DOI: http://dx.doi.org/10.26483/ijarcs.v15i2.7070

ISSN No. 0976-5697

Volume 15, No. 2, March-April 2024

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Classification method for Insects using Data Augmentation and Deep Neural Networks

Bui Hai Phong
Hanoi Architectural University,
Hanoi, Vietnam

Pham Hoang Quan
Hanoi-Amsterdam High School,
Hanoi, Vietnam

Nguyen Thi Hong Thuy
Trade Union University,
Hanoi, Vietnam

Nguyen Duc Manh
HUS High School,
Hanoi, Vietnam

Truong Manh Dung
HUS High School,
Hanoi, Vietnam

Abstract: In the nature, there exists a huge number of species of insects. Insects have caused damage for human and crops. Traditional identification methods of insects require expert knowledge and time consuming. Therefore, the automatic identification and classification have been more and more necessary. In recent year, one of the efficient approaches to classify insects is the application of Deep neural networks. The paper presents the improvements for the classification of insect images. Firstly, we apply the data augmentation to improve the number of insect images. Then, we applied various deep neural networks to improve the classification accuracy of insect classification. Obtained classification accuracy of 98% on public insect datasets shows the efficiency of the proposed method.

Keywords: Insect classification, Machine learning, Deep Neural Networks, Feature extraction, Data augmentation

I. INTRODUCTION

In the nature, there exists a million species of insects. Insects have caused damage for human and crops. The identification of insects allows us to prevent them [1]. Traditional classification methods of insects require expert knowledge (e.g., knowledge of biologist). Several appearance properties of insects can be considered to classify them (e.g., wing, body shape, head shape, number of legs) [2]. However, the traditional classification is a time-consuming task and the classification errors can be found. So far, we have been able to identify about 750 thousand of species of insects. A huge number of insects have disappeared and we cannot identify them [2], [3].

There are several difficulties for the automatic classification of insects. The size and shapes of insects are different. The sizes of insects may be tiny, medium and large. The insects may have different color. Moreover, in nature, the colorful background causes many difficulties for the classification of insects. Therefore, the classification of insects in natural images is a challenging task [3].

In recent years, with the rapid development of computing technology, the automatic identification of insects has been well researched. Based on collected data, machines are able to discriminate insects automatically. The classification approach of insects based on computer vision and machine learning has been considered as an efficient way [4]. Figure 1 shows examples of various species of insects.

The paper investigates the improvements of the classification of insect images. The contributions of the paper are two folds:

(1) The paper analyzes and compares various approaches of the classification of insects in images. (2) The paper applies the augmentation strategies to improve the classification accuracy of insects.

Figure 1 Different species of insects

II. RELATED STUDIES

This section reviews and analyzes existed solutions for the classification of insects. The classification methods can be divided into two categories. The first one applied the visual features and machine learning classifiers. The second one applied the deep neural networks to improve the accuracy of the classification [3].

A. Traditional classification method of insects based on handcrafted feature extraction and machine learning classifiers.

The work in [5] proposed a method to classify insects using image processing. The work extracted low-level features such as color, shapes and sizes of insect images. The work in [6] combined several feature extraction methods and classifiers to classify orchard insects. Firstly, a region feature and a scale invariant feature transform (SIFT) descriptor are applied to extract features of images. Then, six classifiers (e.g., K nearest neighbors, support vector machine, linear classifier) are applied to classify 5 classes of insects. Other features of insect images such as texture, geometry, contour and color [7] have been extracted to classify insects. The properties and characteristics of insects have been applied to classify them [8].

B.Deep neural networks for the classification of insects.

The works in [9] applied the CNN based on VGG-19 to detect and classify insects. The work obtained the detection accuracy of 88% on the private dataset. The work faced many errors when detecting tiny insects. The work in [7] optimized the Densenet-201 to recognize 19 classes of insects. The work obtained the classification accuracy of 87%. The work in [10] provides a comparison of different neural network structures for the classification of insect images. The work in [11] applied the transfer learning of Inception v3 network to detect crop pests. The work also applied the Unsupervised Auto-encoder to detect pests in small datasets. The work in [12] applied the data augmentation and the transfer learning of Convolutional Neural Networks (CNN) (e.g., VGG-16, ResNet-50) to classify butterflies. There are several approaches for the classification of insects, however, the accuracy of the classification needs to be improved.

