

Effective classification of birds' species based on transfer learning

Mohammed Alswaitti¹, Liao Zihao¹, Waleed Alomoush², Ayat Alrosan², Khalid Alissa³

¹School of Electrical and Computer Engineering (ICT), Xiamen University Malaysia, Bandar Sunsuria, Selangor, Malaysia

²School of Information Technology, Skyline University College, Sharjah, United Arab Emirates

³Saudi Aramco Cybersecurity Chair, Department of Networks and communication, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Article Info

Article history:

Received Feb 3, 2021

Revised Mar 29, 2022

Accepted Apr 18, 2022

Keywords:

Birds' classification

Deep learning

Machine learning

Transfer learning

ABSTRACT

In recent years, with the deterioration of the earth's ecological environment, the survival of birds has been more threatened. To protect birds and the diversity of species on earth, it is urgent to build an automatic bird image recognition system. Therefore, this paper assesses the performance of traditional machine learning and deep learning models on image recognition. Also, the help-ability of transfer learning in the field of image recognition is tested to evaluate the best model for bird recognition systems. Three groups of classifiers for bird recognition were constructed, namely, classifiers based on the traditional machine learning algorithms, convolutional neural networks, and transfer learning-based convolutional neural networks. After experiments, these three classifiers showed significant differences in the classification effect on the Kaggle-180-birds dataset. The experimental results finally prove that deep learning is more effective than traditional machine learning algorithms in image recognition as the number of bird species increases. Besides, the obtained results show that when the sample data is small, transfer learning can help the deep neural network classifier to improve classification accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mohammed Alswaitti

School of Electrical and Computer Engineering (ICT), Xiamen University Malaysia

Bandar Sunsuria, 43900 Sepang, Selangor, Malaysia

Email: alswaitti.mohammed@xmu.edu.my

1. INTRODUCTION

Artificial intelligence (AI) has gradually come into the public's vision with the explosion of data. There is no doubt that artificial intelligence will replace some people in the future but it also creates some jobs. Although it will temporarily bring unemployment risk, in the long run, it can achieve industrial upgrading and social productivity progress. Artificial intelligence was first proposed in the 1950s. After experiencing the first high tide, it was limited by the hardware computing power at that time and fell into a low ebb in the 1970s. In [1], Hinton and Salakhutdinov came up with deep learning (DL) neural network, which made a breakthrough in artificial intelligence technology. Algorithms based on DL sprang up like weeds. From the end of 2016 to the beginning of 2017, Google's intelligent computer program Alpha Go defeated the world champions Li Sedol and Ke Jie successively, which triggered a wide discussion on artificial intelligence in the public domain and became another major milestone in the development of AI.

The ultimate goal of AI is to endow machines with the capacity to think and deal with problems as humans do. Image identification is basic and core area in the domain of computer vision. Its goal is to identify and understand the contents in the picture and categorize or classify its embedded objects [2]–[6].

Considering the image content, image recognition can be separated into two classes: general image recognition and fine-grained visual categorization. The objective of general image recognition is to distinguish objects [7]–[12], such as people, cars, and aeroplanes, which belong to coarse-grained image recognition tasks. Fine-grained image recognition is dedicated to the classification of subclasses under the broad category. For example, dogs of different breeds, such as Shiba Inu, Akita, and Golden Retriever, are classified and recognized. Common fine-grained image recognition tasks include bird, flower, insect, food, aeroplane, and automobile. Because subclasses often differ only in subtle ways, fine-grained image recognition is harder than general image recognition and has a wider scope of research requirements and usage scenarios in both academia and industry. Among all the fine-grained image recognition tasks, bird classification is one of the most typical and intricate recognition tasks on account of its large outside-class difference and small inside-class difference, and its recognition difficulty brings higher academic value to the subject research. Specifically, the difficulties of bird image recognition are: i) the diversity of attitude, illumination and shooting perspective as shown in Figure 1 makes the bird images have a large intra-class difference, and the biological dependence between different bird subclasses makes the bird images have a small inter-class difference; ii) the changeable posture and complex background also lead to the complex structure of the bird image, and it is always troublesome to express the complex information with simple image features; and iii) the collection and annotation of bird images are difficult and require a lot of expert knowledge, so the database norm is usually small. For deep learning, where the model is mostly complex, small data volume will bring the risk of overfitting, which is often a difficult problem faced by most fine-grained image identification.



Figure 1. Intra-class and inter-class variance of birds

As one of the most typical and complex fine-grained recognition tasks, bird recognition has high academic significance as well as strong practical significance. In the process of ecological environmental protection, efficient recognition of different species is an important precondition for ecological research. If the low-cost fine-grained image recognition can be realized with the help of a computer vision technique, it will be of great meaning to research and business. In the last few years, with the continuous deterioration of the earth's ecological environment, the survival of birds has been more threatened than ever before. The construction of an automatic bird identification system can not only strengthen the management and research of birds in ecological protection areas but also realize the monitoring and tracking of birds, especially endangered birds. In today's deteriorating ecological environment, image recognition of birds is more meaningful for protecting species diversity and maintaining ecological balance.

To protect birds and the diversity of species on earth, it is urgent to build an automatic bird image recognition system. In this research work, the classification of bird dataset by different machine learning algorithms is studied. Also, an analysis of the reasons for the different accuracy of these methods is elaborated. The dataset was retrieved from the Kaggle-230-birds-dataset (180 American Bittern) [13]. From the repository, 3, 6, 10, and 20 species of birds were selected and classified on these classifiers respectively, and the accuracy of these algorithms in the test set was calculated. In this research work, the classification algorithms are separated into 3 groups. The first one is the traditional machine learning algorithms, which includes: The support vector machine (SVM), linear discriminant analysis (LDA), adaptive boosting (AdaBoost), the k-means, multilayer perceptron (MLP), the random forest (RF), quadratic discriminant analysis (QDA), gaussian naive Bayes theorem, bagging classifier, the decision tree (DT), k-neighbors, and gradient boosting. The second group is the convolutional neural network (CNN) algorithms, which includes: ResNet-50 [14] (unpretrained), GoogLeNet [15] (unpretrained), DensNet-121 [16] (unpretrained), and AlexNet [17] (unpretrained). The third group is the convolution neural network algorithms based on transfer learning, which includes: Resnet-50 (pretrained), GoogLeNet [15] (pretrained), Densnet-121 (pretrained), and AlexNet (pretrained).

