International Journal on Emerging Technologies **11**(3): 303-312(2020)
ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Comparative Analysis of Applying Object Detection Models with Transfer Learning for Flower Species Detection and Classification

Isha Patel[1] and Sanskruti Patel[2]
[1]Research Scholar, Faculty of Computer Science and Applications,
Charotar University of Science and Technology (CHARUSAT), Changa-388421 (Gujarat), India.
[2]Assistant Professor, Faculty of Computer Science and Applications,
Charotar University of Science and Technology (CHARUSAT), Changa-388421 (Gujarat), India.

(Corresponding author: Isha Patel)
(Received 22 February 2020, Revised 12 April 2020, Accepted 14 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Flower species identification refers to a process of comparing defined characteristics of a given flower to allocate a particular species to a known taxonomic group. Flowers can be identified and classified by observing certain distinguishing basic and morphological characteristics. Classifying flower species is challenging for people and needs in-depth specialist knowledge as some flower species look similar, whereas some look differently despite of being in the same species. Traditional computer vision methods remain inefficient and less accurate while considering environmental complexity and similarity and difference between flowers species. Deep CNN, an emerging field of machine learning and artificial intelligence, has grown rapidly and widely applied in computer vision applications with promising results, especially in the field of object detection from visual images. In this paper, we have performed a comparative analysis of the performance of various object detection models. We have compared; SSD quantized model (8-bit) using MobileNet V1 and MobileNet V2, Atrous model using Faster R-CNN with Inception ResNet V2, low proposals model using Faster R-CNN with ResNet 50 and NAS, atrous and low proposals models using Faster R-CNN with ResNet 101 and Inception ResNet V2 with the proposed NAS-FPN with modified Faster R-CNN model. Based on the results obtained during experiment, the proposed NAS-FPN with modified Faster R-CNN model achieved good performance and highest mAP score of 87.6% on F102 flower class and 96.2% on J30 flower class datasets.

Keywords: object detection, transfer learning, faster r-cnn.

I. INTRODUCTION

Computer vision is a field of study that makes computer enable to understand the content of digital images. Deep learning has recently been successfully used for the automated analysis of various types of images. Deep leaning offers a variety of architecture that successfully employed for different computer vision tasks like image segmentation, object detection and classification. Convolutional neural network (CNN) is one of the variants of artificial neural networks (ANN) and it is a part of wider family of deep learning architectures. CNN is heavily used in the field of computer vision specifically for classification, detection, and segmentation tasks [1]. Moreover, in computer vision, object detection is problems that identifies and localize an object of specific classes in an image. Object detection can be single class object detection or multiclass object detection. In case of presence of only one object in an image, it is termed as single class object detection. When there is availability of more than one object belong to different classes, it is known as multi class object detection. There are varieties of object detection models available based on deep convolutional neural networks such as YOLO (You Only Look Once), Single Shot Detector (SSD) and Faster R-CNN (Regional Convolutional Neural Network) have been proposed for object detection with high accuracy [2]. For training and tuning of the model, deep learning algorithms always need large datasets and powerful resources and both the requirements have been satisfied in the field of agriculture.

The flower species identification indeed useful for professionals, example are farmers, botanists, architects, foresters, ecologists, and public. It is highly difficult to identify any flower with its specifications by a common person. Many activities like flower species conservation and monitoring, impact of climate change on these species and diseases & growth monitoring are highly dependent on correct identification and classification of flower species [3]. Taxonomists and specialists are striving to identify effective techniques for flower species identification and classification. Recently, deep CNN, an emerging field of machine learning and artificial intelligence, has grown rapidly and widely applied to many domains with promising results, especially in the field of object detection from visual images [4]. Flower species detection comprises with two problems; image localization and image classification. It is required to detect a flower in the image as an object and then recognize which species it belongs to [5]. Motivated by the research outcomes obtainable for applying object detection using deep CNN models in various computer vision tasks, in this research, we are aimed to apply and evaluate deep CNN based object detection models for flower species classification with localization.

There are variety of deep CNN based models are available for object detection. This paper presented the different methodologies and architectures available for object detection from flower species images. In this paper, we presented and evaluated the performance of the four different object detection models; a) quantized, b) atrous, c) low proposals and d) proposed NAS-FPN with modified Faster R-CNN. As a base architecture, we have taken SSD and Faster R-CNN. For transfer learning, we need to use pre-trained CNN models. For that, Inception ResNet V2, ResNet 101, NAS, ResNet 50, MobileNet V1 and V2 are considered during experiment and TensorFlow library is used for implementation. During our experiment, we have found that proposed NAS-FPN with modified Faster R-CNN provides promising results on flower species dataset including normal flowers and wild flowers. The model was able to extract low level and high level features efficiently and from generated feature map, the localization and classification performed more accurately.

The rest of the paper is organized as follows. Section II presented a literature survey of the related work. The flower detection models; quantized model, atrous model, lowproposals model, and the proposed NAS-FPN with modified Faster R-CNN model discussed in section III. Section IV contains the details regarding the experimental setup and flower datasets used. Section V described the training loss for different object detection models and it also discussed experimental results and performance evaluation metrics.

II. LITERATURE REVIEW

Aminur *et al.,* [6] performed a skin cancer detection using the atrous convolution with the transfer learning technique. They used the HAM10000 dataset with seven classes and 10015 total dermoscopic images. They have applied MobileNet, InceptionV3, VGG19, and VGG16 deep learning architectures and achieved 102.42%, 89.81%, 85.02%, and 88.22% accuracy. Anup *et al*., [7] applied DWT technique using fusion method. For the medical diagnosis, they used MRI and PET scan variants. For the performance measures, they used PSNR and MSE with values 50.10 and 0.012 respectively. Hongshan *et al*., [8] presented three categories of methods that are based on hand-engineered features, learned features and weakly supervised learning. Also, they described new segmentation algorithms for aiming datasets. For detection and classify the malaria disease, Laxmi *et al*., [9] proposed deep CNN onthe images of blood samples. They carry out an experiment on 13,779 Parasitized and Uninfected blood cell image dataset with accuracy of 97%.S. Siddesha *et al.,* [10] represented their work for detecting and computing the disease area of raw coconut. For extraction of ROI, they used color threshold and K-means clustering. To conduct an experiment, they used 100 self-created disease coconut image dataset. The segmentation method achieved 59.55% affected area of coconut. Prateeth *et al.,* [11] presented the quantization model using large datasets like ImageNet. That achieved as low as 3-bit precision without affecting accuracy. They have shown object class clustering with lower bit-precision quantization. They used an 8-bit and 6-bit quantized model and achieved the 28.22 and 28.09 accuracy of SSD MobileNet and 46.10 and 45.52 accuracy for tiny YOLO V2. Liang-Chieh *et al.* [12] presented semantic image segmentation with Deep Learning and make three main contributions which are; atrous convolution, atrous spatial pyramid pooling (ASPP), and combined methods from DCNNs and probabilistic graphical models. The proposed “DeepLab” model achieved 79.7% mIOU. Darshana *et al*., [13] detected the intruders using reinforcement learning (Qlearning) techniques. They have achieved 99.1008, 95.0773, 98.5016, and 99.0996 values for sensitivity, specificity and accuracy respectively. Zhishuai *et al*., [14] proposed a novel SSD model with Enriched Semantics for image segmentation. The experimental results on both PASCAL VOC and MS COCO detection datasets demonstrated the effectiveness of the proposed method. They applied VGG16 based DES and achieved 81.7 and 32.8 mAP on both datasets. Liang- Chieh *et al.,* [15] presented an atrous convolution model of semantic image segmentation. The proposed ‘DeepLabv3’ system significantly improves over our previous DeepLab versions without DenseCRF post-processing and achieved 79.7 % accuracy.

