A Transfer Learning Approach to Pollinator Recognition

Mid-Project Report

Unit Coordinator: Dr Greg Falzon

Unit Supervisor: Dr Romina Rader

Co-Supervisor: Dr Tobias Smith

Unit: COSC592 Master's Information Technology Project

Student Name: Sharifah Aldossary

Abstract

Using a transfer learning approach, we used a pre-trained convolutional neural network, AlexNet, to recognise and classify hundreds of images of three insect pollinator groups: bees, beetles, and flies. To achieve this, we fine-tuned AlexNet's layers to learn the specific features of our dataset. Though the recognition model is still in development, this mid-project report describes how we built and fine-tuned the algorithm over the past four weeks and presents our results thus far. The primary project focus up until this stage was to achieve a coarse-grained classification. After training and testing the developed model, we have generated promising results, especially considering the short working period and the nature of our image dataset. At the writing of this report, the model has achieved approximately 70% accuracy in classifying pollinator images.

Introduction

Many applications use automatic identification technologies. To understand how these technologies are used, Kärkkäinen and Ala-Risku (n.d.) proposed four categories of automated identification applications: authentication, tracking, process effectiveness, and information management.

The first application, object authentication, is commonly used for access control and relies on methods such as biometric identification, including finger and handprint recognition, and voice identification. The second application, item tracking, includes all implementations that collect knowledge about items' route and location; the most well-known tracking technology is the Global Position System (GPS) satellite system. In process effectiveness applications, automatic identification is used to automate data inputs and processes in order to reduce errors and increase time efficiency, for example, in industrial applications such as warehousing and retail applications such as sales-point scanning. Another example of this is optical character recognition (OCR), which is commonly used to automate data inputs and collection in office environments – for example, when registering cheques at banks. Finally, in information management applications, attached identifiers are used to access information about an object, and this data can be stored in either the identifier or a database. In cases where the stored information is located in a database, the identifier can be used to access it. Bar coding is the most commonly used technology for item information management. Another example is recognition systems, whereby images of objects are automatically captured and analysed, and then identified by their physical dimensions (Kärkkäinen & Ala-Risku, n.d.).

More recently, automated techniques have been applied in recognising and identifying insect pollinators. This data is of great importance for a range of fields, including environmental and rural sciences, agricultural engineering, human food production, and public health. Whilst insect pollinators and other insects have traditionally been identified through visual identification by trained experts, more recently, a number of novel approaches to identification using automated techniques have emerged. Most of these approaches have considered automating insect identification using a machine learning or deep learning approach (Glick & Miller, 2016; Ding & Taylor, 2016; Mortensen et al., 2007; Wen & Guyer, 2012; Wang, Lin, Ji, & Liang, 2012; Weeks, O'Neill, Gaston, & Gauld, 1999; Larios et al., 2007). The present project aims to apply convolutional neural network (CNN) transfer learning to develop a pollinator recognition (PR) model using an image dataset.

Ding and Taylorn (2016) proposed using a sliding window-based detection pipeline to classify images. In this system, which adopts the CNN presented by Lecun, Bottou, Bengio, and Haffner (1998), the classifier scans the input image in patches (Figure 1). Model evaluation metrics and charts – for example, a confusion matrix, which shows the number of positive and negative predictions made by the model compared to the target data output, or the receiver operating characteristic (ROC) curve, which is a graphical plot that illustrates the performance of a binary classifier system – are then used to assess the model's quality and fitness. Ding and Taylor (2016) achieved a precision-recall area under the curve (AUC) and log-average miss rate of 0.934 and 0.0916, respectively.

Figure 1. As Deshpande (2016) explained in his blog, to explain a convolutional layer is to picture a flashlight that is shining over the top left of the image. This light covers n by n area. This flashlight, called a filter, is also an array of numbers, which are called weights or parameters. The filter starts sliding across all the areas of the input image, and the region that it is shining over at any given time is called the receptive field.

Mortensen et al. (2007) took another approach: they segmented the image dataset by applying a Bayesian matting process.[1] Part of their approach followed the general region-based methodology and used a scale-invariant feature transform (SIFT) descriptor to detect regions.[2] By applying hierarchical deep CNNs Glick & Miller (2016), achieved a misclassification rate of 14.01%. In another study, Wen and Guyer (2012) developed three different models: 1) an invariant local features model using a Hessian affine detector and SIFT descriptors; 2) a global features model using numbers of features, such as counters and geometric colour features;[3] and 3) a combined model for global and local features. The final model achieved classification rate of 86.6%.

