SindhUniv. Res. Jour. (Sci. Ser.) Vol. 52 (01) 93-102 (2020)
http://doi.org/10.26692/sujo/2020.03.14
SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Automated Flower Classification using Transfer Learning and Meta-Learners in Deep Learning Framework

P. KHUWAJA*, S. A. KHOWAJA++, B. R. MEMON, M. A. MEMON, G. LAGHARI**, K. DAHRI

Institute of Information and Communication Technology, University of Sindh, Jamshoro, Pakistan

Received 10th May 2019 and Revised 18th November 2019

Abstract: The classification of flowers is a challenging task due to the wide variety of flowers along with inter- and intra-variations amongst the flower categories. Furthermore, the information such as the grass and leaves does not help in providing context to the recognition system. Researchers have extensively used deep learning frameworks for improving the classification accuracy but there is still room for improvement in terms of the recognition performance. In this paper, we use the transfer learning aspect to fine-tune the existing pre-trained networks which provide us an edge for the improved classification accuracy. We then apply various decision-level fusion strategies to combine the class probabilities from the individual pre-trained networks for further boost in recognition performance. Our method has been validated on two well-known flower datasets. The experimental results show that the proposed method achieves the best performance i.e. 99.80 % and 98.70 % on Oxford-17 and Oxford-102 datasets, respectively, which is better than the state-of-the-art methods.

Keywords: Flower Classification, Transfer Learning, Deep Learning, Meta-learners, Decision-Level Fusion

1. INTRODUCTION

Flower classification plays a vital role in a wide variety of fields which includes forestry, agriculture, fragrance, and medical industries. The automated system for flower identification helps to eliminate the manual query of flower information from large scale databases. The automation also improves the efficiency of the flower information retrieval system while reducing labor costs and human error at the same time. In terms of research, flower classification is a more challenging task in comparison to the object classification due to similar characteristics such as color, appearance, and shape. Moreover, the contextual information in terms of surrounding leaves, grass, and so forth, does not account for better recognition performance (Hiary, *et al.*,2018). Another reason for increased complexity in flower recognition systems is the number of species needed to be classified. A study reveals that there are more than 250K known flower species which can be categorized into 350 families (Kenrick, 1999). A wide variety of applications such as flower taxonomy, live plant identification (Chi, 2003), floriculture industry, plants monitoring system (Larson, 1992), and content-based image retrieval for flower indexing and representation (Das, *et al.*, 1999), heavily relies on the accurate flower classification system. Although manual systems based on field experts are in use they are time-consuming, and prone to human error when dealing with large scale datasets. In this regard, the demand for automated flower classifications systems is increasing and is of great value to the associated applications.

Traditional flower classification systems heavily rely on the shape, texture, color, and statistical features extracted from the flower images (Khan, *et al.*, 2012; Maria-Elena, *et al.*, 2008; Xie, *et al.*, 2017; Yang, *et al.*, 2014). The recognition performance of such systems depends on the quality and the number of features being used for the classification of flower species (Maria-Elena*et al.*, 2008; Yuning Chai, *et al.*, 2011). Besides, human interaction has also been used to improve recognition performance (Hsu, *et al.*,2011; Mottos and Feris, 2014). The most commonly used classification algorithms for automated flower recognition are support vector machines (SVM), logistic regression, and shallow classification methods (Chai, *et al.*, 2012; Khowaja, *et al.*, 2015; Khowaja, *et al.*, 2019; Khowaja, *et al.*, 2018; Khowaja, *et al.*, 2017; Khuwaja, *et al.*, 2019). These methods learn the representation derived from extracted features to classify the flower into their respective category.

As discussed, the existing methods rely on the handcrafted features therefore, the recognition performance was as good as the quality of information represented in the features themselves. The feature extraction method includes, speed up robust features (SURF), scale-invariant feature transform (SIFT), the histogram of oriented gradients (HoG), and local binary patterns (LBP) (Khowaja *et al.*, 2015). However, the

++Corresponding Author: sandar.ali@usindh.edu.pk
*Institute of Business Administration, University of Sindh, Jamshoro, Pakistan
**Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan

handcrafted features exhibit two main problems, the first is the scalability and flexibility of the feature extraction method and the second is the generalization. As the size or the categories of the flower are increased in the employed dataset the characteristics of the feature extraction method need to be changed which is a tedious task. The generalization problem refers to the variance in the recognition performance as the scalability and flexibility of the dataset are altered. These two problems motivate the researchers to move towards artificial neural networks (ANN) and deep neural networks (DNN) which automatically extracts the features understandable by the learning networks, therefore, copes with the scalability and the flexibility problem quite nicely. Moreover, the depth of the DNN allows generalizing similar performance on multiple datasets with varying characteristics.

Deep learning techniques have gained a lot of interest from the computer vision research community due to their superior recognition performance in comparison to the shallow learning algorithms. The convolutional neural network (CNN) is a kind of deep learning technique that is extensively used for different computer vision applications. Also, these deep learning techniques are compatible with graphical processing units (GPUs) which speeds up the processing i.e. training and testing time, due to their transformation in tensors (Krizhevsky, *et al.,* 2012). The GPUs indeed reduce the training time quite significantly yet the training of CNNs from scratch is takes a lot of time. The current trends are to use transfer learning approaches i.e. pre-trained networks, to reduce the training time of very deep network architecture. Many studies try to fine-tune an existing pre-trained network for the specified classification task which significantly reduces the training time while achieving improved recognition performance.