From the literature review conducted, it has observed that several researchers conducted study to apply the digital image processing, machine learning algorithms and deep learning networks for flower species classification. As a result, this research aimed to develop an object detection model that provides better detection of the flower species with optimum accuracy and localization with multi-class classification.

III. OBJECT DETECTION MODELS FOR DETECTING AND CLASSIFYING FLOWER SPECIES

In this section, we have mentioned the details of all four models that can be used for object detection. First; the details are provided for SSD object detection with the MobileNet V1 and MobileNet V2 based on the quantized model. Second; the atrous convolutional model is explained that is available in Faster R-CNN model using the pre-trained Inception ResNet V2. Third; the low proposals model is explained that is available with pre-trained models ResNet 50 and NAS with Faster R-CNN. Also, the combination of atrous and low proposals using pre-trained ResNet 101 and Inception ResNet V2 with Faster R-CNN is evaluated. And fourth; we have proposed an integrated model based on NAS-FPN and Faster R-CNN using pre-trained ResNet 50 V1. Based on the experiment carried out and the results obtained, the proposed integrated methodology is able to detect, locate and classify flower species with optimum accuracy.

A. Quantized Model

In recent years, the success of deep neural network in various applications like natural language processing, computer vision, speech translations, etc. is due to train large datasets on a large computational model. Also, different variety of learning tasks like; object detection, segmentation, instance segmentation (masking) has performed with better accuracy, while depending on high computational memory power [16]. This makes it difficult to deploy on embedded devices with limited resources. This problem can be solved through the quantized model. In this model, deep convolutional neural network can be transformed from a 32-bits FP (floating point) to an 8-bits INT (integer point) model [17]. The important use of converted deep CNN of the low bit is, it needed small memory with minimal loss. The quantized model is applied to one-stage detection models as the quantization is mainly used with a single-shot convolutional network, not for the region-based convolutional network. Therefore, in this paper, we applied the SSD (single shot detector) object detection model with MobileNet V1 and MobileNet V2 quantized model [18 and 19]. An SSD-based object detection model mainly consists of four modules; pre-processing, feature extractor, bounding box predictor and post-processing. The standard (32-bits FP) SSD with MobileNet V1 and V2 models is converted into quantized 8-bit integer models of pre-trained SSD MobileNet V1 and V2 on the TensorFlow COCO dataset. This quantized model is consistent with the overall design of backbone architecture (MobileNet V1 and V2) and it is more computationally efficient as 32-bits Floating Point is more time-consuming than 8-bits Integer. The following figure 1 illustrated the SSD quantized model.

The SSD quantized model preserves the computation process for the pre-processing, the bounding box predictor and the post-processing parts with 32-bits for maintaining the accuracy, which is illustrated in figure 1. It only changed the feature extractor part. As per integer quantization scheme, the feature extractor part is quantized from 32-bits to 8-bits. After partial quantization scheme is applied, a typical building block in feature extractor is appeared like figure 1.

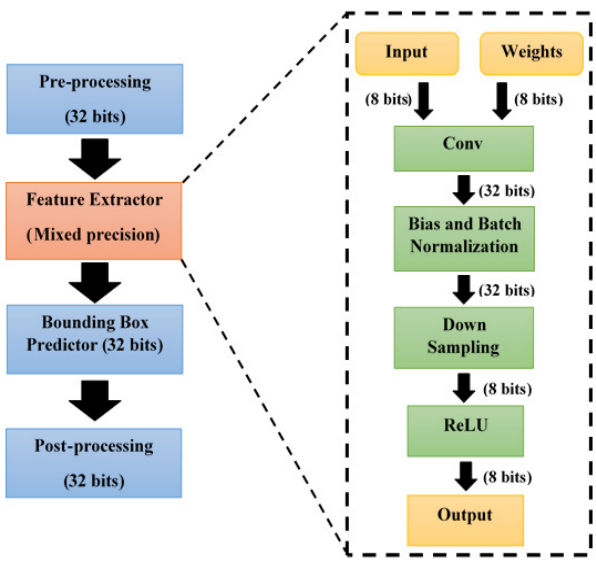

Fig. 1. The Components of SSD Quantized Model.

B. Atrous Model

Based on the framework of Faster R-CNN, another approach called Atrous Region Proposals is considered to enhance the efficiency of multi-scale region proposal search in the feature extraction layer [20 and 21]. In object detection, atrous convolution is also known as dilated convolutional. Specifically for pixel-wise prediction tasks, such as image segmentation, object detection, and semantic segmentation, atrous convolution has demonstrated its significant performance [22]. The aim of dilated convolution is to insert hole or "zeros" between the matrix and kernels of convolutional layers to improve the resolution of image in deep CNN [23 and 24]. The following figure 2 described the scenario of convolutional layers for atrous convolutional.

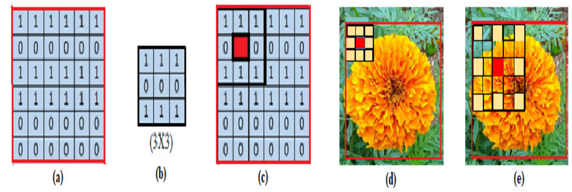

Fig. 2. A scenario of convolutional layers for atrous convolutional kernel size 3X3, stride is 1; (a) binary feature map convolutional matrix 6 × 6 (b) kernel matrix size 3 × 3 (c) baseline CNN with kernel size = 3 × 3, rate = 1 (d) scenario of flower image convolutional with the rate = 1, kernel size = 3 × 3 and (e) scenario of atrous convolutional flower image with kernel size = 3 × 3, rate = 2.