III. PROPOSED METHOD FOR THE CLASSIFICATION OF INSECTS

The overall framework of our proposed system is described in Figure 2. The framework consists of two steps. Firstly, the data augmentation based on image processing is applied to increase the number of insect images. Then, various deep neural networks (DNN) are fine-tuned to improve the classification accuracy of insects.

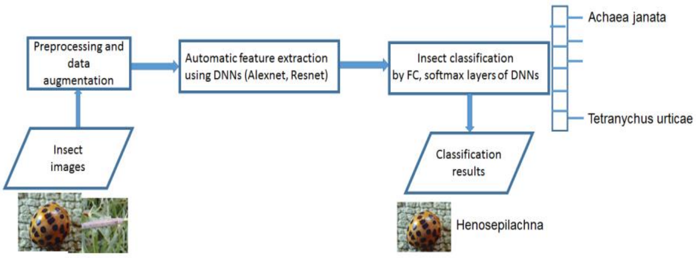

Figure 2 Overall steps of the classification of insects.

A.Data augmentation of insect images.

The augmentation using image processing is applied to increase the number of insect images. The data augmentation allows to train DNN models efficiently. The image processing techniques including rotation and noise addition are applied to increase the number of training images. Figure 3 demonstrates the application of image processing to increase the number of images. The images are rotated with the angle from 15 to 30 degrees. The Gaussian and pepper noises are added to the original image [13].

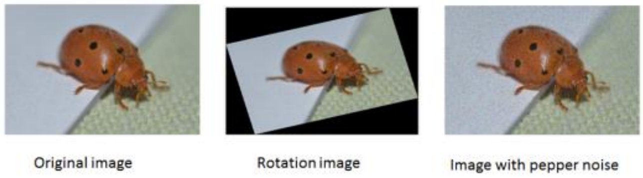

Figure 3 The augmentation techniques using the rotation and noise addition image processing.

B. The classification of insects using deep neural networks

In the paper, advanced DNNs have been applied and fine-tuned to classify insect images in an end-to-end way. The feature extraction and classification are performed by using DNNs including Alexnet [14], Resnet-50 [15] and VGG-16 [16]. Detail information of the DNNs is described in Table I. Figure 4 and 5 demonstrate the accuracy and the loss values of the Resnet-50 during the training process. The Figures show that the accuracy increases and the loss values decrease during the training DNNs. Input insect images are resized and normalized as the requirement of the DNNs. The implementation of the DNNs is supported by the Matlab 2021b environment with the 8GB RAM and core-i5 processor. Moreover, the implementation on the GPU GTX 980 is performed to compare the computing time.

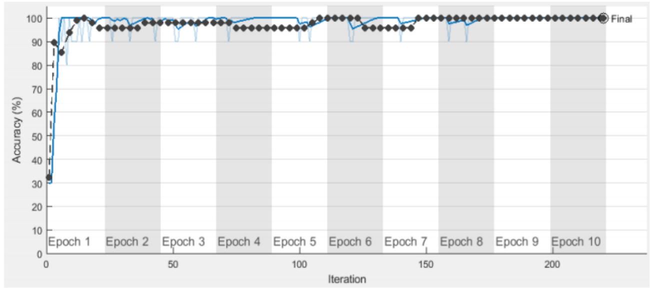

Figure 4 Accuracy values of the training Resnet-50 for the classification of insect images.

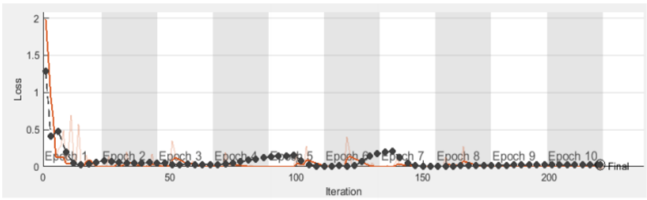

Figure 5 Loss values Accuracy values of the training Resnet-50 for the classification of insect images.

IV. EXPERIMENTAL RESULTS

A.Datasets and evaluation metrics

Datasets: The proposed method has been evaluated on the insect dataset [17]. Detail information of the dataset is described in table II. There are four species of insects that are: Achaea janata, Henosepilachna, Tetranychus urticae and Xylotrechus quadripes. After applying the data augmentation, number of classes of insects increase compared to original dataset. The number of insect images that is used for training and testing is described in table III.

