrading representative security, hindering vandalism and illicit exercises. People of intrigue, permit plates, vehicles or different objects should be recognized and followed in those applications.

1.4 Problem Statement

Every category of the object has specific shape and color. ObDC contains the four steps which are preprocessing, feature extraction after segmentation, feature reduction, features fusion, and classification.

Some significant gaps found in the ObDC systems are:

- ObDC accuracy found affected by the noise, distortion and less/high illumination effect.
- Many irrelevant features cause the less accuracy of ObDC.
- After relevant feature extraction, selection of the best features is also the main reason for less accuracy of ObDC.
- After extraction of any two types of features descriptors, descriptor unbalancing is an issue.
- After generating the final codebook for each descriptor, the fusion of all feature vectors is another challenge in the domain of ObDC.

1.5 Contribution

The objective of this proposition is to dissect, propose, create and assess approaches for object detection. The primary center is to create quick and precise strategies by depending on logical data, competitor age, and clear highlights. To accomplish this objective, we depend on the accepted object detection pipeline. To acquire a quick detection, a decent district of intrigue or hopeful age module is required. This progression guarantees that unessential zones from the picture are disposed of as quick as would be prudent. Additional tedious classification and thinking should be possible on a set number of troublesome cases.

- Initially, data normalization is performed to make the data ready to use for DCNN as far as the DCNN requires the RGB images while the dataset consists on some gray images.
- Two different pre-trained DCNN features used to extract the features and perform activation on FC layers. This activation on FC layer extracts deep features, which are later tuned by max-pooling to remove the features causing noise.
- SIFT point features are extracted for object localization by extracting the local key points first, secondly assigning the orientation to those key points.
- The features, which are obtained after max-pooling are fused in a vector and selection of best features is performed. The best features are selected by employing Renny-Entropy on the fused vector, which returned the feature by their score values in descending order.

1.6 Thesis Layout

The proposition is sorted out into five parts, each concentrating on various highlights of the examination work. Coming up next is an outline of the substance of every section.

Chapter 1 gives a review of Object Detection and Classification (ObDC) and a more point by point examination of the issues. The examination points and goals of this investigation are likewise displayed. What's more, the inspirations which have prompted this examination, the commitments of the proposal and issue explanation are displayed in this early on the section.

Chapter 2 depicts the foundation and the writing survey notwithstanding fundamental data on ObDC. It likewise gives data on the dimension of headway in the region of picture handling for ObDC frameworks.

Chapter 3 clarifies the improvement of the proposed framework, utilizing the recently portrayed datasets and picture handling procedures, for example, unique pre-preparing strategies, distinctive element parameters and diverse classifiers in the ObDC framework.

Chapter 4 shows the general outcome examination for the programmed created frameworks. The section talks about the examination performed on the master determination. With the end goal to create a difference of framework testing results and framework execution, the general examination of the created frameworks is introduced in two different ways.

Chapter 5 condenses the achievements of the exploration work. It finishes up the substance of the postulation and furthermore features a few suggestions for future research work. It likewise gives data concerning the examination commitments, which have profited various territories.

Chapter 2

Literature Review

2. Literature Review

This chapter explains and discusses the existing techniques related to object recognition classification and few core concepts helpful to understand the classification technique proposed in Chapter 3. In machine learning and computer vision domain, different techniques are presented for image classification. Object detection and classification is an evolving field of computer vision based on their developing application such as video surveillance. Recently, in the field of machine learning to handle large data, D-CNN gains much attraction in this domain. The existing machine learning algorithms such as SVM, single layer neural network, decision tree, and KNN have degraded the classification accuracy and take much computation time when huge number of images are utilized for validation.

Object detection [47-50] belongs to principal machine vision undertakings. The exploration from this area has delivered various methodologies. Despite the fact that much advancement has been made, most detection frameworks are a long way from prepared for genuine applications. A vital test to survive is giving hearty outcomes notwithstanding when the information picture is uproarious. It is likewise hard to create quick calculations that are equipped for detection continuously. There are a few applications where object detection is required. Therapeutic applications, for example, distinguishing dangerous cells are of a colossal significance. Programmed path stamping detection and movement sign detection have been progressively presented in current vehicles. Bits of knowledge into object detection assist us with understanding the manner in which our minds decipher visual data. Organically motivated framework structures have demonstrated to give great outcomes despite the fact that eventually, it is important to split far from impersonation. The human visual framework is naturally multilayered with each layer relating to a specific channel. The objects are perceived at the most abnormal amount.

Face detection is apparently the most looked into the region in this field. Basic applications include programmed photograph labeling for web-based life applications and programmed center for computerized cameras. Increasingly explicit acknowledgment systems can be utilized to distinguish crooks or to check the character of people consequently at visa checkpoints. The trouble with the issue lies in covering diverse face introductions and treating incomplete impediment because of glasses, hands or different objects.

The person on foot detection is another greatly explored point in this field. Expanding worry for walker wellbeing in the most recent year has brought about the prospering of the person on foot detection calculations. These are fundamental in Advanced Driving Assistance Systems for avoiding mishaps including people on foot. Vehicle organizations are thinking about consolidating such frameworks into their models. For instance, Volvo wants to discharge autos that accompany a person on foot and cyclist detection module which will have the capacity to stop the vehicle naturally in the event of a fast approaching impact. Despite the fact that the subject of distinguishing people on foot was handled by numerous specialists, it remains to a great extent unsolved because of a few challenges: the different visual appearance and fluctuated garments of people on foot, distinctive conceivable stances and verbalizations, swarmed scenes where fractional impediment forestalls detection and the extensive scope of scales. The issue is as yet open to inquire about, with frameworks that meet continuous necessities being particularly hard to create.

2.1. Dataset Normalization and Balancing

Dataset normalization and augmentation [51-54] is performed if the availability is not enough and the deficiency of images in each class which may cause the under-fitting and over-fitting. Balancing is performing by augmenting data using different sources, i.e., augmentation can be performed using differently available datasets or maybe internet sources. Salman et al. [55] mainly focused on data balancing by introducing a cost-sensitive algorithm which works by the back propagation method. In this approach, 16 layers of the convolutional neural network were used. Simply, by adding two fully connected layers before the output layer by which backpropagation was performed, the approach performed well for data balancing and classification as well. By adding two new layers, a new architecture was made consisting of the 18-layer model. On Caltech-101, the best accuracy achieved was 87.1% and 90.8% by using 15 and 30 training samples respectively. The proposed approach is presented in Figure 2.1 which can be used for the brief understanding of the workflow of proposed technique.

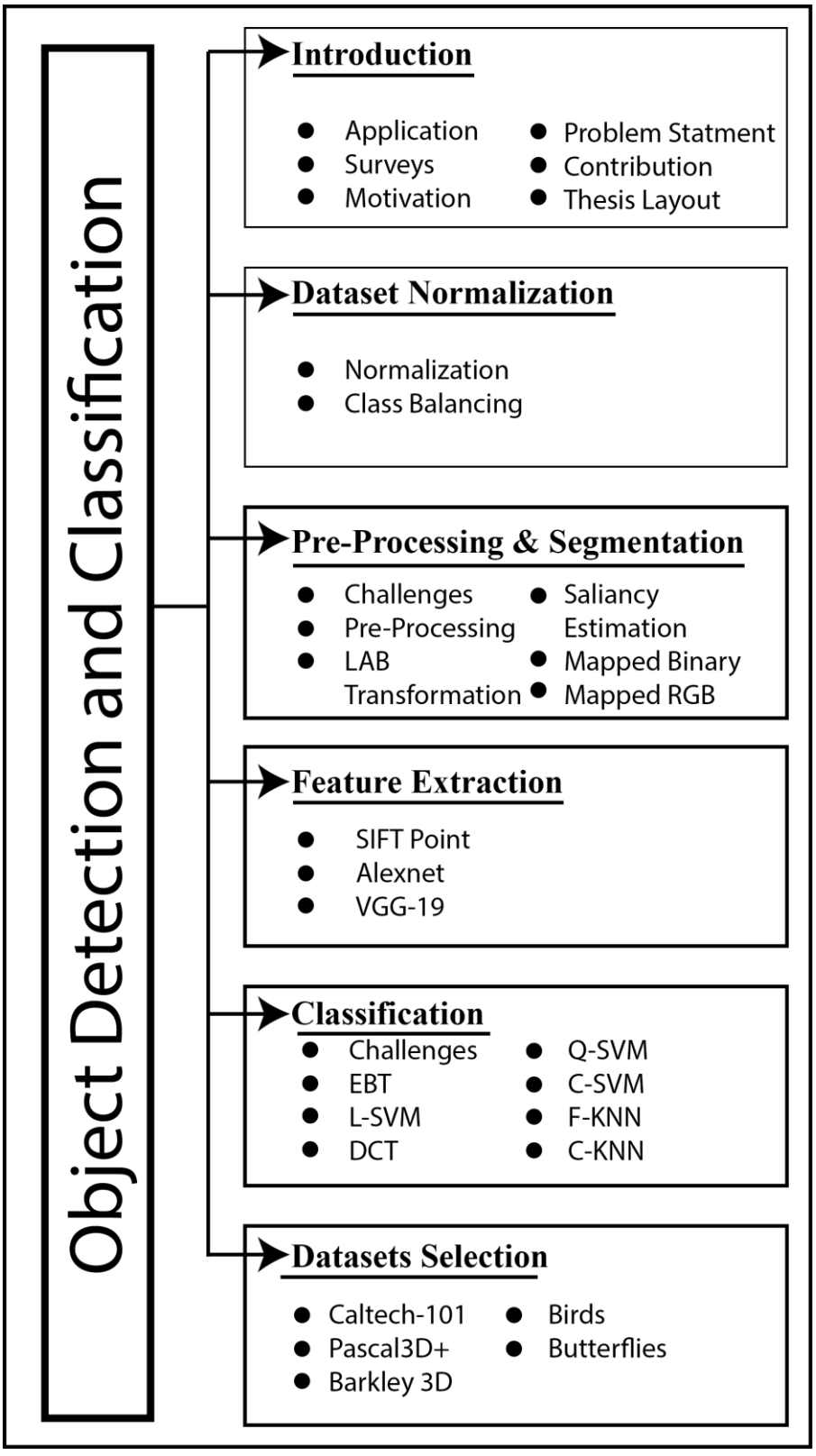


Figure 2.1 Flow Diagram of ObDC

2.2. Segmentation

Segmentation is the way toward dividing a computerized picture into multiple sections (sets of sub-regions, called super-pixels) [56, 57]. The motive behind segmentation is to rearrange and additionally change the portrayal of pixels to a substance that is progressively significant and less needed to analyze [58]. Segmentation is normally used to trace objects and limits (bends, lines, etc.) from pictures. Saliency map based segmentation can be illustrated be Figure 2.2.

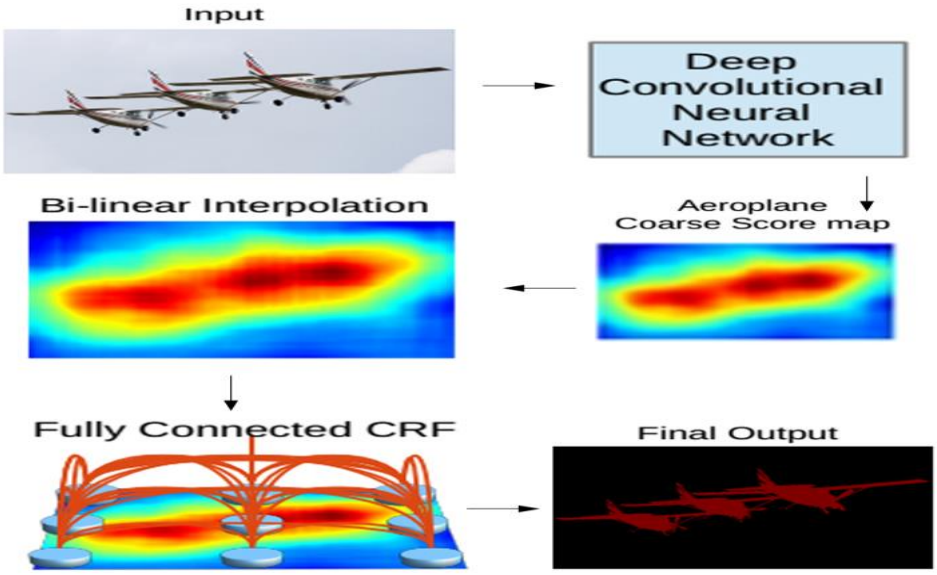


Figure 2.2 Reconstructed Segmentation Technique proposed by [59]

Ultimately, segmentation is the procedure to relegating a mark to each pixel in a digital picture to a level such that homogenous pixels share some certain qualities [60]. The conclusion of segmentation is an adjustment of regions that by and large cover the whole picture, or an arrangement of shapes separated from the picture (see edge detection). Connected areas are altogether unique as for the equivalent characteristic(s). [1] When connected to a heap of pictures, ordinary in medicinal imaging, the shapes after image segmentation may be utilized to adopt 3D redesigning with the help of insertion calculations including Marching solid shapes. Bharath et al. [61] presented segmentation technique by employing hypercolumn based pixels. Figure 2.3 shows their presented results for segmentation.

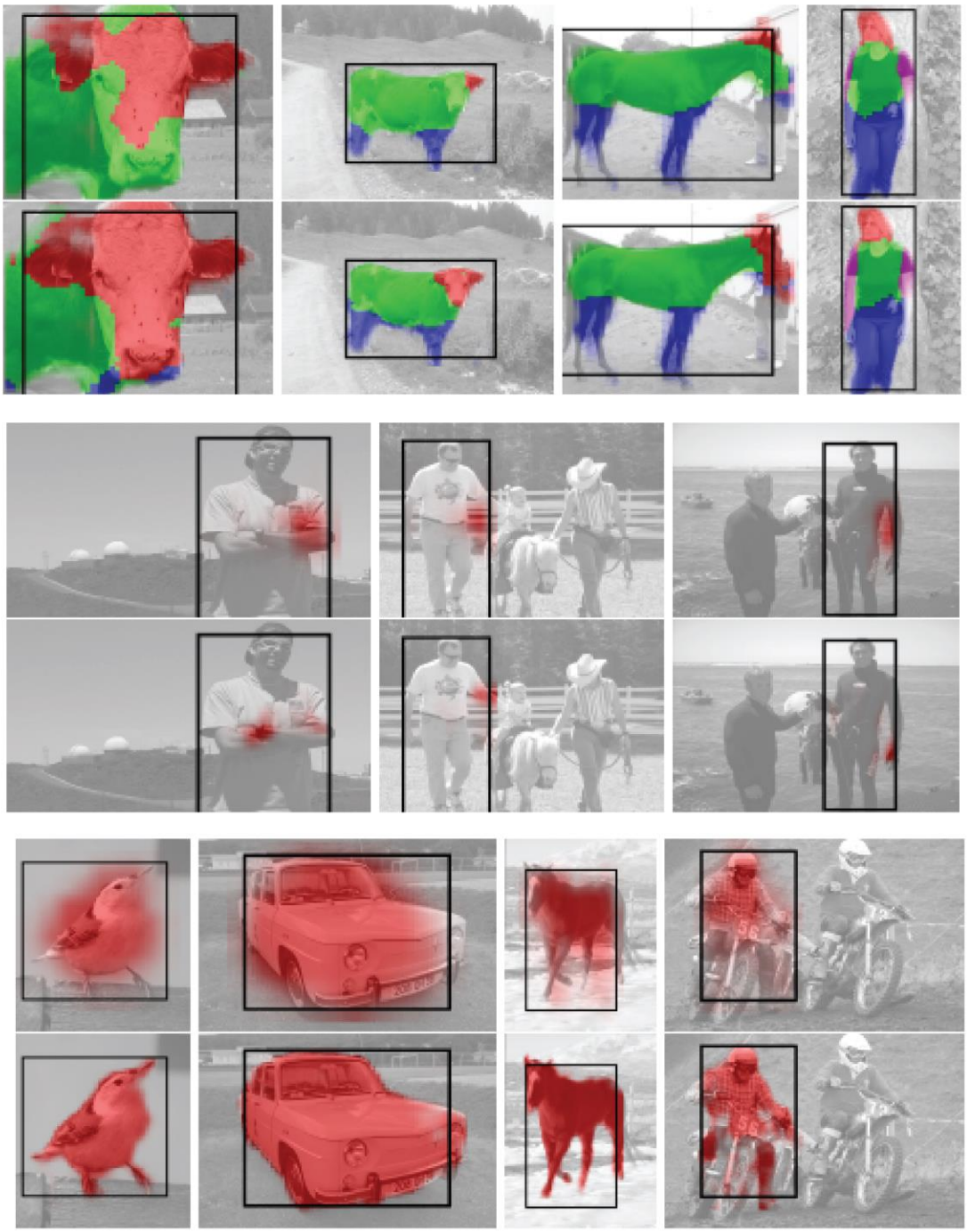


Figure 2.3 Proposed segmentation results by [61]

For the reasons previously mentioned, frontal area objects and foundation districts ought to be divided consequently, effectively and precisely before recovery. There are numerous strategies to section the frontal area from the foundation. Saliency-Cut [62], a Saliency-Segmentation strategy [63-67] is referred to cropping the image with respect to visually prominent objects.

Figure 2.4 shows the Saliency measure estimation based segmentation using RC saliency map and then segmenting the image into binary image.

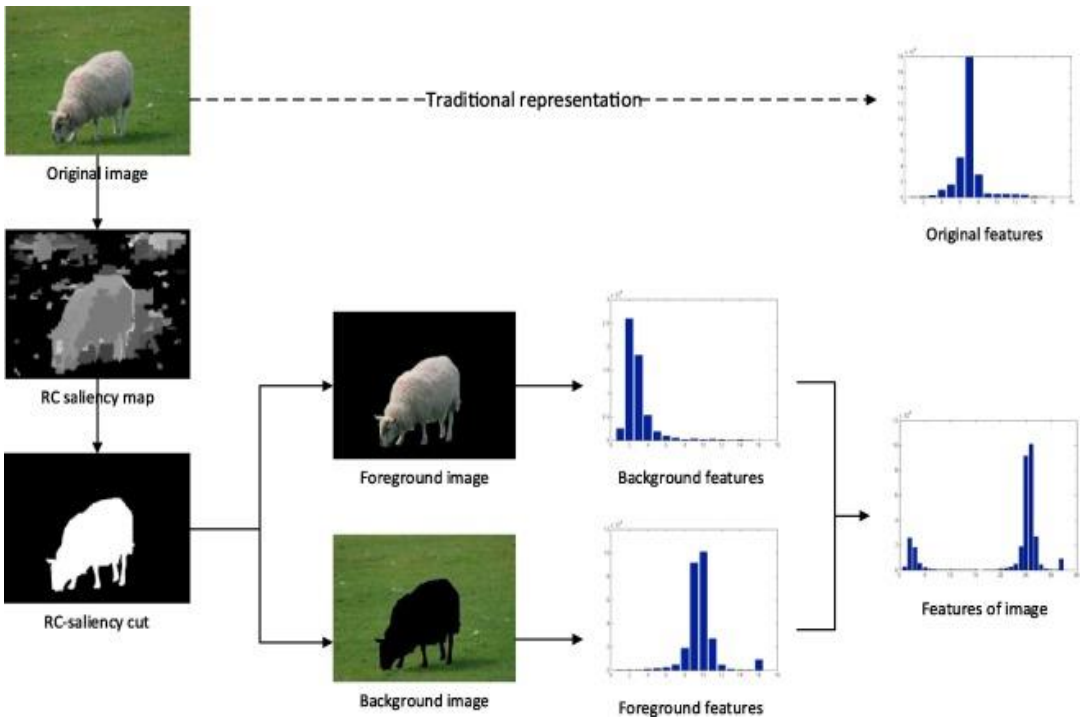


Figure 2.4 Saliency measure segmentation proposed in [68].

Moreover, this segmentation strategy joining with saliency detection technique using Region-based Contrast (RC) [62] may accomplish unrivaled execution contrasted and cutting-edge unsupervised remarkable object extraction strategies [69].

2.3. Feature Extraction

Image features are pertinent data separated from pictures for an explicit undertaking. They are basic for any abnormal state undertaking, for example, detection, acknowledgment and following. Highlights can be anything from a basic point to a huge descriptor characterized on a district. Picture highlights are more often than not gotten via a procedure known as highlight extraction, Where the highlights are separated can be controlled by utilizing an element indicator which gives a rundown of intrigue focuses or on the other hand by examining thickly from the entire picture. Figures 2.5 presents the system proposed by [70] which works on the basis of neural network based clustered features using fuzzy logics.