The rest of the paper is organized: section 2 is the literature review, primarily introduces the current research situation as well as relative research works. Section 3 is the basic knowledge of related technologies,

mainly introduces the related technologies involved in the research work, including the basic concepts and basic theories of deep learning, as well as the overview of support vector machines and integrated learning. Sections 4 and 5 elaborate on the bird image classification algorithm in deep learning proposed in this research. Finally, section 6 provides the summary, analyzes the work described in this thesis, and points out its deficiencies and future improvement directions.

2. RELATED WORK

In recent years, fine-grained image identification, especially bird recognition, is a very hot topic in computer vision [18]. Due to the difference between a subclass object mainly reflected on the local details, so how to effectively identify the target, and to discover some small area is the key in a fine-grained image recognition algorithm. At present, fine-grained image recognition research is separated into two types, the classification model based on weak supervision and strong learning supervision depending on the amount of supervision information

2.1. Fine-grained image classification model based on strong-supervised learning

Most early bird image recognition techniques were based on multi-level classification models of artificial features. In the technical report for the standard bird image database CUB200-2011 [18], Wah *et al.* [18] implemented the classification using local areas and a word package model based on traditional characteristics, but this method only gave 10.3% of the benchmark test results on the validation set of CUB200-2011 database. After that, Berg and Bellhumeur [19] proposed the part-based one-vs.-one feature (POOF) feature, which is a feature coding algorithm based on local areas. It can automatically find image information that plays a vital part in classification but requires high positioning accuracy of key points. When using the database to precisely annotate the information, the algorithm achieved a recognition accuracy of 73.3% in CUB200-2011. But when using a location algorithm to determine key points, the algorithm accuracy was only 56.8%. Although the recognition accuracy is not high, this was the first work that proposes a two-step classification strategy of location-recognition, which becomes one of the datum methods in the field of bird recognition.

In addition, Yao *et al.* [20] and Yang *et al.* [21] tried to replace the sliding window in the localization algorithm with the template matching method to reduce the algorithm complexity. The above studies showed that local area information is crucial for bird recognition, while stronger feature expression and feature encoding formula also have a great influence on classification accuracy. However, most studies at this stage depend largely on manual annotation information to obtain more precise local localization or adopt a simple localization algorithm based on artificial features, which has great limitations in practical application.

In 2012, with the appearance of the AlexNet deep learning model [17], deep CNN gained huge success in the domain of general image classification. Donahue *et al.* [22] attempted to migrate CNN to fine-grained image recognition, proving the strong generalization ability of CNN features through experiments, and named the feature DeCAF. This method analyzed the CNN model trained on the ImageNet dataset [23], chose the outcome of the first full connection layer of AlexNet as the image features, and found the features abstracted by CNN possess stronger semantic information and higher differentiation than the artificial features. DeCAF builds a tube in CNN and fine-grained image recognition, which is of great significance.

In 2014, Zhang *et al.* [24] introduced part-based region based convolutional neural networks (R-CNN) which used the CNN architecture to abstract features from local areas. It is the first work based on the multi-level classification model of CNN features. In the first-level model, the classical target detection algorithm, R-CNN [25] is used to detect the key parts of the bird image, and the key parts are set as the head and body. Then, geometric constraints are used to modify and fine-tune the detection frame. The second level model sends the detected local region images into AlexNet [17] to extract features. Finally, several features are combined, and the classification task is realized in the classification decision module. Compared to the previous methods, part-based R-CNN further reduce the degree of the dependency on artificial tagging algorithm, where it is only needed in the training phase. Without any labelling information during the test, the technique can get a recognition accuracy of 73.9% in the standard birds' image database CUB200-2011 [18]. The performance is far better than the classification results obtained by algorithms based on traditional characteristics. Moreover, the position-based algorithms were improved from the detection link of key parts to increase the detection efficiency and accuracy, and the CNN network architecture was replaced and upgraded in the feature extraction link as in the part-stacked CNN presented by Huang *et al.* [26]. In [27], R-CNN in the first-level model was replaced by fully convolutional network (FCN), which could slightly increase the identification accuracy when AlexNet is also used to extract features. Although the above part

detection methods do not need any additional annotation in the testing process, they still need a large amount of annotation information in the training. Therefore, these methods belong to the strong supervision multi-level classification algorithm based on the location.

2.2. Fine-grained image classification model based on weak-supervised learning

The object-part attention model for fine-grained image classification (OPADDL) was proposed by Peng *et al.* [28] that uses class activation map (CAM) [29] to achieve target location and uses target-place spatial constraint model and clustering algorithm to achieve the location. Since the training of category activation diagram only requires category information at the image level and does not need any site annotation, this method belongs to the weak supervised multi-level classification model based on location. Compared to a strong supervised classification algorithm, a weak-supervised algorithm can reduce the use of annotation and is more widely used in real situations. However, at the same time, the reduction of annotation information will lead to a decrease in positioning accuracy and thus the recognition performance will be limited.

The fine-grained classification model of constellation proposed by Simon and Rodner [30] visualized the features abstracted by CNN. Also, it discovered that areas with mighty reactions are just relevant to some critical parts in the input image and extracted local area information based on these critical points. The feature output resolution and input Image were large, which made the area for precise positioning of the input image is difficult. Hence, a gradient map was used to generate a regional location. Specifically, the output is a characteristic of convolution $W \times H \times P$ dimension tensor, where P is the channel number. Each channel can be represented as a $W \times H$ dimension matrix. A strong area of response in the characteristics of the gradient map represents a local region in the input image, and each response is the strongest in a gradient map location as the key part of the original images. The outcome of the convolution layer of P channels, respectively acts in response to the P key positions, by sorting out the most important first. Finally, the key point to extract feature to complete fine-grained recognition.