Evaluation metric: To obtain the clear performance evaluation, the precision (P), recall (R) and F1 score metrics are commonly used for the image classification task. The F1 score is the harmonic mean of precision and recall. The F1 score is defined as follows [3]:

$$F1 \text{ score } = \frac{2\times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

B.Performance evaluation of various flower classification methods

Table V shows the performance comparison of the proposed method. Table IV shows the confusion matrix of the classification of insects. Examples of the classification of insects can be seen in Figure 7 and 8. The classification of Resnet-50 gains the highest scores.

The performance of the DNNs is improved compared to the application of SIFT and SVM classifier [6]. The better in the feature extraction and classification of DNNs allows to obtain higher accuracy. However, the execution time of the SIFT and SVM is better than that of DNNs. Table 6 shows the execution time of the methods on CPU and GPU.

To visualize the classes of insects based on the feature extraction, the t-SNE dimensional reduction technique is applied [18]. The data reduction technique aims to reduce the number of extracted features to display them efficiently. Figure 6 illustrates the distribution of classes of insects. The Resnet-50 network extracted 512 visual features of insect images. Then, the number of extracted features is reduced to three by using the dimensional reduction technique. The Figure shows that we can separate the classes of the insects using the extracted visual features.

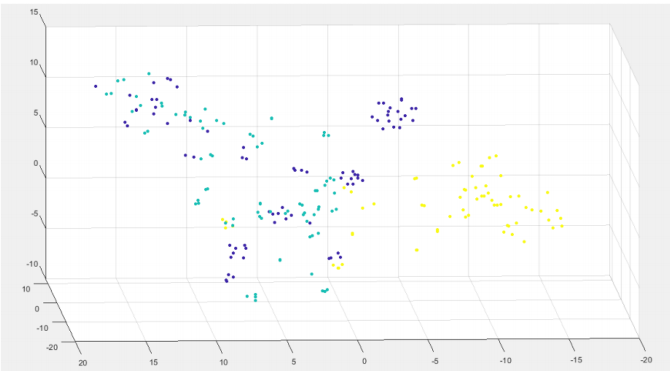

Figure 6 The distribution of extracted features of insect images. The classes of insects are marked by cyan, blue and yellow dots.

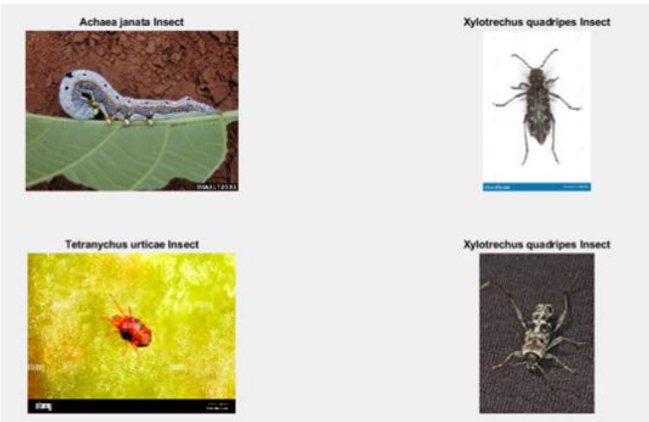

Figure 7 Examples of the classification of insects

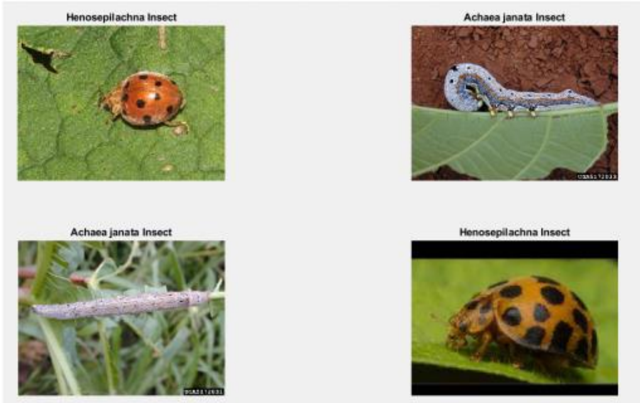

Figure 8 Examples of the classification of insects

Table I. STRUCTURAL INFORMATION OF DNNs FOR THE CLASSIFICATION OF INSECTS

Models	Number of layers	Size of input images	Number of extracted features
Alexnet [14]	7	227x227x3	1024
VGG-16 [16]	16	224x224x3	1000
Resnet-50 [15]	50	224x224x3	512