The Fig. 2 represented atrous convolution where each small square represents a pixel and atrous CNN uses dilated convolutional kernels. In the figure 2 (d) and (e), for convolution operation, red pixel is seen as the center pixel. The yellow pixels are associated with the center pixel and covered by the atrous convolution. It means that the feature of yellow pixels can be sampled at that location. The others are the holes from which the convolution cannot learn the feature of the location [25 and 26]. In this paper, we have set the stride = 1, rate = 2 and the kernel size = 3X3 for dilated convolutional layer for flower species detection. To obtain larger scale feature information, we used InceptionResNet V2 using atrous convolution in the high-level feature extraction process. The formula for atrous convolution is defined as follows;

$$m[i] = \sum_{j=1}^{j} n[i + r.j]\,p[j] \quad (1)$$

Where m[i] is output signal, n[i] is input signal with a filter p[j] of length j, r resembles to the dilation rate to sample n[i], and standard convolution is a special case for the rate r = 1, 2, 4, etc. It allows enlarging the receptive field without increasing the number of parameters.

C. Low proposals Model

One of the other object detection models is the low proposals model that is faster than the standard CNN model. Low proposals is used with the two-stage detection model. In this paper, we have experimented Faster R-CNN with ResNet 50 and NAS low proposals models for flower species detection. In this scenario, the proposals are decreasing in RPN (Region Proposal Network) which is affected by the model performance and accuracy. Based on the backbone architecture, we have decreased the 200 proposals of RPN, finding that it significantly decreases the time to convergence. This modification is consistent with the overall design of backbone architecture and makes them more computationally efficient.

Moreover, we have also experimented the combination of atrous and low proposals. Dilated convolution aims to insert a hole between the kernels of convolutional layers to enhance the resolution of the image in deep CNNs. Furthermore, low proposals and dilated convolution used to increase the field of observation of the matrix with an identical amount of computational cost. So, it is very much useful with some images which cannot afford multiple convolutions or larger matrix. In this paper, InceptionResNet V2 and ResNet 101 backbone architectures are used for the atrous convolutional and low proposals combination models with Faster R-CNN [27]. It changed all the standard convolutions of backbone architecture in the Faster R-CNN prediction layers [28].

D. NAS-FPN with Modified Faster R-CNN Model

In this research, we have evaluated the proposed NAS-FPN (Neural Architecture Search-Feature Pyramid Network) with modified faster R-CNN model using a transfer learning approach for flower classification and localization [29]. It is based on the TensorFlow pre-trained object detection model on COCO (Common Objects in Context) dataset. The following Fig. 3 illustrated the proposed model that aims to find the low-level feature resolution and fine-tuned with the high-level feature maps to find better accuracy, optimum detection of the flower with other significant details that includes flower division, class, subclass, order, family and herb flower or not.

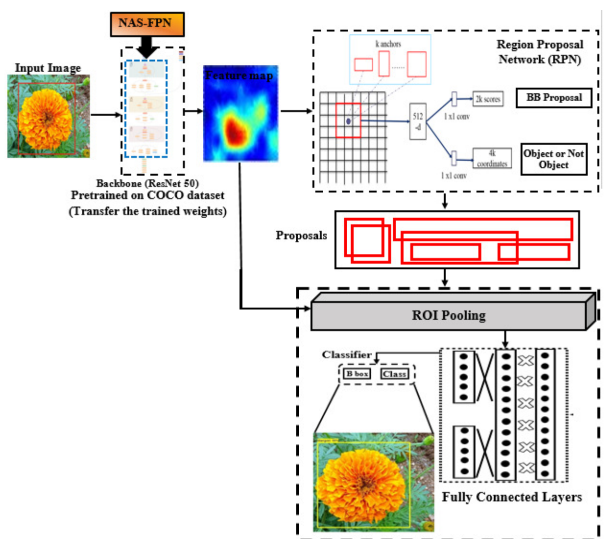

Fig. 3. A Proposed NAS-FPN with Modified Faster R-CNN Model for Flower Species Localization and Classification.

The proposed flower object detection model is divided into two major parts which are described as follow;

Transfer Learning with NAS-FPN: After the pre-processing step, it is required to use a pre-trained image classification network with transfer learning that is an easier and faster approach to train an object detection model network. NAS-FPN (Neural Architecture Search-Feature Pyramid Network) is an automatic neural architecture search algorithm that focuses on finding optimal connections between different layers for pyramidal representations. Feature Pyramid Network (FPN) represents the pyramid architecture for deep convolutional neural network task [30]. FPN is a concept of a pyramid feature extractor network designed to increase speed and accuracy. NAS (Neural Architecture Search) architecture is used as a backbone model for FPN [31]. FPN combines semantically strong and low-resolution with high-quality features, via a lateral connection and top-down pathway. In order to produce the final feature map, FPN prediction gives a 3 × 3 convolution and it is appended on each merged map [32]. This feature map is transfer to the Faster R-CNN backbone architecture (ResNet 50 V1) for training a feature from the flower species dataset. We have modified the ResNet 50 architecture by freezing some convolutional layer and fine-tune the feature maps of NAS-FPN to get the better feature maps of flower objects [33]. Also, ResNet 50 architecture has pre-trained from the COCO dataset. Therefore, it is known as ResNet 50 V1 architecture.

Modified Faster R-CN: The Faster R-CNN is a two-stage detection model as described in Fig. 3. This method is divided into following steps; in the first step, it is known as the backbone network which is used a pre-trained NAS-FPN as a feature encoder that transfers the weights for fine-tuning the Faster R-CNN backbone CNN model. The backbone model of CNN (i.e. ResNet-50 V1) is the pre-trained network of the COCO dataset used to build the RPN and Faster R-CNN network [34]. The process of feature extraction is performed by CNN using fine-tuned convolution neural network backbone architecture. At the end of the last layer, convolution feature maps are produced. In the second step, anchor boxes are generated based on the feature map using a sliding window approach. This step is known as region proposal network (RPN) which generates the different proposal of regions and outputs them to the detection network to realize the identification for the proposed region [35, 36]. RPN consists of three parts: first is anchor window, second is the loss function, and the third is a set of region proposals. At the last, the presence of objects is indicated by refining these anchor boxes in the next step (i.e. RoI Pooling layer). In the third step, ROI pooling generated anchors are refined using a smaller network. These are used to calculate the loss function to select top anchors containing objects. Finally, in the last step, it classifies the image with the prediction of the class (classification) and bounding box (localization). In this research, we have also included the multi-class and multi-labeling [37] of flower image classification and localization for better accuracy and high-level performance of flower detection.

IV. EXPERIMENTAL SETUP AND FLOWER SPECIES DATASETS

The following sections described the details of experiment and the datasets accumulated and used.