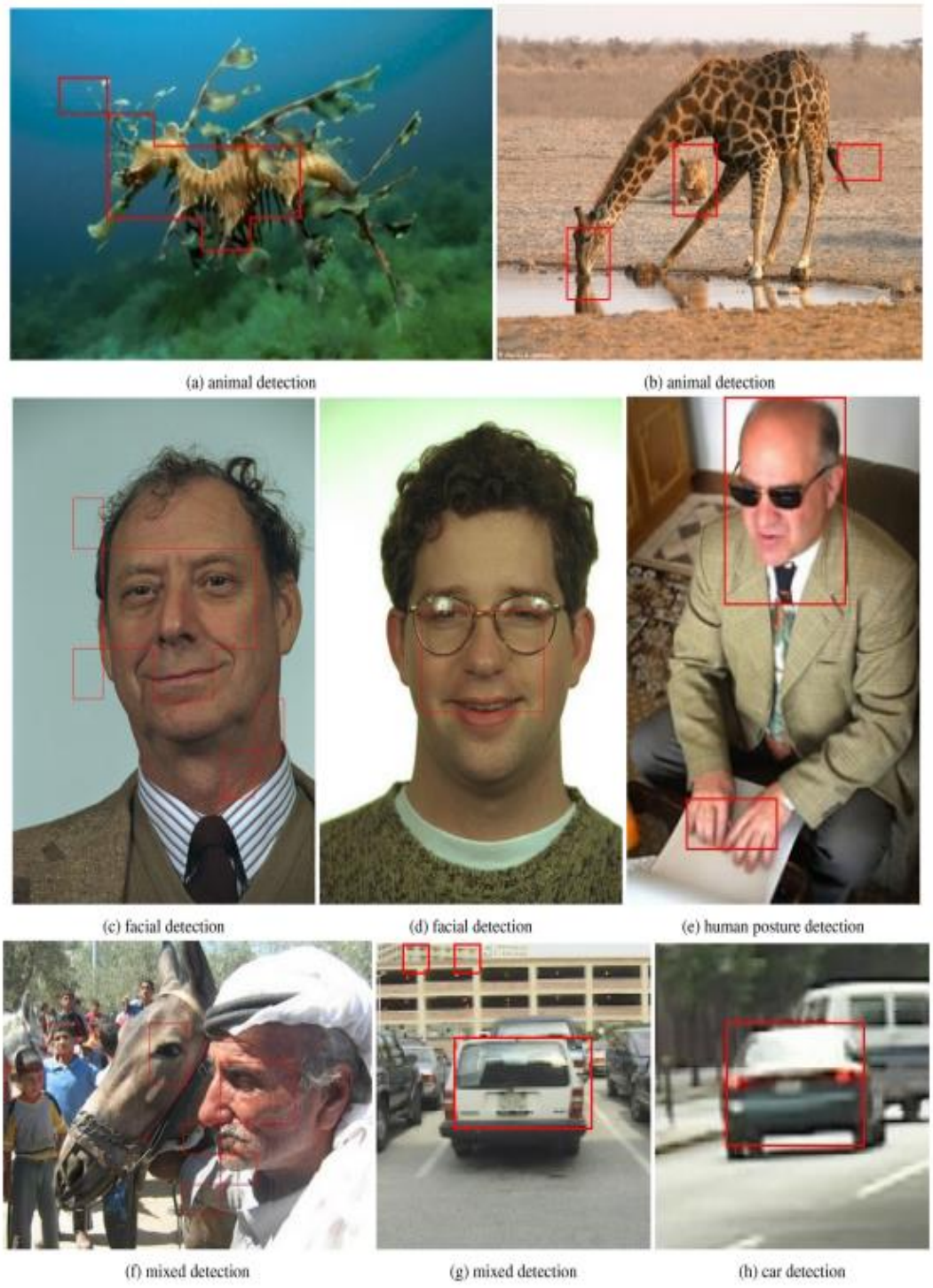


Figure 2.5 Object recognition proposed by using clustered features [70]

The last methodology is rehearsed with extraordinary achievement in present-day machine vision approaches because of advances in PC equipment that makes this requesting approach conceivable. Shading or force at a situation from the picture is a case of likely the most basic

what's more, most minimal dimensions includes. Increasingly mind-boggling ones are determined on a picture district, for example, Histograms of Oriented Gradients [71-73]. At the point when a picture area is depicted by a component vector, we call such a vector a "highlight descriptor" of the district. Highlight descriptors can be gathered into a few classifications. The accompanying area represents, we give a scientific classification and embody every classification with comes closer from the writing. Object detection techniques regularly utilize custom highlights intended for the explicit errand.

Feature extraction is the first mandatory and key step in image recognition which refers to creating a codebook for a visual representation of images. Features are playing an important role for object classification and recognition in computer vision. Many researchers used many types of features in their proposed techniques like shape features, point features, texture features. From past few decades, researches using hand-crafted features were made for object classification [74, 75]. Shape features include Histogram of Oriented Graph (HOG) features [20] and geometric features (GLCM) [76-78].

Color can be seen as the feature of the most reduced dimension since it consolidates just neighborhood data. Indeed, even at this dimension, there is a vast intricacy in view of the huge number of accessible portrayal. Various color schemes may be utilized to speak to colors.

HOG are being widely used since two decades in the domain of object recognition and facial recognition. Using HOG, 8100 features are extracted for any image. Scale Invariant Feature Transformation (SIFT) [79, 80] was introduced to detect the point of interest with reference of Difference of Gauss[81]ian (DoG) [82] based on image patches. Object detection implementation in videos [83] is shown in Figure 2.6.

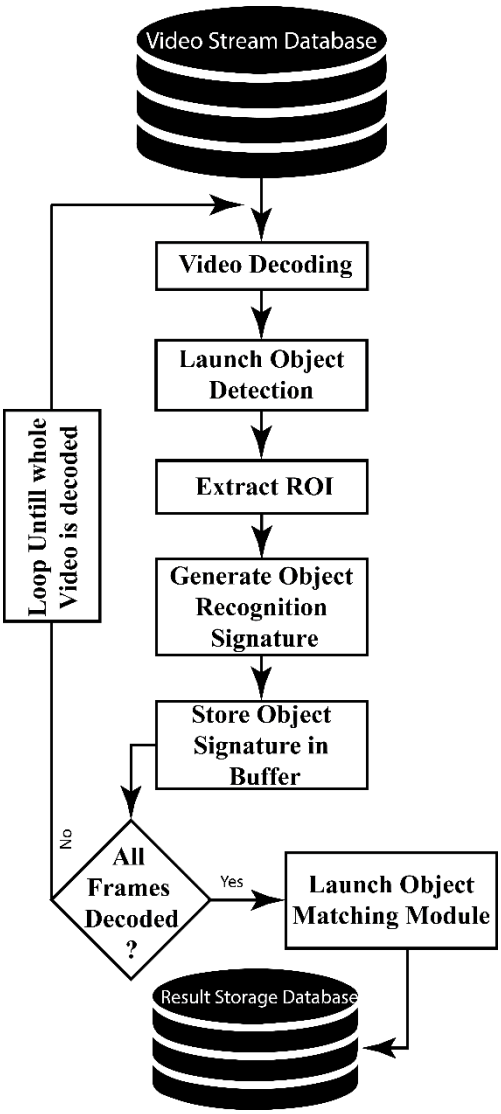


Figure 2.6 Object Detection in video streams redrawn from [83].

SIFT features [84] performed well due to their point of interest which remained the same whether the position of pixels is changed or not. Easiness to implement was the main advantage for these features, but the high dimensional space was a big issue in these features. Speeded Up Robust Features (SURF) [81, 85] inspired by SIFT also performed well in the discussed domain, is a type of local features being used in the domain of object classification. SURF works on the basis of a blob detector and six different types of points of interest. Texture features (HARLICK features) [86-88] having a feature vector of 17 features is extracted using a gray co-matrix. Researched showed that many approaches were proposed to use different features fused in the same vector. Since, concluding the more useful features, was a very

challenging task. Local features were more than enough when there were simple mages with a simple background. With the passage of time, due to less intra-class variability, the local feature was unable to classify the objects hence Deep Convolutional Neural Network (Deep-CNN) was presented. A DCNN has mainly three main categories of layers including first convolutional layers, the pooling layers and at the last, fully connected layers. By using these layers, a Deep-CNN train itself using two types of techniques which are feedforward network and backpropagation as the system proposedd by [89] is shown in figure 2.7.

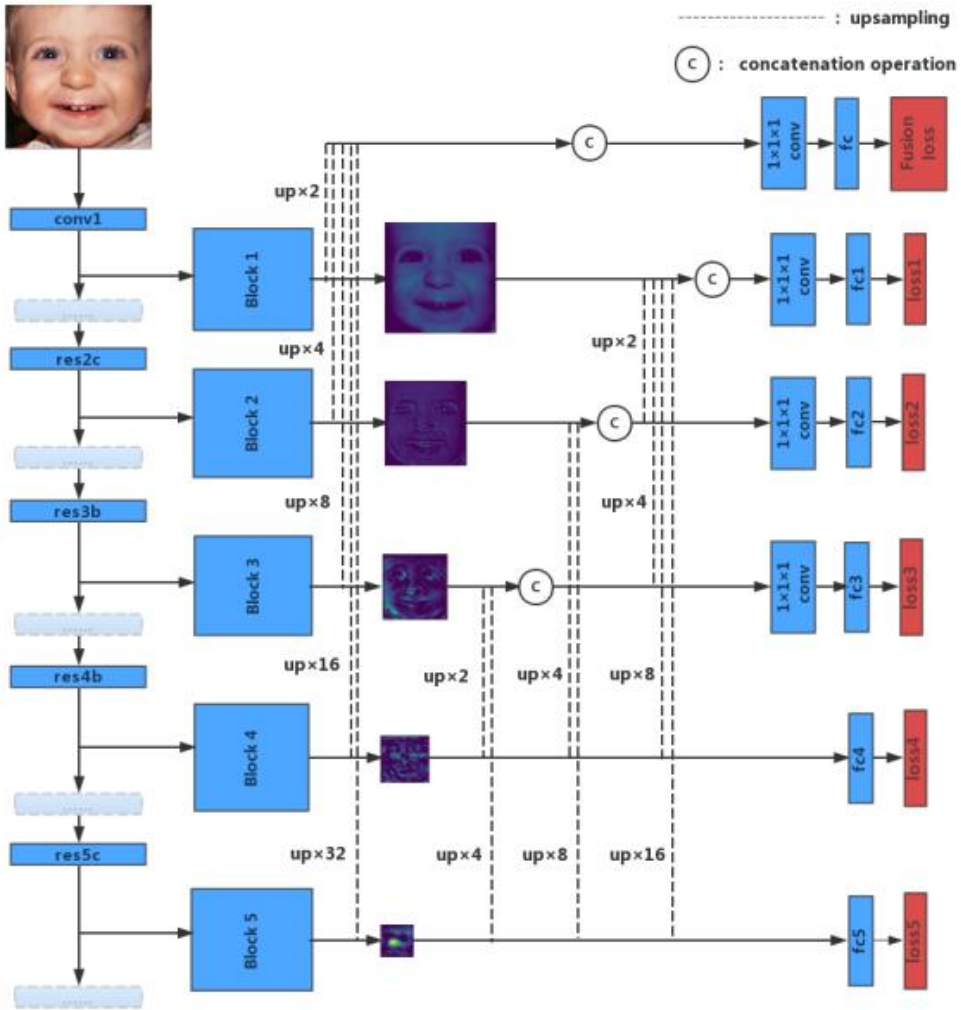


Figure 2. 7 system proposed by [89] for feature extraction using Neural network

In convolutional layer, a Deep-CNN convolutes the whole image by utilizing various size type of kernels. By using a pooling layer, extracted features are reduced, and the reduced features

are passed to network. Researchers introduced many networks having various numbers of layers [90]. AlexNet, VGG [36] (VGG-16, VGG-19), GoogleNet, ResNet (Resnet-50, ResNet-102, and ResNet-152) were introduced with the passage of time to solve the recognition complexities. Pre-trained network refers to utilizing the already initialized weights instead of random weights. We used VGG-VeryDeep-19 [36] pre-trained models which belong to VGG-VD, and they are already trained on ImageNet ILSVRC challenge data. Chunjie et al. [91] introduced a new technique to handle the drawbacks occurred by local features named as Contextual Exemplar. The technique followed by three phases. The first phase combines the regions based image, the second phase was constructing the relationship between those regions, and the third phase was using the relationship of those regions for semantic representation. By using top 1000 features, best accuracy reported was 86.14%. Qing et al. [18] presented an image classification method using Color based Information and intelligent learning algorithm. The algorithms are integrated by a bounded soft-assignment coding method, which is passed to SVM for categorization. The presented approach is validated by Caltech-101 [92] dataset and achieved classification accuracy 73% using YCbCr color space. Wei et al. [93] presented a new method of image classification using intra-class CNN feature pyramid. The advantage of lower level layers is used to get the structural information and the higher-end layers to get the semantic features. The high-level layers resolve the problem of semantic ambiguity created by lower-level layers. The features are extracted using already trained models such as AlexNet and VGG16 and shows improved performance on the Caltech-101 dataset. In [94] Vector of Regionally Aggregated Descriptors (RAD) is stated. In the presented method, pre-trained DCNN models such as VGG-16 and VGG-M are utilized for building Spatial Pyramid Patch (SPM), and later on Principle Component Analysis (PCA) is utulized as dimensionality reduction. Roshan et al. [95] also presented a unique method for object classification.

Weifeng et al. [31] presented an enhanced dimensionality reduction by employing conventional CCA and introduced hessian CCA. They exploited the CCA and hence presented hessian multiset canonical correlations and employed it in the domain of multiview. The main flow of presented enhanced reductio method is visualized in Figure 2.8.

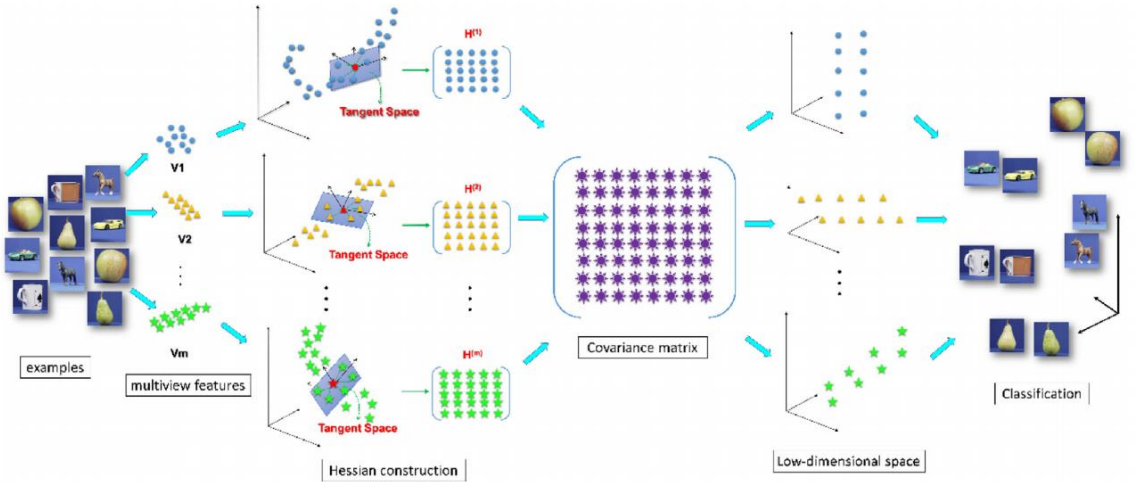


Figure 2.8 Hessian CCA based classification [31].

They apply the proposed algorithm on a VGG-16 architecture and performed training from scratch and further applied the transfer learning on top layers. An accuracy of 91.66% by this algorithm on the Caltech-101 dataset. Image input resolution was used as 224x224x3 in this technique. Qun et al. [96] gave a new technique by feature extraction using a pre-trained network with associative memory banks. They extracted the features using ResNet-50 and VGG-16, later on, K-Means clustering was used on memory banks to perform unsupervised clustering. The proposed approach reported 80.8 % recognition accuracy with VGG-16 model and 91.0% with the ResNet-50 model. Hang et al. [97] introduced a new technique for recognition by using two phases technique. First phase of technique encoded the feature representation, and the second phase was about Texture Encoding Network in which extraction of features, and an end to end classification. The resnet-50 architecture was used to illustrate the proposed technique. Maximum accuracy reported by this method was 85.3% on Caltech-101.

2.4. Feature Reduction

As with the growing complexity of feature descriptors, there comes a need to minimize descriptors concerning their influence on the accuracy and also minimize the computational cost. Many types of research were made for feature reduction to optimize the visual representation including PCA[98, 99], and Linear Discriminant Analysis (LDA)[100]. There are basically two types of feature reduction which are supervised and unsupervised techniques.

Pearson Correlation Coefficient (PCC) [101] assigns a ranking to the features using li1-5near co-relation among features and category label. Unsupervised techniques are the most commonly used techniques due to their easiness to implement. Principal Component Analysis (PCA) [102] is the type of dimensionality reduction technique which first normalizes the features, then perform Eigen decomposition and finally decomposed values are sorted in descending order. PCA is familiar due to its advantage for easy implementation and less computation cost. Also, since it is an unsupervised technique, therefore it doesn't require the class labels. The disadvantage of using PCA is that it requires the manual intervention of human and it may suffer from more non-linear complex features as it constructs the features in linear manners. Independent Component Analysis (ICA) [103] is the technique which may work on broader data of any category using unsupervised source separation. Pros of ICA is that no prior knowledge to experiments is required and the cons of ICA is that it needs high computation cost. Miguel et al. [104] proposed a technique for the pooling method rather than average pooling or max pooling known as Ordered Weighted Pooling (OWP). Which worked on the basis of high activations and low representativeness. Results showed the fair comparison among the results achieved using max pooling, average pooling and OWP. OWP outperformed as compared to other methods. Jinjoo et al. [105] stated a unique technique to solve dimensionality reduction by Structured Sparse Principle Component Analysis (SSPCA). The SIFT features are grabbed from input images and then remove irrelevant features using proposed SSPCA approach sown in Figure 2.9.

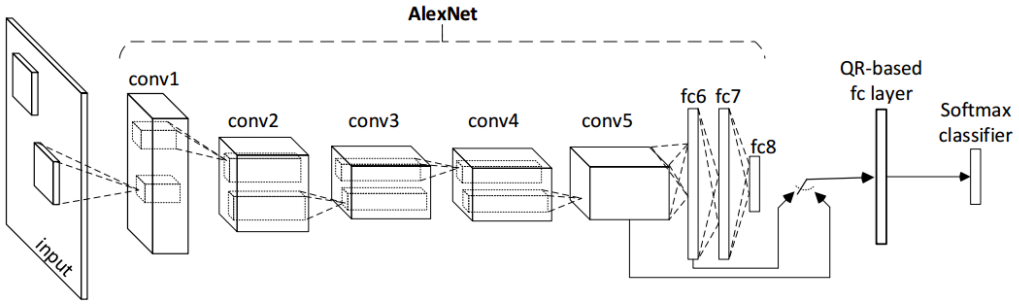


Figure 2.9 Feature reduction Proposed by [106]

Jingjing et al. [107] introduced a clustering based multiple kernel learning (MKL) method to differentiate the inter-class correlation. The introduced method makes groups to divide images into their relevant categories based on their characteristics and achieved maximum

classification accuracy 84.6% on the Caltech-101 dataset. Yongsheng et al. [108] described a decomposition technique, which is used for encryption of frequency and spatial information. using this method, break down input image into sub-regions using Spatial Pyramid Matching (SPM). The SIFT features are extracted from smaller regions in the initial stage and then by using codebook, the global features are extracted. Moreover, irrelevant features are reduced using K-means and obtained maximum classification accuracy 85.78% on the Caltech-101 dataset. Jongbin et al. [109] introduced a new Discrete Fourier transform based technique for feature building by discarding the max pooling or average pooling between fully connected layers and convolutional layer. The technique proceeded by two main modules. The first module known as DFT was performing replacement of max pooling from the architecture by a user-defined size pooling. The second module known as DFT+ was the fusion of multiple layers to get the best classification accuracy. They achieved 93.2% classification accuracy on the Caltech-101 dataset by applying VGG-16 deep CNN model and 93.6% accuracy on Caltech-101 using Resnet-50 model. Wen et al. [110] gave an idea of object detection using the standard Bag of Features (BoF). Principle Component Analysis (PCA) based reduction is employed on the final descriptor. Top 3000 features were selected for the classification purposes. Results showed the maximum accuracy of 72.36% but less time to classify images. Hamayun [111] proved that the most robust features are extracted from fully convolutional layer-6 (FC-6) instead of FC-8. Their proposed approach exploits the deep convolutional network output and modifies it at middle-level layer instead of deepest layer. VGG-16 and VGG-19 are used to illustrate the proposed technique. They extracted 4096 features from the FC-6 layer and then applied reduction using PCA. After multiple experiments, the results showed that they achieved a maximum accuracy of 91.35% using reduced features from FC-6 of VGG-19 on Caltech-101.

2.5. Feature Fusion

Feature fusion [112-117] is the concept of combining two different feature populations into a single feature space. There are two primary techniques for fusion including serial based fusion[118-120] and parallel based fusion [121, 122]. Since classification increases by fusing different features into single feature vector but the computation cost may increase gradually. To solve the computation cost problems, many feature reduction techniques were introduced.

Qing et al. [123] contributed to the area of image recognition by fusing the CNN features and then applying the three different types of coding techniques on to fused vector. VGG-M and VGG-16 models were used in this technique. After feature extraction from 5-Conv-Layer, PCA based reduction was applied and then the features were fused into a final vector using proposed coding techniques. Results showed that the introduced technique reported an improved accuracy of 92.54% by using their third coding technique on Caltech-101 database. Xueliang et al. [124] engineered a new technique for improved image recognition by late fusion technique of three pre-trained networks which are AlexNet, VggNet, and the ResNet-50. Firstly, they evaluated that middle layers of CNN architecture have more robust details for visual representation and then features were extracted from mid-level layers. Feature fusion from these three models showed improved result and reported 92.2% accuracy on Caltech-101 dataset. Yao et al. [125] introduced a novel idea for object recognition in which mid-level features were fused with the local features. In this technique mid-level features were collected from super-pixel based segmentation and local features were SIFT features. The proposed technique achieved 76.20% recognition accuracy on Caltech-101 database. Also multi-level feature fusion was introduced by [126] Qian which was consisting on three different levels of fusion.

Serial based feature fusion is also used from past few decades to get new matrix such as the newly added values are horizontally concatenated to a single feature vector. The serial based combination is a procedure of serial feature blend strategy, and the resultant feature vector is known as a serial melded feature. M. Nasir et al [127] also used serial based fusion in the domain of Dermoscopic classification tasks. They stated the equation which are as below.

$$f_{Fv1\times n} = \left\{ Fv_{1\times 1}, Fv_{1\times 2}, Fv_{1\times 3} \ldots Fv_{1\times n} \right\} \quad 2.1$$

$$f_{Fv2\times n} = \left\{ Fv_{2\times 1}, Fv_{2\times 2}, Fv_{2\times 3} \ldots Fv_{2\times n} \right\} \quad 2.2$$

$$f_{Fv3\times n} = \left\{ Fv_{3\times 1}, Fv_{3\times 2}, Fv_{3\times 3} \ldots Fv_{3\times n} \right\} \quad 2.3$$

$$Fused(FV)_{1\times q} = \sum_{i=1}^{3} \left\{ Fv_{1\times n}, Fv_{2\times n}, Fv_{3\times n} \right\} \quad 2.4$$

Parallel based feature fusion [112] is a strategy of feature vector creation using parallel feature combinations, and the final feature matrix is called parallel based fused feature. Jian et al.

proved that parallel based feature fusion performs well as compared to serial based feature fusion.

