Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID MODEL TO ENHANCE RECOGNITION OF MULTI-CLASSES IMAGES

[1]MANAL A. ABDEL-FATTAH, [2]HAGAR M. El HADAD, [3]AHMED ELSAYED YACOUP, [4]SHAIMAA S. ABDEL-KADER

[1,3] Faculty of Computers and Artificial Intelligence, Department of Information System, Helwan University, Egypt

[2,4] Faculty of Computers and Artificial Intelligence, Department of Information System, Beni-Suef University, Egypt

E-mail: [1]Manal_8@hotmail.com, [2] hmohamed@fcis.bsu.edu.eg, [3] eng_ahmedyakoup@yahoo.com, [4] Shimaa.sayed@fcis.bsu.edu.eg

ABSTRACT

Nowadays, the volume of digital data increasing very rapidly especially the image datasets. The reason behind this increase is the rapid development of digital technologies and platforms such as Facebook and Instagram etc. From this point of view, the researchers started to build applications based on using image classification models. These models used traditional techniques or deep learning techniques to classify the multi-classes images. Most researchers in the field of image classification concluded that there are different problems such as the wrong classification for objects and low accuracy rate value in the case of using many classes that are found as a result of the classification phase. Focusing on the essential problem is recognizing a huge number of images for different classes with a high accuracy rate. This paper presents an improved model in multi-classes images recognition. This model combines traditional techniques with deep learning techniques where the feature vector of these techniques (VGG16+HOG+SURF) or (ResNet50+HOG+SURF) are combined in one feature vector for classification. The fine-tunning method is used to perform classification by the combined feature vectors to classification layers in ResNet50. VGG16 and ResNet50 are examples of deep pre-trained networks while Histogram of Oriented Gradients (HOG) and Speeded Up Robust Features (SURF) are examples of traditional techniques. The experimental results of the presented model in this paper provide an improvement through an excellent accuracy rate when using a combined feature vector of (ResNet50+HOG+SURF) that reached 98.9% for the recognition of the cifar-10 dataset.

Keywords: *Image Recognition, Feature Extraction, SURF, HOG, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, VGG16, ResNet50, Cifar-10.*

1. INTRODUCTION

Computer vision is one of the most important topics in the technology industry as it relates to the ability of a machine to understand and analyze images and videos. Robots, facial recognition, and self-driving cars are the most important applications that rely on computer vision. Image recognition is considered at the core of computer vision in the task of what an image represents.

Recently, with the remarkable progress of technology and increasing data volumes that vary into text, images, videos, etc. The need for automatic recognition techniques have become the most challenging idea, especially in the image field. Previously, researchers were based on using traditional techniques in image recognition such as the artificial neural network (ANN) and genetic algorithms. Researchers face large problems in classifying dataset that contains many images with multi-classes where the accuracy rate decreases with increasing the number of classes. So, dealing with the huge volume of data with different classes and extracting the most important information to recognize it has become one of the most common issues currently in the field of computer vision. That does not give the required accuracy by traditional machine learning techniques as proved by Y. Y. Wang in [1] and S. T. Krishna in [2]. This encourages many researchers to look at using the latest methods such as deep learning techniques to provide solutions to this problem.

In the early years, the reason for the low rate of accuracy in image recognition using traditional machine learning techniques in a large dataset with multiple categories is that these techniques extract features manually with a limited description [3]. Thus, these traditional techniques such Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) and Speeded Up Robust Features (SURF) can't access high-level features that help in recognizing image more accurately. Because the small length of features vector is not effective in the cases that need to classify objects more than twenty class at the same time. This weak point is solved using deep learning techniques.

On the other hand, N. Sharma, et al., found in [5] that deep learning techniques have seen a breakthrough in the image recognition field in the past few years. Convolutional neural network (CNN) is the most widely used deep learning technique in image recognition which combines the process of feature extraction and classification in the same stage and has the capability to learn features with strong descriptions automatically [6]. This is due to the length of features extracted using CNN is larger than the features vector of traditional techniques. So, the more length of the feature vector the greater number of features per image which means the depth model with more classification accuracy rate. Where the classification is based on several features in addition to distinctive features that are available in deep learning techniques.

Furthermore, recent developments in the convolutional neural network which are represented in transfer learning or what is called pre-trained networks. Pre-trained networks focus on increasing depth of the network, which leads to an increase in the accuracy rate for image recognition, as mentioned in previous research [16] [2]. Increasing the number of layers in the network leads to learning and representing a greater number of complex features with a high level of information which leads to a higher accuracy rate. These pre-trained networks such as AlexNet, VGG, GoogleNet, and ResNet50 have several features that reached 100352 in ResNet50. In contrast to the low number of features extracted from traditional techniques such as HOG which are 324 and this makes it not comparable to this huge number of features learned by deep networks. Given these recent advancements in the convolutional neural network [19], researchers believe it is the future of computer vision.

From this standpoint, this paper presented a model to utilize features of traditional techniques and deep learning techniques based on the fine-tuning method in a combined feature vector. By training the model on this combined feature vector, its efficiency has been demonstrated in recognizing more images with high accuracy.

The proposed model in this paper is based on two directions, one of them is the improvement in pre-trained networks of VGG16, Resnet50, and the second is the combination of the features extracted from these improved pre-trained networks and the features extracted from traditional techniques such HOG and SURF. After that, the classification process is done for involving each image to a specific object, and based on identifying the object, the accuracy of the model is specified.

Improvements in VGG16, Resnet50 pre-trained networks are represented in the change in their architecture with additional layers, using optimization methods, and preprocessing techniques such as image normalization and one-hot encoding for images before extracting features. According to these improvements, images are input to layers of VGG16 and Resnet50 networks that are responsible for feature extraction as described in the following sections. The discriminative and robust feature maps for each network of VGG16, Resnet50 are extracted and flattened to one feature vector.

After that, features are extracted using traditional techniques of HOG and SURF. Preprocessing techniques are used before feature extraction phase to prepare data. The algorithms used for removing noise in this paper are Gaussian blur, bilateral filter algorithm and Histogram equalization algorithm is used for increase brightness transformations in images. However, this paper presents a solution to multi-classes image recognition problem through combination between both traditional features and deep features in one feature vector.