A. Setup of Experiment

The experiment is carried out on a machine with Windows 10 Operating system. The prerequisite software and setup are installed on a machine equipped with a processor of Intel® Core™ i7 8th generation, Titan Xp GPU of NVIDIA and 32 GB RAM. The software used are TensorFlow Object Detection API framework, Anaconda virtual environment included CUDNN 7.6, CUDA 10, MS Visual Studio for editing purposes and Python 3.8. Object Detection API also depends on the various libraries like; Python-tk, lxml, Matplotlib, contextlib2, coco API, Protobuf 3.0.0, Cython, Pillow 1.0, Tensorflow 2.0, and Jupyter notebook.

B. Description of Flower Datasets

For conducting experiments of deep convolutional neural network based object detection, various benchmark flower datasets are used. These datasets are publically available and widely used for the task of flower identification and classification. In all, in this research, there are 19,679 different kinds of flower images are used, which are divided into two datasets. We further divided each dataset into 80% as training images, 10% as validation images, and 10% as test images.

Dataset 1: This dataset contains 18,200 general flower species images for 102 flower classes. It has been acquired from Oxford [38, 39]; Kaggle's [40].

Dataset 2: The experiment also performed on Jena flower 30 datasets [41] with 30 classes based on common wild-flowering species found on semi-arid grasslands around the city of Jena in Germany. The dataset consists of 1,479 images. The dataset is challenging since multiple species exhibit large visual similarities and it also covers a variety of blooming stages of the flowers.

V. RESULTS AND PERFORMANCE ANALYSIS

This section explained the performance of different object detection models during training and testing phase. It also described the accuracy obtained by these object detection models.

A. Training Losses of Object Detection Models

In deep learning, training is the stage when the model is gradually optimized, or the model learns the features from the dataset. Loss is used to understand and improve the training of the object detection model. In this paper, we have considered two forms of object detection models; Faster R-CNN and SSD model. The category of Faster R-CNN is a two-stage detector. In the Faster R-CNN, there are four types of losses generated during the training of model; (i) RPN classification loss, (ii) RPN regression loss, (iii) Classification loss and (iv) Regression loss. For the Faster R-CNN training model, total loss is the sum of all four losses mentioned above. Another is the SSD model, which is a one-stage detection model. In the SSD, there are two types of losses generated during the training of model; (i) Classification loss and (ii) Regression loss. For the SSD training model, total loss is the sum of the above mentioned two losses. Total loss has a decreasing tendency. It means that the minimum value of the total loss indicates the well accurate training model. In the end, training loss generates the selected foreground and background ROIs that meet overlap criteria based on ground truth labels and specific class target regression coefficients for the ROIs. In this paper, we have applied the 20k iterations for training all object detection models for the flower datasets. The following Figs. 4 and 5 represented the total loss for different object detection models for F102 and J30 flower datasets. The Fig. 4 (h) and 5 (h) indicates the optimum loss curve generated by the proposed model as compared with the other object detection models for F102 and J30 flower class datasets respectively. This loss is only associated with the ground-truth label of flower class and relies on the classification of flower class to predict the flower category.

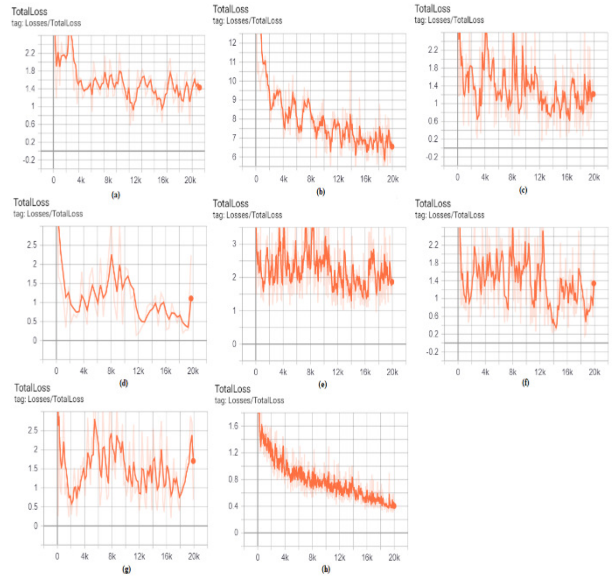

Fig. 4. Performance of Total loss of Object Detection models for F102 flower dataset at training time; Quantized model using SSD with (a) MobileNet V1 and (b) MobileNet V2, Atrous models using Faster R-CNN with (c) InceptionResNet V2, Low proposals model using Faster R-CNN with (d) ResNet 50 and (e) NAS, Atrous and Lowproposals model using Faster R-CNN with (f) InceptionResNet V2 and (g) ResNet 101 and (h) Proposed Model (NAS-FPN with Modified Faster R-CNN).

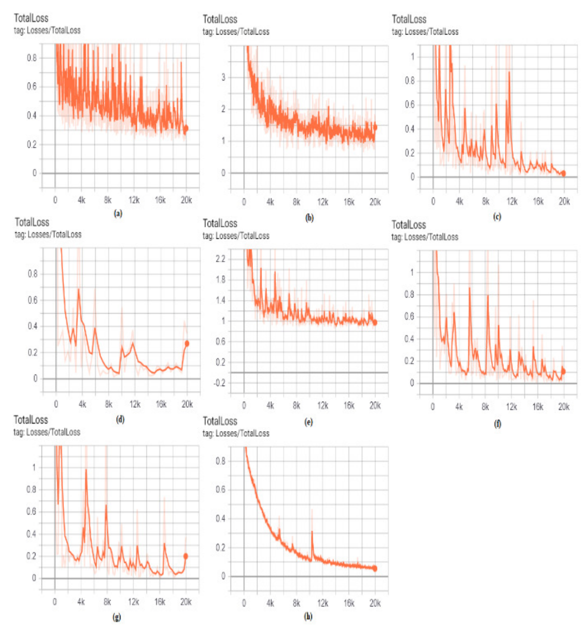

Fig. 5. Performance of Object Detection models for J30 flower dataset; Quantized model using SSD with (a) MobileNet V1 and (b) MobileNet V2, Atrous models using Faster R-CNN with (c) InceptionResNet V2, Lowproposals model using Faster R-CNN with (d) ResNet 50 and (e) NAS, Atrous and Lowproposals model using Faster R-CNN with (f) InceptionResNet V2 and (g) ResNet 101 and (h) Proposed Model (NAS-FPN with Modified Faster R-CNN).