2.6. Classification

Classification is the process of differentiating/estimating data points into relevant categories. Jun et al. [18] presented a Group Sparse Deep Stack Network (GS-DSN) based image classification method. The presented method has two modules. First module acquires the interdependencies amongst hidden units by splitting them into amalgamate groups. In the second module, splitting image description into sub-groups to design for clustering of each sample and then gradient descent is used for estimation of weights. Thereafter, extract features using pre-trained CNN models like VGG and classified by GSNM module. The presented method is validated on a Caltech-101 dataset and obtained maximum classification accuracy 89%. Ridha et al. [128] introduced a DCNN based on multi-resolution model for object classification. The introduced method works on the base of NN design, fast wavelet transform (FWT), and AdaBoost algorithm (AA). The FWT is utilized for extraction of features based on the multi-resolution analysis (MRA), whereas the descriptor is calculated on a hidden layer. Thereafter, effective inputs are selected by the AA and design an Autoencoder by utilizing wavelet network of all input images. Finally, pooling is performed on hidden layers and achieved classification accuracy 75.60%. Shuangshuang et al. [129] stated a new sampling-based method for object classification. The objects are categorized into semantic groups using these sampling methods. Thereafter, a supervised dimensionality reduction approach is provided, which remove the irrelevant features and only select the best features for classification. Classification method by Saliency-based segmentation is employed in [130] shown in Figure 2.10.

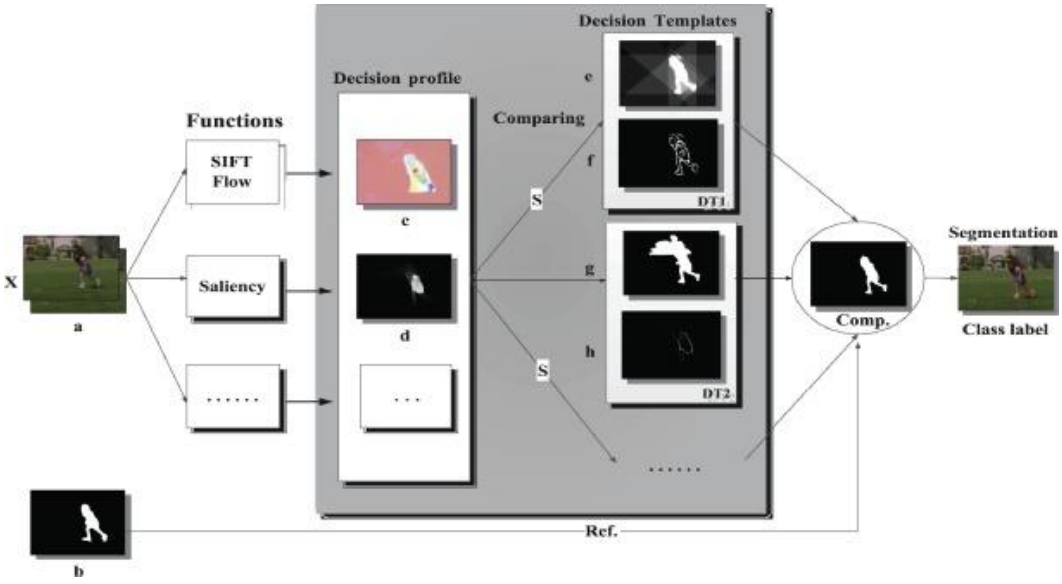


Figure 2.10 Classification technique proposed by [130] using Saliency-Based Segmentation

The introduced method is validated on the STL-10 dataset and achieved classification accuracy 67%. Moreover, recently introduced GLCM features based method for object classification. The extracted GLCM features are classified by Random forest and Naive based method, which gives classification accuracy 89% and 94.2%, respectively on the Caltech-101 dataset [131].

Zhang [130]. Chunjie et al. [132] gave an idea of the object recognition by using the three-component model of Object, Context, and Background (OCB). The technique followed by locating object, recognizing contextual area of the located object and then taking the remaining area as background. A fusion resultant feature vector of local features and the deep CNN features was generated and fed to the classifier. Proposed technique was evaluated on Caltech-101 dataset and achieved 81.68% accuracy with OCB-Sparse Coding technique and 95.5% by using OCB-CNN features technique. All results were achieved by 30 images/class as training purpose. Dongmei et al. [133] introduced a new technique for the overfitting problem caused by CNN models due to smaller datasets. The method was consisting of two modules. First module applied the transfer learning and second module applied the web data augmentation to solve the smaller dataset problem. AlexNet, VGG-16 and ResNet-152 models were used in it. Best accuracy achieved by this technique was 93.8% on Caltech-101 dataset by using ResNet model. A. Mahmood et al. [134] gave an idea to object detection and classification using pre-trained networks (ResNet-50 and ResNet-152). After features extraction, reduction was

applied using PCA. Caltech-101 database was selected to illustrate the proposed method. 30 training samples per class was used for training purpose. The results show that 92.6 % results were achieved using conventional features while 94.7% after optimization on Support Vector Machine classifier. Emine et al. [135] utilized Caffee implementation for object recognition purpose. X-ray imagery based object classification implementation is shown in Figure 2.11 is proposed by [136].

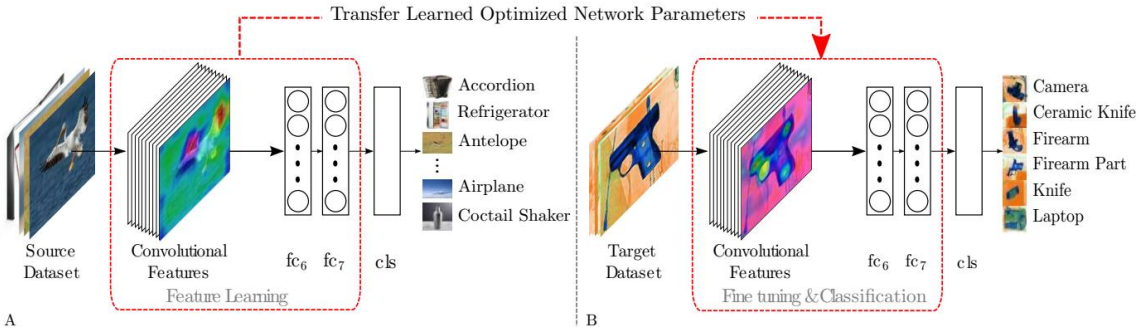


Figure 2.11 Classification Proposed in [136]

Caltech-101 dataset was worn to test the presented technique. For results, 300 images were tested from whole dataset and 260 of them were correctly classified and 40 were misclassified. They achieved almost ~86.0% recognition accuracy. The all above discussed methods focused on DCNN methods for object detection and classification, however, these methods do not work well when it deals with large datasets which include hundreds of classes. we noticed that the above studies never be performed preprocessing step to improve the classification accuracy. The preprocessing step for in the initial step is important due to background factors such as illumination. Moreover, we noticed that in [105] [108] SIFT features are extracted from input images but they achieved maximum accuracy 85.78% on the Caltech101 dataset, which does not show improved performance. To inspire with above methods, in this research we introduced a model which is depending on the fusion of DCNN and SIFT point features. Moreover, an entropy-controlled best feature selection method is also implemented.

2.6.1 Classification Algorithms

There is plenty of classification algorithms accessible now [137], however, it is unimaginable to expect to finish up which one is better than other. It relies upon the application and nature of the accessible informational collection. For instance, if the classes are straightly divisible,

the direct classifiers like a Logistic relapse, Fisher's straight discriminant can beat refined models and the other way around.

Decision tree [138-140] constructs classification or relapse models as a tree structure. It uses an on the off chance that standard set which is fundamentally unrelated and comprehensive for classification. The tenets are found out successively utilizing the preparation information each one in turn. Each time a standard is found out, the tuples secured by the principles are evacuated. This procedure is proceeded on the preparation set until meeting an end condition. Tree is built in a best down recursive partition and-overcome way. Every one of the characteristics ought to be all out. Else, they ought to be discretized ahead of time. Traits in the highest point of the tree contains more effect in classification and they are recognized utilizing the data gain idea.

Naive Bayes [141] is a classifier which is inspired by the Bayes theorem keeping an assumption whose attributes are independent conditionally [142].

$$F\left(X \mid Bb_j\right) = \prod_{k=1}^{n} Fv\left(x_k \mid Bb_j\right) = Fv\left(x_1 \mid Bb_j\right) \times Fv\left(x_2 \mid Bb_j\right) \times ..... \times F\left(x_n \mid Bb_j\right) \quad [142]\ 2.5$$

The classification is led by determining the greatest back which is the maximal P(Ci|X) with the above presumption applying to Bayes hypothesis. This presumption incredibly lessens the computational expense by just checking the class dispersion. Despite the fact that the presumption isn't legitimate as a rule since the characteristics are needy, shockingly Naive Bayes has ready to perform stunningly. Naive Bayes is an extremely straightforward calculation to actualize and great outcomes have acquired by and large. It very well may be effectively versatile to bigger datasets since it requires direct investment, instead of by costly iterative estimation as utilized for some different kinds of classifiers.

k-Nearest Neighbor [143] is an apathetic learning calculation which stores all occurrences relate to preparing information focuses in n-dimensional space. At the point when an obscure discrete information is gotten, it dissects the nearest k number of examples spared (closest neighbors)and restores the most well-known class as the expectation, and for genuine esteemed information, it restores the mean of k closest neighbors. Out there weighted closest neighbor calculation, it weights the commitment of every one of the k neighbors as indicated by their

separation utilizing the accompanying inquiry giving more prominent weight to the nearest neighbors.

Support Vector Machines [144], are a standout amongst the most incredible classifiers. Right off the bat, amid preparing the choice limit with the greatest edge is chosen. This guarantees the unmistakable detachment of the precedents and lessens the classification blunder when the classes are detachable. Furthermore, piece strategies utilize bit works that empower examination between features in a higher dimensional (perhaps vast dimensional) feature space. This dispenses with the requirement for direct detachability which does not generally hold. Thirdly, the idea of a delicate edge allows a few exceptions by presenting a heedlessness term in the objective capacity.

2.6.2 Neural Network

Artificial Neural Networks [145] proposed to emulate neural system from the mind by utilizing a streamlined scientific model called "edge rationale.". The system is worked from basic measures called perceptron having different sources of info and a solitary yield. The yield is enacted when a direct blend of the info outperforms a specific edge. A key advancement in this domain was the presentation of the recursion calculation which permitted the preparing of multiple layered network that could take in more mind-boggling issues, (for example, the xor work). System proposed by [146] detected the 2D shapes of objects by using local curvature with aspect of patches and then by applying the fuzzy logic on them which is presented in Figure 2.12.

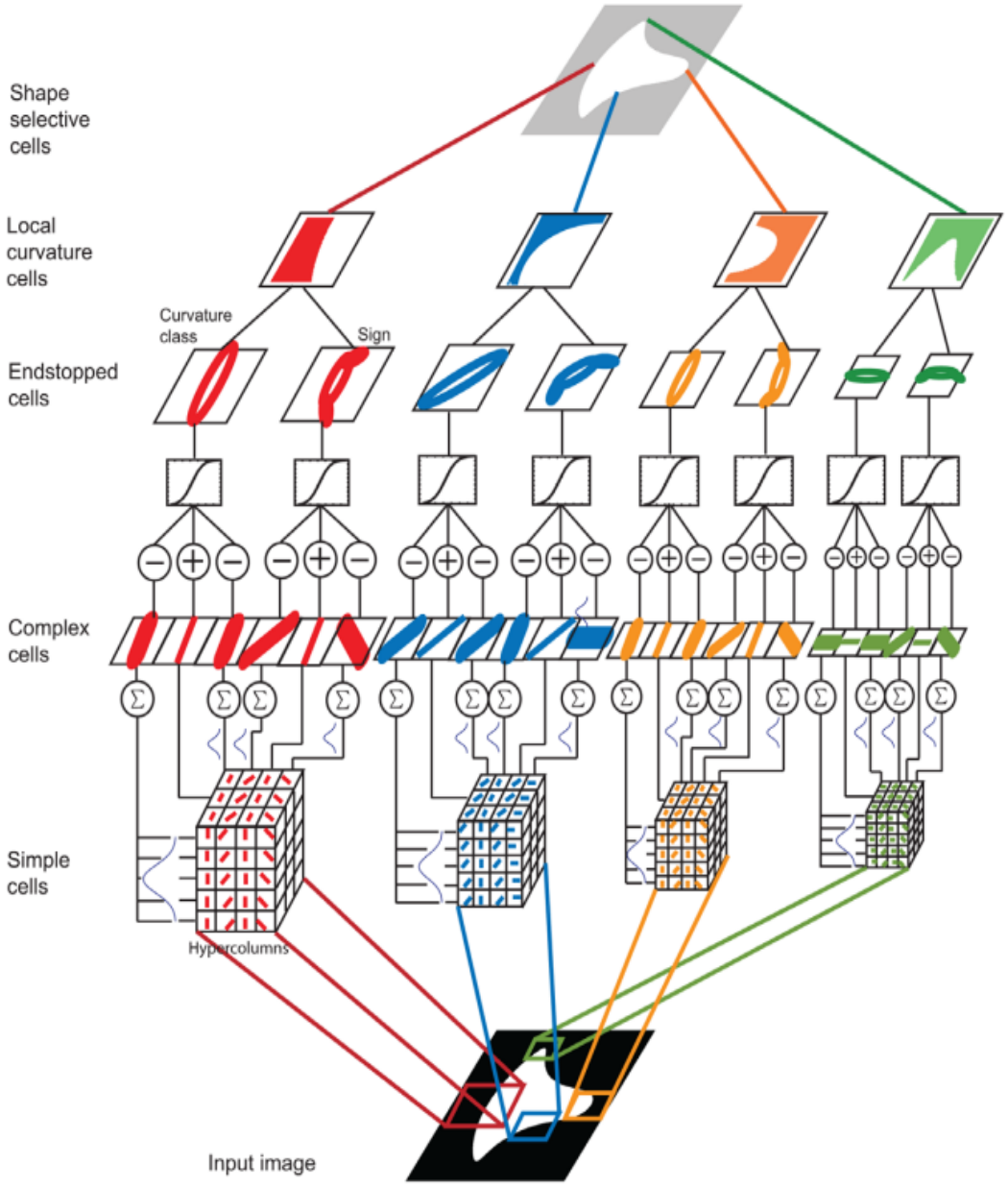


Figure 2.12 2D contour-based object estimation Proposed by [146].

2.7. Datasets

The previous techniques were validated on multiple publically available datasets such as Caltech 101 [92], Birds [147] and Butterflies [148], PASCAL 3D+ [149], and Barkley 3D [150] dataset. Statistical information of five publicly available datasets is presented in Table 2.1.

Table 2.1 Available Dataset Statistics

Dataset	Classes	Total Images	Range
Caltech-101 **[92]**	101	9144	31-800
Birds **[147]**	6	600	100-100
Butterflies **[148]**	7	619	42-134
Pascal3D+ **[149]**	12	22394	536-6704
Barkley 3D Dataset **[150]**	10	6604	474-721

Many Others datasets are also available like Caltech-256 and more. Table 2.1 shows that the Caltech-101 dataset is having maximum classes among other datasets.

2.7.1 Caltech-101

The Caltech-101 [92] database consists of 102 distinct object classes having 9144 images. Each class consists of approximately 31~800 images. However, this dataset consists of both RGB and gray images, which is a major issue of this dataset. Because, if objects are recognized by their color, then color features are not performed well on grayscale images. The three sample images from each class are displayed in Figure 2.13.

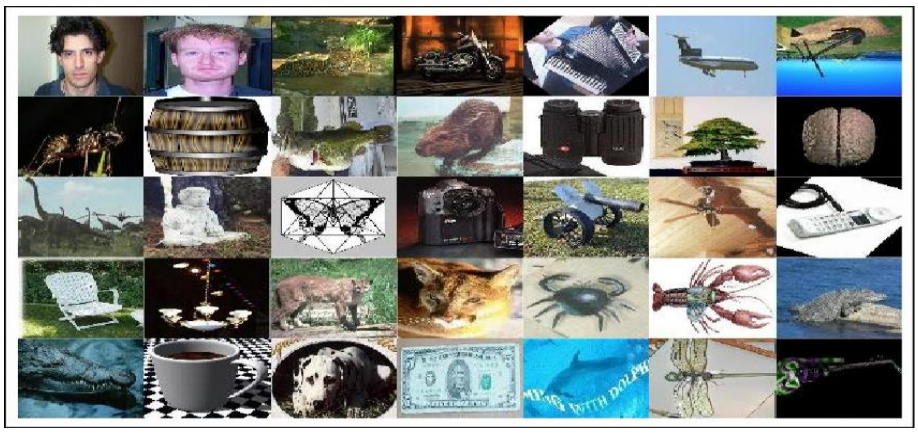


Figure 2.13 Sample images from Caltech-101 [92] (3 images per class)

2.7.2 Pascal 3D+

Pascal 3D+ dataset [149] is another challenging database which is used for object classification. This dataset is the combination of Pascal VOC 2012 and ImageNet. It contains

total 22394 images of 12 unique classes. The classes which are common between PASCAL VOC 2012 and ImageNet are merged into a new database, called Pascal 3D+. The dataset was initially collected from flicker website. The three sample images from each class are shown in Figure 2.14.

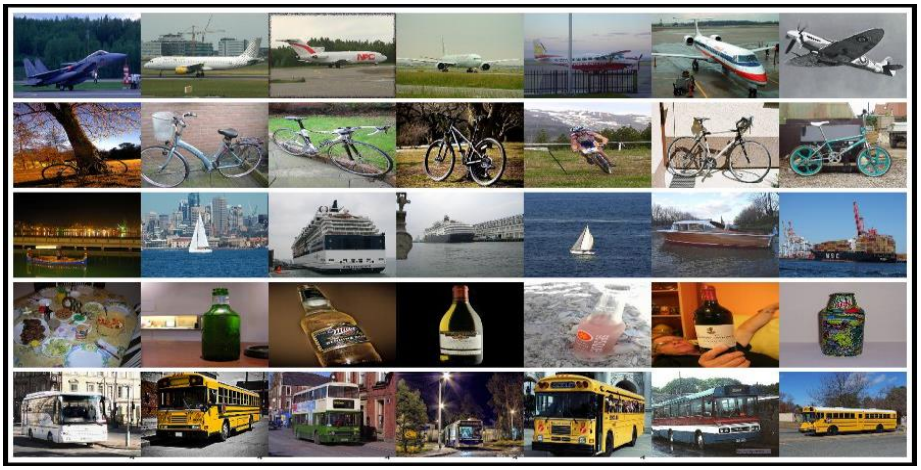


Figure 2.14 Sample Images from Pascal3D+ [149] (3 images per class)

2.7.3 Barkley 3D

Barkley 3D object dataset [150] consists of total 6604 images of 10 categories including bicycle, car, cellphone, head, iron, mouse ,monitor, shoe, stapler, and toaster. The number of images in each class range of 474-721. The seven sample images from each class are shown in Figure 2.15.

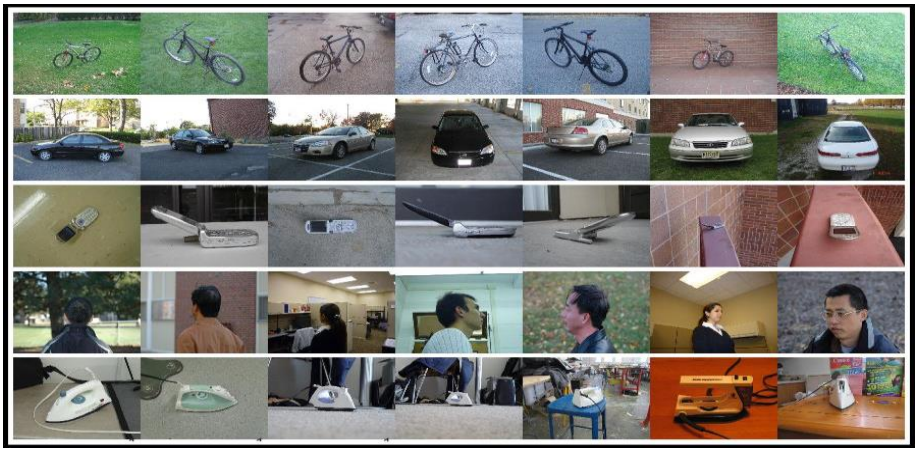


Figure 2.15 Images from Barkley 3D [150] dataset (7 images per class)

The dataset was initially introduced to recognize the 3D objects in video frames and further annotated.

2.7.4 Birds

Birds [147] dataset is also another publicly available dataset used for challenge of object recognition. This database has total images of 600 from six different types of classes of birds. Minimum and maximum number of images in all classes are the same which is hundred. The seven sample images from each class are displayed in Figure 2.16.

Figure 2.16 Images from Birds [147] database (7 images per class)

2.7.5 Butterflies

Butterflies [148] is a publically available dataset. Dataset is having 7 different classes with 619 images/class having 42 images as minimum and 134 as maximum images. The seven sample images from each class are shown in Figure 2.17.

Figure 2.17 Sample Images from Butterflies [148] (7 image per class)

Details of all the above-mentioned datasets is presented in tabular form in Table 2.1 below which includes the number of images, number of classes and category description of dataset.

2.8. Evaluation

Distinctive classification results can be spoken to in confusion matrices, for example, the general one spoke to in Table 2.2. The capacity of the grid is to demonstrate the quantity of things that were arranged accurately (TP and TN) and dishonesty (FP and FN). The lattice in Table 2.2 is a disarray grid for a two-class classification task. The total of the quantity of grouped things in each cell is the aggregate number of the arranged things. Helpful estimation rates can be figured from the numbers in the cells of the confusion matrices:

Table 2.2: Confusion Matrix

		Predicted Class	
		Yes	No
True Class	Yes	True Positive (TrPo)	False Negative (FaNe)
	No	False Positive (FaPo)	True Negative(TrNe)

$$FNR = \frac{FaNe}{Po} = \frac{FaNe}{FaNe + TrPo} \qquad 2.6$$

$$Accu = \frac{TrPo + TrNe}{Po + Ne} = \frac{TrPo + TrNe}{TrPo + TrNe + FaPo + FaNe} \qquad 2.7$$

Where Po represent positive cases, Ne is the number of negative outcomes in the data, also error rate and accuracy rate sum to 1.