In this work, we tackle the flower classification problem with existing pre-trained CNNs on large-scale datasets. Most of the existing studies fine-tune a single pre-trained network for the said classification task. The proposed study employs three popular pre-trained networks and combines its classification results using multiple fusion strategies to improve recognition performance. The analysis also explores the use of meta-learners for combining the classification results from multiple pre-trained network architectures. This study provides an extensive comparative analysis for multiple fusion strategies along with meta-learners and reports state-of-the-art accuracy on popular flower classification dataset (Nilsback and Zisserman, n.d.; Maria-Elena *et al.,* 2008).

The rest of the paper is structured as follows: Section 2 consolidates the relevant existing studies for flower classification. Section 3 provides the details of pre-trained network architecture, network parameters, and the fusion strategies employed for the recognition task. Section 4 presents the experimental results and comparison with the existing results. Section 5 concludes the finding of the paper and presents the possible future directions of this study.

2. RELATED WORK

In this section, we present various works that address the flower classification problem. This section first summarizes the work using hand-crafted features with shallow learning methods and then recapitulates the works using variants of CNNs.

The earlier works in the field of flower classification focused on extracting meaningful representations in the form of hand-crafted features to improve the classification performance. Nilsman and Zisserman (Maria-Elena *et al.,* 2008) proposed to extract the color values using the HSV model as features along with HoG and SIFT. (Yuning Chai *et al.,* 2011) used the existing features and bi-level co-segmentation (BiCoS) with a multi-tasking approach (BiCoS-MT) to recognize a large number of flower species. Chai et al. (Chai *et al.,* 2012) extended their work by extracting fisher vector (FV) and principal component analysis (PCA) coefficients in addition to the existing features. They also extended their previous approach to a higher abstraction level by proposing tri-level co-segmentation (TriCoS) to improve the classification accuracy. The features play an important role to improve recognition performance. Keeping this in view, some works proposed the feature extraction while allowing the users to manually interact with the data. For instance, Zou and Nagy (Jie *et al.,* 2004) presented a computer-assisted visual interactive recognition (CAVIAR) method which allows a user to extract multiple features related to the shape and curvature of the flower. (Hsu *et al.,* 2011) proposed the use of weighted Euclidean distance from the feature space of existing representations along with the center area and boundary shape of the flowers. Although, the results are promising the use of interaction to extract the features loses the essence of end-to-end automation of flower classification.

Following the trend, many other researchers instead of using classical features proposed the new extraction techniques to represent the flower data. These extracting techniques include graph-regularized robust late fusion (Guangnan *et al.,* 2012), Haar-features (Zhang, *et al.,* 2013), bag-of-words using color attention (Khan *et al.,* 2012), bag-of-frequent local histograms (FLH) (Fernando, *et al.,* 2014), dictionary learning based on Fisher vectors and FLH (Yang *et al.,* 2014), local saliency map using generalized hierarchical matching (Qiang *et al.,* 2012), co-occurrence features (Ito and

Kubota, 2010), visual adjectives (VAs) along with improved FV and SIFT (Xie, *et al.*, 2016), power normalization and FV with generalized max-pooling (Murray and Perronnin, 2014), and local binary patterns with pairwise rotation invariant co-occurrence features (Qi *et al.,* 2014). All these studies with hand-crafted features use shallow learning methods such as support vector machines, random forests or logistic regression to classify a flower image.

The hand-crafted features which represent the data well achieved good recognition performance, however, the features do not generalize even on the same kind of data across different datasets suggesting that features need to be designed specifically for the dataset in hand. Moreover, designing a method for extracting hand-crafted features is not an easy task. In light of the above limitations, many researchers use deep learning techniques to tackle the recognition problem. Amongst many, CNN has gained a lot of attention due to its capability of extracting high-level features in an automated way and achieving superior accuracy than the shallow learning algorithms. Few works specifically employ CNNs for solving a flower classification problem. (Song, *et al.,* 2016)proposed a two-level hierarchy for flower classification. They used a pre-trained network on the target dataset for extracting high-level features and used those features to train a shallow classifier for classifying flowers. This is a classic example of transfer learning approach to increase the classification accuracy. A similar kind of approach was also proposed by (Razavian, *et al.,* 2014) and (QiQian, *et al.*,2015) where they used the features from CNN architecture to train a shallow classifier. (Xie, *et al.,* 2015)focused on the image retrieval problem using classification as an integrated task. The said study extracted the features from CNN and used nearest-neighbor estimation for computing the similarity from the feature space of the queried image and the candidate image. It was based on the simple assumption that shorter the distance the queried image would be of the same label as the candidate image and vice versa. (Xie *et al.,* 2017) tried to increase the capacity of the CNN network by introducing reverse-invariant features and CNN layers (RI-Deep) and (RI-Conv) while keeping the almost same number of model parameters. They show that the classification accuracy can be improved by increasing the capacity of the network.