B. Experimental Results and Discussion

In this paper, eight object detection models are implemented over two types of flower datasets. They are namely; SSD quantized model (8-bit) using MobileNet V1 and MobileNet V2, an Atrous model using Faster R-CNN with InceptionResNet V2, lowproposals model using Faster R-CNN with ResNet 50 and NAS, atrous and low proposals model using Faster R-CNN with ResNet 101 and InceptionResNet V2 against the proposed NAS-FPN with modified Faster R-CNN model. The experimental results are divided into two parts; (i) Quantitative and (ii) Qualitative performance analysis was taken to evaluate the flower detection.

For analysing quantitative performance of implemented object detection models, AP (Average Precision), AR (Average Recall), and mAP (mean Average Precision) were used as the evaluation metrics [42, 43]. In this paper, we have chosen 0.50:0.95, 0.75 and 0.50 for the AP IoU value. And for AR we have chosen 1, 10 and 100 IoU values. The evolutions of AP and AR across scales (area) are small, medium, and large objects. AP and AR are two most commonly used evaluation metrics. Moreover, the correctness of a detected object is also evaluated by the IoU (intersection-over-union) overlap with the corresponding ground truth bounding box. If the IoU is greater than the threshold value, it was considered as a true positive (TP). In the case of non-matching of a flower detected object with the ground truth bounding box, it was considered to be a false positive (FP). Moreover, in the occurrence of the missed ground truth bounding boxes, a false negative (FN) is determined. The overall detection performance was measured with the mean average precision (mAP) score, which is the average AP value over all classes. The higher the mAP is, the better the overall detection performance of the flower dataset [44, 45].

AP and AR for F102 Flower Dataset: The following graphs described the performance of different object detection models on F102 and J30 flower datasets respectively.

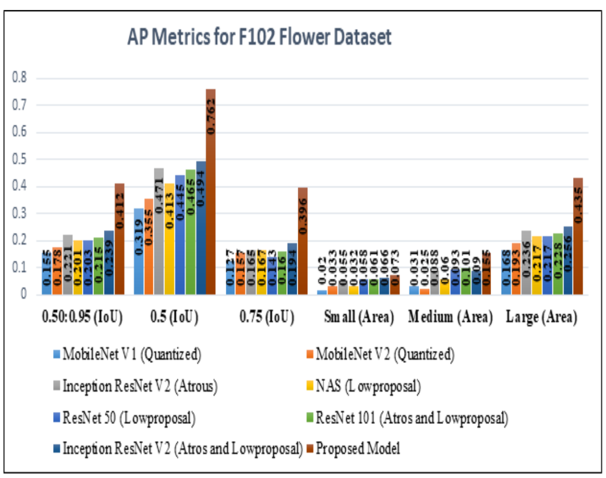

Fig. 6. The Performance of Average Precision (AP) IoU and Area for different object detection models for F102 flowers dataset.

Figs. 6 and 7 represented the result of AP and AR for the F102 flower class dataset. The high AP indicates the high correctness of flower detection. And high recall indicates the fewer targets are missed during the flower detection.

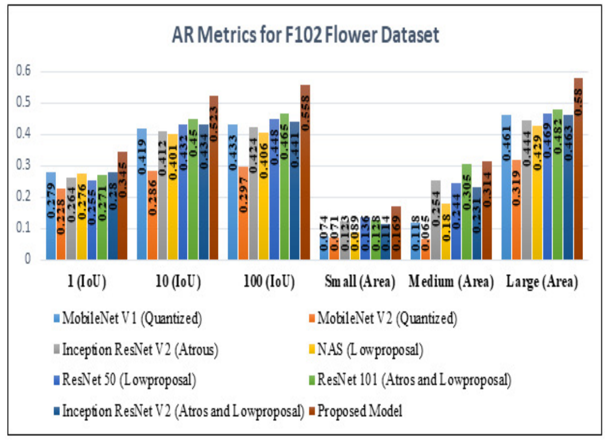

Fig. 7. The Performance of Average Recall (AR) IoU and Area for different object detection models for F102 flower dataset.

The Figs. 6 and 7 described the performance of the AP and AR of different object detection models on the F102 flower class dataset. Fig. 6 described the average precision (AP) IoUs and Areas for F102 flower class dataset and the proposed model has achieved the highest precision values of different AP IoU (0.5:0.95, 0.50, and 0.75) i.e. 0.412, 0.762, and 0.396. Accordingly, AP Areas (small, medium, and large) are 0.073, 0.155, and 0.435, which are the highest among all other models. While figure 7 described the result of average recall (AR) IoUs and Areas. Here, the proposed model has achieved the high recall values among all other object detection models and the values of different AR IoUs (1, 10, and 100) of the 102 classes are 0.345, 0.523, and 0.558. In addition, for AR areas (small, medium, and large), the values are 0.169, 0.314, and 0.580 respectively.

AP and AR for J30 Flower Dataset: The following figures 8 and 9 represented the result of AP and AR for J30 flower class dataset.

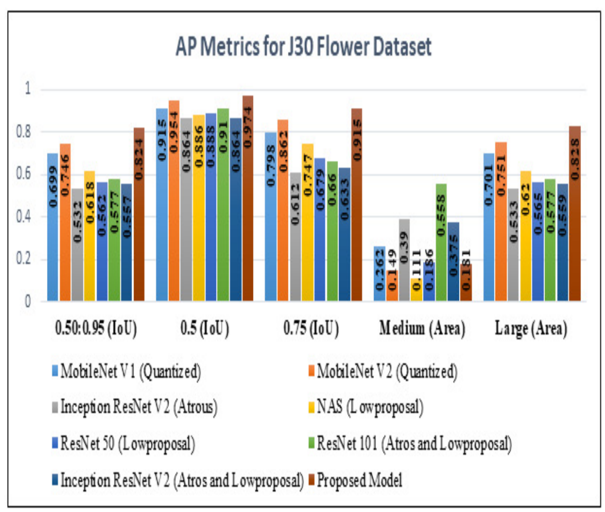

Fig. 8. The Performance of Average Precision (AP) IoU and Area for different object detection models for J30 flower dataset.

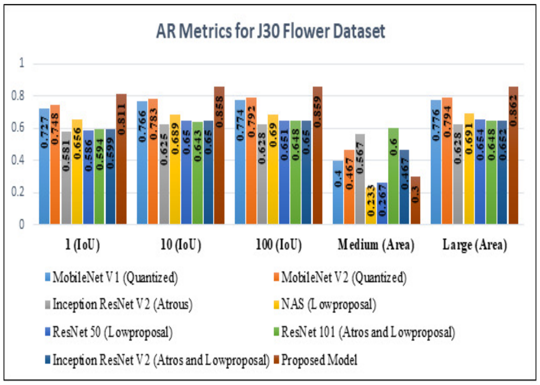

Fig. 9. The Performance of Average Recall (AR) IoU and Area for different object detection models for J30 flower dataset.