The error rate or the accuracy rate measures how well a model effectively orders things. When preparing the model, one tries to get the accuracy rate as high as would be prudent and picks the model with the best precision on the approval set.

Accuracy measures the rightness of classification and review measures its handiness. For instance, in spam messages it is vital that the accuracy rate is very high, in light of the fact that a client would be irritated if a vital email were to be characterized erroneously as spam and hence left new. It is less irritating to get some spam into one's Inbox occasionally. It is up to

every classification task to choose whether it is more hurtful to have False Negatives than False Positives. In the event that False Negatives are not as hurtful as False Positives, at that point the choice model would incline toward high accuracy rate over a high review rate. This applies additionally the different way: Positives are not as unsafe as False Negatives, a high review rate is of significance as opposed to a high accuracy rate.

Chapter 3

Proposed work

3. Overview of Work

Object recognition is a meaningful work in the domain of artificial intelligence and it gains much attention from last two decades based on their useful implementation like video surveillance and pedestrian detection. In this research, we deal with complex object detection and classification using three famous datasets such as Caltech101, PASCAL 3D, and 3D dataset. These datasets contain hundreds of object classes and thousands of images. To instigate with these datasets challenges, we introduced a novel technique for object recognition and classification depending on DCNN features extraction along with SIFT points. The proposed framework consists of two advanced steps, which are executed using parallel stream. In the first step, SIFT point features are extracted from mapped RGB segmented object. Secondly, DCNN features are obtained from pre-trained DCNN models like AlexNet and VGG. The both SIFT point and DCNN features are combined into a one matrix by a fusion method, which will be later employed for classification. The description of each step is explained below in section 3.2 to 3.4. Also, workflow of presented method is displayed in the Figure 3.1.

3.1. Improved Saliency Method

An improved saliency method is employed by utilizing existing saliency approach name HDCT, for single object detection. In this step, we extracted a single object from an image by an existing saliency method namely HDCT saliency estimation. The idea behind the improvement of saliency method is to implement the color spaces before gives the input image to saliency method. The LAB color transformation is utilized for this purpose, which identifies color in 3 dimensions consisting of L for lightness, A and B are used for color ingredient, green-red and blue-yellow The three components L is brighter white at 100 and darker black at 0, whereas ‘A’ and ‘B’ channels show the natural values for the RGB image. By Figure 3.1 proposed methodology workflow is described. This transformation is defined as follows:

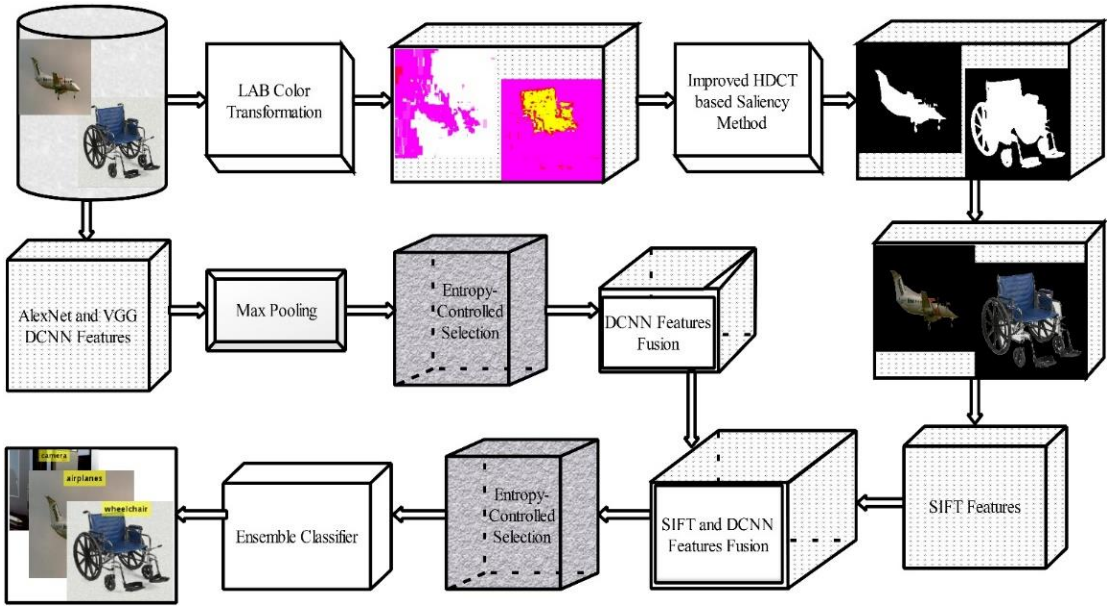


Figure 3.1 Flow diagram of proposed object classification method [151].

Let U(i, j) is an RGB image with length N × M, then for RGB to LAB conversion, first RGB to XYZ conversion is performed as:

$$\begin{bmatrix} \varphi(P) \\ \varphi(Q) \\ \varphi(R) \end{bmatrix} = [M \times N] \begin{bmatrix} \varphi^r \\ \varphi^g \\ \varphi^b \end{bmatrix} \qquad 3.1$$

Where, φ(X), φ(Y), and φ(Z) denotes the P, Q, and R channels, which are extracted from red (φ^r), green (φ^g), and blue channel (φ^b). The φ^r, φ^g, and φ^b channels are defined as:

$$\varphi^r = \sum_{t=1} \frac{\varphi t}{\triangle_t}, t = \mathrm{Re}\, d \qquad 3.2$$

$$\varphi^g = \sum_{t=2} \frac{\varphi t}{\triangle_t}, t = Green \qquad 3.3$$

$$\varphi^b = \sum_{t=3} \frac{\varphi t}{\triangle_t}, t = Blue \qquad 3.4$$

Then LAB conversion is defined as:

$$\left(\varphi^L = \beta_1 \times \left(f_y - 16\right)\right), \beta_1 = 116 \qquad 3.5$$

$$\left(\varphi^{*A} = \beta_2\left(f_x - f_y\right)\right), \beta_2 = 500 \qquad 3.6$$

$$\left(\varphi^{*B} = \beta_3\left(f_y - f_z\right)\right), \beta_3 = 200 \qquad 3.7$$

Where, f_x, f_y, and f_z are linear functions which are computed as:

$$f_x = \left\{\sqrt[3]{x_r} \mid \frac{kx_r + 16}{116}, \rightarrow x_r > \in \mid otherwise\right\}, x_r = \frac{X}{Xr} \qquad 3.8$$

$$f_y = \left\{\sqrt[3]{y_r} \mid \frac{ky_r + 16}{116}, \rightarrow y_r > \in \mid otherwise\right\}, y_r = \frac{Y}{Yr} \qquad 3.9$$

$$f_z = \left\{\sqrt[3]{z_r} \mid \frac{kz_r + 16}{116}, \rightarrow z_r > \in \mid otherwise\right\}, z_r = \frac{Z}{Zr} \qquad 3.10$$

Thereafter, we employed a saliency approach for salient object detection. Salient region detection technique detects the salient region from an image by utilizing high dimensional color transform. In this work, the superpixel saliency features are used to identify the initial salient sections of the dermoscopic images. The superpixels of the LAB image are given as:

$$Y = \left\{p_1, \ldots . p_N\right\} \qquad 3.11$$

For low computational cost and better performance, we utilized the SLIC superpixel [152] with a total number of superpixels N=400. The color features are computed from LAB color space. The parameters which are used for color features extraction from LAB color space are mean, variance, standard deviation, and skewness. These color features are concatenated with the histogram features because the histogram features are effective for saliency approach. The Euclidean distance is calculated between extracted color features as:

$$\vec{D} = \vec{D}\left(A\right) = \| l_i - l_j \|_2^2 \qquad 3.12$$

Where, l_i and l_j denote the ith and jth features in the given matrix A. In this work, the global contrast/color statistics of objects are used to define the saliency values of the pixels by using a histogram-based method. The saliency amount of pixel defined as:

$$S(\varphi_k) = \sum_{\forall \varphi_i \in I} \vec{D}(A) \qquad 3.13$$

Where $\vec{D}(A)$ is the color distance between the features l_{i_i} and the l_j in the LAB color space. By rearranging the above equation, we get the saliency value for each color as:

$$S(\varphi_k) = \sum_{l=1}^{n} f_l D(c_j, c_l) \qquad 3.14$$

Where n, c_j, f_l denotes the total number of the different pixel color, the color value of pixel φ_k, and the frequency of the pixel color respectively. For shape and texture features, we utilized the HOG and the SFTA texture features. After the calculation of feature vector for every superpixel, the random forest-based regression is used to estimate the salient degree of each region. Further to recognize the very salient regions calculated from basic saliency map the Trimap is constructed by using adaptive thresholding. First, the input images divided into 2x2, 3x3, and 4x4 patches and then apply the Otsu thresholding on each patch individually. Finally, the Trimap is obtained by using global thresholding as:

$$Th(i) = \begin{cases} 1 \rightarrow Th(i) \geq \tau \\ 0 \rightarrow Th(i) \leq \tau \\ unknown...else \end{cases} \qquad 3.15$$

Where τ denotes the global threshold value. After getting the optimal coefficient α (estimate for the saliency map) manages saliency map as:

$$Sal_{fi}(X_u) = \sum_{u=1}^{n} K_{uv}\alpha_v, u = 1, 2, ..., N \qquad 3.16$$

Where K shows the enriched vector to present the color for input image. The final map is obtained by adding the color saliency map (CSM) and spatial saliency map (SSM) as:

$$Sal_{fi}(X_u) = Sal_{fi}(X_u) + S_S(X_u), u = 1, 2, ..., N \qquad 3.17$$

The SSM is illustrated as:

$$S_S(X_i) = \exp\left(-K \frac{\min_j \in f\left(d\left(P_i, P_j\right)\right)}{\min_j \in \beta\left(d\left(P_i, P_j\right)\right)}\right) \qquad 3.18$$

Where the K= 0.5, and $\min_j \in \beta\left(d(P_i, P_j)\right)$ and $\min_j \in f\left(d(P_i, P_j)\right)$ are the Euclidian distance from the ith pixel to definite background pixel and to definite foreground pixel respectively. The improved saliency method effects are displayed in Figure 3.2. In Figure 3.2, the 1st row shows input images, second rows present LAB transformation, third row defines improved saliency image in a binary form, and the last row depicts the mapped RGB image.

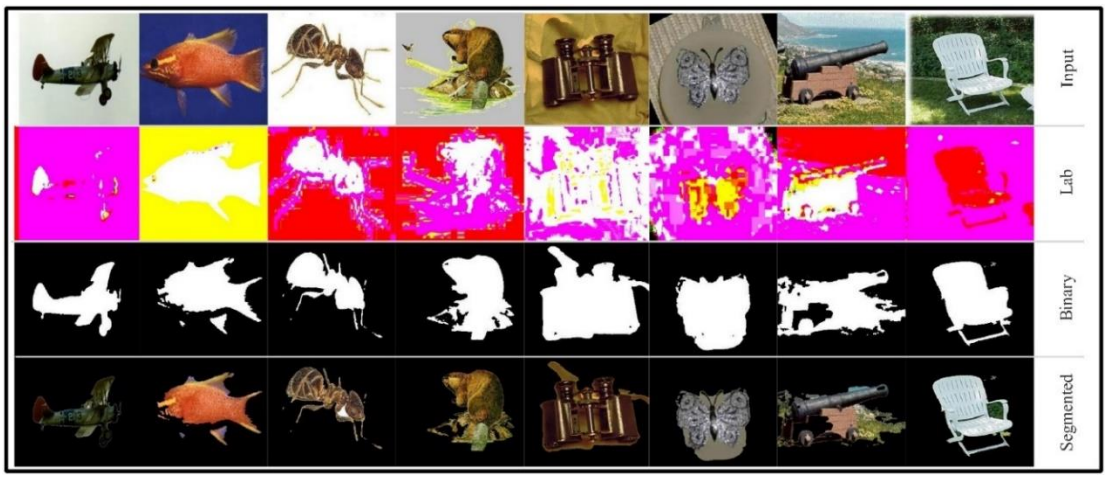


Figure 3.2 Proposed improved saliency method results [151]

3.2. SIFT Features

SIFT is originally designed in 2004 by [153] and have appeared as a strength descriptors for object detection and recognition. The SIFT features are computed in four steps. In the first step, local key points are determined that are important and stable for given images. Then features are extracted from each key point that explains the local image region samples, which are related to its scale space coordinate image. In the second step, weak features are discarded by a specific conditional value. Thirdly, orientations are allocated to each pixel depending on local gradient rotations. Finally, 1×128 dimensional feature vector is obtained and perform bi-linear interpolation to improve the robustness of features. The above theory is defined as follows:

$$\xi(\mu,\nu,\sigma) = \psi_G(\mu,\nu,\sigma) \otimes S_{final}(X_i) \quad 3.19$$

$$\psi_G(\mu,\nu,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\left(\frac{\mu^2+\nu^2}{2\sigma^2}\right)} \quad 3.20$$

$$D(\mu,\nu,\sigma) = \left(\psi_G(\mu,\nu,k\sigma) - \psi_G(\mu,\nu,\sigma)\right) \otimes S_{final}(X_i) = \quad 3.21$$
$$\xi(\mu,\nu,k\sigma) - \xi(\mu,\nu,\sigma)$$

Where, $\xi(u, v, \sigma)$ is scale space of an image, $\psi_G(u, v, k\sigma)$ denotes the variable-scale Gaussian, k is a multiplicative factor, and $D(u, v, \sigma)$ denotes the difference of Gaussian obtained by a segmented output.

3.3. Deep CNN Features

Recently, in the domain of computer vision, machine, and pattern recognition, deep learning have shows improved performance for image classification on large datasets [154]. The deep learning designs such as deep CNN and recurrent NN have been employed to human action recognition, speech recognition, document classification, agricultural plants, medical imaging, and many more and shows superior performance. In object classification, CNN shows much attention due to their ability to automatically determine appropriate contextual features in image categorization problems. A simple CNN model consists of four types of layers. Initially, an input image is passed and computes its neurons by convolution layer, which are connected to local regions of the input. Th each neuron are computed by dot product between their small regions and weights, which are connected to in the input volume. Thereafter, activation is performed using ReLu layer. The ReLu layer never changed the size of an input image. Then, pooling layer is performed to downsample the noise effects in the obtained features. Finally, high-level features are calculated by fully connected (FC) layer.

In this article, we employed two DCNN models which are VGG19 and AlexNet, which are utilized for feature descriptor building. These models incorporate convolution layer, pooling layer, normalization layer, ReLu layer, and FC layer. As discussed above that convolution layer extract local features from an image, which is formulated as:

$$g_i^{(L)} = b_i^{(L)} + \sum_{j=1}^{m_1^{(L-1)}} \psi_{i,j}^{(L)} \times h_j^{(L-1)} \quad 3.22$$

Where, $g_i^{(L)}$ denotes the output layer L, $b_i^{(L)}$ is base value, $\psi_{i,j}^{(L)}$ denotes the filter connecting the jth feature map, and h_j denotes the $L-1$ output layer. Then, pooling layer is defined which extract maximum feedback from end convolutional having a motive of downsampling irrelevant descriptors. The max pooling also resolves the problem of overfitting and mostly 2×2 polling is performed on extracted matrix. Mathematically, max pooling is illustrated as:

$$o_1^{(K)} = o_1^{(K-1)} \quad 3.23$$

$$o_2^{(K)} = \frac{o_2^{(K-1)} - F(K)}{S^K} + 1 \quad 3.24$$

$$o_3^{(K)} = \frac{o_3^{(K-1)} - F(K)}{S^K} + 1 \quad 3.25$$

Where, S^K denotes the stride, $o_1^{(K)}$, $o_2^{(K)}$ and $o_3^{(K)}$ are defined filters for feature map such as 2×2, 3×3. The other layers such as ReLu and fully connected (FC) are defined as:

$$\mathrm{Re}_i^{(l)} = \max\left(h, h_i^{(l-1)}\right) \quad 3.26$$

$$Fc_i^{(l)} = f\left(z_i^{(l)}\right) with\ z_i^{(l)} = \sum_{j=1}^{m_1^{(l-1)}} \sum_{r=1}^{m_2^{(l-1)}} \sum_{s=1}^{m_3^{(l-1)}} w_{i,j,r,s}^{(l)} \left(Fc_i^{(l-1)}\right)_{r,s} \quad 3.27$$

Where, $\mathrm{Re}_i^{(l)}$ denotes the ReLu layer, $\mathrm{Fc}_i^{(l)}$ denotes the FC layer. The FC layer follows the convolution and pooling layers. The FC layer is similar to convolution layer and most of the researchers perform activation on FC layer for deep feature extraction.

3.3. Pre-Trained Deep CNN Networks

In this research, we used two DCNN networks that are VGG and AlexNet as deep features codebook. AlexNet DCNN model designed by Krizhevsky et al. [37] using ImageNet dataset. This network contains 5 convolution layers, three pooling layers, and 3 FC layers along with softmax classification function. This network trained on input image size 227x227x3.

VGG-19 CNN network is proposed by Zisserman et al. [154] which contains 16 convolution layers, 19 learnable weights layers, 3 FC layers along with softmax function. This network trained on ImageNet dataset and shows exceptional performance. The size of training input images is 227x227x3.

3.4. Features Extraction and Fusion

Our proposed feature extraction and fusion strategy is shown in this section. The features are obtained from pre-trained deep CNN models by different number of layers. In this work, two pre-trained models are used such as VGG19 and AlexNet for features extraction. The major aim of deep CNN features extraction from two models is to improve the classification accuracy. Because each model has distinct characteristics and gives different features. Therefore, using this advantage we extract features by performing activation on the FC7 layer and perform max pooling to remove the noise factors. Thereafter, an entropy-controlled method is implemented for best feature reduction. The proposed feature extraction and reduction architecture are shown in Figure 3.3.

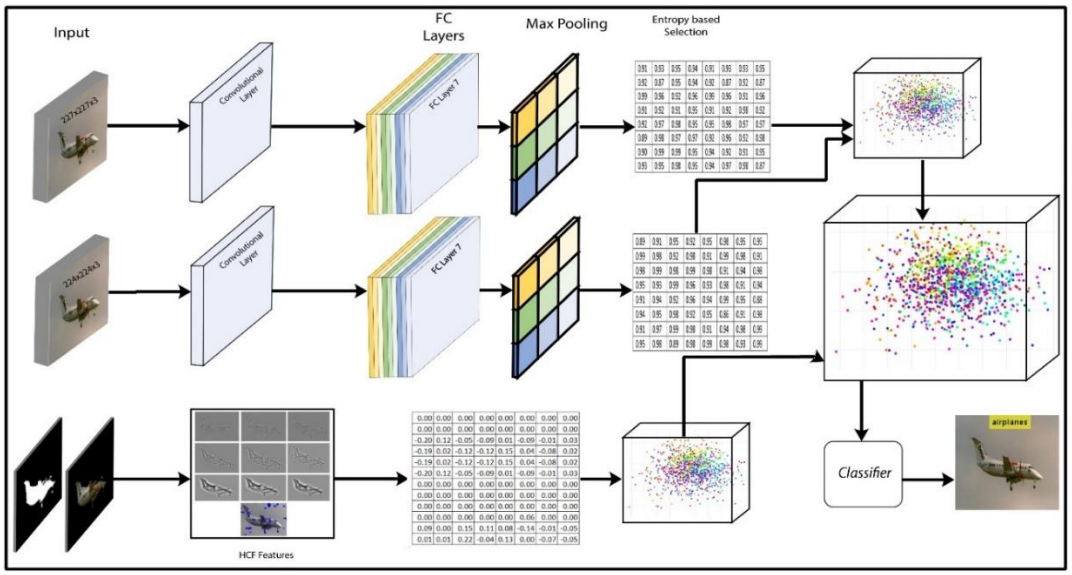


Figure 3.3 Proposed deep CNN and SIFT features fusion and reduction method for object classification [151]

As shown in Figure 3.3, three types of features are obtained like AlexNet deep CNN, VGG19 CNN. For AlexNet and VGG19, convolution layer is employed as an input layer. Then perform activation on FC7 layer for both networks to extract deep CNN features. The size of deep CNN

features for output layer FC7 is 1×4096 for both networks. The feature size of both output layer is higher, therefore we performed pooling of filter size 2×2, which removes the noise effects and select the maximum value feature of given filter. Example of max-pooling is drwan in Figure 3.4.