A few works have been carried out which modify the intrinsic characteristics of the CNN network to improve the classification accuracy. (Xie, *et al.,* 2015) proposed the use of a task-driven pooling layer instead of average or max-pooling to improve the flower image representation. (Zheng, *et al.,* 2016) extended their work by introducing multi-task driven pooling to make the feature maps much smoother and less noisy. (Zhang ,*et al.,* 2017) proposed the extraction of semantic representation combined with contextual modeling within the deep architecture pipeline. (Liu, *et al.,*2016) presented a way to extract the convolutional features with multiple scales that could target different regions of the flowers. Their method was a compromise between the accuracy and the network parameters as scaling the maps would eventually increase the model complexity and the testing time. However, it was noticed that such networks lead to the over fitting problem, thus do not improve the accuracy ofthe testing/validation set. Most of the CNN network architectures focus on extending the depth (number of layers) which in turn increases the network capacity. Another way to increase the capacity is by extending the width of the network such as Inception networks. These networks not only increase the capacity but also reduce the model complexity at the same time. (Xiaoling *et al.,* 2017) proposed the use of Inception networks for the flower classification problem. (Wei, *et al.,* 2017) selectively aggregated the convolutional descriptors to fine-grain the image retrieval process using unsupervised learning. The method was also tested on the flower recognition problem. (Xie *et al.,* 2017) proposed the fusion of two different networks, one focuses on the global geometry of the image while the second considers the local parts. The recognition results are then fused to improve recognition performance. (Wu, *et al.,* 2018) used the transfer learning approach for sharing weights from popular pre-trained networks and applied on the flower recognition dataset. (Hiary *et al.,* 2018) used the segmented flower images for training the CNNs suggesting that the network should only learn the representation of flower region instead of the objects surrounding it such as branches, leaves, grass, and so forth.

In this work, we intend to use popular pre-trained network architectures to extract the visual features and explore different fusion strategies to improve the recognition performance for flower images. As per the available literature, the studies did not explore the fusion of networks with shared weights (transfer learning) using adaptive weighting or meta-learners as performed in our proposed work.

3. PROPOSED METHOD

We present our methodology in three subsections. The first section describes the existing pre-trained models being considered for the flower classification. The second section presents the transfer learning approach which is being used for the said recognition task, and the third section shows the method for fusion strategies that are used to combine the classification results from different CNN architectures.

1.1 CNN Architectures:

As we stated earlier that we will use existing pre-trained models for fine-tuning the network on the flower classification task. In this regard, we employ four pre-trained networks i.e. VGG19 (Simonyan and Zisserman, 2014), ResNet101 (He, *et al.*, 2016), DenseNet161 (Huang, *et al.*, 2017), and GoogleNet (Szegedy *et al.*, 2015). The VGG network is a classic example of CNN which was intended to increase the network capacity through the addition of more layers, thus making the network deeper. The ResNet network is a variant of CNN that uses shortcut connections to add the result from the convolutional layer and the provided input. This results in the superimposition of the feature maps. The ResNets have proved that such operations do improve recognition performance significantly. ResNet architecture uses identity mapping of the feature maps from their immediate preceding layers, if we extend the mapping of feature maps from the preceding layers to all the subsequent layers by concatenating the input compositely, it will result in an architecture which is called DenseNet. The DenseNets have been very successful in image recognition problems as they propagate collective knowledge to all subsequent layers. Finally, GoogleNet uses inception like architecture which not only extends the depth but also the width to increase the network capacity. The inception modules use different sizes to extract the feature maps, therefore the spliced version of the representation is passed to the subsequent layer. Inception-like network architecture has proven that it can achieve not onlybetter recognition results but also can reduce the computational overhead caused by the increased parameters. These pre-trained networks have varying characteristics, they are best suited for our study as fusing the results from similar networks will not contribute to the change in performance instead it will only increase the network complexity and computational overhead of the recognition system.

All of the said architectures focus on multi-class classification and therefore employ the cross-entropy loss function to update the weights. This loss function allows CNN to output the probabilities of the $\mathcal{L}$ classes for each image. The mathematical formulation for cross-entropy loss is given in equation (1):

$$CE = -log\left(\frac{e^{score_{pos}}}{\sum_{k}^{\mathcal{L}} e^{score_k}}\right) \quad (1)$$

where$score_{pos}$ refers to the classification scores for the positive class and k represents the number of classes i.e. $k = 1, \ldots, \mathcal{L}$. The forward propagation step computes the gradient response from the neurons and the loss is computed. Based on the loss, error back propagates throughout the network architecture. We use the same loss function for all the pre-trained networks to back propagate the error.

3.2 Transfer Learning from pre-trained models

There are many types of transfer learning but in this an existing pre-trained network trained on the source dataset to train the target dataset. In simple words, we just employ the pre-trained network and fine-tune it on the target dataset. The transfer learning has shown to achieve better results as compared to the network architecture training from scratch. Moreover, transfer learning reduces the computation time for training a particular network. An example of transfer learning using GoogleNet architecture is shown in **(Fig. 1).** GoogleNet which has been pre-trained on ImageNet having large scale dataset and 1000 categories. We use the same pre-trained model and re-train (fine-tune) the network on small scale dataset such as Oxford 17 and Oxford 102 flower datasets, respectively.

3.3 Fusion Strategies:

A test image when passed through multiple streams (i.e. through each pre-trained network), generates class probabilities for multiple classes. It is essential to