Fine-Tuning method play an important role in this paper where the combined feature vector of (VGG16+HOG+SURF) is entered to classification layers in VGG6 network and froze feature extraction layers. The same steps are followed in combined feature vector of (ResNet50+HOG+SURF) to be entered to classification layers of ResNet50. Accordingly, the proposed model with feature vector of (ResNet50+HOG+SURF) give satisfactory accuracy rate suitable for recognition of multi-classes images.

The rest of this paper is organized as follows; Section 2 discusses Backgrounds, section 3 offers a literature review of previous works, the proposed image recognition model is presented in Section 4, Section 5 presents the experimental results followed by Conclusion and future work in Section 6.

2. BACKGROUND

2.1 Preprocessing Techniques

This phase is very important in increasing the accuracy rate of the classification model and affects the results. The absence of such phase affects the training phase, as it makes it more difficult and consumes a lot of time, also affects the test phase in giving a poor accuracy rate. As proven in [23], the classification of images without using preprocessing techniques gives lower accuracy than accuracy achieved by applying preprocessing techniques on images. This fact and actual results have shown the importance of preprocessing not only in image recognition but also in all fields that are related to image identification, classification, and so on.

There are a lot of pre-processing techniques that need to be taken into consideration. The techniques used in this paper are represented in:

2.1.1 Resize image

Sometimes, images taken from sensors or captured by a camera that is fed to a machine learning technique vary in size. Thus, it is necessary to standardize the size (width, height, and channel) for all images to ensure the accurate results of the used technique.

2.1.2 Gray scaling

The image will be converted from RGB to grayscale where the computer will set the appreciate degree of dark for each pixel. This process reduces the memory required to store grayscale than RGB. Also, a grayscale image is easier than RGB images in many tasks.

2.1.3 Remove noise (De-noise)

Removing noise is done by Gaussian blur or what is called Gaussian smoothing. This function aims to reduce image noise to improve the viewing of an image at different scales. Also, viewing an image in a transparent state differ from an object under regular brightening. Also, the bilateral filter technique can reduce noise very well and preserving edges sharp and enable the SURF technique to extract features.

2.1.4 Histogram equalization

The histogram equalization technique is widely used in image preprocessing to contrast enhancement due to its performance of Pixel brightness transformations on almost all types of images.

2.1.5 Normalizing image

When normalizing input images (training set, testing set) by dividing the pixel values by 255, the neural networks become more efficient while scaling down values to [0,1]. Additionally, they have ability to learn the optimal parameters more quickly.

2.1.6 One-hot Encoding

This process helps in enhancing the performance of machines and deep learning techniques. Where the categorical variable is encoded to a binary variable for each unique integer value that means 0 refers to a non-existent object while 1 refers to its existent.

2.2 Traditional Feature Extraction Techniques

The preprocessed images will be moved to the second phase which is the feature extraction phase that extracts hand-engineered features in the traditional techniques and automatically extracted features of a deep convolutional neural network. The robust features improve the accuracy rate of a model than poor features [23]. These extracted features are categorized based on the objects present in the image.
The most traditional feature extraction techniques [21] that give good results in image recognition are represented in:

2.2.1 Speeded Up Robust Features (SURF)

The first algorithm presented in this paper for extracting features used in image recognition is SURF algorithm. SURF is based on using the Haar wavelet parameter, this parameter gives SURF a great advantage which helps to minimize both matching time and feature extraction. So, SURF

becomes a speed-up version of SIFT and faster than the SIFT algorithm. There are three main steps in the surf algorithm, the first one is keypoints detection, the second step is keypoints description and the third is matching. Initially, the SURF used determinant of Hessian blob to detect interest points using an integral image. In the step of keypoints description, the Haar wavelet response is used to describe features of interest point as a vector for matching based on making square region related to selected orientation. SURF only changes the filter size without any change in image size. The final step is matching, which is often based on a distance between the different vectors.
Hence, equation 1 is an example of input image I with all pixels in a square region are matched to the integral image I (K).

$$I(k) = \sum_{x=0}^{k} \sum_{y=0}^{n} I(x,y) \quad (1)$$

Suppose we have location x, the Hessian matrix of it is described by equation 2.

$$(\mathrm{K}, \sigma) = \begin{bmatrix} \mathrm{L}(\mathrm{K}, \sigma) & \mathrm{L}(\mathrm{K}, \sigma) \\ \mathrm{L}(\mathrm{K}, \sigma) & \mathrm{L}(\mathrm{K}, \sigma) \end{bmatrix} \quad (2)$$

2.2.2 Histogram of Oriented Gradients (HOG)

HOG is the second presented feature extraction technique in this paper, which is used for object detection through a set of characteristics such as its shape or edge directions without any knowledge about its location. HOG descriptor is implemented by calculating the gradients of each pixel in the image location. The gradients are the change in k and n directions. HOG descriptor of a particular image location is implemented based on split image to a set of cells where each cell is assigned to a part of an image. Inside each cell, the edge orientations of all pixels in this cell are collected by measuring the gradient directions. Then, the features extracted using the HOG descriptor are combined in one vector [24].
Equation 3 shows how to use filtering for a grayscale image I to calculate the gradient.

$$G_k = [-1 \quad 0 \quad 1] \quad then \quad I_k = I * G_k$$

$$G_n = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad then \quad I_n = I * \; G_n \quad (3)$$

2.3 DEEP NEURAL NETWORKS

There are two trends to train image recognition models; the first is to design CNN from scratch [10] [12]. While the second is using pre-trained networks based on transfer learning [5] [7] [28]. Pre-trained networks are preferred based on their efficient recognition accuracy and performance as proved in the previous studies results. Consequently, this paper is focused on utilizing pre-trained networks in the proposed model.