The above figures 8 and 9 described the performance of the AP and AR of different object detection models on the J30 flower class dataset. Fig. 8 described the average precision (AP) IoUs and Areas for J30 flower class dataset respectively and the proposed model has achieved the highest precision values of different AP IoU (0.5:0.95, 0.50, and 0.75) i.e. 0.824, 0.974, and 0.915. Accordingly, AP for medium area, it obtained 0.181 but for large area, the value is 0.828, which is again highest. The figure 9 described the result of average recall (AR) IoUs and Areas. Here also, the proposed model has achieved the highest recall values among all other object detection models. The values of different AR IoUs (1, 10, and 100) of the 30 classes are 0.811, 0.858, and 0.859. And for AR medium and large area are 0.3 and 0.862 respectively. In the J30 flower dataset; there are no small size flower objects available. Therefore, there is no performance value shown in the small area column.

mAP Score for Different Object Detection Models: The overall detection performance was also measured with the mean average precision (mAP) score, which is calculated as the mean of average precision (AP) value for all flower classes. The highest value of mAP is indicated the better performance of the object detection model for flower detection. The mAP values of different object detection models are plotted in the following graph that is included in Fig. 10.

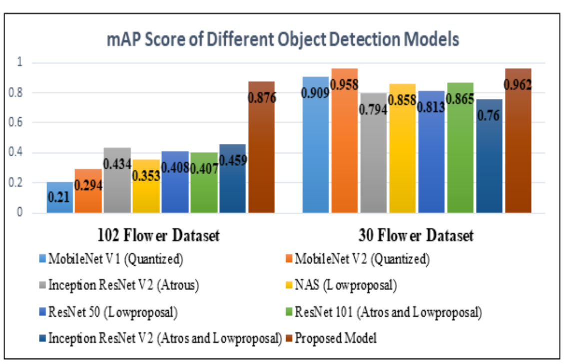

Fig. 10. The mAP Score for different object detection models for F102 and J30 flower datasets.

The proposed model achieved the highest accuracy, 87.6% for F102 flower class dataset and 96.2% for J30 flower class dataset, among all other object detection models. We observed that all other models including quantized and atrous are not performed well compare with the proposed model.

IoU values of Different Object Detection Models: The quality of correctness of flower object detection is also measured by the IoU (intersection-over-union) which is the combination of the ground truth label and the prediction bounding box. The prediction bounding box indicates the output of the object detection model and the ground truth bounding boxes indicates the hand-labeled bounding boxes form the training set that indicates the flower labeling. The following figures represented the IoU values of all the eight object detection models for F102 and J30 flower datasets.

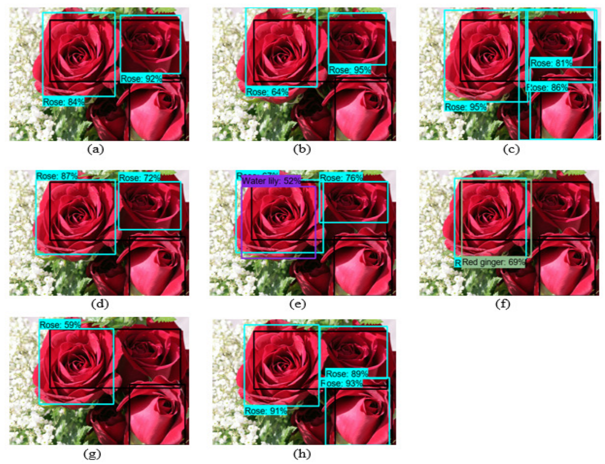

Fig. 11. Performance of Object Detection models for flower F102 dataset for Faster R-CNN using (a) Inception ResNet V2 atrous, (b) Inception ResNet V2 atrous and low-proposal, (c) NAS low-proposal, (d) ResNet 50 low-proposal, (e) ResNet 101 atrous and low-proposal, and using SSD with (f) MobileNet V1 quantized, (g) MobileNet V2 quantized, and (h) proposed NAS-FPN with Modified Dater R-CNN using ResNet 50 V1.

The Fig. 11 represented the results of the qualitative analysis of the different object detection models for the F102 flower class dataset. The two types of boxes are available in this analysis; one is the black colored box (Ground Truth label) and another is colored boxes (Prediction bounding box).

The prediction bounding box also represents the value of IoU. In the Figs.11 (e) and (f), ResNet 101 atrous with low proposals model and MobileNet V1 quantized model are overlapping the prediction bounding box and predicted different flower class as “water lily” and “red ginger”, which is the wrong prediction of flower object with less IoU values. Likewise, among all other models, the proposed NAS-FPN with modified Faster R-CNN provides the highest accuracy with the IoU values 91%, 89%, and 93% as per shown in Fig. 11 (h).