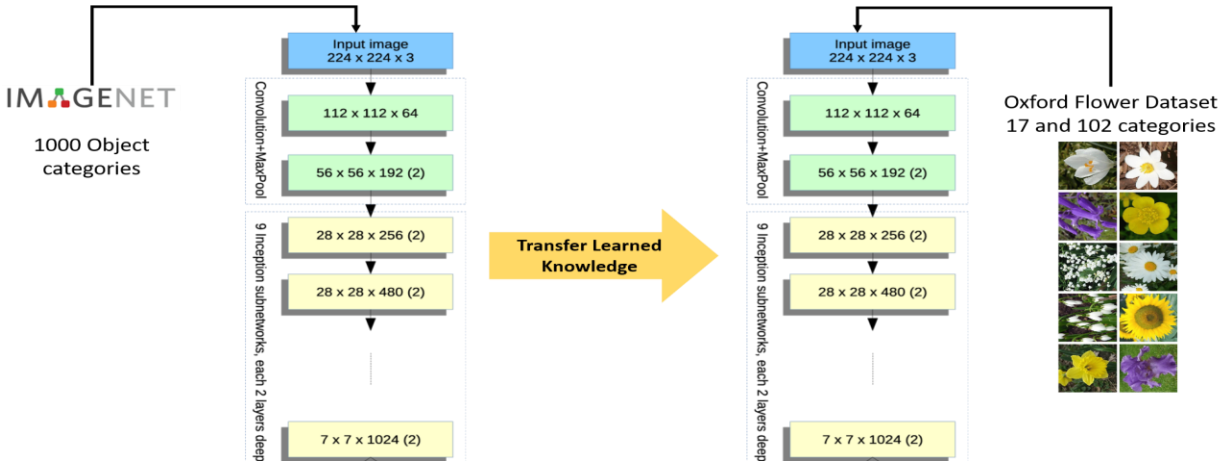

Fig. 1. Example of Transfer Learning using GoogleNet network architecture

provide a way for combining the class probabilities from multiple streams to get an improved classification performance. We assume that each pre-trained network exhibits its strength for specific flower categories. For example, GoogleNet may be associated strongly with classes having a specific color or pattern whereas DenseNet may recognize the class better having unique low-level features. The fusion methods are quite significant for combining the results from multiple streams to improve recognition performance. The most popular fusion methods are feature-level and decision-level fusion. Feature-level fusion is performed by concatenating the feature maps extracting from convolutional layers of CNN whereas the decision-level fusion is performed by combining the class probabilities from the individual streams. In this work, we mainly focus on the fusion strategies performed at the decision-level hierarchy. There are several methods through which the decision-level fusion is performed. The mostly used algorithm is the weighted average fusion methods (Feichtenhofer, *et al.*, 2016; Khowaja and Lee, 2019) where the class probabilities are averaged and the class with maximum probability is selected as the final label. Another strategy is the adaptive weighting fusion mechanism (Khowaja *et al.*, 2017) where the class probabilities are combined and averaged out based on the proportion of the data available for a specific class. It has been shown that the adaptive weighting fusion mechanism performs better when there is an imbalance in the dataset concerning the available images for a specific class. As the employed dataset shows highly imbalance characteristics we employ the adaptive weighting fusion mechanism for this study. The last method is the meta-learning technique where a shallow learning method is employed to train on the output class probabilities obtained from individual streams. This method integrates the predictions adaptively from existing pre-trained networks thus we get optimal fusion weights for specified categories. Methods such as (Khowaja*et al.*, 2019, 2018, 2017; Khuwaja *et al.*, 2019) have extensively used the meta-learning technique to improve the final classification result. Let's denote the probabilities of a specific class with $\mathscr{p}$. We stack the probabilities of each class as a vector presented in equation 2.

$$ML_{score} = \left[\mathscr{p}_n^{1^T}, \dots, \mathscr{p}_n^{k^T}, \dots, \mathscr{p}_n^{\mathcal{L}^T}\right]^T \in \mathbb{R}^{\mathcal{L}n} \quad (2)$$

where$n$ is the number of training samples. Considering the coefficient vector the fusion weights can be learned by any shallow learning classifier such as Bayesian networks, logistic regression, or so forth. An example of learning of the fusion weights is provided in equation 3.

$$W_i = \arg\min_{W,\dots,W_{\mathcal{L}}} \sum_n \log\left(1 + \exp\left[\ \left(1 - 2\mathscr{p}_{n,k}\right) ML_{score}^T W_{i-1}\right]\right) \quad (3)$$

whereW_i is the current weight ought to be updated and $\mathscr{p}_{n,k}$ is the class probability of the *n-th*sample for class k. The final prediction from the individual streams will be obtained using the updated weights optimized using meta-learner (i.e. shallow classifier stacked on the combined output of individual streams).

4. EXPERIMENTAL RESULTS

This section presents the analysis and recognition performance using pre-trained networks and fusion strategies. All of the experiments use PyTorch (Paszke *et al.*, 2017) framework using a PC having a clock rate of 3.20 GHz, core i7, memory 8 GB, and NVIDIA GeForce GT 730 GPU. The dataset and implementation details for pre-trained networks and the fusion strategies are provided in this section followed by the quantitative and qualitative results.

4.1 Flower Datasets:

In this study, we used two flower datasets i.e. Oxford-17 and Oxford-102. The oxford-17 dataset was created by Andrew Zisserman and Maria-Elena at Oxford University in 2006. The dataset comprises of 17 flower categories with 80 images each. The flower categories included in this dataset are common in Britain. The images in this dataset are of large scale having variations in lighting and posture, respectively. The challenging aspect of this dataset is that there is a lot of inter- and intra-variations amongst the flower categories.

The oxford-102 dataset was also created by the same authors in 2005. In this dataset, the flower categories were extended to 102 having 40-258 images. The complexity of inter- and intra-variations amongst the flower categories was also enhanced for this dataset suggesting that many flower categories have a lot of similarities in terms of color and characteristics.

4.2 Implementation Details:

The only pre-processing which was applied to the flower images were the image resizing. Most of the pre-trained networks employed in this study consider the input image size of 224x224x3 except for GoogleNet which considers the input image size of 299x299x3. The resized images were passed through the pre-trained networks and fine-tuned on the given dataset. All the networks are pre-trained on ImageNet, therefore, the last layer was trained to classify 1000 categories. However, in our case we have to classify between 17 or 102 categories, in this regard, the last layer was removed and was replaced by a new layer that was configured to classify either number of categories as per

the employed dataset. We used the stochastic gradient descent (SGD) optimizer with the learning rate of 0.001 and decrease by the factor of 0.01 for each subsequent layer or block depending on the network architecture. The weight attenuation was set to 0.005 and the default value of momentum i.e. 0.9 was used for fine-tuning the pre-trained networks. We only used 10 epochs for fine-tuning each of the pre-trained networks.