It is observed from the literature that deep learning has been applied to bird recognition. Almost none of the related works has employed transfer learning in bird recognition. Hence, it seems to be worth trying to evaluate all these techniques on bird recognition and analyze the differences in their performances to recommend or guide the future enhancements for an optimal bird's recognition system.

2.3. Transfer learning

Transfer learning widely exists in-class activities. In case two different sharing the more factors, the easier it to transfer learning, otherwise, the more difficult, or even "negative transfer" happens [31]. For instance, learning to ride bicycles don't adapt to learn tricycle, because of the two different models of the centre of gravity position. There are three main problems to be studied in transfer learning, which are "what to transfer", "how to transfer" as well as "when to transfer". As the name implies, "what to transfer" refers to what knowledge can be transferred in a cross-domain or task. This knowledge can be separated into two parts, one is particular to a task or domain called specific knowledge; The other part is shared knowledge between the source field and the target field, called shared knowledge, and try to identify those parts of shared knowledge can improve the target locality energy. According to the research question "what to transfer", transfer learning methods can be divided into following the four categories including instance-based transfer learning, feature-based transfer learning, parameter-based transfer learning and relation-based transfer learning.

2.3.1. Instance-based transfer learning method

The main idea of the instance-based transfer learning approach is that if the source domain sample plays a very important role in the training of the target domain model, the weight of the source domain sample will be increased; otherwise, the weight of the source domain sample will be reduced. The ultimate goal is to maximize the help of the information obtained from the source field to the learning of the target field. An iterative learning method called TrAdaBoost was proposed in [31], which is very helpful for solving the problem of negative transfer learning. On the one hand, the weight of "bad" data was reduced, on the other hand, the weight of "good" data was increased, and generalization error's upper bound in the model was deduced based on the theory of probability approximately correct (PAC). This transfer learning approach on the basics of Boosting is widely used, but it also has some disadvantages, including weight mismatch; the first half of the classifier is ignored; the sample imbalance and the decrease of the source domain weight are too fast. Other research proposed an instance-weight framework to settle the matter of transfer learning in natural language processing scenarios [32]. Although the weight method based on instances is not hard to obtain the generalization error's upper bound and has good theoretical support [33], this kind of method is usually applicable to scenarios with small differences between source field distribution and target field distribution and does not apply to some complex computer vision tasks

2.3.2. Feature-based transfer learning

The main idea of feature-based transfer learning is to obtain the new feature representation of samples in the new space through feature transformation technology learning, and reduce the distribution difference between the source field and the target field in the new subspace, even if the domain sharing features are enhanced while the exclusive features are weakened, and then directly use the source domain data samples for training. This method has a better transfer capability than the instance weight method. Feature-based methods can be divided into feature selection and feature extraction. Pan *et al.* [34] proposed maximum mean discrepancy embedding (MMDE) method, whose idea is to map the data features between domains to a new space and reduce the original feature differences between domains in the new feature space. Due to the high computational complexity of MMDE, transfer component analysis (TCA) [35] algorithm was proposed to reduce the computational burden. Raina *et al.* [36] proposed a sparse feature coding representation that mainly deals with the source domain data and obtains a feature dictionary. Characteristics of the dictionary is a matrix; the matrix of the base vector holds the essential feature of the data. Data between domain are expressed with base vector coding, as long as with the characteristics of the base vectors can greatly reduce differences between domain. In [37], a maximum interval straight push migration method of study, the method of thinking is derived from the regular risk minimization and maximum average differences. Under the framework of classification regularization, the authors tried to find the new feature representation between the source domain and the target domain in the feature mapping space, and then reduce the difference between the domains, and finally realize the knowledge transfer.

2.3.3. Parametric transfer learning

The idea of this method is to assume that there exists a gap between the source domain and the target domain or some prior distributions or model parameters are available to share. These prior distributions or model parameters are then transferred during training and testing. In [38], the authors used prior knowledge in the gauss process to establish connections between multiple tasks. Also, other researchers proposed to model the prior knowledge between tasks using the covariance matrix between the source domain and the target domain [39]. Finkel and Manning [40] proposed a hierarchical Bayesian prior knowledge transfer learning algorithm wherein [41] the authors proposed a transfer learning method based on the regularization framework. The transferred knowledge is the main parameter of the SVM. The idea is to divide SVM parameters into special parameters and general parameters.

2.3.4. Transfer learning method based on the structural relationship

The transfer learning method based on the structure relation is the most difficult learning scenario in the transfer learning method. Mainly because there is no similarity between the source domain sample and the target domain sample in the instance. On the other hand, it is difficult to map the features into the new subspace by feature transformation to reduce the differences between domains. The only useful knowledge that can be transferred between the source and target domains is the structural relationship between the data, such as the distance relationship between the data. The transfer learning method based on the structural relationship takes the structural relationship as the knowledge of the transfer and shares the structural relationship between the source domain and the target domain, finally improving the performance of the model.

3. EXPERIMENTAL DESIGN OF THE EXISTING CLASSIFICATION TECHNIQUES

3.1. Dataset description

Kaggle is a dataset website that provides various kinds of datasets. In this work, Kaggle-180-birds-dataset [13] which includes a total of 20,000 images of 180 species of birds has been chosen for the evaluation of the existing classification techniques. All images are cropped coloured images with $224 \times 224 \times 3$ size where the body of the bird occupies more than 50% of the pixels of the image. These settings were applied to improve the accuracy of the CNN classifier.