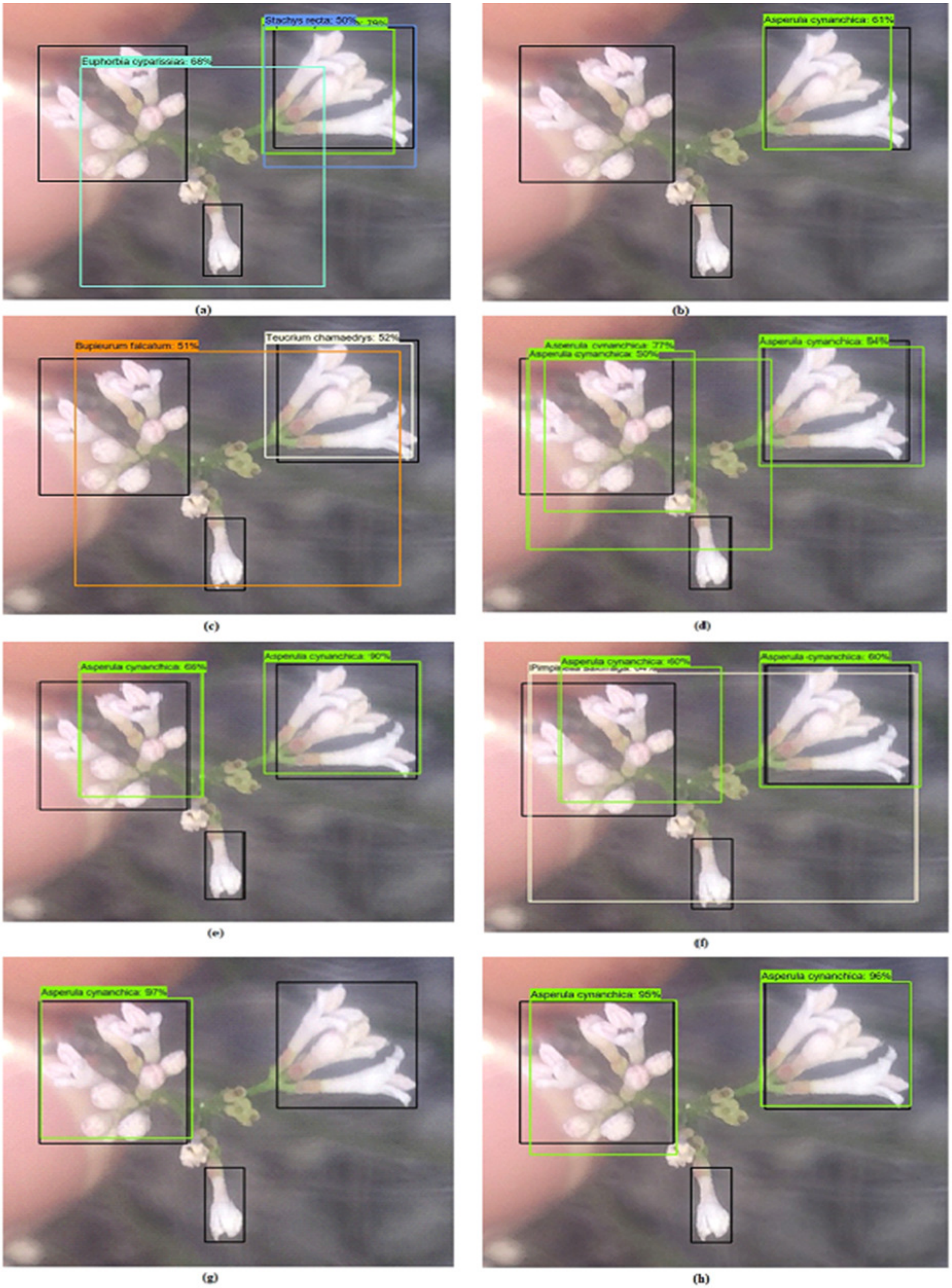

Fig. 12. Performance of Object Detection models for flower J30 dataset for Faster R-CNN using (a) Inception ResNet V2 atrous, (b) Inception ResNet V2 atrous and low-proposal, (c) NAS low-proposal, (d) ResNet 50 low-proposal, (e) ResNet 101 atrous and low-proposal, and using SSD with (f) MobileNet V1 quantized, (g) MobileNet V2 quantized, and (h) proposed NAS-FPN with Modified Faster R-CNN using ResNet 50 V1.

The Fig. 12 represented the results of qualitative analysis of the different object detection models for J30 flower class dataset. In the Fig. 12 (a) and (c), Inception ResNet V2 atrous model and NAS low proposals models are overlapping the prediction bounding box and predicted the different flower class which is wrong prediction of flower object with less IoU values. Among all other models, the proposed NAS-FPN with modified Faster R-CNN provides highest accuracy with the IoU values 95% and 96%.

VI. CONCLUSION

In this paper, different small scale object detection models i.e. quantized model, atrous convolutional, lowproposals model, combination of atrous with lowproposals model and proposed NAS-FPN with modified Faster R-CNN are explained and experimented on flower species datasets. For experiment, two types of flower datasets are used; one has 102 flower classes and second has 30 flower classes that contain 18200 and 1479 flower images respectively. Values of different evaluation parameters and training losses are being measured. From the results obtained, it has been shown that the proposed model provides better accuracy than other object detection models for flower species detection and classification. The proposed model has achieved the mAP of 87.6% and 96.2% for F102 and J30 flower species datasets respectively.

VII. FUTURE SCOPE

The proposed object detection model for flower species detection and classification can become more generalized with the learned parameters. The dataset can be accumulated with more number of flower species images and related classes. Moreover, visualization techniques can be added to interpret the intermediate results.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

Conflict of Interest. The authors don't have anyconflicts of interest.

REFERENCES

[1]. Wei, Q., Xue, C., Chang-An, Y., Xiao, Q., Li, S., Zhi-Kai, H., & Si-Zhe, W. (2019). Flower Species Recognition System Combining Object Detection and Attention Mechanism. *Springer Nature Switzerland AG*. 1-8.

[2]. Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Science Direct, *Computers and Electronics in Agriculture*, *163*, 104-846.

[3]. Dat, T. T., Toke, T. H., Moncef, G., & Alexandros, I. (2018). Automatic Flower and Visitor Detection System. *26th European Signal Processing Conference* (EUSIPCO).

[4]. Keiller, N., Otavio, A., B., P., & Jefersson, A., S. (2016) Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification. *arXiv: 1602.01517v1 [cs.CV]*.

[5]. Musa, C., Umit, B., Yanhui, G., M., Cevdet. I., & Abdulkadir, S. (2019). Efficient deep features selections and classification for flower species recognition. *Science Direct Measurement, 137*, 1-8.

[6]. Md., Aminur, R., R., Mohammad, H., M., Enea, P., & WonSook, L. (2019). Atrous Convolution with Transfer Learning for Skin Lesions Classification. bioRxiv.

[7]. Anup, K., & Anil, K. (2019). Analysis of Multimodal Medical Image Fusion using Discrete Wavelet Transform. *International Journal of Electrical, Electronics and Computer Engineering,* 8(2), 01-05.

[8]. Hongshan, Y., Zhengeng, Y., Lei, T., Yaonan, W., Wei, S., Mingui, S., & Yandong, T. (2018). Methods and datasets on semantic segmentation: A review. *ScienceDirect, Neurocomputing, 304*, 82–103.

[9]. E., Laxmi, L., G., Jose, M., N., Sharmili, K., Shankar, & Andino, M. (2019). Image Classification using Deep Neural Networks for Malaria Disease Detection. *International Journal on Emerging Technologies, 10*(4), 66-70.

[10]. Siddesha, S., & Niranjan, S., K. (2020). Detection of Mite Disease and Computation of Affected Area in Raw Coconut. *International Journal on Emerging Technologies*,*11*(2), 08-12.

[11]. Prateeth, N., David, Z., & Sek, C. (2019). Bit Efficient Quantization for Deep Neural Networks. *EMC2- NeurIPS Workshop*, Vancouver, Canada, 1-5.

[12]. Liang-Chieh, C., George, P., Iasonas, K., Kevin, M., & Alan, L., Y. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv: 1606.00915v2 [cs.CV]*.

[13]. Darshana, K., & Prof. Ram, R. A. (2018). NIDSRL: Network Based Intrusion Detection System Using Reinforcement Learning, *International Journal of Electrical, Electronics and Computer Engineering*, *7*(1), 42-48.

[14]. Zhishuai, Z., Siyuan, Q., Cihang, X., Wei, S., Bo, W., & Alan, L., Yuille, (2018). Single-Shot Object Detection with Enriched Semantics. *The Center for Brains, Minds and Machines (CBMM)*.