The fusion strategies play a vital role in improving the recognition performance in this study. We used four decision-level fusion strategies i.e. Average weighting, adaptive weighting, random forest as a meta-learner, and Naïve Bayes as a meta-learner. We used the same parameters as proposed in (Khowaja *et al.*, 2017) for average and adaptive weighting and the same settings as proposed in (Khowaja and Lee, 2019) for the meta-learners, respectively. All the images are augmented in terms of rotation, scaling, and translation to increase the data magnitude.

4.3 Result and Analysis

In this subsection, we first present the quantitative and qualitative analysis using multiple pre-trained networks and their fusion on Oxford-17 and Oxford-102 datasets and then we compare our results with the existing studies to show the increase in recognition performance. Table 1 shows the classification accuracy on both the datasets using individual pre-trained networks as well as the fusion strategies. It is apparent from the results that Dense161 performs better than all other pre-trained networks on an individual basis. Average weighting improves the recognition performance for Oxford-17 but decreases the recognition performance for Oxford-102 in comparison to DenseNet161. An interesting fact to be noticed is that the meta-learners perform better than the weighting score methods and the best classification accuracy was obtained using Naïve Bayes (meta-learner) which is 99.8% and 98.7% for Oxford-17 and Oxford-102, respectively. The results support our assumption that the fusion strategies can improve the recognition performance significantly and considering the best results from Naïve Bayes and the least accuracy from VGG19 it can be noticed that the performance was improved by 12.2% and 11.9% on Oxford-17 and Oxford-102. We performed the tests between the accuracies of each flower for VGG19 and Naïve Bayes and found that the improvement in accuracy is significant with $p < 0.01$ which proves that the fusion strategy using Naïve Bayes significantly improves the flower recognition performance. We also present the qualitative results for flower classification **(Fig. 2–4).**

Table 1 Classification Accuracy using pre-trained networks and fusion strategies

Method/Fusion Strategy	Classification Accuracy	
	Oxford-17	Oxford-102
VGG19	87.6%	86.8%
GoogleNet	88.4%	89.6%
ResNet101	95.2%	96.6%
DenseNet161	96.8%	97.1%
Fusion (Average Weighting)	97.3%	95.8%
Fusion (Adaptive Weighting)	98.1%	97.3%
Fusion (Random Forest)	98.7%	97.6%
Fusion (Naïve Bayes)	**99.8%**	**98.7%**

We intend to compare the classification accuracy of the proposed work to that of the existing studies. It is a necessary step to showcase whether the proposed method only improves the performance from the base convolutional pre-trained network or the improvement is in general. We compare the results of Oxford-17 and Oxford-102 with the existing studies in **(Table 2 – 3),** respectively.

The results convey the importance of the fusion strategies as the best results on both datasets have been obtained using the proposed work. The second-best accuracy was achieved using FCN-CNN on both the datasets which are 1.3 % and 1.6 % less on Oxford-17 and Oxford-102 than the proposed work, respectively. It should also be noted that we only used 10 epochs for fine-tuning the individual pre-trained networks on the flower dataset for fusing the class probabilities which itself is plausible as many existing works fine-tune or train for atleast 30 epochs.

Table 2 Classification accuracy of existing works on Oxford-17 Datasets and its comparison with the proposed work

Method	Classification Accuracy
(Maria-Elena Nilsback and Zisserman, 2008)	88.33 %
(Ito and Kubota, 2010)	94.19 %
(Yuning Chai et al., 2011)	91.10 %
(Guangnan Ye et al., 2012)	91.70 %
(Qiang Chen et al., 2012)	93.50 %
(Khan et al., 2012)	95.00 %
(Fernando et al., 2014)	94.00 %
(Weiming Hu, Ruiguang Hu, Nianhua Xie, Haibin Ling, and Maybank, 2014)	91.39 %
(Yang et al., 2014)	97.80 %
(G.-S. Xie et al., 2015)	94.80 %
(Xiaoling Xia et al., 2017)	95.00 %
(Zhang, Li, et al., 2017)	87.10 %
(Zhang, Huang, and Tian, 2017)	93.70 %
(Hiary et al., 2018)	98.50 %
(Wu et al., 2018)	95.29 %
(Tian, Chen, and Wang, 2019)	90.50 %
Proposed Method	**99.80 %**

Table 3 Classification accuracy of existing works on Oxford-102 Datasets and its comparison with the proposed work

Method	Classification Accuracy
(M.-E. Nilsback and Zisserman, n.d.)	72.80 %
(Chai et al., 2012)	85.20 %
(Razavian et al., 2014)	86.80 %
(Murray and Perronnin, 2014)	84.60 %
(Qi et al., 2014)	84.20 %
(Qi Qian et al., 2015)	89.45 %
(Chakraborti, McCane, Mills, and Pal, 2016)	94.80 %
(Zheng et al., 2016)	95.60 %
(G.-S. Xie et al., 2017)	96.60 %
(Wei et al., 2017)	92.10 %
(Xiaoling Xia et al., 2017)	94.00 %
(Bakhtiary, Lapedriza, and Masip, 2017)	83.20 %
(L. Xie et al., 2017)	94.01 %
(Xu, Zhang, and Wang, 2018)	93.51 %
(Hiary et al., 2018)	97.10 %
Proposed Work	**98.70 %**