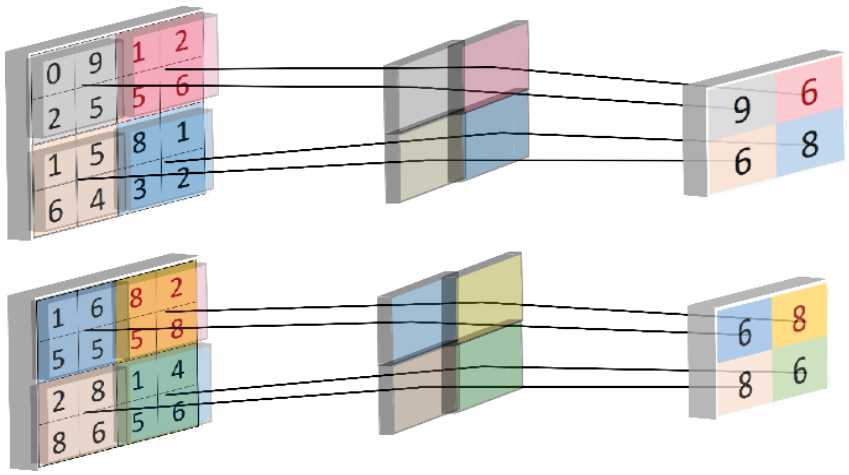


Figure 3.4 An example of max-pooling operation

After max-pooling, the new feature vectors of size 1×2048 are obtained, which is further improved by entropy-controlled feature reduction method. As extracted feature vectors can produce better results but they increase the execution time. Therefore, our focus is to enhance the accuracy and decrease the execution time. This sort of problem is resolved by an entropy controlled method. The entropy gives the knowledge about flexibilities in a signal by showing the flow disorder [155]. Due to its capacity to describe system behavior, entropy gives the noticable information which may be employed in descriptor design [156]. Amongst several, we used Renyi-entropy based feature reduction. Amongst many, the Reyni entropy method is applied for feature reduction. In the circumstances of fractal dimension estimation, the Reyni entropy method determines the basis of the theory of generalized dimensions. The fractal dimension estimates the change patterns of the given feature space. The Reyni entropy is defined as follows:

Let $f_1, f_2, \ldots, f_n$ denotes the A feature space after max-pooling, $g_1, g_2, g_3, \ldots g_n$ denotes the B feature space after max-pooling, and $\xi_1, \xi_2, \ldots, \xi_n$ denotes the ξ feature space, where $A \in$ AlexNet DCNN features, $B \in$ VGG19 DCNN features, and $\xi \in$ SIFT point feature vector,

where the dimension of each features space is 1×2048, 1×2048, and 1×128. The entropy is formulated as:

$$E_{\alpha}(X) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^{a} \right) \qquad 3.28$$

Where, $\alpha \geq 0 \,\&\, < 1$, $X \in (f_n, g_n, \xi_n)$, and p_i denotes the probability value of extracted feature space A, B and ξ which is defined by $p_i = Pr(X = i)$ and denotes the length of all feature spaces. The entropy function gives a new $N \times M$ feature vector, which controls the randomness of each feature space. Then, sort each $N \times M$ feature vector into ascending order and select the top 1000 features from A and B vectors and 100 features from ξ vector.

$$E(A) = \Phi(f_n, \varrho),\ E(B) = \Phi(g_n, \varrho),\ E(\xi) = \Phi(\xi_n, \varrho) \qquad 3.29$$

Where, $E(A)$ denotes the entropy information of feature space A, $E(B)$ denotes the entropy information of feature space B, $E(\xi)$ denotes the entropy information of feature space ξ, Φ denotes sorting function, and ϱ denotes the ascending order operation. Thereafter, fused both $E(A)$ and $E(B)$ entropy information features in one matrix by the simple serial based method, which returns a feature vector of size 1×2000, which further fused with SIFT point feature by the serial-based method as shown in the above Figure 3 and below expression:

$$\prod(Fused) = (N \times 1000) + (N \times 1000) + (N \times 100) \qquad 3.30$$

$$\prod(Fused) = N \times f_i \qquad 3.31$$

The size of the final feature vector is 1×2100, which feed to ensemble classifier for classification. The ensemble classifier is a supervised learning method, which needs to training data for prediction. Ensemble method combines several classifiers data to produce a better system. The formulation of ensemble method is given below.

Let we have extracted features and their corresponding labels $((f_1, y_1), (f_2, y_2), \ldots, (f_n, y_n))$, where f_i denotes the extracted features which are typically vectors of form $(f_{i+1}, f_{i+2}, \ldots, f_{i+n})$, then the unknown function is defined as $y = f(x)$. An ensemble classifier is a bunch of

classifiers whose separate kernels are merged to a classifier by typical weights and voting. Hence the ensemble classifier is formulated as:

$$Y = Sign\left(\sum_{k=1}^{K} w_k l_k(x)\right) \qquad 3.32$$

Where $l_k(m) = l_1(m), l_2(m), ..., l_k(m)$ and $\hat{w}_k = \hat{w}_1, \hat{w}_2, \ldots \hat{w}_k$

The proposed method is tested on five datasets such as Caltech101, PASCAL 3D+ dataset, 3D dataset, Birds dataset and Butterflies dataset. The sample labeled outputs are shown in the Figure 3.5 and 3.6.

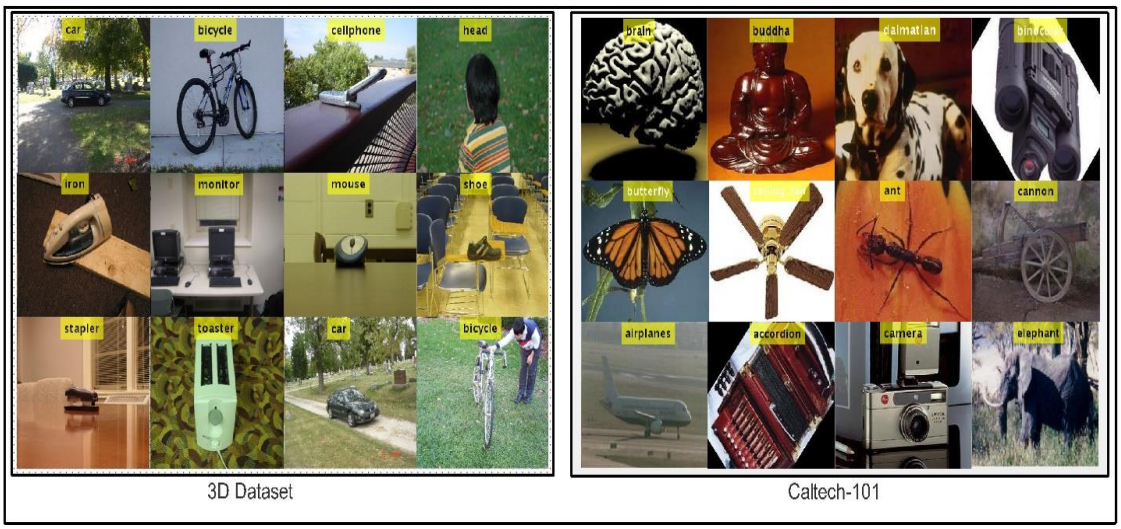


Figure 3.5 Proposed labeled classification results for 3D dataset and Caltech101 dataset.

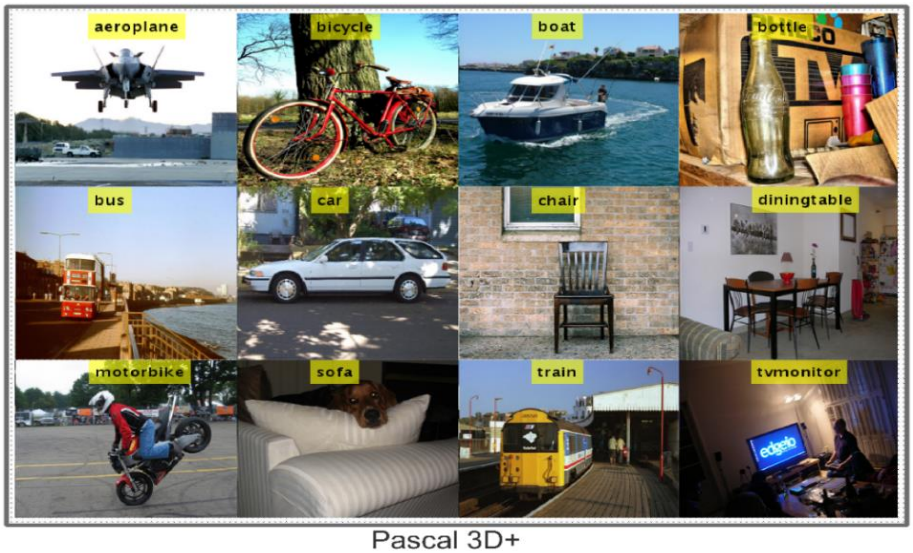


Figure 3.6 Proposed labeled classification results for PASCAL 3D+ dataset.

Chapter 4

Experimental Results and Analysis

4. Experimental Results and Analysis

In this chapter, proposed experimental results are presented in the form of numerical and graphical plots. For classification, we used Ensemble classifier and also test its performance with Linear SVM, Quadratic SVM, Cubic SVM, Fine KNN, Cubic KNN, decision tree, and weighted KNN. Each classifier's performance is calculated by three measures including accuracy, FNR, and execution time. All results are evaluated on 3.4 Gigahertz Corei7 7th generation Laptop with a RAM of 16 Gigabytes and a GPU of Radeon R7 (4GB, 128 bit) having Matlab-2018a.

4.1. Experimental Results

The proposed method is validated on five publically available datasets such as Caltech 101, PASCAL 3D+, Barkley 3D dataset, Birds dataset and butterflies dataset. The Caltech-101 [92] dataset consists of total 102 distinct object classes of 9144 images. Each class consists of approximately 31~800 images. However, this dataset consists of both RGB and gray images, which is a major issue of this dataset. Because, if objects are recognized by their color, then color features are not performed well on grayscale images. Pascal 3D+ dataset [149] is another challenging database which is used for object classification. This dataset is the combination of Pascal VOC 2012 and ImageNet. It contains total 22394 images of 12 unique classes. The classes which are common between PASCAL VOC 2012 and ImageNet are merged into a new database, called Pascal 3D+. Barkley 3D object dataset [150] consists of total 6604 images of 10 object classes including bicycle, car, cellphone, head, iron, monitor, mouse, shoe, stapler, and toaster. The number of images in each class range of 474-721. Birds [147] dataset is also another publicly available dataset used for challenge of object recognition. This database has total images of 600 from six different types of classes of birds. Minimum and maximum number of images in all classes are the same which is hundred. Butterflies [148] is a publically available dataset. Dataset is having 7 different classes with 619 images per class with a minimum of 42 images per class and a maximum of 134.

4.1.1 Classification Results on Caltech-101 Dataset

In this section, we discussed detailed results of our method on selected datasets. For classification results on Caltech-101 dataset, we define multiples experiments on distinct classes such as 20, 34, 50, and 102. The experiments are a) classification of selected classes

using AlexNet DCNN features with entropy-based selection method; b) Classification of selected classes using a VGG-19 DCNN model with entropy-based features selection; c) Fusion of deep CNN and SIFT features along with entropy-controlled selection method.

Table 4.1: Summary of Experiments on Caltech-101 dataset

Model	Experiments	Classes				Classifier	Performance Measures		
		20	34	50	102		Accuracy	FNR	Time
AlexNet	Experiment #1	✓				Ensemble Boosted Tree	86.5	13.5	105.00
	Experiment #2		✓				84.6	15.4	172.63
	Experiment #3			✓			83.5	17.0	193.00
	Experiment #4				✓		71.7	28.3	620.42
VGG-19	Experiment #5	✓					92.0	8.0	88.129
	Experiment #6		✓				**87.5**	12.5	198.480
	Experiment #7			✓			86.0	14.0	168.660
	Experiment #8				✓		73.8	27.2	454.270
Proposed	Experiment #9	✓					**86.5**	13.5	75.70
	Experiment #10		✓				**93.8**	6.2	**289.90**
	Experiment #11			✓			**93.5**	6.5	178.20
	Experiment #12				✓		**89.7**	10.3	302.50

The classification is performed on each class and finally compared the performance of all 102 classes regarding accuracy and execution time with 20, 32, and 50 number of classes. Table 4.1 shows the overview to the experiments made on Caltech-101 dataset.

4.1.1.1 Classification Results on Caltech-101 Dataset Using AlexNet DCNN

In the first experiment, we select 20 object classes randomly and perform validation. For validation of the proposed algorithm on Caltech-101 dataset, we divide each class into 50:50. The chosen ratio explain that 50% images are selected for training from each class and the remaining 50% for testing the proposed method on selected classifiers. For testing results, we used 10-fold cross-validation and obtain classification results.

In the first step, we extract DCNN features of 20 object classes by pre-trained AlexNet model and select the best features using an entropy-based method. The selected features are feed to classifiers and achieved best classification accuracy is 86.5%, which is achieved on ensemble classifier. The classification accuracy of ensemble classifier is given in Table 2. The testing time of ensemble classifier is 105.00 seconds which is best as compare to other state of the art

methods. The second-best tranining time is 114.25 seconds for quadratic SVM, which achieved classification accuracy 83.70% and FN rate is 16.30%. In second experiment, classification is performed on 34 classes and obtained maximum classification accuracy 84.6 % on ensemble classifier with FN rate is 15.4%. Also, the classification is performed on some other classification methods and second highest accuracy 78.2% for cubic SVM as given in Table 4.2. The best execution time on the classification of 34 classes is 172.63 seconds which shows that the execution time is increased with the addition of more number of classes.

Table 4.2 Classification Results for Caltech-101 Dataset Using AlexNet Deep CNN Features.

Method	No of Classes				Performance Measures		
	20	34	50	100	Accuracy (%)	FNR (%)	Time (seconds)
Ensemble Boosted Tree	✓				**86.5**	13.5	105.00
		✓			**84.6**	15.4	172.63
			✓		**83.5**	17.0	193.00
				✓	**71.7**	28.3	620.42
Linear SVM	✓				82.0	18.0	197.71
		✓			75.6	24.4	266.74
			✓		78.6	21.4	859.90
				✓	67.9	32.1	6270.00
QSVM	✓				83.7	16.3	114.25
		✓			76.3	23.7	332.70
			✓		81.7	28.3	1325.00
				✓	70.3	29.7	20355
CSVM	✓				83.5	16.5	118.77
		✓			78.2	21.8	339.92
			✓		81.7	18.3	1879.00
				✓	70.4	29.6	12105.00
FKNN	✓				79.8	20.2	144.54
		✓			68.7	31.3	327.07
			✓		77.2	32.8	270.76
				✓	65.2	34.8	714.21
CKNN	✓				82	18	138.16
		✓			70.8	29.2	251.37
			✓		78.3	21.7	379.01
				✓	65.7	34.3	2038.00
Decision tree	✓				78.5	21.5	136.35
		✓			67.3	32.7	267.00
			✓		74.3	25.7	434.82

				✓	58.9	41.1	949.13
WKNN	✓				79.8	20.2	232.06
		✓			69.9	30.1	264.89
			✓		76.4	23.6	378.13
				✓	65.3	34.7	845.73

In the third phase, classification is performed on 50 number of classes and obtained maximum classification accuracy 83.5% on ensemble classifier, which decreases 1% as compared to 20 and 34 number of classes. This problem is caused, when an increase in some more object classes. Moreover, the best execution time for 50 object classes is 193.00 seconds which is better than other classification methods as shown in Table 2 but it increases as compared to a classification of 20 and 34 classes. Finally, classification is performed on 100 classes and obtained maximum correct classification rate 71.7% on ensemble classifier. However, the FN rate is increased up to 28.3%, which is higher than 20, 34, and 50 classes object classification. Moreover, the execution time of ensemble classifier on 100 classes is 620.42 seconds, which is better as compared to other classification methods as given in Table 4.2 but the overall execution time for 20 object classes is better, which shows that increases in the number of classes' effects on both classification accuracy and execution time.

4.1.1.2 Classification Results on Caltech-101 Dataset Using VGG-19 DCNN

In this phase, Classification is performed on 20 classes using VGG-19 DCNN features. The D-CNN features are extracted by performing activation on FC layer and extract 4096 features for each sample, which are later passed to an entropy-controlled method. Through the entropy-controlled method, relevant features are selected, and passed to classifiers for classification. For classification, ten-fold validation is adopted for each group of object classes. In the first group, 20 object classes are randomly selected and performed classification. The best-achieved classification accuracy for 20 object classes is 92.0% with FN rate is 8.0% on ensemble classifier. The classification results of ensemble classifier are also compared with other supervised learning methods and obtained second best accuracy 91.1% with FN rate is 8.9% on CSVM as presented in Table 4.3. The execution time of ensemble classifier is also calculated and obtained best testing time 88.129 seconds, which is efficiently well as compared to other classification methods as LSVM, QSVM, and few more in Table 4.3. In the second group, 34 object classes are selected randomly and performed classification. For testing, 10-

fold x-validation is adopted and achieved best testing recognition accuracy is 84.6% with FN rate 15.4% on ensemble classifier as presented in Table 4.3. The classification performance of ensemble classifier is compared with seven other supervised learning methods and achieved second best accuracy is 78.2% on CSVM. Moreover, achieved best execution time for classification of 34 object classes is 172.63 seconds on ensemble classifier, which is significantly good as compared to other techniques. However, the worst training time for classification of 34 object classes is 327 seconds on Fine KNN.

Table 4.3 Classification Results for Caltech-101 Dataset Using VGG-19 Deep CNN Features

Method	No of Classes				Performance Measures		
	20	34	50	100	Accuracy (%)	FNR (%)	Time (seconds)
Ensemble Boosted Tree	✓				**92.0**	8.0	88.129
		✓			**87.5**	12.5	198.480
			✓		**86.0**	14.0	168.660
				✓	**73.8**	27.2	454.270
Linear SVM	✓				86.7	13.3	180.709
		✓			81.0	19.0	256.700
			✓		78.9	21.1	944.080
				✓	54.8	45.2	1122.200
QSVM	✓				90.6	9.4	147.250
		✓			83.0	17.0	299.100
			✓		80.5	19.5	1205.900
				✓	54.6	45.4	820.010
CSVM	✓				91.1	8.9	191.395
		✓			82.9	17.1	311.200
			✓		81.0	19.0	1624.00
				✓	52.5	47.5	795.360
FKNN	✓				84.3	15.7	114.719
		✓			78.0	22.0	23.620
			✓		75.3	24.7	127.55
				✓	59.4	40.6	1119.490
CKNN	✓				82.6	17.4	315.130
		✓			77.5	22.5	99.800
			✓		77.6	22.4	160.780
				✓	59.2	40.8	81.680
	✓				85.5	14.5	351.162

Decision Tree		✓			78.9	21.1	65.790
			✓		77.7	22.3	106.90
				✓	73.3	26.7	431.22
WKNN	✓				83.3	16.7	215.49
		✓			78.7	21.3	72.39
			✓		75.8	24.2	201.6
				✓	65.6	24.4	341.83

After that, 50 object classes are selected randomly and performed classification. The increase in the number of categories, effects on the classification accuracy and execution time. However, using VGG deep CNN features with entropy-controlled method achieved best classification accuracy is 86.0%, which is increased up to 2.55% as compared to AlexNet deep features. Moreover, the performance on VGG deep features is also improved and achieved best computation time 168.66 seconds which is better as compared to AlexNet deep features and other supervised learning methods as given in Table 4.3. Finally, classification is performed on all object classes and achieved best classification accuracy 73.8% on an ensemble classifier, which is executed in 454.270 seconds. The classification accuracy of ensemble classifier is increased up to 1.8% as compared to performance on AlexNet model but the execution time of ensemble classifier is lower than the WKNN, which is 341.83 seconds as presented in Table 4.3. The above discussion, it is clear that entropy-controlled selection method performs well along with VGG-19 deep CNN features as compared to AlexNet deep features along with entropy-controlled selection approach. Moreover, the execution time for object recognition on VGG features is improved as compared to AlexNet features on 20, 50, and 101 object classes.

4.1.1.3 Classification Using Proposed Technique

Features fusion is an important step in the domain of machine learning because each feature extraction technique has their characteristics. Therefore, in this study, we used two pre-trained deep CNN for features descriptor and selected the best features from each model by an entropy-controlled method.

Table 4.4 Classification Results for Caltech-101 Dataset Using Proposed Features.

Method	No of Classes	Performance Measures

	20	34	50	100	**Accuracy** (%)	**FNR** (%)	**Time** (seconds)
EBT	✓				**86.5**	13.5	**75.70**
		✓			**93.8**	6.2	**289.90**
			✓		**93.5**	6.5	**178.20**
				✓	**89.7**	10.3	**302.50**
L-SVM	✓				82.0	18.0	97.71
		✓			87.2	12.8	495.22
			✓		87.8	12.2	1457.60
				✓	60.7	39.3	5613.70
Q-SVM	✓				83.7	16.3	114.25
		✓			88.2	11.8	654.53
			✓		88.9	11.1	1655.10
				✓	60.0	40	9846.00
C-SVM	✓				83.5	16.5	118.77
		✓			88.8	11.2	1010.60
			✓		88.9	11.1	1754.10
				✓	55.0	45	10322.00
F-KNN	✓				79.8	20.2	14.540
		✓			82.2	17.8	71.486
			✓		82.6	17.4	71.950
				✓	71.3	28.7	66.749
C-KNN	✓				82.0	18	38.160
		✓			82.9	17.1	110.560
			✓		81.0	19	327.660
				✓	70.7	29.3	595.820
Decision tree	✓				78.5	21.5	36.350
		✓			81.3	18.7	99.080
			✓		85.1	14.9	106.200
				✓	88.2	11.8	470.750
WKNN	✓				79.8	20.2	32.060
		✓			86.3	13.7	18.110
			✓		80.8	19.2	68.810
				✓	65.7	34.3	321.97

After that, we extracted SIFT features from RGB silhouette image and fused along with deep CNN selected features by the parallel approach. Finally, the fused features are feed to classifiers for recognition accuracy. For testing, we perform 10-fold cross validation for each

group of object classes. The best-achieved classification accuracy for 20, 34, 50, and 100 classes is 86.5%, 93.8%, 93.5%, and 89.7% on ensemble classifier, as shown in the Table 4.4.

The recognition accuracy of ensemble classifier on 34, 50, and 100 classes is significantly improved as compared to individual AlexNet, and VGG-19 deep CNN features with an entropy-based selection. However, we notice that the execution time of proposed method on ensemble classifier is increased as compared to Table 4.2 and 4.3.

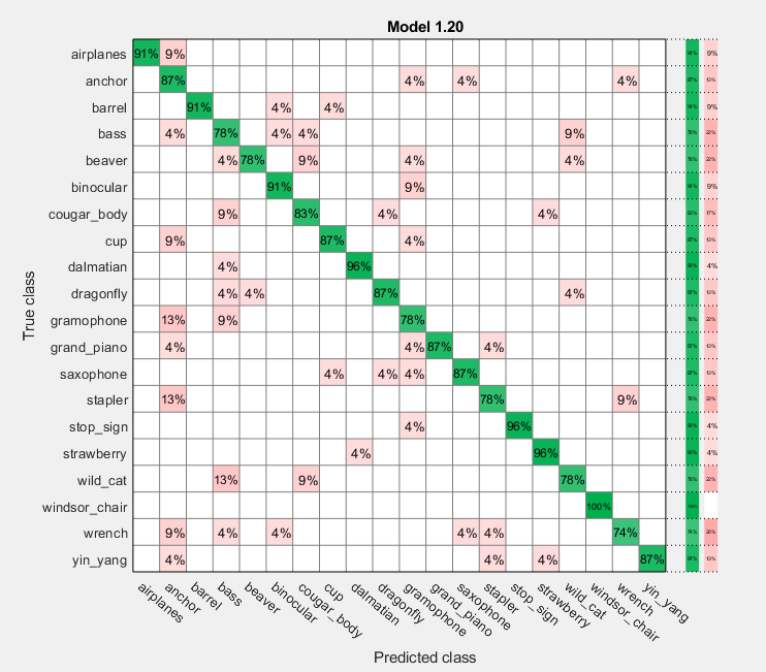


Figure 4.1 Confusion Matrix for 20 Classes Using Caltech-101 Dataset on Proposed Method.