Weeks, O'Neill, Gaston, and Gauld (1999) proposed a novel automated identification system prototype called the digital automated identification system (DAISY), which they used to develop an algorithm for identifying five wasp species. They presented different calculations for the system's accuracy based on the number of classifiers. More than 98% of test images were correctly identified using just two classifiers. The proportion decreased to 86% when using all classifiers presented in the study. Two other systems – one proposed by Wang, Lin, Ji, and Liang (2012) and another proposed by Larios et al. (2007) – also demonstrated high performance, achieving accuracies of over 90%.

However, unlike these various approaches, which use deep learning or computer image analysis technologies, our project develops an automated PR model using a transfer learning approach, or in other words, the knowledge of a pre-trained deep neural network. Furthermore, we employ the CNN approach to extract, represent and classify, as well as fine-tune the last three layers in the pre-trained CNN AlexNet, which was designed by Krizhevsky, Sutskever, and Hinton (2012). AlexNet trained for 1,000 classes and 'consists of five convolutional layers, some followed by max-pooling layers, and three fully connected layers with a final 1,000-way softmax'. To the best of our knowledge, this is the first project to apply transfer learning to the PR problem.

[1] This process was presented by Chuang, Curless, Salesin and Szeliski (2001).
[2] We refer readers to their approach for a more detailed explanation (Mortensen et al., 2007).
[3] For more details, refer to the study (Wen & Guyer, 2012).

Materials and Methods

Dataset

Two sourcing methods were used to collect the colour image dataset. The first sourcing method involved using images provided by experts,[4] while for the second method, we used images available under a creative commons licence.[5] It is important to mention that our preference was to standardise the resolution, the distance between the source and the subject, the camera orientation, and the background as much as possible for later development.

As most of the data were collected before the project started and from different sources, there was no way to standardise or control the surrounding environment. However, as the project aimed to achieve a coarse-grained classification (e.g. course groupings of insects, such as bees, flies, and beetles) was desired, this was not a major concern. Many of the collected images had a high-quality resolution. The dataset items were then labelled and refined by an expert,[6] and the image sizes were adjusted to match the size of AlexNet's input layer, which is 277 by 277 pixels.

To implement transfer learning, a source and target problem must be defined. To serve this purpose, the number of images for each class was balanced (the number of images was around 700s in each class), and then the entire dataset was split into two datasets, wherein 80% of the images were used for training and the rest were used for testing.

Framework

As explained earlier, the algorithm for this model was developed based on a pre-trained CNN called AlexNet. The layers were transferred to the new task, except for the last three layers, which consisted of a fully connected layer, a softmax layer, and a classification output layer. It is important to mention that this project used the 23-layer version of AlexNet, rather than the 25-layer version, and the training option parameters were modified several times and adjusted to smaller numbers; this was appropriate for the transferred learning approach. In general, developing and training a neural network algorithm that uses a neural network and a computer vision toolbox in MATLAB[7] requires a high-performance graphics processing unit (GPU) and CUDA-enabled NVIDIA with a computing capability of 3.0 or

---

[4] Dr Romina Rader and Dr Tobias Smith from the School of Environmental and Rural Science at the University of New England, as well as other researchers.

[5] Online sources included Flickr, Twitter, Google, and Facebook.

[6] Dr Tobias Smith.

[7] This is the integrated development environment (IDE) used to deploy the algorithm.

higher, the number 3.0 indicate to the version of the GPU in use. This project used a GPU with a computing capability of 3.7.

Results

The following confusion matrices and ROC charts show the results of a trained multiclass support vector machine (SVM) classifier that uses a fast linear solver for three classes. Different solvers were tested; however, thus far, the linear solver has returned the best results.

In Figure 1, the numbers on the axes show the classes used: ‘1’ refers to bees, ‘2’ refers to beetles, and ‘3’ refers to flies. The accuracy was calculated as the sum of predicted labels compared to the tested labels divided by the number of elements of all predicted labels:

- Accuracy = sum (predicted labels = test labels) / number (predicted labels);
- Accuracy = TP + TN / TP + TN + FP + FN.

After repeating the algorithm multiple times, the accuracy fluctuated between 60% and 80%, and the highest percentage reached was 74%. The correct classification percentages for all classes ranged between 19% and 27%. Moreover, misclassifications accounted for approximately 7% of the results.

The ROC charts in Figure 2 compare the classification models. The false positive rate (1 – specificity) is illustrated on the X-axis against the true positive rate (sensitivity) on the Y-axis. The curve climbs towards the top left, indicating that the classes were correctly predicted by the model, and that any sensitivity increases are accompanied by a decrease in specificity. The closer the curve is to the Y-axis and the top of the border, the more accurate the test.