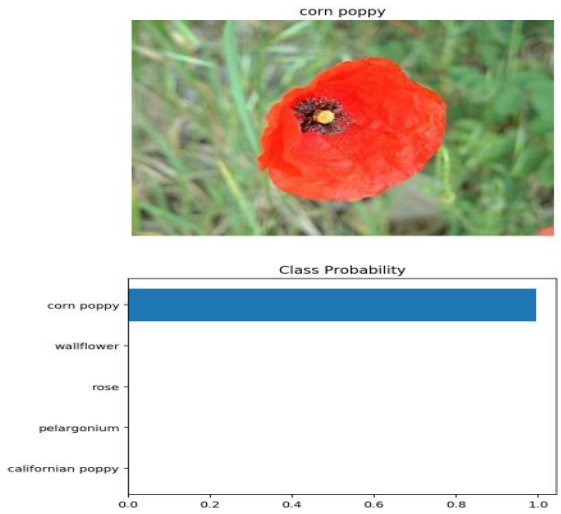

Fig. 2. An example of flower classification using Naive Bayes (meta-learner) fusion strategy using multiple pre-trained networks

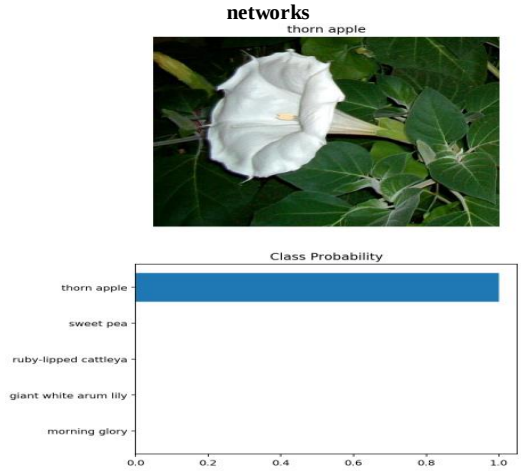

Fig. 3. An example of flower classification using Naive Bayes (meta-learner) fusion strategy using multiple pre-trained networks

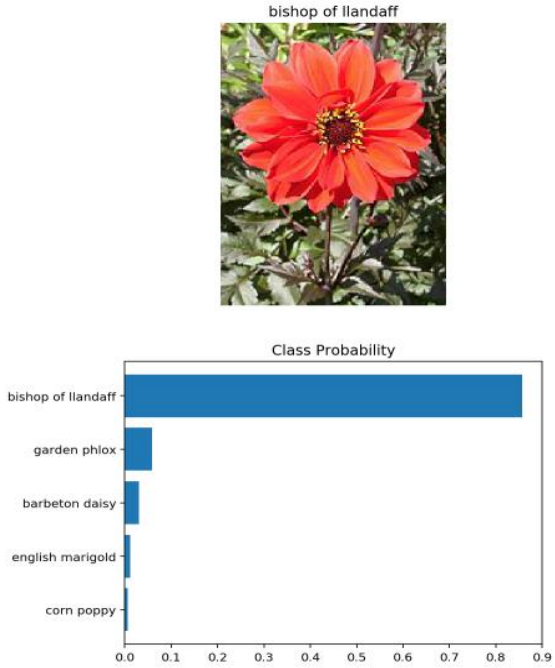

Fig. 4. An example of flower classification using Naive Bayes (meta-learner) fusion strategy using multiple pre-trained networks

5. CONCLUSION

In this paper, we present a flower classification method using transfer learning and fusion strategies. We used individual pre-trained networks such as VGG19, GoogleNet, ResNet101, and DenseNet161 and fused their class probabilities using average weighting, adaptive weighting, and meta-learning techniques. The use of fusion strategies specifically using meta-learning with Naïve Bayes shows significant improvement in recognition performance not only concerning the individual pre-trained networks but also in comparison to the existing studies. The best classification accuracy on both Oxford-17 and Oxford-102 has been achieved using the proposed work.

Although we have achieved the best results on both flower classification datasets so far, this achievement is accomplished at the cost of increased computational power. As the existing studies only train individual network architecture, we fine-tune four existing networks which increase the number of parameters to be optimized. Furthermore, adding a meta-learner to learn the class probabilities is another computational overhead that needs to be considered. Although, the test time is not so significant as compared to the recognition from individual pre-trained networks it can be reduced by applying the fusion strategies at the start of pre-trained networks. This will not only reduce the testing time but also the computational overhead for training as the number of parameters will be significantly reduced. We intend to apply the feature-level fusion strategies on flower classification datasets as our future work to

provide a trade-off between computational complexity and classification accuracy.

REFERENCES:

Bakhtiary, A. H., A. Lapedriza, and D. Masip, (2017). Winner takes all hashing for speeding up the training of neural networks in large class problems. *Pattern Recognition Letters*, *93*, 38–47. https://doi.org/10.1016/j.patrec.2017.01.001

Chai, Y., E. Rahtu, V. Lempitsky, L. Van Gool, and A. Zisserman, (2012). TriCoS: A Tri-level Class-Discriminative Co-segmentation Method for Image Classification. *European Conference on Computer Vision*, 794–807. https://doi.org/10.1007/978-3-642-33718-5_57

Chakraborti, T., B. McCane, S. Mills, and U. Pal, (2016). Collaborative representation based fine-grained species recognition. *International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 1–6. https://doi.org/10.1109/IVCNZ.2016.7804421

Chi, Z. (2003). Data Management for Live Plant Identification. *Multimedia Information Retrieval and Management*, 432–457. https://doi.org/10.1007/978-3-662-05300-3_20

Das, M., R. Manmatha, and E.M. Riseman, (1999). Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems*, *14*(5), 24–33. https://doi.org/10.1109/5254.796084.