Using a Pre-trained Deep Learning Models for image recognition:

Recently, a convolutional neural network has been developed to present more deep pre-trained networks that aim to transfer learning in the image recognition field. The transfer learning approach is preferred in new deep learning applications. Pre-trained networks are a developed type of CNN and have different architectures where each network has its architecture that is represented in the internal layer (number, order) and used techniques [4]. The idea of this approach is to use existing pre-trained networks such as VGG16 or ResNet50 in this paper, and feed new data with unknown classes. One of the most important uses of pre-trained networks is that they are developed to implement on a dataset with large size and give more accurate results in the last years with the development in computational power. The goal behind this development is to decrease time-consuming and increasing accuracy results [18]. This is due to the model has trained on thousands or millions of images such as the ImageNet dataset with 1000 classes. There are a lot of other recognition tasks that can benefit from these networks largely [2].

Pre-trained networks focus on using fine-tuning method by training the network on a new dataset. fine-tuning makes transfer learning faster and easier than constructing a network from scratch because the model has been trained on thousands or millions of images [5]. also, features specific to a new dataset are learned by the network rather than a dataset that has already been learned.

In this section, the most popular pre-trained models such as VGG16 and ResNet50 are some state-of-the-art and are widely used in image recognition tasks [20]. There are other deeper pre-trained models than VGG16 and ResNet50, but these models are chosen in this paper according to available memory resources.

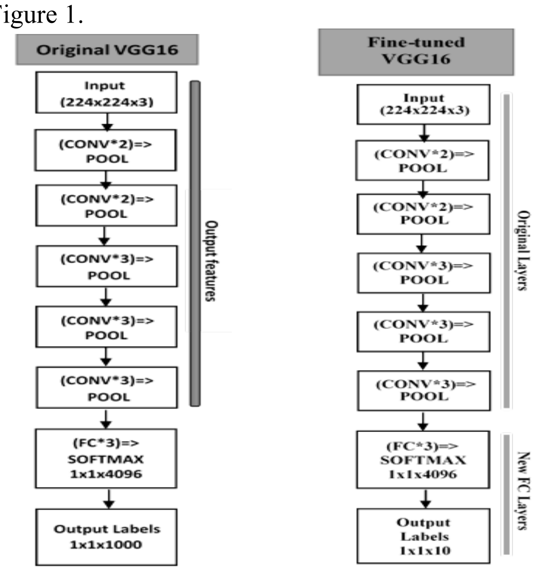
3.1 VGG16

The VGG16 is one of the most popular pre-trained

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

models for image recognition. It was significantly applied by researchers and the industry for their image classification tasks.

The original architecture of VGG16 model [11] and after performing fine-tuning are shown in Figure 1.

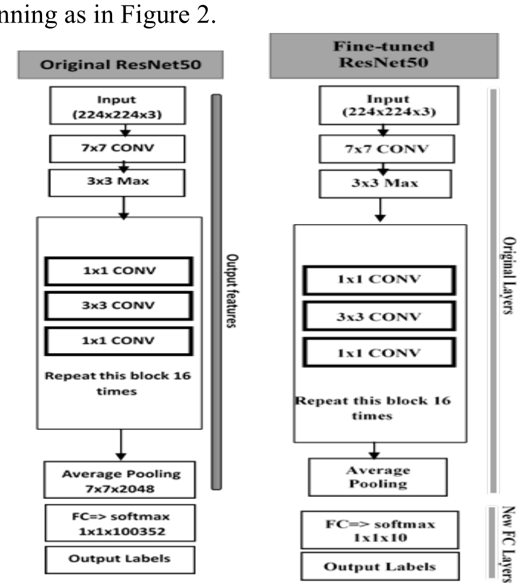


a) VGG16 before fine-tuning b) VGG16 after fine-tuning

Figure 1: Architecture of VGG16 before and after fine-tuning

3.2 ResNet50

The main motivation behind ResNet50 model was to become deeper to solve VGG16 problems of poor accuracy to improve network performance. Additionally, the ResNet50 model aimed to avoid Vanishing Gradient issue. The following is the architecture of ResNet50 before and after fine-tunning as in Figure 2.



a) ResNet50 before fine-tuning b) ResNet50 after fine-tuning

Figure 2: Architecture of ResNet50 before and after fine-tuning

3. LITERATURE REVIEW

Table 1: State of the Art Techniques used in Image Recognition in Last Five Years

S. No	Year	Authors	Methodology	Dataset	Accuracy
1	2018	S. Loussaief, et al.,[8]	Feature extraction techniques (SIFT, SURF, Histogram, BoF) Classifiers (SVM, KNN, BRT) The best techniques are SURF and SVM	Caltech101 With 4 categories	- Cubic SVM: 90% - Quadratic SVM: 88.8% - Bagged Trees: 85% - Weighted KNN:67.5%
2	2018	Y. Y. Wang [1]	nearest neighbor, ANN, CNN, ResNet	Cifar-10	23.9%, 50%, 80%, 92.5%