The AUC was based on the traditional academic point system.[8] The model fell between being a poor system and a fair system.

[8] The categories are as follows: 0.90–1, excellent (A); 0.80–0.89, good (B); 0.70–0.79, fair (C); 0.60–0.69, poor (D); and 0.50–0.59, fail (F).

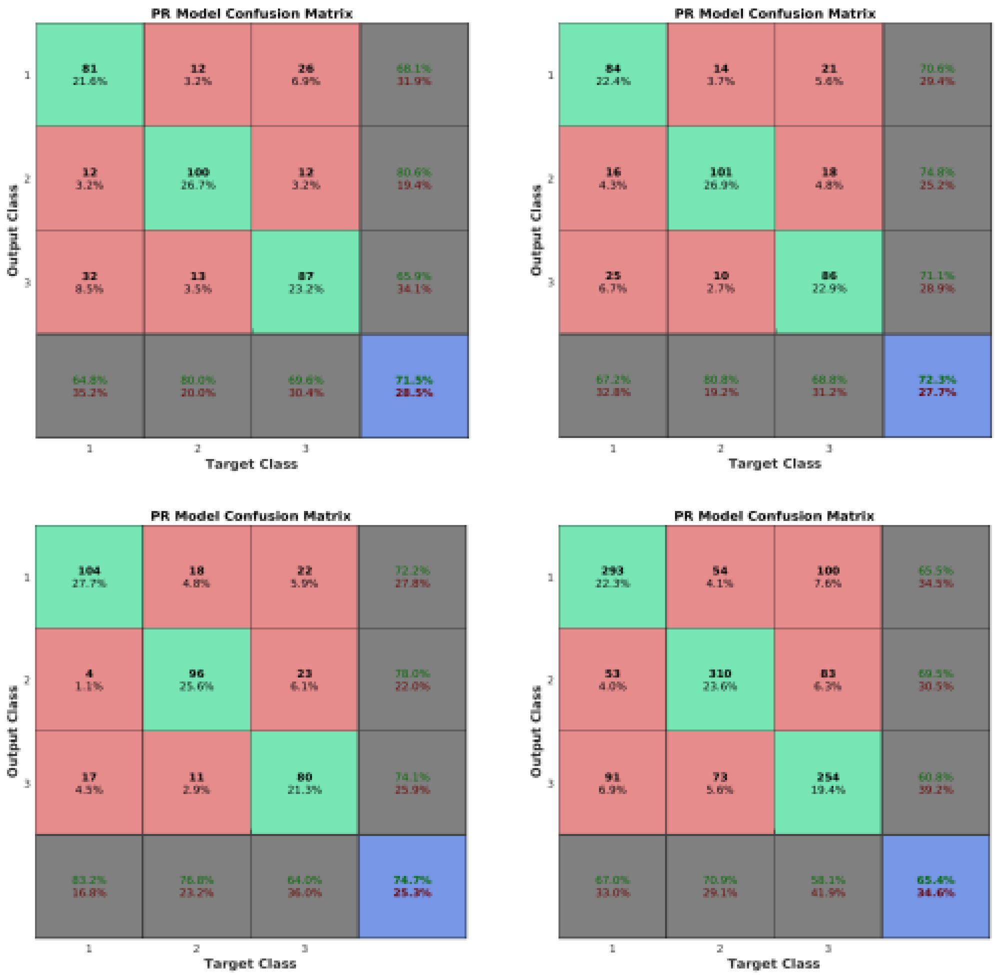

Figure 1. Confusion matrixes for some of the runs performed using the PR model.