Feichtenhofer, C., A. Pinz, and A. Zisserman, (2016). Convolutional Two-Stream Network Fusion for Video Action Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1933–1941. https://doi.org/10.1109/CVPR.2016.213.

Fernando, B., E. Fromont, and T. Tuytelaars, (2014). Mining Mid-level Features for Image Classification. *International Journal of Computer Vision*, *108*(3), 186–203. https://doi.org/10.1007/s11263-014-0700-1

Guangnan Ye, L. Dong, I-Hong Jhuo, and Shih-Fu Chang. (2012). Robust late fusion with rank minimization. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3021–3028. https://doi.org/10.1109/CVPR.2012.6248032

He, K., X. Zhang, S. Ren, and J. Sun, (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hiary, H., H. Saadeh, M. Saadeh, and M. Yaqub, (2018). Flower classification using deep convolutional neural networks. *IET Computer Vision*, *12*(6), 855–862. https://doi.org/10.1049/iet-cvi.2017.0155

Hsu, T.-H., Lee, C.-H., and Chen, L.-H. (2011). An interactive flower image recognition system. *Multimedia Tools and Applications*, *53*(1), 53–73. https://doi.org/10.1007/s11042-010-0490-6

Huang, G., Z. Liu, L. Maaten, van der, and K.Q. Weinberger, (2017). Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Healthcare Internet of Things, Services, and People (HIoTSP): An architectural framework for healthcare monitoring using wearable sensors. *Computer Networks*, *145*, 190–206. https://doi.org/10.1016/j.comnet.2018.09.003

Ito, S., and S. Kubota, (2010). Object Classification Using Heterogeneous Co-occurrence Features. *European Conference on Computer Vision*, 701–714. https://doi.org/10.1007/978-3-642-15555-0_51

Jie Zou, and G. Nagy, (2004). Evaluation of model-based interactive flower recognition. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 311-314 Vol.2. https://doi.org/10.1109/ICPR.2004.1334185

Kenrick, P. (1999). Botany: The family tree flowers. *Nature*, *402*, 358–359.

Khan, F. S., J. van de Weijer, and M. Vanrell, (2012). Modulating Shape Features by Color Attention for Object Recognition. *International Journal of Computer Vision*, *98*(1), 49–64. https://doi.org/10.1007/s11263-011-0495-2

Khowaja, S. A., K. Dahri, M. A. Kumbhar, and A.M. Soomro, (2015). Facial expression recognition using two-tier classification and its application to smart home automation system. *International Conference on Emerging Technologies (ICET)*, 1–6. https://doi.org/10.1109/ICET.2015.7389223

Khowaja, S. A., P. Khuwaja, and I.A. Ismaili, (2019). A framework for retinal vessel segmentation from fundus images using hybrid feature set and hierarchical classification. *Signal, Image and Video Processing*, *13*(2), 379–387. https://doi.org/10.1007/s11760-018-1366-x

Khowaja, S. A., and S.L. Lee, (2019). *Semantic Image Networks for Human Action Recognition*. Retrieved from http://arxiv.org/abs/1901.06792

Khowaja, S. A., A.G. Prabono, F. Setiawan, B. N. Yahya, and S. L. Lee, (2018). Contextual activity based

Khowaja, S. A., B. N. Yahya, and S.L. Lee, (2017). Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems. *Expert Systems with Applications*, *88*, 165–177. https://doi.org/10.1016/j.eswa.2017.06.040

Khuwaja, P., S. A. Khowaja, I. Khoso, and I.A. Lashari, (2019). Prediction of stock movement using phase space reconstruction and extreme learning machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 1–21. https://doi.org/10.1080/0952813X.2019.1620870

Krizhevsky, A., I. Sutskever, and G.E. Hinton, (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Larson, R. (1992). *Introduction to Floriculture* (2nd ed.). San Diego, CA, USA: Academic Press.

Liu, Y., F. Tang, D. Zhou, Y. Meng, and W. Dong, (2016). Flower classification via convolutional neural network. *IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA)*, 110–116. https://doi.org/10.1109/FSPMA.2016.7818296

Mottos, A. B., and R.S. Feris, (2014). Fusing well-crafted feature descriptors for efficient fine-grained classification. *IEEE International Conference on Image Processing (ICIP)*, 5197–5201. https://doi.org/10.1109/ICIP.2014.7026052

Murray, N., and F. Perronnin, (2014). Generalized Max Pooling. *IEEE Conference on Computer Vision and Pattern Recognition*, 2473–2480. https://doi.org/10.1109/CVPR.2014.317

Nilsback, M. E., and A. Zisserman, (n.d.). A Visual Vocabulary for Flower Classification. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, *2*, 1447–1454. https://doi.org/10.1109/CVPR.2006.42

Nilsback, Maria-Elena, and A. Zisserman, (2008). Automated Flower Classification over a Large Number of Classes. *Sixth Indian Conference on Computer Vision, Graphics and Image Processing*, 722–729. https://doi.org/10.1109/ICVGIP.2008.47

Paszke, A., S. Gross, S. Chintala, G. Chanan, E.Yang, Z. DeVito and A Lerer, (2017). Automatic Differentiation in PyTorch. *NIPS Autodiff Workshop*.