The proposed classification performance is proved by the confusion matrix given in Figure 4.1. The figure shows that wrench has the lowest TPR and has a higher FNR which is causing less accuracy.

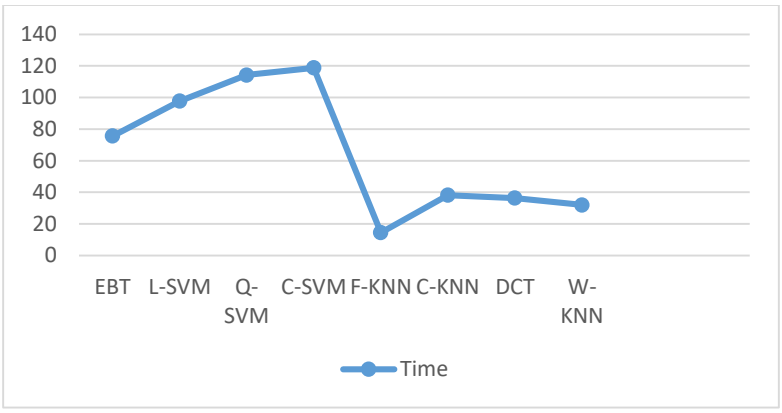


Figure 4.2 Execution Time for Each Classifier for 20 Classes of Caltech-101 on Proposed Method

Figure 4.2 shows the time consumed for training in case of 20 classes of Caltech-101 dataset on proposed technique. The figure shows that Fine KNN took lowest time and the Cubic SVM took maximum time to train. Weighted KNN performed as second best classifier in case of training time taken. Proposed algorithm can be validated on Caltech-101 dataset using confusion matrices shown in Figure 4.1, Figure 4.3, Figure 4.5 and Figure 4.7. Training time can be compared for all classifiers shown in Figure 4.2, Figure 4.4, Figure 4.6 and Figure 4.8.

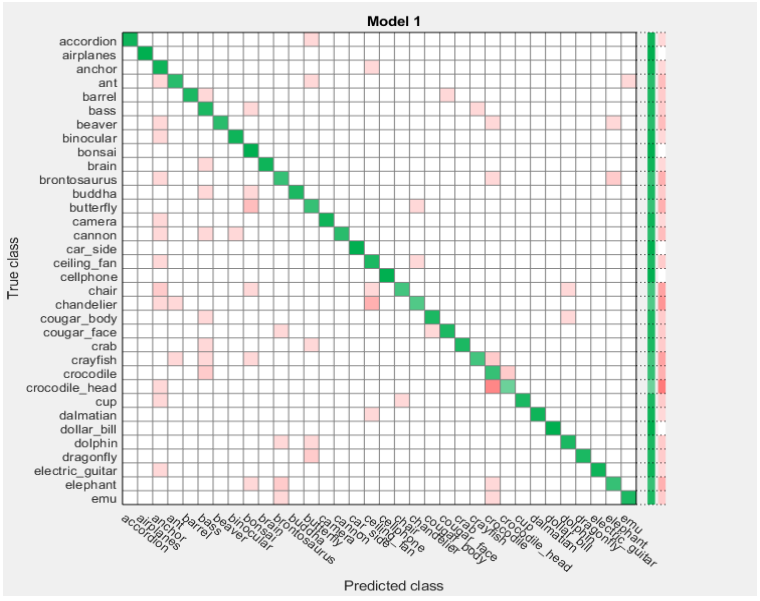


Figure 4.3 Confusion Matrix for 34 Classes Using Proposed Method on Caltech-101 Dataset

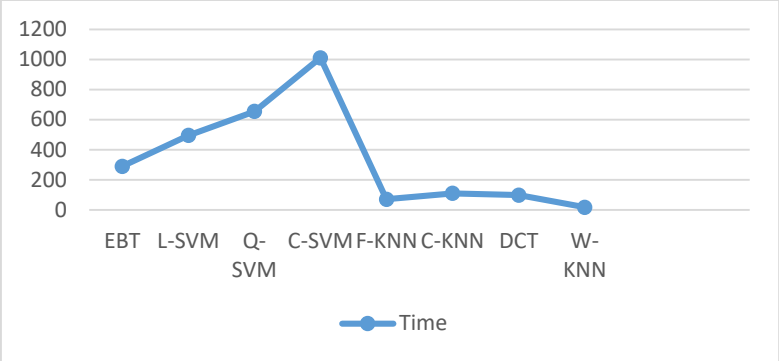


Figure 4.4 Execution Time for Each Classifier for 34 Classes of Caltech-101 on Proposed Method

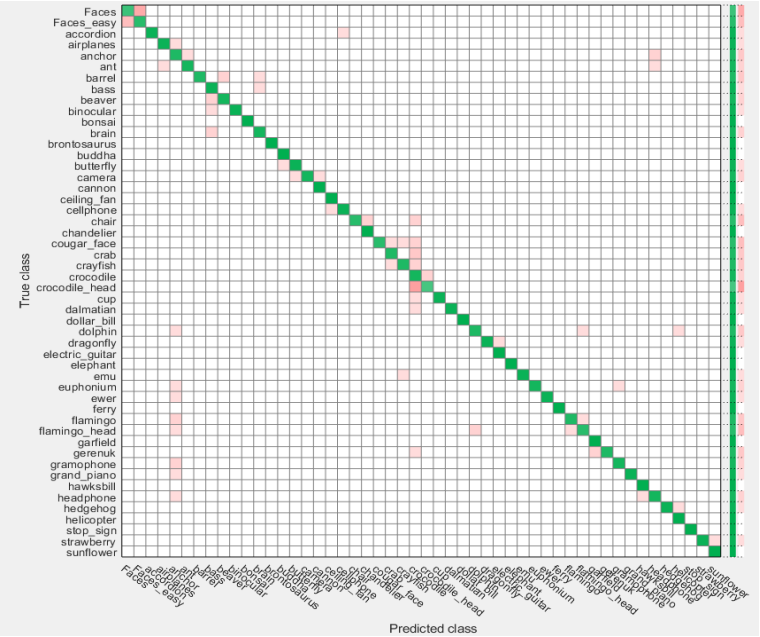


Figure 4.5 Confusion Matrix for 50 Classes Using Caltech-101 Dataset on Proposed Method

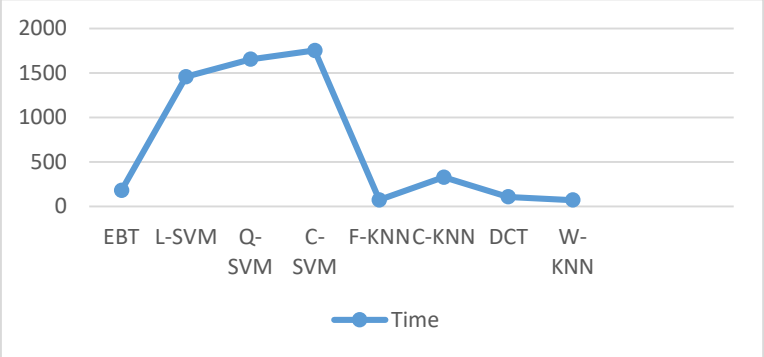


Figure 4.6 Execution Time for Each Classifier for 50 Classes of Caltech-101 on Proposed Method

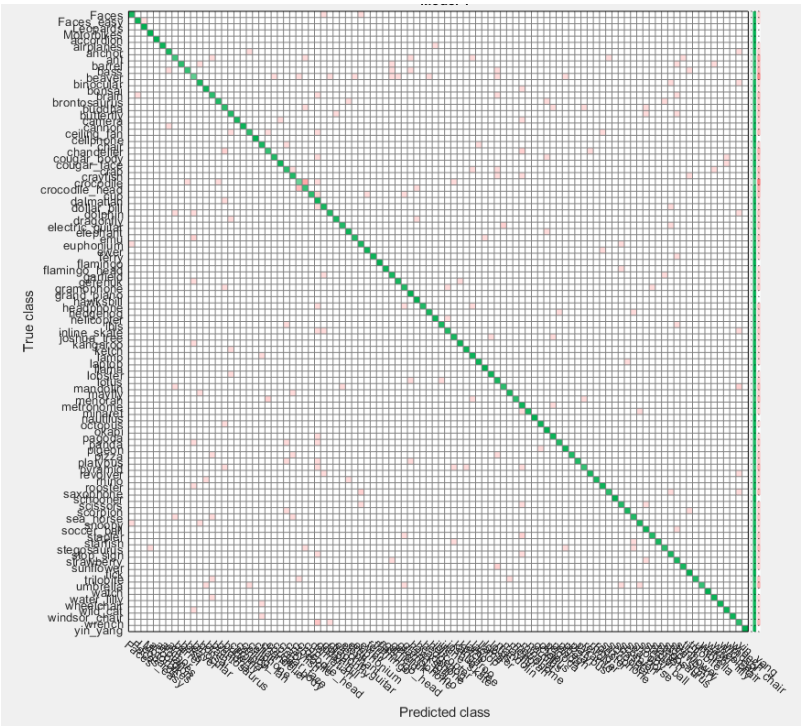


Figure 4.7 Confusion Matrix for 101 Classes of Caltech-101 Dataset on Proposed Method

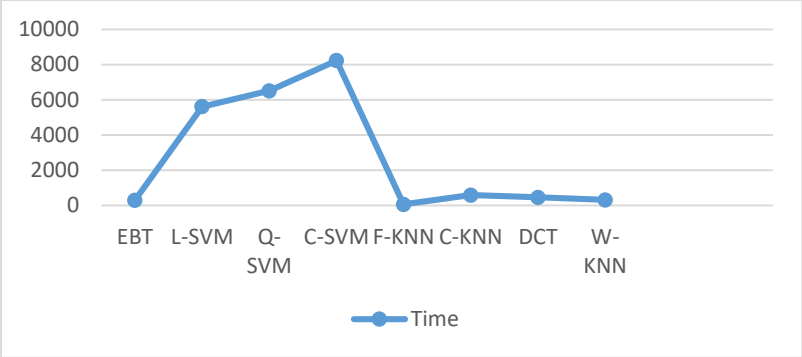


Figure 4.8 Execution Time for Each Classifier for 101 Classes of Caltech-101 on Proposed Method

4.1.2 Classification Results on Pascal3D+ Dataset

In this section, we present the proposed algorithm results on PASCAL 3D+ dataset. The results are calculated in four different steps: a) AlexNet D-CNN features along with entropy-controlled feature selection, b) VGG features extraction and entropy-controlled selection, c) fusion of VGG and AlexNet D-CNN features along with selection method, and d) fusion of SIFT and D-CNN features along with entropy-controlled method. Three parameters (i.e., accuracy, FNR, and time) are used to analyze the performance of each classifier. As we discussed above, this dataset consists of total 22394 images of 12 unique object classes. For validation of the proposed method on this dataset, we opt an method of 50:50 for training and testing. This approach is followed for each step. The achieved best classification accuracy for AlexNet D-CNN features along with entropy-controlled selection method is 76.8% on ensemble classifier. The FN rate on ensemble classifier is 23.2% and testing execution time is 154.5 seconds. The recognition results of an ensemble classifier are also compared with other state-of-the-art classification methods as presented in Table 4.5.

Table 4.5 Classification Results for PASCAL 3D+ dataset

Method	No of Classes				Performance Measures		
	AlexNet	VGG-19	Fused	Proposed	Accuracy (%)	FNR (%)	Time (seconds)
Ensemble Boosted Tree	✓				76.8	23.2	154.5
		✓			81.8	18.2	304.8
			✓		**87.4**	12.6	**230.2**
				✓	**88.6**	**11.4**	**111.99**
	✓				71.0	29.0	437.18
		✓			78.1	21.9	626.3

Linear SVM			✓		56.6	43.4	834.9
				✓	82.8	17.2	175.26
QSVM	✓				75.6	24.4	641.2
		✓			80.6	19.4	698.2
			✓		86.9	13.1	1821.7
				✓	81.3	18.7	210.39
CSVM	✓				73.6	26.4	600.57
		✓			79.1	20.9	765.77
			✓		86.6	13.4	1211.2
				✓	81.4	18.6	217.5
FKNN	✓				64.0	36.0	195.068
		✓			70.8	29.2	222.43
			✓		75.0	25.0	198.71
				✓	23.4	76.6	134.934
CKNN	✓				71.5	28.5	2149.1
		✓			78.9	21.1	5277.1
			✓		82.6	17.4	4133.5
				✓	26.5	83.5	659.35
Decision tree	✓				70.6	29.4	164.36
		✓			78.0	22.0	240.86
			✓		82.6	17.4	398.78
				✓	73.85	26.15	185.585
WKNN	✓				64.7	35.3	1087.5
		✓			71	29	2457.3
			✓		74.8	25.2	2020.2
				✓	78.2	21.8	409.78

In the second step, the classification is performed by using VGG-19 deep CNN features and achieved maximum classification accuracy 81.8%, which is improves as compared to AlexNet features. But the execution time on VGG-19 deep features along with selection method is increased on ensemble classifier and best-achieved execution time is 240.86 seconds on decision tree as given in Table 4.5. In the third step, fused selected AlexNet DCNN and VGG DCNN features by a serial-based method and perform classification. The best-achieved classification accuracy is 87.4% on ensemble classifier, which is significantly improved after fusion of DCNN features. The proposed algorithm can be validated using confusion matrix presented in Figure 4.9.

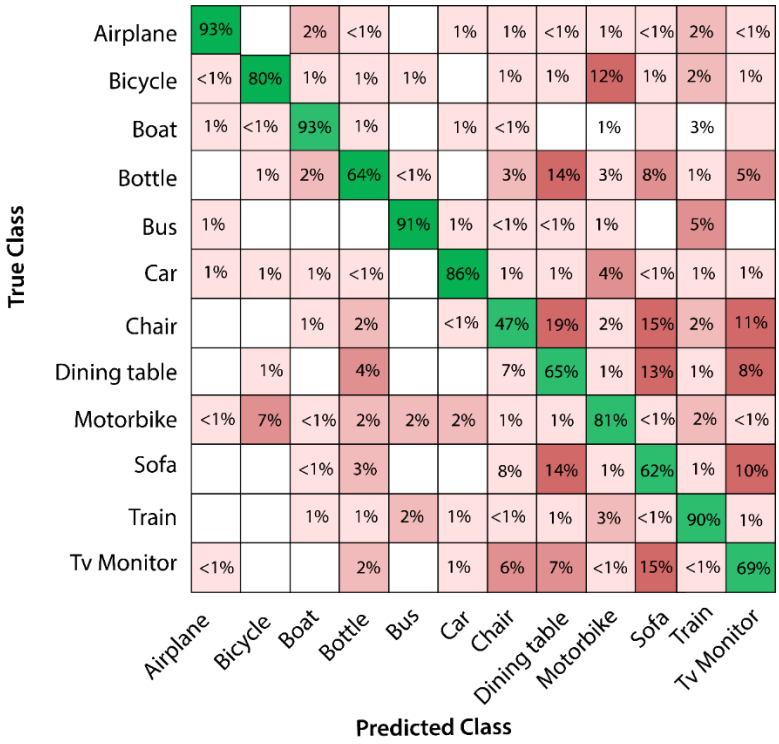


Figure 4.9 Confusion Matrix for PASCAL 3D+ Dataset using Proposed Method

The execution time of ensemble classifier for step 3 is 230.2 seconds, which is higher than the FKNN as presented in Table 4.5. Finally, on fused DCNN features, we employed entropy features and selected best 7000 features. For classification, ensemble classifier is used and obtained maximum accuracy 88.6% and FN rate is 11.4%, which is significantly improved as compared to step 1, 2, and 3. Moreover, best execution time is 111.99 seconds for ensemble classifier as given in Table 4.5. From Table 4.5, the performance of ensemble classifier is compared with several other supervised learning methods such as LSVM, WKNN, FKNN, and few more. These supervised learning methods also perform well using proposed features fusion and selection method, which gives its authenticity.

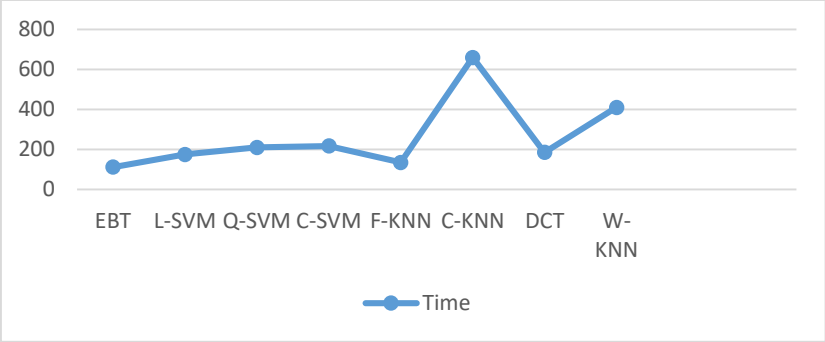


Figure 4.10 Execution Time for Each Classifier using Proposed Method on PASCAL 3D Dataset

Moreover, the classification performance of ensemble classifier is confirmed by confusion matrix in Figure 4.10. Also, Figure 4.10 shows that EBT performed best among all classifiers. EBT took minimum time among all the classifiers which can be verified using Figure 4.10.

4.1.3 Classification Results on Barkley 3D Dataset

The 3D dataset consists of total 6604 images of 10 object classes including bicycle, car, cellphone, head, iron, monitor, mouse, shoe, stapler, and toaster. The number of images in each class range of 474-721. For validation of the proposed method on a 3D dataset, 50:50 approach is opted for training and testing of classifier. To analyze the performance of a proposed method, we employ four distinct experiments. To resolve this issue, in experiment 4 we fused SIFT features along with deep CNN features and achieved classification accuracy 99.7% with FN rate is 0.3%. The training time of the proposed method is also reduced on ensemble classifier and achieved testing time is 177.49 seconds. Moreover, the classification accuracy of ensemble classifier for experiment 4 is validated by confusion matrix from Figure 4.11.

In the first experiment, Alexnet D-CNN features are acheived and most relavant features using the entropy method. The best-achieved classification accuracy for the first experiment is 97.90% on ensemble classifier with FN rate is 2.1%. The execution time on ensemble classifier for experiment 1 is 978.00 seconds, which is higher than the other classifiers. The best execution time for experiment 1 is 245.68 seconds as given in Table 4.6.

Table 4.6 Classification Results for Barkley 3D dataset

Method	No of Classes				Performance Measures		
	AlexNet	VGG-19	Fused	Proposed	Accuracy (%)	FNR (%)	Time (seconds)
Ensemble Boosted Tree	✓				97.9	2.1	978.00
		✓			97.5	2.5	900.5
			✓		98.8	1.2	5342.2
				✓	**99.7**	**0.3**	**177.49**
Linear SVM	✓				96.1	3.9	220.29
		✓			96.2	3.8	132.01
			✓		98.7	1.3	453.66
				✓	99.5	0.5	94.242
QSVM	✓				97.5	2.5	231.44
		✓			97.3	2.7	338.24
			✓		99	1.0	566.5
				✓	99.7	0.3	120.52
CSVM	✓				97.5	2.5	245.68
		✓			97.3	2.7	363.6
			✓		99	1	651.6
				✓	99.7	0.3	124.92
FKNN	✓				96.9	3.1	55.29
		✓			97.3	2.7	113.15
			✓		97.9	2.1	141
				✓	62.6	37.4	19.102
CKNN	✓				95.4	4.6	1133.6
		✓			94.8	5.2	1167.8
			✓		97	3	2248.7
				✓	34	66	359.51
Decision tree	✓				92.3	7.7	101.08
		✓			92.8	7.2	128.71
			✓		94.5	5.5	151.44
				✓	85.8	14.2	153.816
WKNN	✓				97	3	596
		✓			97.2	2.8	1862.9
			✓		98	2	1163.1
				✓	98.6	1.4	215.24

In the second experiment, the VGG D-CNN features are obtained and select the top features by an entropy-controlled method. 10-fold x-validation is adopted for testing the recognition performance and achieved the best accuracy 97.5% with FN rate is 2.5%.

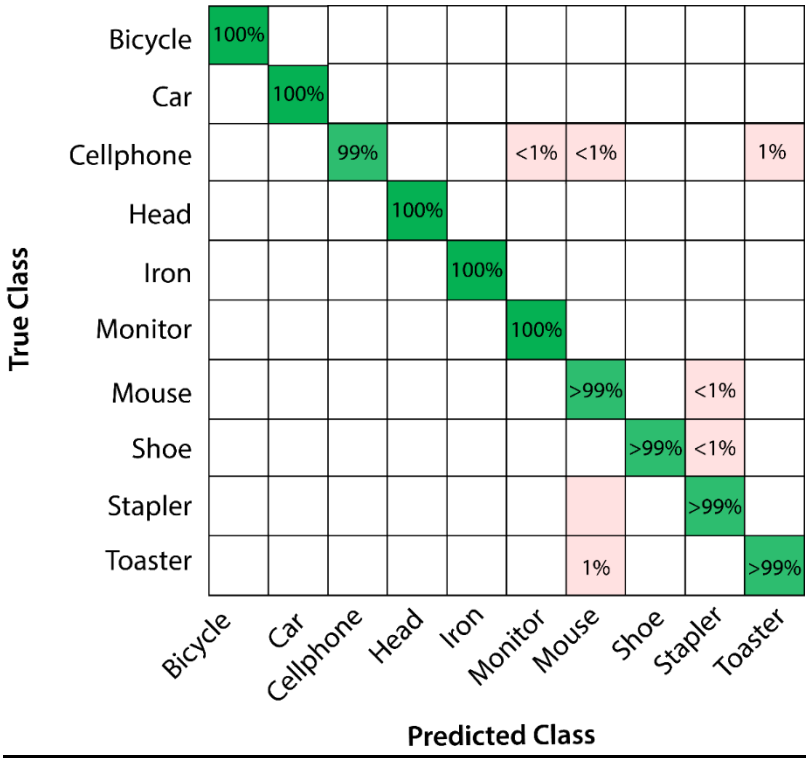


Figure 4.11 Confusion matrix of proposed method results on Barkley 3D dataset

The execution time of ensemble classifier for VGG features is 900.5 seconds as given in Table 4.6, which shows that FNN performs fast and execute in 113.5 seconds. In the third experiment, to improve the classification accuracy and execution time, we fused both VGG and AlexNet deep CNN features and achieved classification accuracy 98.8% on ensemble classifier. The fused matrix improves the classification accuracy as compared to an individual selected deep CNN features as presented in Table 4.6.