3	2018	S. Loussaief and A. Abdelkrim [25]	Feature extractor approaches: - BoF (SURF) - CNN(AlexNet) Classifiers (SVM, KNN, DT)	Caltech101	-SVM, Deep learning: 95.2%, BoF: 65.9% -KNN, Deep learning: 96.8%, BoF: 57.1% -DT, Deep learning:76.2% BoF: 36.5%
4	2019	A. D. Akwaboah[12]	neural network, CNN with 8 layers, CNN with 7 layers	Cifar-10	38.18%, 71.81%, 75.43%
5	2018	X. Yang, et al.,[7]	VGG, ResNet , Inception	1719 ALB fish 67 LAG fish	72%
6	2018	N.S. Lele[10]	CNN with 4 layers CNN with 6 layers	Cifar-10	74.57% 75.31%
7	2018	S. Lee, et al.,[15]	CNN + Adaboost	Cifar-10	88%
8	2018	K. Chauhan, et al.,[9]	CNN	cat and dog	90.54%
9	2019	,M. Xin, et al.,[13]	CNN, SVM, NB, KNN, RF, DT, GBDT	Minist Cifar-10	83.67%, 87.63%, 89.36% 72.59%, 79.31%, 69.47%, 76.23%
10	2019	Y. Huang, et al.,[14]	CNN + GPipe	imagenet-2012 cifar100 food101	84.4% 91.3% 93%
11	2018	N. Sharma, et al.,[5]	identifying objects in videos AlexNet, GoogLeNet, ResNet50	ImageNet, cifar-10, cifar100	- Cifar-100 44.10%, 64.6%, 59.82% - Cifar-10 36.12%, 71.67%, 78.1%
12	2018	M. Hussain, et al.,[11]	CNN	Cifar-10 -Test A (10000 training image) - Test B (1000 training image	70.1% in test A 66.1% in test B
13	2017	S. Srivastava, et al.,[26]	Sif t+ deep ensemple	Cifar-10	91.1%
14	2018	M. T. Islam, et al.,[27]	CNN, inception v3	1664food images	82%, 92.86%
15	2018	N. Jmour, S. Zayen and A. Abdelkrim[28]	AlexNet + fine tuning	traffic sign images	93%
16	2018	M. M. Alzorgani and H. Ugail[29]	SVM, KNN, DA BDT, NB	2608 image with 3 classes	99.98 %, 99.98%, 99.81% 97.32%, 93.68%
17	2019	M. A. Abu, et al.,[31]	DNN	3670 flower images with 5 classes	90%
18	2020	P. Wan et al.,[30]	SVM CNN	Mnist COREL1000	- Minist dataset: SVM 88% CNN 98% -COREL1000 SVM 86% CNN 83%
19	2019	R. Madan, et al.,[22]	Classifier features (CNN, HOG), (CNN, SURF) (SVM, HOG)	GTSRB	91.09%, 77.41%, 96.93%
20	2020	Giuste FO, Vizcarra JC [35]	VGG+Inception+HOG	Cifar-10	94.6%

According to the studies that were carried out in the field of image recognition, as mentioned in Table 1, it became clear that the accuracy rate of deep learning techniques is superior to machine learning techniques, especially in the case of recognizing large datasets with multi-classes images. Therefore, although machine learning techniques of SVM and KNN achieved a high accuracy rate reached 99.9%, as mentioned by M. M. Alzorgani and H. Ugail [29] as the size of the dataset in this study is small and the number of categories is few. But in case of increasing number of categories and number of features, the accuracy rate decreased as mentioned by S. Loussaief, et al., [8] and M. Xin, et al., [13] and P. Wan et al.,[30] on Minist and cifar-10 dataset.

Compared to machine learning techniques, increasing the volume of data is an advantage in deep learning techniques, where the more training the model on the images, the greater its accuracy in recognizing, and this is what has been proven by M. Hussain, et al.,[11] and K. Chauhan, et al.,[9]. Thus, most studies began to apply deep learning on a large dataset of multi-classes images, represented by CNN, which achieved a high accuracy rate as P. Wan et al.,[30] showed on the Minist dataset.

M. T. Islam, et al., [27], N. Jmour, S. Zayen and A. Abdelkrim[28], and Y. Y. Wang [1] used pre-trained networks of AlexNet, Inception, and Resnet50 networks to develop deep learning techniques in image recognition. Through their studies, it was noted that ResNet achieves a remarkable accuracy rate reached 92.5% on cifar-10 compared to the rest of the networks and CNN from scratch.

Some researchers presented other models for image recognition that depends on the combination of machine learning techniques and deep learning techniques, as mentioned in Table 1 above. Some of these models provided improvement and increased accuracy rate such as (VGG + Inception + HOG) that presented by Giuste FO, Vizcarra JC [35] and score 94.6% on the cifar-10 dataset, CNN + Adaboost presented by S. Lee, et al.,[15], S. Loussaief and A. Abdelkrim [25] and (CNN, HOG), (CNN,SURF) (SVM, HOG) presented by R. Madan, et al.,[22] in addition to S. Srivastava, et al.,[26].

According to the study of the previous researches result, this paper seeks to present an improved model to help in solving the rest problems in image recognition. This model combines traditional techniques of HOG, SURF to deep neural networks of ResNet50 or VGG16 based to R. Madan et al., [22] that concluded the combining SURF and HOG techniques gives features for better accuracy in addition to the best results achieved by ResNet50 [1]. furthermore, this paper presents some improvements in deep neural networks that impact their performance with a positive effect before using it in the proposed model.

The previous studies mentioned above in Table 1 provided multiple and advanced models that keep pace with the rapid spread of large datasets with multi-classes images and provided solutions to most of the problems to identify these images. However, in recent times there are still limitations in the field of classification of multi-classes images despite noteworthy developments. Table 2 shows the developments that have taken place and the limitations that still need a lot of research.

Table 2: Improvements and limitations in Image Recognition Field

Improvements	Limitations
- The popularity of deep learning to recognize large datasets with different numbers of classes due to the availability of huge datasets and recent development of algorithms.	- Recognition of multi-classes images still needs to improvement to increase accuracy rate.
- With the evolution of computational power and GPUs, deep learning could advance in real-world applications.	- The models are not scalable and can easily misclassify images with unusual objects.
- The ensemble learning that combines more than technique whether machine learning or deep learning models has achieved noticeable progress in image recognition.	- Deep learning algorithms require huge datasets for training the model, but data can already be sparse or unavailable.
- GPipe library solved memory problem for training complex neural networks with large hardware.	- Complex training deep learning models are very costly for the users. Sometimes, they require many resources and expensive GPUs.
- Increasing accuracy rate using pretrained networks.	- The model may be good at mapping inputs to outputs but may not be good at understanding the context of the data.
	- Black Box is challenging for both developers to identify fed of data to the system and users to understand the reason behind any fail.

4. PROPOSED MODEL

In this section, the proposed model shown in Figure 3 consists of three basic phases to perform the multi-classification process.

4.1 Pre-processing Phase

The first critical phase in this section is considered as collecting a dataset and partitioning it into two parts one used for training and the other used for testing. Then, the preprocessing phase will be applied to the training and testing dataset. It is necessary to prepare input images used in the training process by performing the preprocessing techniques whether in traditional machine learning or deep learning models for increasing image contrast and removing noise from the original image in addition to normalizing images.

The input images are preprocessed by converting them to grayscale images in traditional techniques and are resized to 224*224*3 instead of 32*32*3 in deep learning techniques. After that, using morphological techniques such as Gaussian blur and bilateral Filter to remove noise from images that entered to histogram equalization for brightness transformations. The normalization of images impacts significantly in deep learning models to learn networks with suitable values ranges between 0 and 1.