Table II. NUMBER OF IMAGES OF INSECTS BEFORE AND AFTER APPLYING THE DATA AUGMENTATION

Insects	Before	After
Achaea janata	650	650
Henosepilachna	600	650
Xylotrechus quadripes	400	650
Tetranychus urticae	250	650

Table III. NUMBER OF IMAGES FOR TRAINING AND TESTING DNNS FOR THE CLASSIFICATION OF INSECTS

Insects	Training images	Testing images
Achaea janata	450	200
Henosepilachna	450	200
Xylotrechus quadripes	450	200
Tetranychus urticae	450	200

Table IV. PERFORMANCE COMPARISON OF CLASSIFICATION METHODS OF INSECTS

Model	P	R	F1 score
SIFT and SVM	72%	71%	71.50%
Alexnet	92%	90%	90.99%
VGG-16	93%	91.5%	92.24.%
Resnet-50	98%	96%	96.99%

Table V CONFUSION MATRIX OF THE CLASSIFICATION METHODS OF INSECTS

Real insects	**Prediction results**			
Achaea janata	**200**	0	0	0
Henosepilachna	0	**195**	0	0
Xylotrechus quadripes	0	5	**200**	5
Tetranychus urticae	0	0	0	**195**

Table VI COMPARISON OF EXECUTION TIME OF VARIOUS METHODS OF THE CLASSIFICATION METHODS OF INSECTS (IN MINUTES)

Methods	CPU	GPU
SIFT and SVM	15	2
Alexnet	22	10
VGG-16	28	12
Resnet-50	30	15

V. CONCLUSION AND FUTURE WORKS

The paper investigated and applied the data augmentation and deep neural networks to classify insect images. Resnet-50 allows to obtain the highest results of 98%. The augmentation data plays an important role to improve the performance of the deep neural networks. Compared to the application of SIFT and SVM, the DNNs obtain higher accuracy. The implementation on GPU reduces the computational time compared to that of CPU. In the future, the results can be extended to classify a large number of insect in nature. The results can be developed to support users to classify insects in an efficient way.

VI. REFERENCES

[1] R. Fox *et al.*, "Comparison of deep learning techniques for classification of the insects in order level with mobile software application," in IEEE Access, 2022.

[2] D. OZDEMIR et al., "Insect population trends and the IUCN red list process," in J. Insect Conservation, 2019.

[3] W.Li et al., "Classification and detection of insects from field images using deep learning for smart pest management: A systematic review," in Sensors, 2021.

[4] C.Martineau et al., "A survey on image-based insect classification," in Pattern Recognition, 2017.

[5] J.Lim et al., "Development of a classification algorithm for butterflies and ladybugs," in Pattern Recognition, 2006.

[6] C.Wen et al., "Local feature-based identification and classification for orchard insects," in Biosystems Engineering, 2009.

[7] R. Akter et al., "Insect recognition and classification using optimized densely connected convolutional neural network," in 12th International Conference on Information Systems and Advanced Technologies, 2019.

[8] Y. S. Chang and E. S. Jun, "An intelligent insect search system based on observation of the insect's structure," in Expert Systems with Applications, 2015.

[9] D.Xia et al., "Insect detection and classification based on an improved convolutional neural network," in Sensors, 2018.

[10] S.Lim et al., "Performance effect analysis for insect classification using convolutional neural network," in 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2017.

[11] S. B. others, "Deep learning and statistical models for detection of white stem borer disease in arabica coffee," in Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 2019.

[12] R. P. P et al., "A novel method for the classification of butterfly species using pre-trained cnn models," in Electronics, 2022.

[13] H.Dou et al., "Patchmask: A data augmentation strategy with gaussian noise in hyperspectral images," in Remote Sens., 2022.

[14] A. Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012.

[15] K. He et al., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[16] K. S. and A. Zisserman, "Very deep convolutional networks for largescale image recognition," in https://arxiv.org/, 2014.

[17] "Insect dataset, url = https://plantwiseplusknowledgebank.org, doi=10.1079/pwkb.species.57204, note = Accessed: 2023-09-30."

[18] van der Maaten et al., "Visualizing data using t-sne," in J. Machine Learning Research, 2008.