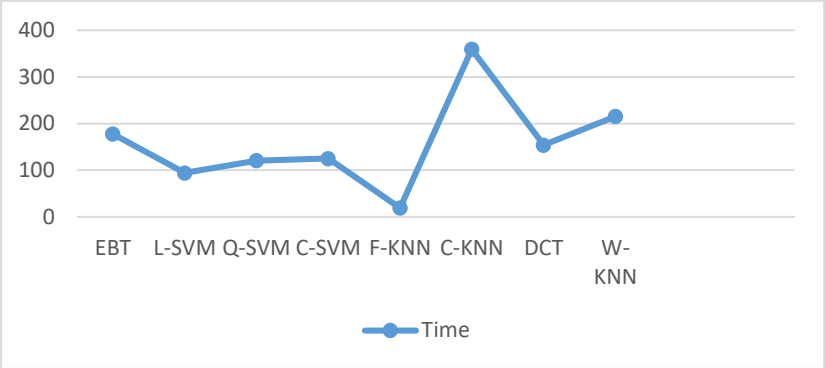


Figure 4.12 Execution Time of Each Classifier using Proposed method on Barkley 3D Dataset

By the execution time of fused approach is increased up to 5342 seconds on ensemble classifier to resolve this issue, in experiment 4 we fused SIFT features along with deep CNN features and achieved classification accuracy 99.7% with FN rate is 0.3%. The execution time of proposed method is also reduced on ensemble classifier and achieved testing time is 177.49 seconds. Moreover, the classification accuracy of ensemble classifier for experiment 4 is confirmed by confusion matrix in Table 4.6. Figure 4.12 shows the time comparison among all the classifiers which shows that the Fine KNN took minimum time and Cubic KNN took the maximum time.

4.1.4 Classification Results on Birds Dataset

In this phase, we employed the proposed technique on another open available Birds dataset in three different phases which are AlexNet features, b) VGG-VD-19 and d) fusion of both CNN features along with entropy-based feature optimization. Performance measures like accuracy, FNR and time are used to analyze the performance of proposed technique. As discussed earlier, the dataset consists on a total of 600 images with six unique butterfly classes. We selected 70:30 portion for training and testing for each experiment. Results showed that by using VGG-VG-19 features, best accuracy reported by Quadratic SVM is 99.0% with FN rate of 1.0% and a training time of 51.03 seconds. Since Cubic SVM also reported 99.0% accuracy but the training time was a bit higher which was 54.59 seconds shown in Table 4.7.

Table 4.7 Results on Birds dataset using Alexnet features, VGG19 features, and Proposed Features

Classifier	Alexnet	VGG19	Fused	Accuracy (%)	FNR (%)	Time (sec)
ESD	✓			**99.0**	1.0	75.09
		✓		**99.5**	0.5	88.31
			✓	**100.0**	0.0	72.45
E-S-KNN	✓			96.7	3.3	45.09
		✓		97.6	2.4	38.31
			✓	97.4	2.6	55.54
LD	✓			98.0	2.0	48.39
		✓		99.0	1.0	11.11
			✓	100.0	0.0	33.92
L-SVM	✓			97.9	2.1	45.36
		✓		99.0	0.5	20.00
			✓	100.0	0.0	47.66
Q-SVM	✓			84.5	1.0	51.03
		✓		99.3	0.7	24.06
			✓	100.0	0.0	45.25
Cub-SVM	✓			99.0	1.0	54.59
		✓		99.5	0.5	73.32
			✓	100.0	0.0	41.29
F-KNN	✓			96.2	3.8	41.47
		✓		97.4	2.6	9.58
			✓	99.5	0.5	24.89
M-KNN	✓			97.6	2.4	32.30
		✓		98.8	1.2	7.31
			✓	100.0	0.0	25.82
W-KNN	✓			97.9	2.1	23.96
		✓		99.3	0.7	13.10
			✓	100.0	0.0	35.16
Cos-KNN	✓			95.7	4.3	31.08
		✓		99.0	1.0	12.00
			✓	99.8	0.2	29.11

In second experiment, classification is performed by using VGG- 19 features which reported best accuracy of 99.5% with an FN rate of 0.5% and training time of 88.31 seconds on Ensemble SD classifier. In last experiment, best accuracy reported was 100% with 72.45 seconds on ESD classifier as shown in Table 4.7. In Table 4.7, the classification results are compared with other supervised learning based classifiers. The achieved accuracy can be validated using confusion matrix shown in Figure 4.13.

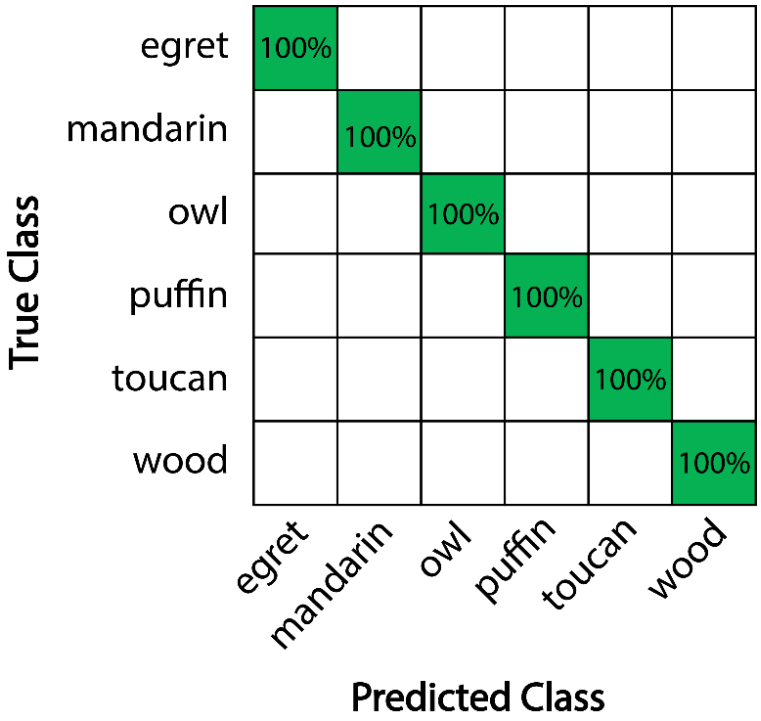


Figure 4.13 Confusion Matrix for Birds dataset

Moreover, the classification performance of ensemble classifier is confirmed by confusion matrix in Figure 4.16. Also, Figure 4.16 shows that EBT performed best among all classifiers also each class gives the equal results.

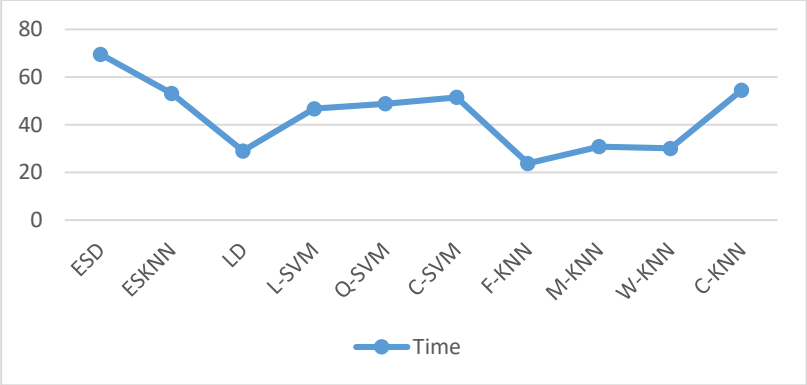


Figure 4. 14 Execution Time of Each Classifier using Proposed method on Birds Dataset

Figure 4.14 shows the time comparison among all the classifiers which shows that the Fine KNN took minimum time and Cubic KNN took the maximum time.

4.1.5 Classification Results on Butterfly Dataset

In this section, classification is performed on publicly available butterfly dataset. Butterfly dataset consists on a total of 619 images from 7 different classes of butterfly with 42-134 range. Classification results are achieved using three techniques, which are a) AlexNet features, b) VGG-VD-19 features and c) fusion of both CNN features along with entropy-based feature optimization. Performance measures like accuracy, FNR and time are used to analyze the performance of proposed technique. We selected 70:30 portion for training and testing for each experiment. Results in Table 4.8 shows that by using AlexNet features, best accuracy reported by Quadratic SVM is 95.1% with FN rate of 4.9% and a training time of 46.05 seconds.

Table 4.8 Results on Butterflies dataset using Alexnet features, VGG19 features, and proposed features.

Classifier	Alexnet	VGG19	Proposed	Accuracy (%)	FNR (%)	Time (s)
ESD	✓			**95.1**	9.4	46.05
		✓		**95.6**	5.9	31.95
			✓	**98.0**	2.0	69.53
ESKNN	✓			85.7	14.3	28.56
		✓		87.7	12.3	18.27
			✓	88.7	11.3	53.08
LD	✓			70.9	29.1	48.44
		✓		94.1	4.6	12.42
			✓	96.6	3.4	29.01
L-SVM	✓			91.6	8.4	40.02
		✓		94.6	5.4	29.65
			✓	96.6	3.4	46.72
Q-SVM	✓			94.1	5.9	39.46
		✓		94.1	5.9	24.58
			✓	96.6	3.4	48.8
C-SVM	✓			90.6	4.9	44.23
		✓		93.6	6.4	19.41
			✓	97.0	3.0	51.51
F-KNN	✓			85.7	14.3	20.82
		✓		89.2	10.8	8.70
			✓	94.1	5.9	23.79
M-KNN	✓			82.3	19.7	19.29
		✓		85.2	14.8	8.30
			✓	92.1	7.9	30.83
W-KNN	✓			85.2	14.8	15.06
		✓		87.2	12.8	15.96
			✓	94.6	5.4	30.12

C-KNN	✓			81.8	18.2	16.02
		✓		85.7	14.3	8.54
			✓	94.1	5.9	54.55

In second experiment, classification is performed by using VGG-VG-19 features which reported best accuracy of 95.6% with an FN rate of 4.4% and training time of 31.95 seconds on Ensemble SD classifier. In last experiment, best accuracy reported was 98.0% with 69.53 seconds on ESD classifier as shown in Table 4.8. In Table 4.8, the classification results are compared with other supervised learning based classifiers.

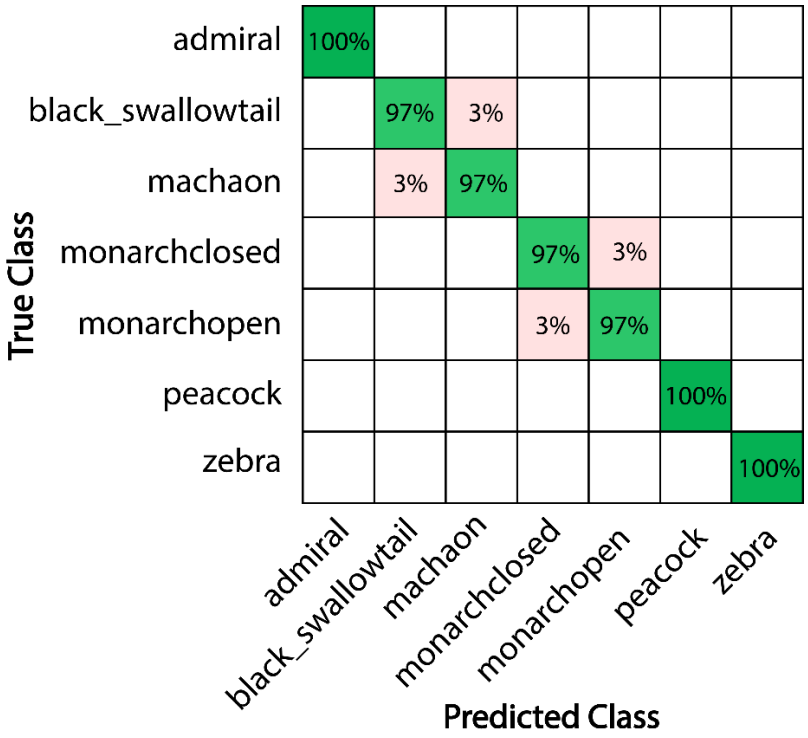


Figure 4.15 Confusion Matrix for Birds dataset

Results shows that the ESD outperformed among all the classifiers and the achieved accuracy can be justified by the confusion matrix presented in Figure 4.15.

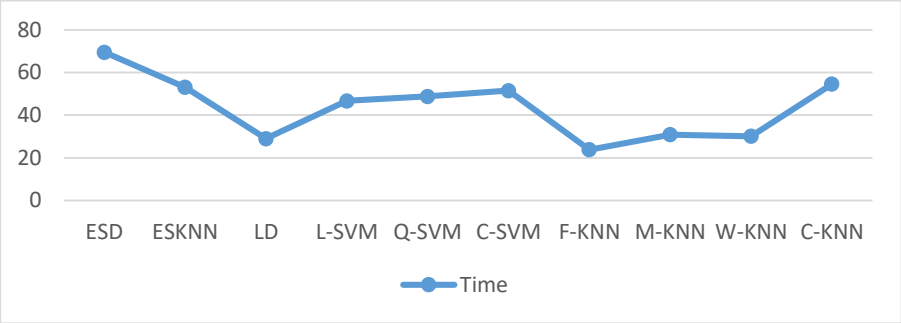


Figure 4.16 Execution Time of Each Classifier using Proposed method on Butterflies Dataset

Figure 4.16 shows the time comparison among all the classifiers which shows that the Fine KNN took minimum time and ESD took the maximum execution time.

4.2. Results Analysis

In this section we performed an analysis of our results with the existing state of the art techniques with respect to all selected datasets.

4.2.1 Caltech-101 Analysis

At the end, our proposed results are compared with existing methods in Table 4.9. Jun et al. [18] introduced deep stack network (DSN) for object classification and achieved classification accuracy 89%. In. [105] sparse structure PCA method is presented for object classification, which is based on SIFT features and SVM classifier. The presented method reported classification accuracy 83.9% on a Caltech-101 dataset. Qing et al. [18] used a combination of YCbCr transformation and Extreme Learning (EL) for object classification and obtained accuracy of 78% on the Caltech-101 dataset. Yongsheng et al. [108] used the K-means based reduction on the SIFT descriptors and achieved 85.78% accuracy.

Table 4.9 Results Analysis for Caltech-101 Dataset

Paper	Year	Features	Technique	Accuracy (%)	Time (s)
Jun et al. **[18]**	2018	PCA Based 2000 features from 4096 Features	Deep Stack Network	89	

Jinjoo et al. **[105]**	2018	Applied SVM on SIFT	SSPCA	83.9	
Qing et al. **[18]**	2018	Fused YCbCr and SIFT	Applied LSC+ELM on fused feaures	78	
Yongsheng et al. **[108]**	2018	SIFT	Reduction using K Means	85.78	
Xiaozhao et al. **[157]**	2018	Salient features using unsupervised	Selected 1500 features using PCA	76	
Ridha et al. **[128]**	2017	Deep CNN	Fast wavelet	75.6	
Proposed	--	Deep CNN and SIFT Features	Fused DCNN and SIFT Features along with entropy-controlled selection method	20 Classes: 86.5	75.70
				34 Classes: 93.8	289.90
				50 Classes: 93.5	178.20
				100 Classes: 89.7	**302.50**

However, in this research, our proposed method showed improved performance in both accuracy and execution time. The proposed method achieved classification accuracy 86.5%, 93.8%, 93.5%, and 89.7% for 20, 34, 50, and 100 object classes on Caltech-101 dataset. The execution time of the presented method is also plotted in Figure 13. The above results illustrates that introduced method performed significantly well and gave authenticity for classification of maximum object classes.

4.2.2 Pascal 3D Analysis

The proposed method results on PASCAL 3D dataset are compared with existing methods as presented in Table 4.10. In Table 4.10, Chi et al. [158] extract deep CNN features for object classification and reported classification accuracy 81.8%. In [159] CNN based features are extracted for object classification and perform experiments on PASCAL 3D dataset and achieved accuracy 83.92%. However, our proposed method shows improved performance on PASCAL 3D dataset and achieved classification accuracy 88.60%.

Table 4.10 Results Analysis for Pascal3D+ dataset

Author	**Year**	**Features**	**Technique**	**Accuracy** (%)	**Time** (s)
Chi Li [158]	2018	CNN	Deep Supervision Object Reconstruction	81.8	--
Alexander [159]	2018	CNN	CNN based multi-view learning	83.92	--
Proposed	2018	SIFT and Deep Features	Fusion of point and DCNN features along with selection method	**88.6**	**111.9**

Table 4.10 shows that most of the research made on this dataset was CNN based classification and our proposed method achieves significantly improved results.

4.2.3 Butterflies Analysis

The proposed method results are compared with the state of the art technique and the summary of results are tabulated in Table 4.11. Svetlana [148] achieved the maximum of 90.4% accuracy on this dataset using their technique.

Table 4.11 Results Analysis for Butterflies dataset

Paper	**Year**	**Accuracy (%)**
Svetlana [148]	2004	90.4
Proposed	2018	**98.0**

Table 4.11 shows that most of the research made on this dataset was CNN based classification and our proposed method achieves significantly improved results.

Chapter 5

Conclusion and Future Work

5. Conclusion and Future Work

5.1 Conclusion

A DCNN and SIFT point features fusion and selection-based approach is proposed in this article. The proposed method proceeds in two parallel steps. In the first step, improved saliency method is implemented and SIFT point features are extracted from RGB mapped image. Secondly, DCNN features are gained using pre-trained CNN models. The max-pooling is performed on extracted features matrices to remove the noisy information. After that, a Reyni entropy-controlled method is proposed which control the randomness of extracted features and select the top features. The best selected features are finally passed to ensemble classifier for object classification. The proposed method automatically detects and labeled object from a large number of sample images with minimum human intervention. The proposed approach performed classification under the supervised method and achieved the maximum classification accuracy 93.8%, 88.6%, 99%, 100%, and 98.0% on Caltech-101, PASCAL 3D Plus, and Barkley 3D dataset, Birds dataset, and Butterflies dataset, which shows exceptional performance as compared to existing methods. Moreover, the proposed method efficiently reduces the computation time, which shows the importance of selection methods. In the future, we implement a new generic method for multiple object detection and classification using deep learning. Moreover, we apply method on real-time object classification.

5.2 Future Work

The proposed method can be utilized as a future augmentation for the classification errands, which includes the steady and equivalent number of test classes. The proposed strategy was introduced after a profound thought that just separating basic, visual and printed highlights don't ensure the best execution and results. These confinements were reapplied utilizing Caltech-101 dataset. In any case, after removing highlights, additionally preparing gave enhanced outcomes. If this handling of highlight combination and highlight choice can be connected on different areas after choosing the required highlights, the outcomes may enhance and also the execution in term of viability and effectiveness. The proposed procedure isn't just constrained to object classification. It may be applied to whatever the domain is, for feature extraction, fusion and reduction.

Chapter 6

References

6. References

1. Liu, W., et al. *An Unsupervised Domain Adaptation Method for Multi-Modal Remote Sensing Image Classification*. in *2018 26th International Conference on Geoinformatics*. 2018. IEEE.
2. Thenkabail, P.S., *Urban Image Classification: Per-Pixel Classifiers, Subpixel Analysis, Object-Based Image Analysis, and Geospatial Meth*, in *Remote Sensing Handbook-Three Volume Set*. 2019, CRC Press. p. 253-264.
3. Xie, J., et al., *Bag of Tricks for Image Classification with Convolutional Neural Networks.* arXiv preprint arXiv:1812.01187, 2018.
4. Kamavisdar, P., S. Saluja, and S. Agrawal, *A survey on image classification approaches and techniques.* International Journal of Advanced Research in Computer and Communication Engineering, 2013. **2**(1): p. 1005-1009.
5. Zhao, B., et al., *A survey on deep learning-based fine-grained object classification and semantic segmentation.* International Journal of Automation and Computing, 2017. **14**(2): p. 119-135.
6. Parekh, H.S., D.G. Thakore, and U.K. Jaliya, *A survey on object detection and tracking methods.* International Journal of Innovative Research in Computer and Communication Engineering, 2014. **2**(2): p. 2970-2978.
7. Prokop, R.J. and A.P. Reeves, *A survey of moment-based techniques for unoccluded object representation and recognition.* CVGIP: Graphical Models and Image Processing, 1992. **54**(5): p. 438-460.
8. Yilmaz, A., O. Javed, and M. Shah, *Object tracking: A survey.* Acm computing surveys (CSUR), 2006. **38**(4): p. 13.
9. Zhang, G.P., *Neural networks for classification: a survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2000. **30**(4): p. 451-462.
10. Cheng, G. and J. Han, *A survey on object detection in optical remote sensing images.* ISPRS Journal of Photogrammetry and Remote Sensing, 2016. **117**: p. 11-28.
11. Phyu, T.N. *Survey of classification techniques in data mining*. in *Proceedings of the International MultiConference of Engineers and Computer Scientists*. 2009.
12. Khan, M.A., et al., *An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification.* BMC cancer, 2018. **18**(1): p. 638.
13. Liu, L., L. Wang, and X. Liu. *In defense of soft-assignment coding*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.
14. Lazebnik, S., C. Schmid, and J. Ponce. *Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories*. in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*. 2006. IEEE.
15. Li, Q., et al., *A Local Neighborhood Constraint Method for SIFT Features Matching*, in *Recent Developments in Data Science and Business Analytics*. 2018, Springer. p. 313-320.
16. Firuzi, K., et al., *Partial Discharges Pattern Recognition of Transformer Defect Model by LBP & HOG Features.* IEEE Transactions on Power Delivery, 2018.
17. Qin, C., M. Sun, and C.-C. Chang, *Perceptual hashing for color images based on hybrid extraction of structural features.* Signal Processing, 2018. **142**: p. 194-205.
18. Li, Q., et al., *Improving Image Classification Accuracy with ELM and CSIFT.* Computing in Science & Engineering, 2018.