3.2. The implementation of the experiments

3.2.1. Data preprocessing

The bird dataset used in this experiment has 180 folders, each folder contains at least one hundred pictures of birds. Before the experiment began, a Python script was used to move about 16.7% of the images in each folder in the data set to the test set folder, making the number of test samples and training samples approximately with a 1:5 ratio. It is worth mentioning that each folder will have the same bird in it, with the same tag.

3.2.2. Computing platform

All experiments of all algorithms have been implemented on the same CPU+GPU heterogeneous platform for a fair comparison. The selected CPU was Intel Core i7-7700@3.6 GHz with 32.00 GB memory. Also, to boost the experiment's speed, an NVIDIA GeForce GTX 1080Ti was configured for that purpose. The specific parameters are shown in Table 1.

Table 1. The specification of the computing platform

CPU	Type	Intel Core i7-7700@3.6 GHz
	Memory	32.00 GB
GPU	Type	NVIDIA GeForce GTX 1080Ti
	Core frequency	1582 MHz
	Memory frequency	11000 MHz
	Memory	11 GB
	Bit width	352 bit
	CUDA core	3584
	CUDA running version	9.0
CUDA driver version	9.0	

3.2.3. Software environment

Experiments were carried out in Windows 7 operating system. The used ResNet [14], AlexNet [17], GoogLeNet [15], DenseNet [16] techniques are based on Python+Pytorch packages. Besides, SVM, DT, RF, AdaBoost, k-means, LDA, QDA, k-neighbors, MLP, naive Bayes, bagging classifier, gradient boosting are based on Python+sklearn library. Similarly, image reading and conversion are done based on Python+NumPy+Pil. Image libraries.

3.2.4. Dimension reduction process

In the static image clustering system, each image is a data sample and each pixel on the image is a feature dimension. Since the colours of the adjacent pixels of the image are always very close in most cases, there must be a lot of redundant features in the system. For an image of $224 \times 224 \times 3$ pixels, its dimension will be $224 \times 224 \times 3 = 150528$. Having so many dimensions during classifier construction can make the program run very slowly and consume a lot of computational resources. Principal component analysis (PCA) can be used to reduce the number of dimensions of such a system from hundreds of thousands to hundreds or even tens, and then cluster on the new dimension. This will speed up the clustering process. In this experiment, the sklearn Python package was used for the decomposition by setting $n_components=0.95$. Hence, the final dimensionality reduction results retain 95% of the original information.

3.2.5. Parameter settings

During the training process, the model can be trained more accurately by reducing the learning rate and increasing the training epochs, although this may be more time-consuming. Adam optimizer [42] was chosen because it combines the advantages of other optimization algorithms and performs well in most cases. In the process of deep learning, the learning rate is set to 0.0005 and the number of learning epochs is 100.

4. EXPERIMENT RESULTS AND ANALYSIS

4.1. The implementation of the experiments

After 100 rounds of training, the experiment has obtained a large amount of data that can be used for analysis. It is worth mentioning that the PCA has been used for dimensionality reduction purposes. Table 2 lists the pre-and post-used dimensions in the experiments.

4.1.1. Results of PCA dimension reduction

In the static image clustering system, each image is a data sample and each pixel on the image is a feature dimension. Since the colors of the adjacent pixels of the image are always very close in most cases, there must be a lot of redundant features in the system. For a picture of $224 \times 224 \times 3$, its dimension is equal to 150528. Having so many dimensions during classifier construction can make the program run very slow and consume a lot of computer resources. Principal component analysis (PCA) can be used to reduce the number of dimensions of such a system from hundreds of thousands to hundreds or even tens, and then cluster on the new dimension. This will speed up the clustering process. In this experiment the Python package sklearn. Decomposition has been used. PCA to the data dimensionality reduction was set $n_components=0.95$ so that the final dimensionality reduction results retain 95% of the original information.

4.1.2. Classification results of traditional machine learning algorithms

In Figure 2, the horizontal axis represents different classifiers and the vertical axis represents classification accuracy. As can be observed from the chart, the classification accuracy of these 12 classification algorithms in the bird image set is up to 82%. Moreover, with the increase of classified species, the classification accuracy becomes lower and lower. When the species is 20, the classification accuracy is only 53%. Both SVM and LDC classifiers showed more promising results than other classifiers.

Table 2. Results of the PCA dimension reduction

Number of bird species	Original dimension	Dimension after reduction
3	150528	228
6	150528	439
10	150528	662
20	150528	992

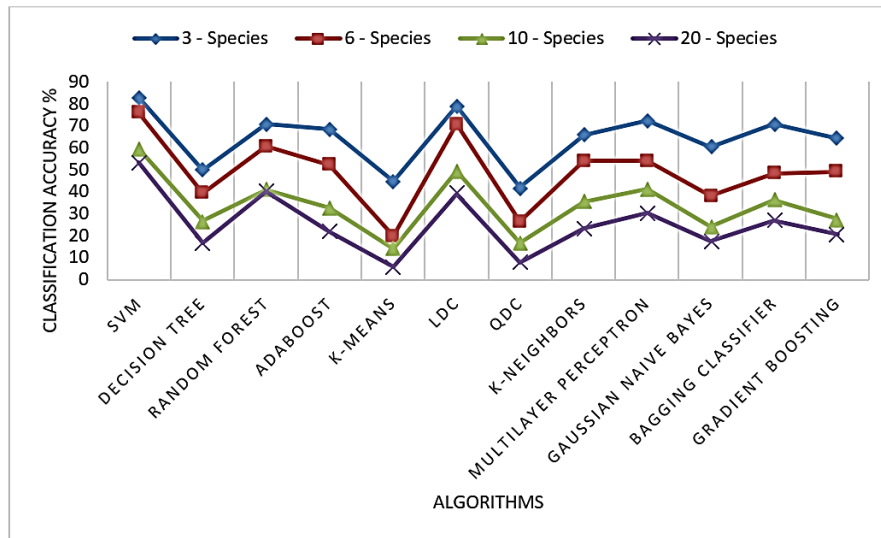


Figure 2. The effect of bird species on the classification accuracy of traditional machine learning algorithms

4.1.3. Classification results of CNN models

This section compares the performance of the second set of chosen algorithms (AlexNet, DenseNet-121, GoogLeNet, and ResNet-50) which has been discussed previously as being the most recent and innovative CNN models for image recognition. The models have been used in two different schemes as unpretrained (without prior training) and pretrained. In Figures 3(a)-(h), the horizontal axis represents the number of training rounds, and the vertical axis represents the classification accuracy on the test set. The summary of the achieved numerical classification accuracies of the CNN models are tabulated in Table 3. Their performance was analyzed on 4 different species sets. The best results on each number of species are bolded as shown in Table 3.