[15]. Liang-Chieh, C., George, P., Florian, S., &Hartwig, A. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv: 1706.05587v3 [cs.CV]*.

[16]. Jun, F., Ali, S., Abdel-Aziz, H., David. T., Georgios, G., & Joseph, H. (2020). Near-Lossless Post-Training Quantization of Deep Neural Networks via a Piecewise Linear Approximation. *arXiv :2002.00104v1 [cs.CV]*.

[17]. Xiao, S., Jungwook, C., Chia-Yu, C., Naigang, W., Swagath, V., Vijayalakshmi, S., Xiaodong, C., & Wei, Z. (2019). Hybrid 8- bit Floating Point (HFP8) Training and Inference for Deep Neural Networks. *33rd Conference on Neural Information Processing Systems (NeurIPS),* Vancouver, Canada.

[18]. Benoit, J., Skirmantas, K., Bo, C., Menglong, Z., Matthew, T., Andrew, H., Hartwig, A., & Dmitry, K. (2018). Quantization and Training of Neural Networks for Efficient Integer- Arithmetic-Only Inference. *IEEE*, 2704-2713.

[19]. Jose, A., C., Javier, B., Manuel, B., Jesús, Fernández, R., & Ginés G., (2019). Machine Learning for Solar Trackers. *AIP Conference Proceedings,* 21-26.

[20]. Hoanh, N., (2019). Improvement Of Detecting Small-Sized Traffic Signs Based On Deep Learning. *Journal of Theoretical and Applied Information Technology,* 97(19), 5012-5022.

[21]. Hoanh, N., (2019). Improving Traffic Sign Detection with Deep Convolutional Neural Networks. *Journal of Theoretical and Applied Information Technology,* 97(13), 3730-3742.

[22]. Amir, R., Richard, Z., & John K., T. (2018). The Elephant in the Room. *arXiv: 1808.03305v1 [cs.CV].*
[23]. Dong, X., Hongzong, L., Chenyi, L., & Qifei, H. (2019). Large-Truck Safety Warning System Based on Lightweight SSD Model. *Hindawi-Computational Intelligence and Neuroscience.*
[24]. Tongfan, G., & Hao, Z. (2017). Atrous Faster R- CNN for Small Scale Object Detection. *2nd International Conference on Multimedia and Image Processing.*
[25]. You, L., Qingxuan, Z., & Yulei, Z. (2019). Multi-layer Feature Fusion Network with Atrous Convolution for Pedestrian Detection. *IOP Conf. Series: Journal of Physics: Conf. Series 1267*, AIACT.
[26]. Junjie, M., Yaping, D., & Yap-Peng, T. (2019). Atrous Convolutions Spatial Pyramid Network for Crowd Counting and Density Estimation. *ScienceDirect, Neurocomputing*, 350, 91–101.
[27]. Zhao, R., Qiuqiang, K., Jing, H., Mark, D., P., & Bj¨Orn, W., S. (2019). Attention-Based Atrous Convolutional Neural Networks: Visualisation and Understanding Perspectives of Acoustic Scenes. *ICASSP*.
[28]. Wei, T., Dongsheng, Z., Su, Y., Jing, S., Jingpei, D., & Guowu, S. (2020). A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab. *Neural Computing and Applications, Springer.*
[29]. Isha, & P., Sanskruti, P. (2020). An Optimized Deep Learning Model For Flower Classification Using NAS-FPN And Faster R-CNN. *International Journal of Scientific & Technology Research, 9*(03), 5308-5318.
[30]. Tsung-Yi, L., Piotr, D., Ross, G., Kaiming, H., Bharath, H., & Sere, B. (2017). Feature Pyramid Networks for Object Detection. *IEEE*, 2117- 2125.
[31]. Ioannis, A., Panagiotis, M., & Loukas, P. (2018). A Framework of Transfer Learning in Object Detection for Embedded Systems. *arXiv: 1811.04863v2 [cs.CV].*
[32]. Tsung-Yi, L., Piotr, D., Ross, G., Kaiming, H., Bharath, H., & Serge, B. (2017). Feature Pyramid Networks for Object Detection. *IEEE*, 2117-2125.
[33]. Golnaz, G., Tsung-Yi, L., Ruoming, P., & Quoc, V., L. (2019). NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. *arXiv: 1904.07392v1[cs.CV].*
[34]. Ross, G. (2015). Fast R-CNN. *arXiv: 1504.08083v1 [cs.CV].*
[35]. Shaoqing, R., Kaiming, H., Ross, G., & Jian, S. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv: 1506.01497v3 [cs.CV].*
[36]. Zhenwei, H., & Lei, Z. (2019). Multi-adversarial Faster-RCNN for Unrestricted Object Detection. *arXiv: 1907.10343v1 [cs.CV].*
[37]. Yongcheng, L., Lu, S., Jing, S., Junjie, Y., Shiming, X., & Chunhong, P. (2019). Multi-Label Image Classification via Knowledge Distillation from Weakly-Supervised Detection. *arXiv: 1809.05884v2 [cs.CV].*
[38].http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html
[39].http://www.robots.ox.ac.uk/~vgg/data/flowers/17/index.html
[40].https://www.kaggle.com/alxmamaev/flowers-recognition
[41]. Marco, S., Michael, R., Nedal, A., Jana, W., & Patrick, M. (2017). Plant species classification using flower images - A comparative study of local feature representations. *PLOS ONE*, 1-29.
[42]. Bernardo, A., Godinho, D. O., Flávia, M., F., F., Carlos, A., P. & Da, S., M. (2018). Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices. *IEEE*, 6.
[43]. Jun, S., Xiaofei, H., Xiao, G., Xiaohong, W., Jifeng, S., & Yingying, S. (2018). Detection of Key Organs in Tomato Based on Deep Migration Learning in a Complex Background. *MDPI Agriculture*, *8*(196), 1-15.
[44]. Shuyang, S., Jiangmiao, P., Jianping, S., Shuai, Y., & Wanli, O. (2019). FishNet: A Versatile Backbone for Image, Region, and Pixel Level Prediction. *arXiv: 1901.03495v1 [cs.CV].*
[45]. Kaiwen, D., Song, B., Lingxi, X., Honggang, Q., Qingming, H., & Qi, T. (2019). CenterNet: Keypoint Triplets for Object Detection. *arXiv: 1904.08189v3 [cs.CV].*

How to cite this article: Patel, I. and Patel, S. (2020). A Comparative Analysis of Applying Object Detection Models with Transfer Learning for Flower Species Detection and Classification. *International Journal on Emerging Technologies*, *11*(3): 303–312.