19. Roy, P.K. and H. Om, *Suspicious and Violent Activity Detection of Humans Using HOG Features and SVM Classifier in Surveillance Videos*, in *Advances in Soft Computing and Machine Learning in Image Processing*. 2018, Springer. p. 277-294.
20. Bhargava, A., et al., *Computer Aided Diagnosis of Cervical Cancer Using HOG Features and Multi Classifiers*, in *Intelligent Communication, Control and Devices*. 2018, Springer. p. 1491-1502.
21. Verma, A., S. Sharma, and P. Gupta. *RNN-LSTM Based Indoor Scene Classification with HoG Features*. in *International Conference on Advanced Informatics for Computing Research*. 2018. Springer.
22. Wei, G., et al., *Content-based image retrieval for Lung Nodule Classification Using Texture Features and Learned Distance Metric.* Journal of medical systems, 2018. **42**(1): p. 13.
23. Akcay, S., et al., *Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-ray Baggage Security Imagery.* IEEE Transactions on Information Forensics and Security, 2018.
24. Juuti, M., F. Corona, and J. Karhunen, *Stochastic Discriminant Analysis for Linear Supervised Dimension Reduction.* Neurocomputing, 2018.
25. Li, K., et al., *Multi-modal feature fusion for geographic image annotation.* Pattern Recognition, 2018. **73**: p. 1-14.
26. Jolliffe, I., *Principal component analysis*. 2011: Springer.
27. Ghose, U. and R. Mehta, *Attribute Reduction Method Using the Combination of Entropy and Fuzzy Entropy*, in *Networking Communication and Data Knowledge Engineering*. 2018, Springer. p. 169-177.
28. Dong, H., et al., *A Novel Hybrid Genetic Algorithm with Granular Information for Feature Selection and Optimization.* Applied Soft Computing, 2018.
29. Naeini, A.A., et al., *Particle Swarm Optimization for Object-Based Feature Selection of VHSR Satellite Images.* IEEE Geoscience and Remote Sensing Letters, 2018. **15**(3): p. 379-383.
30. Agrawal, S., et al., *A comparative study of fuzzy PSO and fuzzy SVD-based RBF neural network for multi-label classification.* Neural Computing and Applications, 2018. **29**(1): p. 245-256.
31. Liu, W., et al., *Multiview dimension reduction via Hessian multiset canonical correlations.* Information Fusion, 2018. **41**: p. 119-128.
32. Singh, C., E. Walia, and K.P. Kaur, *Enhancing color image retrieval performance with feature fusion and non-linear support vector machine classifier.* Optik-International Journal for Light and Electron Optics, 2018. **158**: p. 127-141.
33. Fondón, I., et al., *Automatic classification of tissue malignancy for breast carcinoma diagnosis.* Computers in biology and medicine, 2018.
34. Arel, I., D.C. Rose, and T.P. Karnowski, *Deep machine learning-a new frontier in artificial intelligence research [research frontier]*. IEEE computational intelligence magazine, 2010. **5**(4): p. 13-18.
35. Liu, W., et al., *A survey of deep neural network architectures and their applications.* Neurocomputing, 2017. **234**: p. 11-26.
36. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
37. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
38. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
39. Szegedy, C., et al. *Going deeper with convolutions*. 2015. Cvpr.

40. Rastegari, M., et al. *Xnor-net: Imagenet classification using binary convolutional neural networks*. in *European Conference on Computer Vision*. 2016. Springer.
41. Li, B., et al. *Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN*. in *Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on*. 2017. IEEE.
42. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks.* Nature, 2017. **542**(7639): p. 115.
43. Longadge, M.R., M.S.S. Dongre, and L. Malik, *Multi-cluster based approach for skewed data in data mining.* Journal of Computer Engineering (IOSR-JCE), 2013. **12**(6): p. 66-73.
44. Li, Y., G. Sun, and Y. Zhu. *Data imbalance problem in text classification*. in *Information Processing (ISIP), 2010 Third International Symposium on*. 2010. IEEE.
45. Noce, L., I. Gallo, and A. Zamberletti, *Combining Textual and Visual Features to Identify Anomalous User-Generated Content.* Int. J. Comput. Linguistics Appl., 2015. **6**(2): p. 159-175.
46. Noce, L., I. Gallo, and A. Zamberletti. *Query and Product Suggestion for Price Comparison Search Engines based on Query-product Click-through Bipartite Graphs*. in *WEBIST (1)*. 2016.
47. Gehler, P. and S. Nowozin. *On feature combination for multiclass object classification*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009. IEEE.
48. Javed, O. and M. Shah. *Tracking and object classification for automated surveillance*. in *European Conference on Computer Vision*. 2002. Springer.
49. Socher, R., et al. *Convolutional-recursive deep learning for 3d object classification*. in *Advances in neural information processing systems*. 2012.
50. Chen, Q., et al., *Contextualizing object detection and classification.* IEEE transactions on pattern analysis and machine intelligence, 2015. **37**(1): p. 13-27.
51. DeVries, T. and G.W. Taylor, *Dataset augmentation in feature space.* arXiv preprint arXiv:1702.05538, 2017.
52. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-193.
53. Goyal, M., et al. *Dataset augmentation with synthetic images improves semantic segmentation*. in *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 6th National Conference, NCVPRIPG 2017, Mandi, India, December 16-19, 2017, Revised Selected Papers 6*. 2018. Springer.
54. Liu, X., et al. *Data Augmentation via Latent Space Interpolation for Image Classification*. in *2018 24th International Conference on Pattern Recognition (ICPR)*. 2018. IEEE.
55. Khan, S.H., et al., *Cost-sensitive learning of deep feature representations from imbalanced data.* IEEE transactions on neural networks and learning systems, 2018. **29**(8): p. 3573-3587.
56. Morar, A., F. Moldoveanu, and E. Gröller. *Image segmentation based on active contours without edges*. in *2012 IEEE 8th International Conference on Intelligent Computer Communication and Processing*. 2012. IEEE.
57. Pal, N.R. and S.K. Pal, *A review on image segmentation techniques.* Pattern recognition, 1993. **26**(9): p. 1277-1294.
58. Fulkerson, B., A. Vedaldi, and S. Soatto. *Class segmentation and object localization with superpixel neighborhoods*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009. IEEE.

59. Chen, L.-C., et al., *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs.* IEEE transactions on pattern analysis and machine intelligence, 2018. **40**(4): p. 834-848.
60. Blaschke, T., C. Burnett, and A. Pekkarinen, *Image segmentation methods for object-based analysis and classification*, in *Remote sensing image analysis: Including the spatial domain*. 2004, Springer. p. 211-236.
61. Hariharan, B., et al. *Hypercolumns for object segmentation and fine-grained localization*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
62. Cheng, M.-M., et al., *Global contrast based salient region detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015. **37**(3): p. 569-582.
63. Li, G. and Y. Yu. *Visual saliency based on multiscale deep features*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
64. Sarkar, R. and S.T. Acton, *SDL: Saliency-Based Dictionary Learning Framework for Image Similarity.* IEEE Transactions on Image Processing, 2018. **27**(2): p. 749-763.
65. Sharma, A. and J.K. Ghosh. *A Bottom-Up Saliency-Based Segmentation for High-Resolution Satellite Images*. in *Proceedings of 2nd International Conference on Computer Vision & Image Processing*. 2018. Springer.
66. Ngoc, M.Ô.V., J. Fabrizio, and T. Géraud. *Saliency-based detection of identy documents captured by smartphones*. in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. 2018. IEEE.
67. Kumar, R.K., et al., *Constraint saliency based intelligent camera for enhancing viewers attention towards intended face.* Pattern Recognition Letters, 2018.
68. Bai, C., et al., *Saliency-based multi-feature modeling for semantic image retrieval.* Journal of Visual Communication and Image Representation, 2018. **50**: p. 199-204.
69. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
70. Woźniak, M. and D. Połap, *Object detection and recognition via clustered features.* Neurocomputing, 2018. **320**: p. 76-84.
71. Lim, S.H. and J. Shin, *Configurable histogram-of-oriented gradients (HOG) processor*. 2018, Google Patents.
72. Žemgulys, J., et al., *Recognition of basketball referee signals from videos using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM).* Procedia computer science, 2018. **130**: p. 953-960.
73. Yan, J., et al., *Classification accuracy improvement of laser-induced breakdown spectroscopy based on histogram of oriented gradients features of spectral images.* Optics express, 2018. **26**(22): p. 28996-29004.
74. Hodas, N.O., *Data object classification using feature generation through crowdsourcing*. 2018, Google Patents.
75. Zhao, L., et al., *Real-time moving object segmentation and classification from HEVC compressed surveillance video.* IEEE Transactions on Circuits and Systems for Video Technology, 2018. **28**(6): p. 1346-1357.
76. Kuang, H., et al. *Defect detection of bamboo strips based on LBP and GLCM features by using SVM classifier*. in *2018 Chinese Control And Decision Conference (CCDC)*. 2018. IEEE.
77. Naganjaneyulu, G., C.S. Krishna, and A. Narasimhadhan. *A Novel Method for Logo Detection Based on Curvelet Transform Using GLCM Features*. in *Proceedings of 2nd International Conference on Computer Vision & Image Processing*. 2018. Springer.
78. Vallabhaneni, R.B. and V. Rajesh, *Brain tumour detection using mean shift clustering and GLCM features with edge adaptive total variation denoising technique.* Alexandria Engineering Journal, 2018.

79. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic acids research, 2003. **31**(13): p. 3812-3814.
80. Guo, F., et al. *Research on image detection and matching based on SIFT features*. in *Control and Robotics Engineering (ICCRE), 2018 3rd International Conference on*. 2018. IEEE.
81. Ibrahim, Z., N. Sabri, and N.N.A. Mangshor, *Leaf Recognition using Texture Features for Herbal Plant Identification.* Indonesian Journal of Electrical Engineering and Computer Science, 2018. **9**(1): p. 152-156.
82. Basu, M., *Gaussian-based edge-detection methods-a survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2002. **32**(3): p. 252-260.
83. Yaseen, M.U., et al., *Cloud-based scalable object detection and classification in video streams.* Future Generation Computer Systems, 2018. **80**: p. 286-298.
84. Lowe, D.G. *Object recognition from local scale-invariant features*. in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. 1999. Ieee.
85. Nosseir, A. and R. Roshdy. *Extraction of Egyptian License Plate Numbers and Characters Using SURF and Cross Correlation*. in *Proceedings of the 7th International Conference on Software and Information Engineering*. 2018. ACM.
86. Haralick, R.M. and K. Shanmugam, *Textural features for image classification.* IEEE Transactions on systems, man, and cybernetics, 1973(6): p. 610-621.
87. Brynolfsson, P., et al., *PV-0527: Gray-level invariant Haralick texture features.* Radiotherapy and Oncology, 2018. **127**: p. S279-S280.
88. Webel, J., et al., *A new analysis approach based on Haralick texture features for the characterization of microstructure on the example of low-alloy steels.* Materials Characterization, 2018. **144**: p. 584-596.
89. Fan, Y., J.C. Lam, and V.O. Li. *Video-based emotion recognition using deeply-supervised neural networks*. in *Proceedings of the 2018 on International Conference on Multimodal Interaction*. 2018. ACM.
90. Guo, Y., et al., *Deep learning for visual understanding: A review.* Neurocomputing, 2016. **187**: p. 27-48.
91. Zhang, C., Q. Huang, and Q. Tian, *Contextual Exemplar Classifier-Based Image Representation for Classification.* IEEE Transactions on Circuits and Systems for Video Technology, 2017. **27**(8): p. 1691-1699.
92. Fei-Fei, L., R. Fergus, and P. Perona, *One-shot learning of object categories.* IEEE transactions on pattern analysis and machine intelligence, 2006. **28**(4): p. 594-611.
93. Yu, W., et al., *Hierarchical semantic image matching using CNN feature pyramid.* Computer Vision and Image Understanding, 2018.
94. Li, Q., Q. Peng, and C. Yan, *Multiple VLAD encoding of CNNs for image classification.* Computing in Science & Engineering, 2018.
95. Gopalakrishnan, R., Y. Chua, and L.R. Iyer, *Classifying neuromorphic data using a deep learning framework for image classification.* arXiv preprint arXiv:1807.00578, 2018.
96. Liu, Q. and S. Mukhopadhyay, *Unsupervised Learning using Pretrained CNN and Associative Memory Bank.* arXiv preprint arXiv:1805.01033, 2018.
97. Zhang, H., J. Xue, and K. Dana. *Deep ten: Texture encoding network*. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. IEEE.
98. La, A., J. Salmon, and J. Ellingson, *Identifying Mode Shapes of Turbo-Machinery Blades Using Principal Component Analysis and Support Vector Machines*, in *Structural Health Monitoring, Photogrammetry & DIC, Volume 6*. 2019, Springer. p. 23-26.
99. Aït-Sahalia, Y. and D. Xiu, *Principal component analysis of high-frequency data.* Journal of the American Statistical Association, 2018: p. 1-17.

100. Paredes, R.K., A.M. Sison, and R.P. Medina, *Developing an Artificial Neural Network Algorithm for Generalized Singular Value Decomposition-based Linear Discriminant Analysis.* Int. J. on Adv. Sc. Eng. Info. Tech, 2018. **8**(3): p. 963-969.2018.
101. Mwangi, B., T.S. Tian, and J.C. Soares, *A review of feature reduction techniques in neuroimaging.* Neuroinformatics, 2014. **12**(2): p. 229-244.
102. Jombart, T., S. Devillard, and F. Balloux, *Discriminant analysis of principal components: a new method for the analysis of genetically structured populations.* BMC genetics, 2010. **11**(1): p. 94.
103. Hyvärinen, A., J. Karhunen, and E. Oja, *Independent component analysis*. Vol. 46. 2004: John Wiley & Sons.
104. Pagola, M., et al. *Use of OWA operators for feature aggregation in image classification*. in *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on*. 2017. IEEE.
105. Song, J., et al., *Structure preserving dimensionality reduction for visual object recognition.* Multimedia Tools and Applications, 2018: p. 1-17.
106. Bui, H.M., et al. *Randomized dimensionality reduction of deep network features for image object recognition*. in *Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), 2018 2nd International Conference on*. 2018. IEEE.
107. Yang, J., et al., *Group-sensitive multiple kernel learning for object recognition.* IEEE Transactions on Image Processing, 2012. **21**(5): p. 2838-2852.
108. Pan, Y., et al., *Locality constrained encoding of frequency and spatial information for image classification.* Multimedia Tools and Applications: p. 1-17.
109. Ryu, J., M.-H. Yang, and J. Lim. *DFT-based Transformation Invariant Pooling Layer for Visual Classification*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
110. Tang, W., et al., *Structured Analysis Dictionary Learning for Image Classification.* arXiv preprint arXiv:1805.00597, 2018.
111. Khan, H.A., *DM-L Based Feature Extraction and Classifier Ensemble for Object Recognition.* Journal of Signal and Information Processing, 2018. **9**(02): p. 92.
112. Yang, J., et al., *Feature fusion: parallel strategy vs. serial strategy.* Pattern recognition, 2003. **36**(6): p. 1369-1381.
113. Sun, Q.-S., et al., *A new method of feature fusion and its application in image recognition.* Pattern Recognition, 2005. **38**(12): p. 2437-2448.
114. Han, J. and B. Bhanu. *Statistical feature fusion for gait-based human recognition*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.
115. Chen, J., et al., *Facial expression recognition in video with multiple feature fusion.* IEEE Transactions on Affective Computing, 2018. **9**(1): p. 38-50.
116. Charte, D., et al., *A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines.* Information Fusion, 2018. **44**: p. 78-96.
117. Yu, Y., et al., *An Unsupervised Convolutional Feature Fusion Network for Deep Representation of Remote Sensing Images.* IEEE Geoscience and Remote Sensing Letters, 2018. **15**(1): p. 23-27.
118. LIAQAT, A., et al., *AUTOMATED ULCER AND BLEEDING CLASSIFICATION FROM WCE IMAGES USING MULTIPLE FEATURES FUSION AND SELECTION.* Journal of Mechanics in Medicine and Biology, 2018: p. 1850038.
119. Sharif, M., et al., *A framework for offline signature verification system: Best features selection approach.* Pattern Recognition Letters, 2018.
120. Khan, M.A., et al., *CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features.* Computers and Electronics in Agriculture, 2018. **155**: p. 220-236.

121. Mangai, U.G., et al., *A survey of decision fusion and feature fusion strategies for pattern classification.* IETE Technical review, 2010. **27**(4): p. 293-307.
122. Nakayama, K. and G.H. Silverman, *Serial and parallel processing of visual feature conjunctions.* Nature, 1986. **320**(6059): p. 264-265.
123. Li, Q., Q. Peng, and C. Yan, *Multiple VLAD encoding of CNNs for image classification.* Computing in Science & Engineering, 2018. **20**(2): p. 52-63.
124. Liu, X., et al., *On fusing the latent deep CNN feature for image classification.* World Wide Web, 2018: p. 1-14.
125. Zhang, H. and Y. Lu. *Image classification by combining local and mid-level features*. in *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence*. 2018. ACM.
126. Qian, Q., et al. *Recognition of pavement damage types based on features fusion*. in *Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on*. 2018. IEEE.
127. Nasir, M., et al., *An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach.* Microscopy research and technique, 2018.
128. Ejbali, R. and M. Zaied, *A dyadic multi-resolution deep convolutional neural wavelet network for image classification.* Multimedia Tools and Applications, 2018. **77**(5): p. 6149-6163.
129. Chen, S., et al., *Local Patch Vectors Encoded by Fisher Vectors for Image Classification.* Information, 2018. **9**(2): p. 38.
130. Zhang, P., et al., *Saliency flow based video segmentation via motion guided contour refinement.* Signal Processing, 2018. **142**: p. 431-440.
131. Gomathi, D. and K. Seetharaman, *Object Classification Techniques using Tree Based Classifiers.*
132. Zhang, C., et al., *Image class prediction by joint object, context, and background modeling.* IEEE Transactions on Circuits and Systems for Video Technology, 2018. **28**(2): p. 428-438.
133. Han, D., Q. Liu, and W. Fan, *A new image classification method using CNN transfer learning and web data augmentation.* Expert Systems with Applications, 2018. **95**: p. 43-56.
134. Mahmood, A., et al. *Resfeats: Residual network based features for image classification*. in *Image Processing (ICIP), 2017 IEEE International Conference on*. 2017. IEEE.
135. Cengil, E., A. Çınar, and E. Özbay. *Image classification with caffe deep learning framework*. in *Computer Science and Engineering (UBMK), 2017 International Conference on*. 2017. IEEE.
136. Akçay, S., et al., *Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery*. 2016, IEEE.
137. Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques.* Emerging artificial intelligence applications in computer engineering, 2007. **160**: p. 3-24.
138. Safavian, S.R. and D. Landgrebe, *A survey of decision tree classifier methodology.* IEEE transactions on systems, man, and cybernetics, 1991. **21**(3): p. 660-674.
139. Friedl, M.A. and C.E. Brodley, *Decision tree classification of land cover from remotely sensed data.* Remote sensing of environment, 1997. **61**(3): p. 399-409.
140. Quinlan, J.R., *Induction of decision trees.* Machine learning, 1986. **1**(1): p. 81-106.
141. Friedman, N., D. Geiger, and M. Goldszmidt, *Bayesian network classifiers.* Machine learning, 1997. **29**(2-3): p. 131-163.
142. Asiri, S. *Machine Learning Classifiers*. Available from: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623.

143. Keller, J.M., M.R. Gray, and J.A. Givens, *A fuzzy k-nearest neighbor algorithm.* IEEE transactions on systems, man, and cybernetics, 1985(4): p. 580-585.
144. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.
145. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity.* The bulletin of mathematical biophysics, 1943. **5**(4): p. 115-133.
146. Tsotsos., A.J.R.ı.-S.a.a.J.K., *The roles of endstopped*

and curvature tuned computations in a hierarchical representation of

2D shape. In: PloS One 7.8 (Jan. 2012), 2012(1932-6203.).
147. Lazebnik, S., C. Schmid, and J. Ponce. *A maximum entropy framework for part-based texture and object recognition*. in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. 2005. IEEE.
148. Lazebnik, S., C. Schmid, and J. Ponce. *Semi-local affine parts for object recognition*. in *British Machine Vision Conference (BMVC'04)*. 2004. The British Machine Vision Association (BMVA).
149. Everingham, M., et al., *The pascal visual object classes challenge: A retrospective.* International journal of computer vision, 2015. **111**(1): p. 98-136.
150. Janoch, A., et al., *A category-level 3d object dataset: Putting the kinect to work*, in *Consumer Depth Cameras for Computer Vision*. 2013, Springer. p. 141-165.
151. Rashid, M., et al., *Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features.* Multimedia Tools and Applications, 2018: p. 1-27.
152. Achanta, R., et al., *SLIC superpixels compared to state-of-the-art superpixel methods.* IEEE transactions on pattern analysis and machine intelligence, 2012. **34**(11): p. 2274-2282.
153. Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International journal of computer vision, 2004. **60**(2): p. 91-110.
154. Hu, F., et al., *Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery.* Remote Sensing, 2015. **7**(11): p. 14680-14707.
155. Sankar, A.S., et al., *Wavelet sub band entropy based feature extraction method for BCI.* Procedia Computer Science, 2015. **46**: p. 1476-1482.
156. Cheng, G., et al., *Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition.* Measurement, 2016. **91**: p. 140-154.
157. Fang, X., et al., *Approximate Low-Rank Projection Learning for Feature Extraction.* IEEE Transactions on Neural Networks and Learning Systems, 2018.
158. Li, C., et al., *Deep Supervision with Intermediate Concepts.* arXiv preprint arXiv:1801.03399, 2018.
159. Grabner, A., P.M. Roth, and V. Lepetit, *3D Pose Estimation and 3D Model Retrieval for Objects in the Wild.*