4.2. Result analysis

4.2.1. Classification effect on traditional classifiers

The obtained results showed the following conclusions about the performance of the traditional classifiers. When the number of species is 3, the classification accuracy ranges in 50-80%. On the contrary, when the number of species is 20, the classification accuracy is as low as 6-50%. With the increase in the number of species to be classified, the classification accuracy becomes worse and worse. The main reason is that there are too many classification dimensions and varieties. Although the number of dimensions has been reduced from 150,000 to several hundred using PCA dimensionality reduction, the number of dimensions is still too large for traditional machine learning. The traditional classification principle of machine learning is feature extraction for image classification and other tasks. A feature is an interesting, descriptive, or informational chunk of an image. This step may involve several computer vision algorithms, such as edge detection, corner detection, or threshold segmentation to extract as many features as possible from the image,

which form the definition of each object class (called a word package). If a large number of features in a word package appear in another image, the image is classified as containing that particular object (e.g., chair, and horse). The difficulty with this traditional approach is that you have to choose which features are important in each given image. With the increase of classes to be classified, feature extraction becomes more and more troublesome. It is up to the computer vision engineer’s judgment and a lengthy trial and error process to determine which features best describe different categories of objects. Besides, each feature definition needs to handle a large number of parameters, all of which must be fine-tuned by the computer vision engineer.

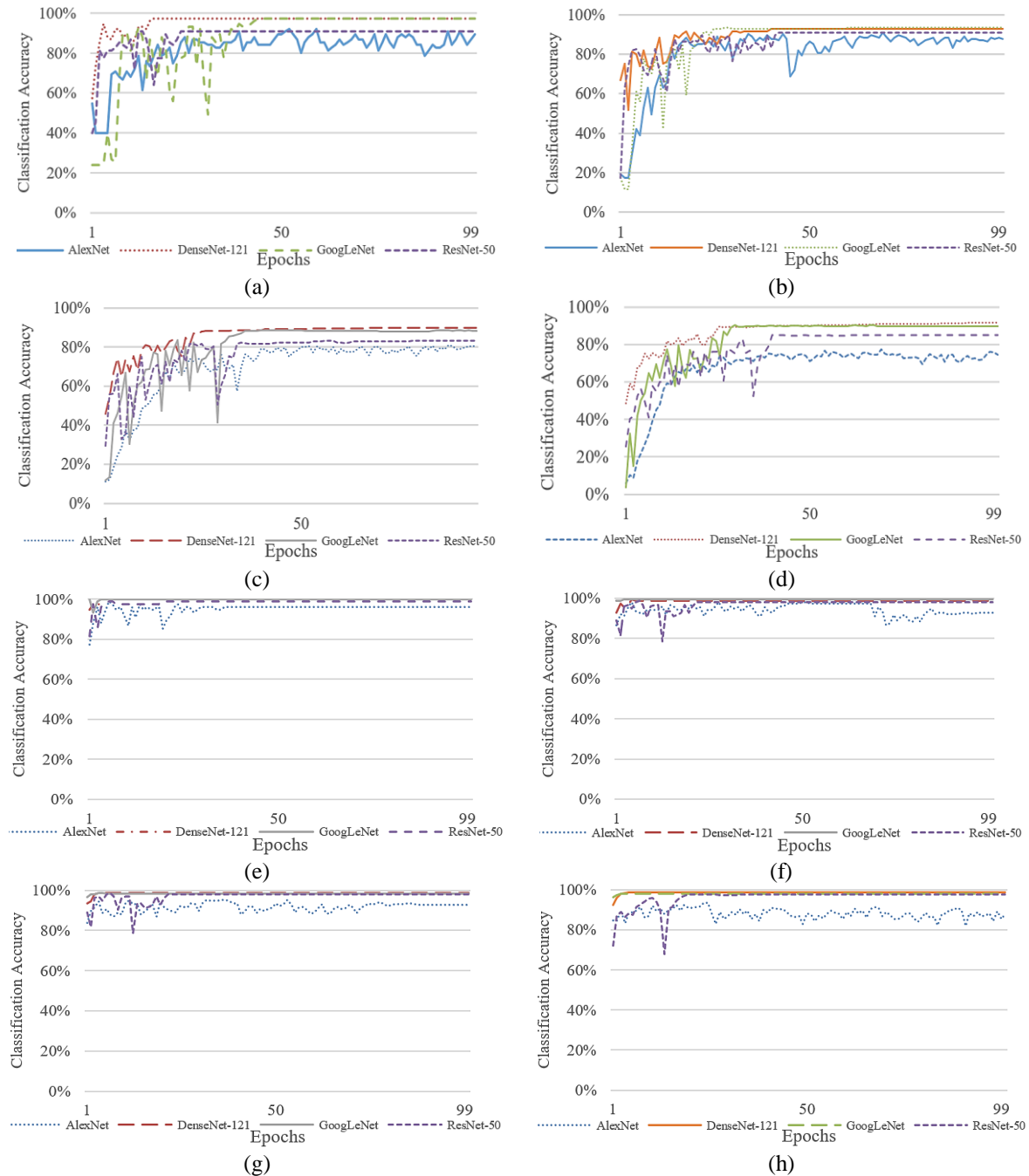


Figure 3. The classification accuracy of the state-of-the-art CNN models AlexNet, DenseNet-121, GoogLeNet, and ResNet-50. a–d, Obtained CA as unpretrained models. e–h (b), Obtained CA as pretrained models (a) 3 species (unpretrained), (b) 6 species (unpretrained), (c) 10 species (unpretrained), (d) 20 species (unpretrained), (e) 3 species (pretrained), (f) 6 species (pretrained), (g) 10 species (pretrained), and (h) 20 species (pretrained)

Table 3. Classification effect comparison of eight CNN models.