4.2 Feature Extraction Phase

4.2.1 Traditional features

The most common used traditional feature extraction techniques in this paper are HOG and SURF [24]. Where each technique being responsible for extracting a specific type of features as follows: HOG is used to represent a shape feature, while SURF is used to represent a local feature. Furthermore, the length of feature vector extracted from HOG is 324 but after performing pre-processing Gaussian blur and histogram equalization the extracted length increased to 756. The same case in SURF features, before applying a median filter such bilateral filter, the SURF technique can't extract features from some images and output none values but bilateral filter solves this problem. The recent research of feature extraction using traditional matching is based on HOG, SURF techniques due to their better performance.

With increasing the number of classes, feature extraction becomes more difficult in traditional methods. On the other hand, deep learning techniques learn the most prominent features automatically ranging from low to high level from the data that provides the best accuracy when combined with hand-engineered featured.

4.2.2 Deep learning features

The pre-trained networks filters contribute to predict target classes of a specific dataset rather than ImageNet the network already trained on. However, network surge is performed to modify the architecture of deep network.

Extracting features in a pre-trained networks is based on the forward propagation of the input image to the model layers until the final max-pooling layer. Those layers are specified for extracting features and are exist before the fully connected layers (FC) as shown in Figure 1 of VGG16 model. Supposedly, the ResNet50 network to extract distinctive and robust features of cifar-10 images utilizes output features with a shape of 7 x 7 x 2048 which flatten in vector with 100,352 as input-dim. Moreover, fine-tuning outperforms feature extraction by network surgery which serves in removing the FC layers of transfer learning architecture and replacing it with new one that can be fine-tuned to a specific dataset.
Performing fine-tuning on multi-classification networks of VGG16 and ResNet50 are represented in these steps:

a. Loaded the VGG16 or ResNet50 network architecture with weights of pre-trained ImageNet.
b. Original fully connected layer which is the head of the deep network were omitted from the architecture.
c. Originally fully connected layers changed with new freshly initialized ones.
d. Froze all convolution layers in VGG16 or ResNet50.
e. New fully connected layer heads only are trained.
f. Unfroze the final set of convolution layer blocks in the deep network.
g. Continued training.

The proposed model is focused on the combination of hand-engineered features and improved pre-trained networks features in one feature vector to pass into the third phase for classifying image to a specific object.

4.3 Classification Phase

The final phase in this paper is the recognition of cifar-10 objects using the new architecture of VGG16 and ResNet50 networks. Which is summarized in 3 up-sampling layers followed by CONV layers of VGG16 or ResNet50 models by adding AveragePooling2D ((7, 7)) as the last layer. The features extracted from these layers is passed to flatten layer in one dimension vector. The original fully connected layers are removed and the new one is added at the head of the model after the change in its architecture. Thus, the classification process is done based on this model.

Consequently, features of these adjusted pre-trained models are combined with HOG and SURF features into one vector and input this vector to a new fully connected neural network that structures in 3 dense layers. Each dense layer of them is previous by normalization layer and followed by dropout layer. Finally, softmax output layer is used to predict image object based on its features which achieve the purpose of this paper. The proposed model in this research used fine-tuning method in the feature extraction part for VGG16 and ResNet50 networks and then compares the accuracy rates of the classification process that depends on combined features with traditional features. This is the continuous contribution of the authors. Then the authors compared this paper model and the previous model accuracy rates, it is observed that paper model are considered one of the best accurate models in image recognition.

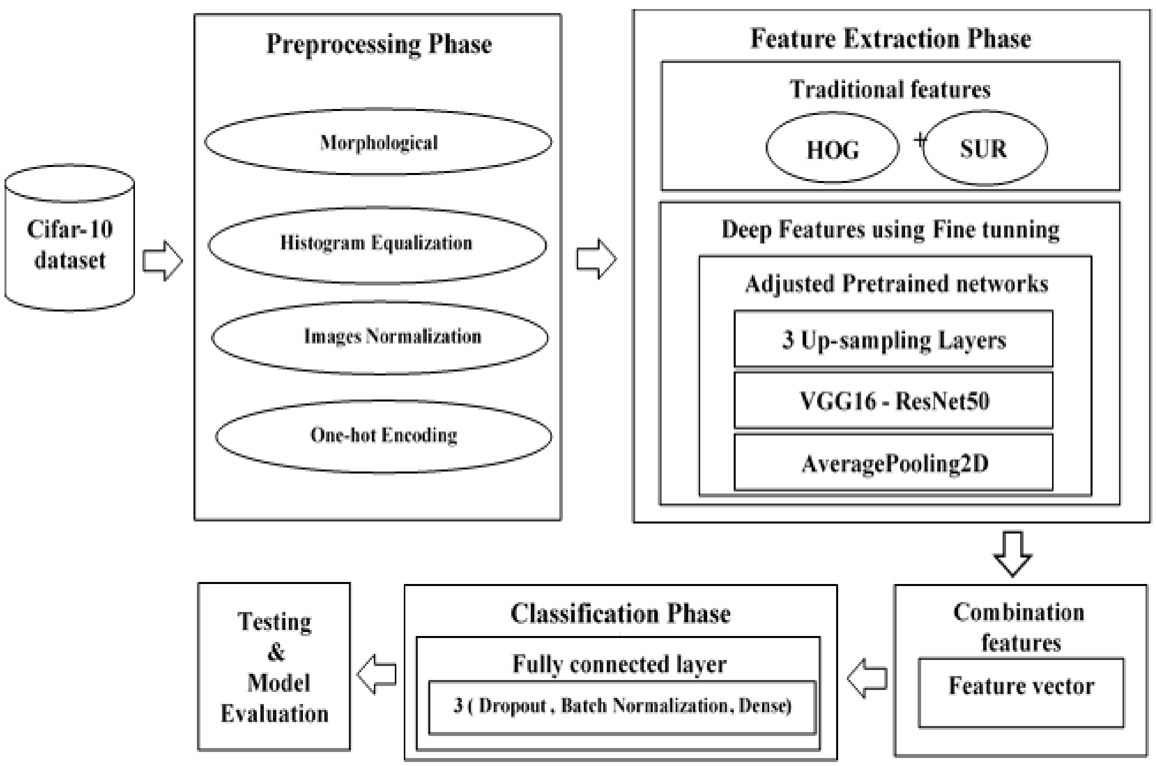


Figure 3: Proposed Hybrid Model Based on Deep and Traditional Features

5. EXPERIMENTAL RESULTS

5.1 Dataset

This paper based on the CIFAR-10 dataset which consists of 60,000 images that are color (RGB) images with small size 32x32 pixels. It is divided into 50000 for the labeled training set and 10000 for the unlabeled testing set. These images are classified into 10 classes of the same size where each class contains 6,000 images and the classes