Qi Qian, R. Jin, S. Zhu, and Yuanqing Lin. (2015). Fine-grained visual categorization via multi-stage metric learning. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3716–3724. https://doi.org/10.1109/CVPR.2015.7298995

Qi, X., R. Xiao, C.G. Li, Qiao, Y., Guo and X. Tang, (2014). Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(11), 2199–2213. https://doi.org/10.1109/TPAMI.2014.2316826

Qiang Chen, Zheng Song, Yang Hua, Zhongyang Huang, and Shuicheng Yan. (2012). Hierarchical matching with side information for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 3426–3433. https://doi.org/10.1109/CVPR.2012.6248083

Razavian, A. S., H. Azizpour, J. Sullivan, and S. Carlsson, (2014). CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 512–519. https://doi.org/10.1109/CVPRW.2014.131

Simonyan, K., and A. Zisserman, (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*. Retrieved from http://arxiv.org/abs/1409.1556

Song, G., X. Jin, G. Chen, and Y. Nie, (2016). Two-level hierarchical feature learning for image classification. *Frontiers of Information Technology and Electronic Engineering*, *17*(9), 897–906. https://doi.org/10.1631/FITEE.1500346

Szegedy, C., Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tian, M., H. Chen, and Q. Wang, (2019). Flower identification based on Deep Learning. *Journal of Physics: Conference Series*, *1237*, 022060. https://doi.org/10.1088/1742-6596/1237/2/022060

Wei, X. S., J. H. Luo, J. Wu, and Z.H. Zhou, (2017). Selective Convolutional Descriptor Aggregation for Fine-Grained Image Retrieval. *IEEE Transactions on*

Image Processing, *26*(6), 2868–2881. https://doi.org/10.1109/TIP.2017.2688133

Weiming Hu, Ruiguang Hu, Nianhua Xie, Haibin Ling, and S. Maybank, (2014). Image Classification Using Multiscale Information Fusion Based on Saliency Driven Nonlinear Diffusion Filtering. *IEEE Transactions on Image Processing*, *23*(4), 1513–1526. https://doi.org/10.1109/TIP.2014.2303639

Wu, Y., X. Qin, Y. Pan, and C. Yuan, (2018). Convolution Neural Network based Transfer Learning for Classification of Flowers. *IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*, 562–566. https://doi.org/10.1109/SIPROCESS.2018.8600536

Xiaoling Xia, Cui Xu, and Bing Nan. (2017). Inception-v3 for flower classification. *2nd International Conference on Image, Vision and Computing (ICIVC)*, 783–787. https://doi.org/10.1109/ICIVC.2017.7984661

Xie, G. S., X.Y. Zhang, X. Shu, S. Yan, and C.L. Liu, (2015). Task-Driven Feature Pooling for Image Classification. *IEEE International Conference on Computer Vision (ICCV)*, 1179–1187. https://doi.org/10.1109/ICCV.2015.140

Xie, G. S., X.Y. Zhang, W. Yang, M. Xu, S. Yan, and C.L. Liu, (2017). LG-CNN: From local parts to global discrimination for fine-grained recognition. *Pattern Recognition*, *71*, 118–131. https://doi.org/10.1016/j.patcog.2017.06.002

Xie, L., R. Hong, B. Zhang, and Q. Tian, (2015). Image Classification and Retrieval are ONE. *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval - ICMR '15*, 3–10. https://doi.org/10.1145/2671188.2749289

Xie, L., J. Wang, W. Lin, B. Zhang, and Q. Tian, (2017). Towards Reversal-Invariant Image Representation. *International Journal of Computer Vision*, *123*(2), 226–250. https://doi.org/10.1007/s11263-016-0970-x

Xie, L., J. Wang, B. Zhang, and Q. Tian, (2016). Incorporating visual adjectives for image classification. *Neurocomputing*, *182*, 48–55. https://doi.org/10.1016/j.neucom.2015.12.008

Xu, Y., Q. Zhang, and L. Wang, (2018). Metric forests based on Gaussian mixture model for visual image classification. *Soft Computing*, *22*(2), 499–509. https://doi.org/10.1007/s00500-016-2350-4

Yang, M., L. Zhang, X. Feng, and D. Zhang, (2014). Sparse Representation Based Fisher Discrimination Dictionary Learning for Image Classification. *International Journal of Computer Vision*, *109*(3), 209–232. https://doi.org/10.1007/s11263-014-0722-8

Yuning Chai, V. Lempitsky, and A. Zisserman, (2011). BiCoS: A Bi-level co-segmentation method for image classification. *2011 International Conference on Computer Vision*, 2579–2586. https://doi.org/10.1109/ICCV.2011.6126546

Zhang, C., Q. Huang, and Q. Tian, (2017). Contextual Exemplar Classifier-Based Image Representation for Classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(8), 1691–1699. https://doi.org/10.1109/TCSVT.2016.2527380

Zhang, C., R. Li, Q. Huang, and Q. Tian, (2017). Hierarchical deep semantic representation for visual categorization. *Neurocomputing*, *257*, 88–96. https://doi.org/10.1016/j.neucom.2016.11.065

Zhang, C., J. Liu, C. Liang, Q. Huang, and Q. Tian, (2013). Image classification using Harr-like transformation of local features with coding residuals. *Signal Processing*, *93*(8), 2111–2118. https://doi.org/10.1016/j.sigpro.2012.09.007

Zheng, L., Y. Zhao, S. Wang, J. Wang, and Q. Tian, (2016). *Good Practice in CNN Feature Transfer*. Retrieved from http://arxiv.org/abs/1604.00133