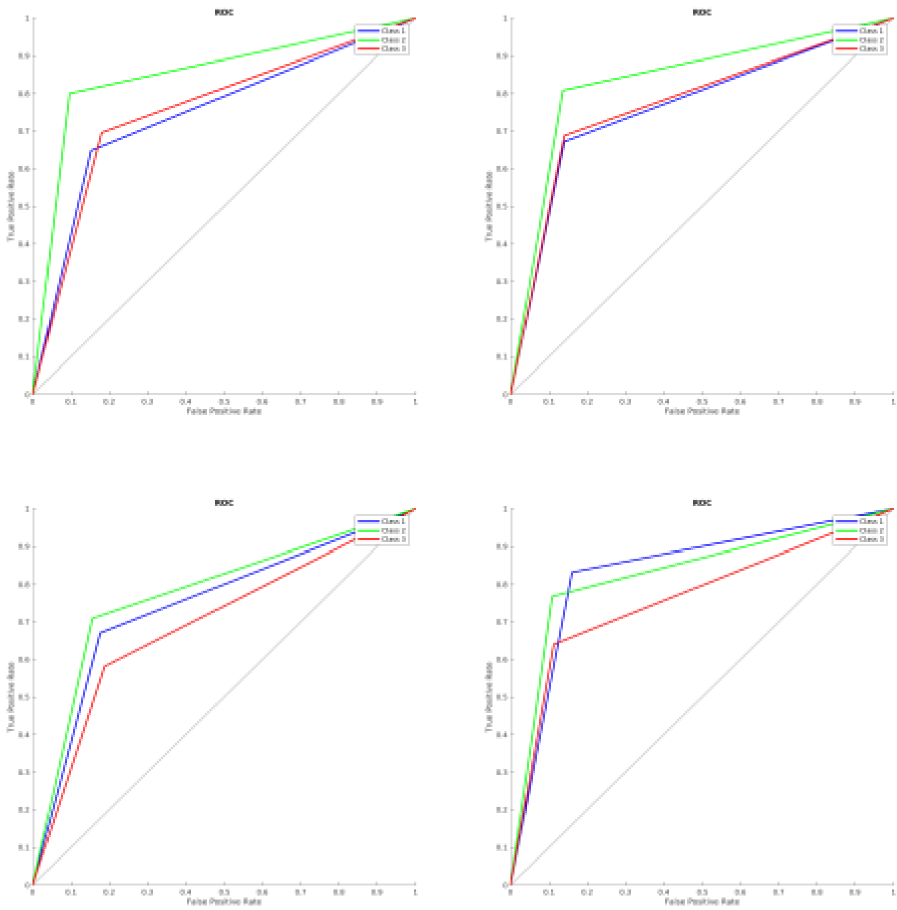

Figure 2. ROC charts for the PR model.

Discussion

We developed a deep transfer learning classification framework for recognising pollinators, and in doing so, we demonstrated that by fine-tuning existing CNNs, we can reuse and build on the knowledge acquired in solving one classification problem to another one.

In general, we observed that the algorithm more accurately classified beetles than bees and flies. To monitor this problem, we generated one folder for the correctly predicted pollinators and another for the falsely predicted ones so that we could identify and analyse common characteristics, and thereby improve the algorithm. In the next step of this research, we intend to modify some of the actual layers of the pre-trained CNN, such as the input layer or add more layers and improve the algorithm based on six classes rather than three.

References

Chuang, Y.-Y., Curless, B., Salesin, D., & Szeliski, R. (2001). A Bayesian approach to digital matting. *Proceedings of IEEE CVPR 2001*, *2*, pp. 264–271. Kauai, Hawaii: IEEE Computer Society.

Deshpande, A. (2016). *A beginner's guide to understanding convolutional neural networks*. Retrieved from adeshpande3.github.io: https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/

Ding, W., & Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Computers and Electronics in Agriculture*, *123*(C), 17–28.

Glick, J., & Miller, K. (2016). *Insect classification with heirarchical deep convolutional neural networks*. Location: Stanford University.

Kärkkäinen, M., & Ala-Risku, T. (n.d). *Automatic identification*: *Applications and technologies*. Location: Helsinki University of Technology, Department of Industrial Engineering and Management.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). *ImageNet classification with deep convolutional neural networks*. Paper presented at the Neural Information Processing Systems Conference.

Larios, N., Deng, H., Zhang, W., Sarpola, M., Yuen, J., Paasch, R., . . . Dietterich, T. (2007). Automated insect identification through concatenated histograms of local appearance features. *Applications of Computer Vision*,.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Mortensen, E., Delgado, E., Deng, H., Lytle, D., Moldenke, A., Paasch, R., . . . Dietterich, T. (2007). *Pattern recognition for ecological science and environmental monitoring: An initial report.*

Moulder, S., Sheridan, T., Cavallo, P., & Rossini, G. (n.d.). *Deep learning with MATLAB and multiple GPUs*. Location: Mathworks.

Wang, J., Lin, C., Ji, L., & Liang, A. (2012). A new automatic identification system of insect images at the order level. *Knowledge-Based Systems*, *33*, 102–110.

Weeks, P., O'Neill, M., Gaston, K., & Gauld, I. (1999). Automating insect identification: Exploring the limitations of a prototype system. *Journal of Applied Entomology*, *123*(1).

Wen, C., & Guyer, D. (2012). Image-based orchard insect automated identification and classification method. *Computers and Electronics in Agriculture*, *89*, 110–115 .

Zeiler, M., & Rob, F. (2014). Visualizing and understanding convolutional networks. In D. A. Fleet (Ed.), *Computer vision – ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, part I* (pp. 818–833). Cham: Springer International Publishing.