match objects such as (ships, cars, airplanes, etc.) and animals such (cats, dogs, frogs, etc.) [3].

5.2 Evaluate Results

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The experimental results demonstrate the effect of preprocessing techniques and the change in the deep neural networks architecture on this paper's model. Fortunately, this paper model has worked efficiently and has taken less time for training. Also, the combination of pretrained models feature with traditional features is practical efficient in learning deep neural networks well.

First Experiment: performing preprocessing on the input images with change in VGG16 and ResNet50 models

Table 4 shows the increase in accuracy rate and decrease in error rate after performing pre-processing techniques in input cifar-10 images as mentioned above. In addition to using the fine-tuning method by loading VGG16 or ResNet with just the convolutional layers and not the dense layers. As new dataset will be trained on the new fully connected dense layers. The architecture of VGG16 and ResNet50 networks is changed to begin with three Upsampling Layers that upscale input image size three times before pass it to VGG16 or ResNet50 models that end with AveragePooling2D ((7, 7)) add to their architecture. After that, the images go through model architecture to be flattened into feature vectors that passed to a fully connected layer with three dense layers of 512, 128, and 64 neurons respectively. each dense layer previous with batch normalization and dropout layer until reach to softmax layer that is responsible to output target classes. This adjusted architecture has contributed significantly to improve the VGG16 and ResNet50 model accuracy rate.

Table 3 shows the difference between previous models that score 89% and 92% when using VGG16 and ResNet50 models respectively while the improved VGG16 and ResNet50 models of this paper provides an enhancement in the accuracy rate to reach 97.7 for VGG16 as shown in Figure 4 and 98.5 for ResNet50 with 5 epochs and batch size 20 with Adam optimizer as shown in Figure 5. Comparing this accuracy rate with the accuracy rate of the authors' previous work in which they use the VGG16 and ResNet50 model can't be compared [33] [34]. Which means that the proposed model made a huge difference in the accuracy rate. Also, the feature vector of this model becomes very robust and distinctive which helps in increasing the model accuracy rate.

Table 3: Accuracy Rate and Error Rate of Original VGG16 and ResNet50 Models

Model	Accuracy Rate (%)	Error Rate (%)
VGG16	89%	0.38
ResNet50	92%	0.32

Table 4: *Accuracy Rate and Error Rate of Adjusted VGG16 and ResNet50 Models with pre-processing*

Model	Accuracy Rate (%)	Error Rate (%)
Adjusted VGG16	97.7%	0.06
Adjusted ResNet50	98.5%	0.04

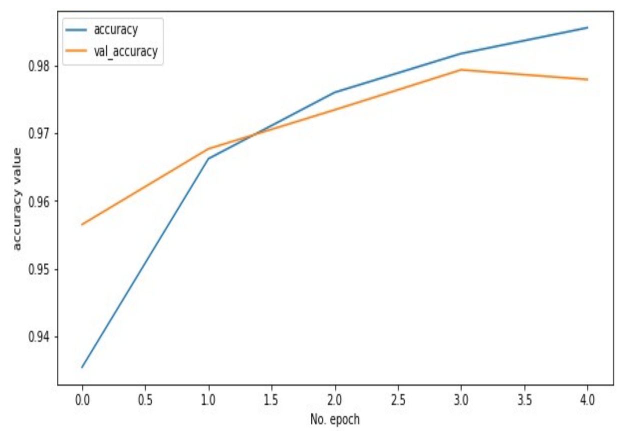


Figure 4: Accuracy Rate of improved VGG16 model

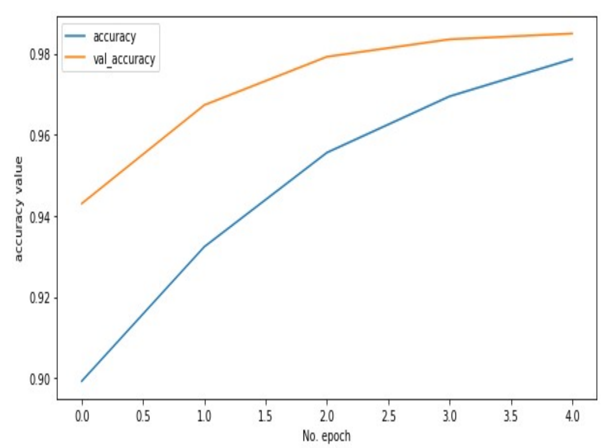


Figure 5: Accuracy Rate of improved ResNet50 model

Second Experiment: combination of improved VGG16 and ResNet50 models features with traditional features as (VGG16+HOG+SURF), (ResNet50+HOG+SURF)

Table 5 comparing between improved VGG16 and ResNet50 models in the first experiment after combining its features with traditional features in one feature vector. Common features that specify image characteristics include color, texture, shape features of one class that distinguish them from other classes. To recognize the category of an image, the features extracted from traditional techniques or deep learning models can not only represent the image well. So, the combined feature vector overcomes this problem and presents discriminative features that help the model to be trained very well and avoid overfitting problems. So, the feature vector of VGG16 which is 512 is combined with 756 features of HOG and 64 features of SURF as shown in Figure 6 to increase the accuracy rate to 98% and reduce loss value than first experiment. At the same time, 100532 features of ResNet50 which combined to HOG and SURF features reached 98.9% accuracy and with less as shown in Figure 7.