CNN Model	3 species CA%	6 species CA%	10 species CA%	20 species CA%
AlexNet	92	92.90	85.50	80.00
DenseNet	97.30	92.90	91.40	90.90
GoogLeNet	97.30	93.50	89.10	90.50
ResNet	90.70	90.90	83.20	84.70
AlexNet (pretrained)	98.70	98.10	95.30	93.40
DenseNet (pretrained)	100	99.40	98.80	98.60
GoogLeNet (pretrained)	100	99.40	98.40	98.10
ResNet (pretrained)	98.70	94.80	98.10	97.70

4.2.2. Image classification effect of CNN

As can be seen from the experimental results, the classification accuracy of CNN models for birds easily reaches more than 90% accuracy. With the increase of species, the accuracy drops slightly and exceeds 80% alone. For some network models, the classification accuracy still reaches 97-98%. Each bird in the dataset used in this experiment contains about a hundred pictures. If each bird contains hundreds or thousands of pictures, the classification effect will be better.

The reason why CNN has a good image classification effect is that CNN introduced the concept of end-to-end learning. The machine only needs to get an image data set, which has annotated the object categories existing in each image. Thus, deep learning models are “trained” on a given set of data, neural networks detect potential patterns in these images, and automatically calculate for each object the most descriptive and salient features associated with each particular class of objects as shown in Figures 4(a) and 4(b). Due to the large number of parameters in the neural network, the increase of the sample number can adjust the parameters more, and the potential patterns in these images become more and more obvious. Therefore, with the increase of the sample number, the classification effect gets better and better.

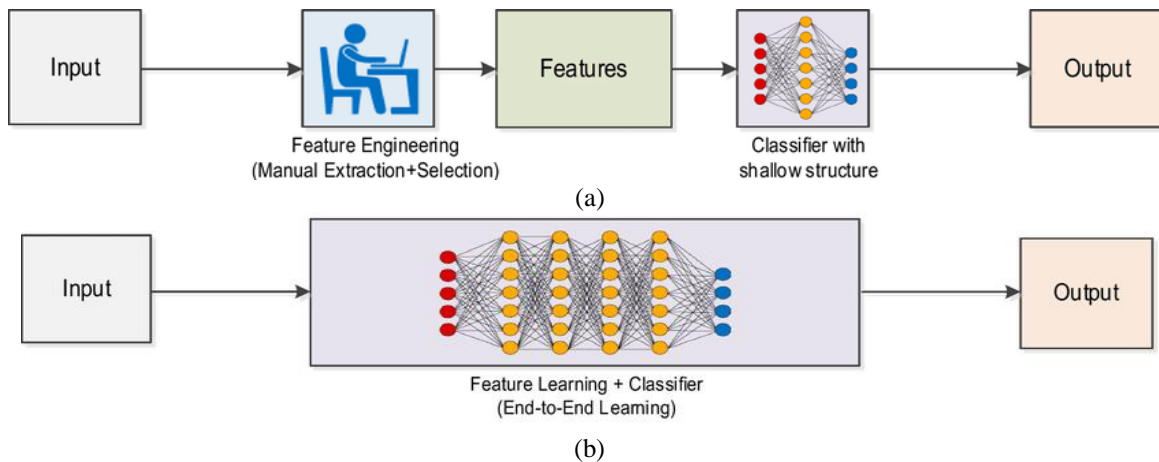


Figure 4. Comparison between two classification techniques in (a) traditional machine learning and (b) deep learning [43]

4.2.3. CNN models based on transfer learning

The pretrained neural network model has a classification accuracy of 90% to 100% and easily reaches 98% to 99%. The CNN based on migration learning has a better training effect than the randomly initialized CNNs, for the reason that deep learning requires a large amount of data. For small data sets, the untrained model can be equipped with image recognition ability by transferring model parameters. The specific method of migration in this experiment is:

- a) Import the CNN model that has been trained on the ImageNet dataset and achieved very good results
- b) Modify the final output layer
 - Num_of_features \rightarrow num_of_outputs
 - The full connection layer of the original model is
 - Num_of_features \rightarrow 1000

- The full connection layer used for bird identification in this experiment is
- Num_of_features → 3/6/10/20
- c) Enter labelled pictures to start the training

Another advantage of transfer learning is that it can reduce training time and improve training speed. The experimental results showed that training the bird dataset with a pre-trained model requires no more than 10 epochs to obtain the best classification accuracy. Some models even require only 2 to 3 epochs, while training with an untrained model requires at least 40 to 50 epochs to achieve the highest classification accuracy. This is a big improvement on deep learning, which requires hardware support and a lot of training time.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

In the field of image recognition, classifiers based on traditional machine learning algorithms (such as SVM) require a lot of manual intervention, like feature selection, data preprocessing, and so on. This makes the classification accuracy of the classifier depends on these human factors to an extent. Deep learning makes use of the depth and breadth of the network layers to enable a deep neural network to have a strong learning ability, which can play a strong role in today's hardware devices. Transfer learning enables the knowledge learned in deep learning to be transferred to other deep learning models, which is of great help for deep learning with a small dataset.