Table 5: Accuracy Rate and Error Rate of Combined traditional and improved deep Features

Model	Accuracy Rate (%)	Error Rate (%)
ImprovedVGG16+HOG+SURF	98%	0.04
ImprovedResNet50+HOG+SURF	98.9%	0.03

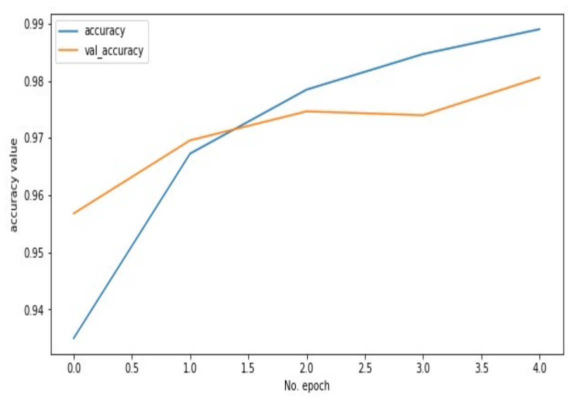


Figure 6: Accuracy and Error Rate of (Improved VGG16+HOG+SURF)

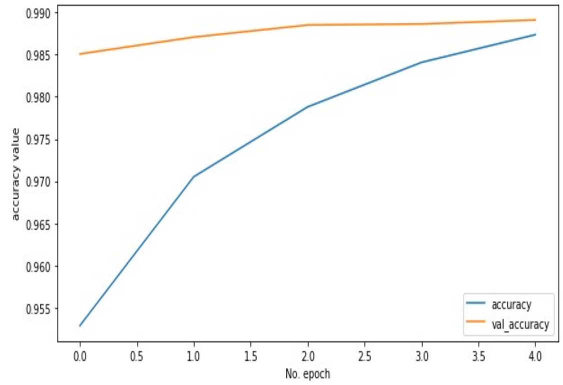


Figure 7: Accuracy and Error Rate of (Improved ResNet50+HOG+SURF)

In a comprehensive comparison between the results of the proposed model and the results of the latest previous studies as shown in Table 6, it turns out that there are some studies that have also combined the features of machine learning techniques and the features of deep learning, but their results are weak compared to the results of this proposed model. This is due to the good effect of other techniques the proposed model relied on in preprocessing and feature extraction, as well as development in deep learning networks architecture, which has a significant role in classification results. This paper is based on VGG16 and ResNet50 networks especially due to their good results in much previous research and because the computational power needed by these networks is within reach.

Table 6: Comparison Between Accuracy Rate of Previous Models and Proposed Model on CIFAR-10 Dataset

Model	Accuracy Rate (%)
CNN + Adaboost [15]	88%
SIFT + deep Ensemble [26]	91.1%
VGG + Inception + HOG [35]	94.6%
Proposed Model	98.9%

6. CONCLUSION AND FUTURE WORK

This paper has presented an enhanced model in the image recognition field. The proposed model has been contributed to present solutions to computer vision problems that focus on increasing the recognition accuracy rate of multi-classes images. The contributions of the proposed model are represented in the higher accuracy rate and good performance in addition to less time required for training. There are two reasons behind these contributions: adjustments in pre-trained networks architecture. After that, the combination of VGG16,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ResNet50 pre-trained network features with traditional features of HOG and SURF in one feature vector that is represented in (VGG16+HOG+SURF) and (VGG16+HOG+SURF). This significantly affects the model performance and accuracy rate. The experimental results have proved that through many experiments that are implemented on the cifar-10 dataset. The accuracy rates achieved by original VGG16 and ResNet50 networks on cifar-10 are 89% and 92 % respectively. The contribution of adjustments in VGG16, ResNet50 pre-trained networks are appeared in the improvement and increasing testing accuracy rate to 97.7% for VGG16 and 98.5 for the ResNet50 network. Moreover, the contribution of the combination between features of adjusted VGG16, ResNet50 networks, and HOG, SURF features improve the model in the last experiment to score 98% for (VGG16+HOG+SURF) and 98.9% for (VGG16+HOG+SURF) respectively using the fine-tuning method. This is the continuous contribution for the authors and one of the best accurate image recognition models if the authors compare this model with the previous model's accuracy rates.

By comparing the experimental results in this paper with the previous research area results, this paper has achieved a certain optimized improvement forward to open the key to improved ideas and solutions in the image recognition field. Therefore, it's recommended to perform the proposed model on a larger dataset with larger classes than the dataset of this paper. Also, there are still many shortcomings that authors forward to completed in the future works such as Combine CNN and LSTM may help in achieving better results and efficiency in image recognition models; Deep learning is expected to focus on unsupervised and reinforcement learning; Developing advanced deep learning models for speech recognition; Optimization in deep neural networks is required for more accurate results and with less running time in addition to Object detection and segmentation in videos.

REFERENCES

[1] Y. Y. Wang, "Image Classification on Cifar-10 Dataset", International Journal of Scientific Research Engineering & Technology (IJSRET), vol. 7, no. 6, 2018, pp. 2278-0882.

[2] S. T. Krishna, and H. K. Kalluri, "Deep learning and transfer learning approaches for image classification", International Journal of Recent Technology and Engineering (IJRTE), vol. 7, no. 5S4, 2019, pp. 2277-3878.

[3] M. M. krishna, M. Neelima, M. Harshali, and M. Venu, " Image classification using Deep learning ", International Journal of Engineering & Technology, vol. 7, no. 2.7, 2018, pp. 614-617.

[4] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Paper of Deep Learning and Its Applications: A New Paradigm to Machine Learning", Archives of Computational Methods in Engineering, Springer, 2019, pp. 1071-1092.

[5] N. Sharma, V. Jain, and A. Mishra, "An Analysis of Convolutional Neural Networks for Image Classification", International Conference on Computational Intelligence and Data Science (ICCIDS), vol. 2018, no. 132, 2018, pp. 377-384.

[6] M. Ramprasath, M.V. Anand, and S. Hariharan, "Image Classification using Convolutional Neural Networks", International Journal of Pure and Applied Mathematics, vol. 119, no. 17, 2018, pp. 1307-1319.