In this research, three different classifiers were used to classify the Kaggle-180-birds dataset, they are traditional machine learning algorithm (SVM, and DT), deep learning algorithm (ResNet50), and transfer learning-based deep learning algorithm (ResNet50-pretrained). The results confirmed that the classifier based on transfer learning has the best classification effect, which can reach 98-100% (based on the classification category). This experiment fully proves the limitations of traditional machine learning in the field of image recognition and the advantages of deep learning in image recognition, as well as the help that transfer learning brings to deep learning.

5.2. Recommendations

Although the research in this article has achieved good results in bird recognition, compared with other classification algorithms, it also has some limitations. However, as a new machine learning method and one of the frontier research directions in artificial intelligence, transfer learning still has many problems to be further studied and explored. Based on this research experience, a series of other related research topics could be extended, and the prospect of future work is listed in the upcoming paragraphs.

In practical applications, migration learning is not necessarily a case of a single source domain and a single target domain, but often a case of multiple source domains. In this case, migration learning will be faced with the problem of how to choose the appropriate source domain as the training data set. In the case that it is not clear which source domain is better, random selection-a source domain is not advisable as the training set. A common solution to this problem is the multi-source migration learning approach. In other words, the source domain with a strong correlation with the target domain has a higher migration priority. Conversely, source domains that are less relevant to the target domain have a lower migration priority. Hence, future research on the transfer learning method of multiple source domains can be further studied.

Transfer learning aims to use the knowledge of the source domain to help the learning of the target domain. Therefore, the performance of the final target domain model largely depends on the degree of correlation between the source domain and the target domain. If there is a strong correlation between the source domain and the target domain, the transfer learning method can make good use of the correlation knowledge and apply it to the training process of the target domain model. If there is a weak correlation between the source domain and the target domain, if the knowledge transfer is carried out in a far-fetched way, it will be counterproductive, not only cannot improve the performance of the target domain model but even may reduce the effect, which will cause the negative transfer problem. This research focused on the positive transfer learning problem, that is, improving learning performance by transferring knowledge from the source domain to the target domain. However, in the practical application scenario, the model established due to the complexity of the problem will not conform to the actual situation, causing negative migration and side effects. If a person can ride a bicycle, he should learn to ride a motorcycle easily. However, sometimes seemingly simple things may not be easy to achieve, which may lead to the "negative transfer" problem. For example, learning how to ride a bicycle may not be suitable for driving a tricycle, because the two kinds of cars have a different centre of gravity. Therefore, how to realize positive transfer and avoid negative transfer is also an important research problem of transfer learning. In the future, the problem of negative transfer worth more attention and investigations.

ACKNOWLEDGEMENTS

This work was supported by Xiamen University Malaysia (XMUM) under the XMUM Research Fund (XMUMRF) received by M. A (Grant No: XMUMRF/2019-C4/IECE/0012).




REFERENCES

- [1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.
- [2] W. Alomoush *et al.*, "A survey: challenges of image segmentation based fuzzy C-means clustering algorithm," *Journal of Theoretical and Applied Information Technology*, 2018.
- [3] W. Alomoush and A. Alrosan, "Review: metaheuristic search-based fuzzy clustering algorithms," *arXiv preprint arXiv:1802.08729*, Jan. 2018.
- [4] W. Alomoush, A. Alrosan, A. Almomani, K. Alissa, O. A. Khashan, and A. Al-nawasrah, "Spatial information of fuzzy clustering based mean best artificial bee colony algorithm for phantom brain image segmentation," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, pp. 4050–4058, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4050-4058.
- [5] W. Alomoush and K. Omar, "Dynamic fuzzy C-mean based firefly photinus search algorithm for MRI brain tumor image segmentation," *Computer science*, 2015.
- [6] A. Alrosan *et al.*, "Automatic data clustering based mean best artificial bee colony algorithm," *Computers, Materials and Continua*, vol. 68, no. 2, pp. 1575–1593, 2021, doi: 10.32604/cmc.2021.015925.
- [7] A. A. Alomoush, A. A. Alsewari, H. S. Alamri, K. Z. Zamli, W. Alomoush, and M. I. Younis, "Modified opposition based learning to improve harmony search variants exploration," in *International Conference of Reliable Information and Communication Technology*, 2020, pp. 279–287, doi: 10.1007/978-3-030-33582-3_27.
- [8] W. Alomoush, A. Alrosan, Y. M. Alomari, A. A. Alomoush, A. Almomani, and H. S. Alamri, "Fully automatic grayscale image segmentation based fuzzy C-means with firefly mate algorithm," *Journal of Ambient Intelligence and Humanized Computing*, Sep. 2021, doi: 10.1007/s12652-021-03430-3.
- [9] W. Alomoush, K. Omar, A. Alrosan, Y. M. Alomari, D. Albashish, and A. Almomani, "Firefly photinus search algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 5, pp. 599–607, Jun. 2020, doi: 10.1016/j.jksuci.2018.06.010.
- [10] A. Alrosan, W. Alomoush, N. Norwawi, M. Alswaiti, and S. N. Makhadmeh, "An improved artificial bee colony algorithm based on mean best-guided approach for continuous optimization problems and real brain MRI images segmentation," *Neural Computing and Applications*, vol. 33, no. 5, pp. 1671–1697, Mar. 2021, doi: 10.1007/s00521-020-05118-9.
- [11] E. H. Houssein *et al.*, "An improved opposition-based marine predators algorithm for global optimization and multilevel thresholding image segmentation," *Knowledge-Based Systems*, vol. 229, Oct. 2021, doi: 10.1016/j.knsys.2021.107348.
- [12] A. Alrosan and N. Norwawi, "Mean artificial bee colony optimization algorithm to improve fuzzy C-means clustering technique for gray image segmentation," Universiti Sains Islam Malaysia, 2017.
- [13] Gerry, "325 bird species-classification." 2022, Accessed: Feb. 10, 2021. [Online]. Available: <https://www.kaggle.com/gpiosenka/100-bird-species>.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [15] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [18] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-UCSD birds-200-2011 dataset." 2011.
- [19] T. Berg and P. N. Belhumeur, "POOF: part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013, pp. 955–962, doi: 10.1109/CVPR.2013.128.
- [20] Bangpeng Yao, G. Bradski, and Li Fei-Fei, "A codebook-free and annotation-free approach for fine-grained image categorization," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 3466–3473, doi: 10.1109/CVPR.2012.6248088.
- [21] S. Yang, L. Bo, J. Wang, and L. Shapiro, "Unsupervised template learning for fine-grained object recognition," in *Advances in Neural Information Processing Systems*, 2012, vol. 25.
- [22] J. Donahue *et al.*, "DeCAF: a deep convolutional activation feature for generic visual recognition," in *Proceedings of the 31st International Conference on Machine Learning*, 2014, vol. 32, no. 1, pp. 647–655.
- [23] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
- [24] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based R-CNNs for fine-grained category detection," in *Computer Vision Textendash ECCV 2014*, Springer International Publishing, 2014, pp. 834–849.
- [25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
- [26] S. Huang, Z. Xu, D. Tao, and Y. Zhang, "Part-stacked CNN for fine-grained visual categorization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1173–1182, doi: 10.1109/CVPR.2016.132.
- [27] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.
- [28] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1487–1500, Mar. 2018, doi: 10.1109/TIP.2017.2774041.
- [29] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2921–2929, doi: 10.1109/CVPR.2016.319.