[7] X. Yang, Z. Zeng, S. G. Teo, L. Wang, V. Chandrasekhar, and S. Hoi, "Deep Learning for Practical Image Recognition: Case Study on Kaggle Competition", Association for Computing Machinery (ACM), 2018, pp. 19–23.

[8] S. Loussaief, and A. Abdelkrim, "Machine Learning framework for image classification", Advances in Science, Technology and Engineering Systems Journal (ASTESJ), IEEE, vol. 3, no. 1, 2018, pp. 1-10.

[9] K. Chauhan, and S. Ram, " Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensorflow and Keras ", International Journal of Advance Engineering and Research Development (IJAERD), vol. 5, no. 2, 2018, pp. 2348-4470.

[10] N.S. Lele, **"**Image Classification Using Convolutional Neural Network**",** International Journal of Scientific Research**,** vol. 6, no. 3, 2018, pp. 22-26.

[11] M. Hussain, J. J. Bird, and D. R. Faria, "A Study on CNN Transfer Learning for Image Classification", In UK Workshop on Computational Intelligence, Springer, Cham, 2018, pp. 191-202.

[12] A. D. Akwaboah, "Convolutional Neural Network for CIFAR-10 Dataset Image Classification", book, 2019.

[13] M. Xin and Y. Wang, "Research on image classification model based on deep convolution neural network ", Springer Open, EURASIP Journal on Image and Video Processing, 2019, pp. 1-11.

[14] Y. Huang, Y. Cheng, A. Bapna, O. Firat, M. X. Chen, D. Chen, H. Lee, J. Ngiam, Q. V. Le, Y. Wu, and Z. Chen, "GPipe: Easy Scaling with Micro-Batch Pipeline Parallelism", 2019, pp. 1-11.

[15] S. Lee, T. Chen, L. Yu, and C. H. Lai, "Image classification based on the boost convolutional neural network", IEEE Access, Vol. 6, PP. 12755-12768, 2018.

[16] W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li, "Development of convolutional neural network and its application in image classification: a paper", Optical-Engineering (Opt. Eng), vol. 58, no. 4, 2019, pp. 040901-19.

[17] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review", Hindawi Computational Intelligence and Neuroscience, vol. 2018, 2018.

[18] S. T. Krishna, and H. K. Kalluri, "Deep learning and transfer learning approaches for image classification", International Journal of Recent Technology and Engineering (IJRTE), vol. 7, no. 5S4, 2019, pp. 2277-3878.

[19] R. Aarthi and S. Harini, "A Paper of Deep Convolutional Neural Network Applications in Image Processing", International Journal of Pure and Applied Mathematics, vol. 118, no. 7, 2018, pp. 185-190.

[20] F. Sultana, A. Sufian, and P. Dutta, "Advancements in Image Classification using Convolutional Neural Network", International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), (IEEE Xplore), 2019.

[21] S.Routray, A. K. Ray and C. Mishra, "Analysis of Various Image Feature Extraction Methods against Noisy Image: SIFT, SURF and HOG", IEEE, 2017.

[22] R. Madan, D. Agrawal, S. Kowshik, H. Maheshwari, S. Agarwal and D. Chakravarty, "Traffic Sign Classification using Hybrid HOG-SURF Features and Convolutional Neural Networks", 8th International Conference on Pattern Recognition Applications and Methods ,2019.

[23] M. M. Alzorgani and H. Ugail, "Comparative Study of Image Classification using Machine Learning Algorithms", The 2nd Annual Innovative Engineering Research Conference (AIERC), 2018.

[24] Y. Kortli, M. Jridi, A. Al Falou and M. Atri, "A comparative study of CFs, LBP, HOG, SIFT, SURF, and BRIEF techniques for face recognition," Proc. SPIE 10649, Pattern Recognition and Tracking, 2018.

[25] S. Loussaief and A. Abdelkrim, "Deep Learning vs. Bag of Features in Machine Learning for Image Classification", IEEE, International Conference on Advanced Systems and Electric Technologies (IC_ASET), 2018.

[26] S. Srivastava, P. Mukherjee, B. Lall and K. Jaiswal,"Object Classification using Ensemble of Local and Deep Features", arXiv:1712.04926v1 [cs.CV] 4 Dec 2017.

[27] M. T. Islam, B.M. Siddique, S. Rahman and T. Jabid, "Image Recognition with Deep Learning", IEEE, ICIIBMS 2018, Track 2: Artificial Intelligent, Robotics, and Human-Computer Interaction, Bangkok, Thailand, 2018, pp. 106-110.

[28] N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional Neural Networks for image classification", IEEE 978-1-5386-4449-2/18/$31.00 ©2018, 2018, pp. 397- 402.

[29] M. M. Alzorgani and H. Ugail, "Comparative Study of Image Classification using Machine Learning Algorithms", The 2nd Annual Innovative Engineering Research Conference (AIERC), 2018.

[30] Wang, Pin, En Fan, and Peng Wang. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning." Pattern Recognition Letters 141, 2021, pp. 61-67.

[31] M. A. Abu, N. H. Indra, A. H. Abd Rahman, N. A. Sapiee and I. Ahmed, "A study on Image Classification based on Deep Learning and Tensorflow" , International Journal of Engineering Research and Technology, ISSN. 0974-3154, vol. 12, no. 4, 2019, pp. 563-569.

[32] Tian, Youhui. "Artificial intelligence image recognition method based on convolutional neural network algorithm." IEEE Access 8, 2020, pp:125731-125744.

[33] Chen, Jianqiu. "Image Recognition Technology Based on Neural Network." IEEE Access8, 2020, pp: 157161-157167.

[34] Ma, Chao, Shuo Xu, Xianyong Yi, Linyi Li, and Chenglong Yu. "Research on image classification method based on DCNN." In 2020 International Conference on Computer Engineering and Application (ICCEA), IEEE, 2020, pp. 873-876.

[35] Giuste FO, Vizcarra JC. "CIFAR-10 Image Classification Using Feature Ensembles". arXiv preprint arXiv:2002.03846. 2020.