- [30] M. Simon and E. Rodner, "Neural activation constellations: unsupervised part model discovery with convolutional networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1143–1151, doi: 10.1109/ICCV.2015.136.
- [31] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proceedings of the 24th international conference on Machine learning-ICML '07*, 2007, pp. 193–200, doi: 10.1145/1273496.1273521.
- [32] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in NLP," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 264–271.
- [33] C. Cortes, Y. Mansour, and M. Mohri, "Learning bounds for importance weighting," in *Advances in Neural Information Processing Systems*, 2010, vol. 23.
- [34] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proceedings of the 23rd National Conference on Artificial Intelligence-Volume 2*, 2008, pp. 677–682.
- [35] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.
- [36] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning," in *Proceedings of the 24th international conference on Machine learning-ICML '07*, 2007, pp. 759–766, doi: 10.1145/1273496.1273592.
- [37] B. Quanz and J. Huan, "Large margin transductive transfer learning," 2009, doi: 10.1145/1645953.1646121.
- [38] A. Schwaighofer, V. Tresp, and K. Yu, "Learning Gaussian process kernels via hierarchical Bayes," in *Advances in neural information processing systems*, 2005, pp. 1209–1216.
- [39] E. V. Bonilla, K. Ming, A. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction." 2008.
- [40] J. R. Finkel and C. D. Manning, "Hierarchical Bayesian domain adaptation," 2009, doi: 10.3115/1620754.1620842.
- [41] T. Evgeniou and M. Pontil, "Regularized multi-task learning," 2004, doi: 10.1145/1014052.1014067.
- [42] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, Dec. 2014.
- [43] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.003.

BIOGRAPHIES OF AUTHORS






Mohammed Alswaitti    received the B.Eng. degree in Computer Engineering from the Islamic University of Gaza (IUG), Palestine, in 2010, and the M.Sc. degree in Electronic Systems Design Engineering from the Universiti Sains Malaysia (USM) in 2011. He worked as a lecturer in Software Engineering and Information Technology Faculties at the University of Palestine and as an Instructor at the Educational Technology Centre at IUG. He pursued his PhD degree in Computational Intelligence under the Malaysian International Scholarship (MIS) scheme and worked as a Graduate Assistant at the Electrical and Electronic Engineering Department, USM, Malaysia. Currently, he is an Assistant Professor at the School of Electrical and Computer Engineering, Xiamen University Malaysia (XMUM). He acts as the Research Coordinator for the school where his research is focused on nature-inspired optimization-based clustering techniques, machine/deep learning applications, and pattern recognition. Recently, Alswaitti managed to publish several articles in the top tier journals in AI field through securing research funds and collaborations with research institutions all over the world. Besides, he is a regular keynote speaker at academic and industry conferences and workshops and does voluntary work as a reviewer and associate editor for journals and language editing services. He can be contacted at email: alswaitti.mohammed@xmu.edu.my.






Liao Zihao    received his Bachelor degree in computer science from Xiamen University Malaysia 2019. He works as a junior data scientist at Xperia industrial for computer vision. His research interest is focused on the applications of deep learning techniques in computer vision, images reconstruction and enhancement. He can be contacted at email: CST1609041@xmu.edu.my.






Waleed Alomoush    received a PhD degree from University Kebangsaan Malaysia (UKM) in 2015. He has published many research papers in International Journals and Conferences of high repute. Currently, he is an assistant professor at School of Information Technology, Skyline University College, Sharjah, United Arab Emirates. His research interest includes, but not limited to, Data clustering, optimization and Image processing. He can be contacted at email: waleed.alomoush@skylineuniversity.ac.ae.



Ayat Alrosan    received a PhD degree from Universiti Sains Islam Malaysia (USIM) in 2017. She has published many research papers in International Journals and Conferences of high repute. Currently, she is an assistant professor at School of Information Technology, Skyline University College, Sharjah, United Arab Emirets. His research interest includes image processing, data clustering, and optimization. She can be contacted at email: ayat.alrosan@skylineuniversity.ac.ae.



Khalid Alissa    received a PhD degree from Universiti Sains Islam Malaysia (USIM) in 2017. He has published many research papers in International Journals and Conferences of high repute. Currently, she is an assistant professor at School of Information Technology, Skyline University College, Sharjah, United Arab Emirets. His research interest includes image processing, data clustering, and optimization. She can be contacted at email: kaalissa@iau.edu.sa.