Butterfly Species Identification Using Convolutional Neural Network

Reema Rodrigues<sup>1</sup>, R. Manjesh<sup>2</sup>, P. Sindhura<sup>3</sup>, Samrddhi N. Hegde<sup>4</sup>, A. Sheethal<sup>5</sup>  
<sup>1,3,4</sup>B.E. Student, Dept. of Computer Science & Engineering, Srinivas Institute of Technology, Mangalore India  
<sup>2</sup>Professor, Dept. of Computer Science & Engineering, Srinivas Institute of Technology, Mangalore India  
<sup>*</sup>Corresponding author: samruddhi.n.hegde@gmail.com

Abstract: Butterfly Species Identification project focuses on identifying different butterfly species based on different categories. The main idea of the project is to consider around ten butterfly species for the classification purpose. The gathered data set consists of nearly 830 images of different classes. The manual approaches for butterfly classification is a tedious task and costly as well as it might include gathering the data, recognition and physically archiving specimen images. Therefore, Convolutional Neural Network (CNN) has been implemented as it is an automated method which doesn’t require any human intervention. The testing carried witnessed 90% overall classification accuracy on consideration of ten classes of 832 images.

Keywords: Butterfly species, Convolutional Neural Network (CNN).

1. Introduction

There exist nearly 17,000 species of butterfly recordings throughout the globe. These variety of species respond to climate change and also interact with plants. Therefore, the identification of species is crucial inspite of complexity. The classification of species requires information on their morphology. The traditional methods of butterfly’s taxonomy differentiates species of butterflies by the analysis of color, texture and size of the wing spot and other anatomical features. The existing system can classify the butterflies by their outer morphological qualities, genetical character which should be carried out manually by preparing genetical slides manually using chemical substances and various processes. Common man doesn’t always have the ability to prepare these slides. Therefore, there is a need to computerize the human effort to recognize the butterfly easily and efficiently. So the main objective of this project is to aim at evaluating a computer vision and machine learning system that correctly identifies butterfly species easier, faster and cheaper than traditional methods.

2. Literature Survey

Butterfly Identification using Convolutional Neural Network (CNN): In this paper firstly classification of images, detection of the objects and segmentation is included. CNN is well known for the power focused low assortment in inputs, they require low pre-processor for execution. ResNet, AlexNet, ZFNet and Google Net, the most pre-trained Convolutional Neural Networks that are available. Uploading the butterfly image is the beginning of the application. Images from the data are resized into 224x224 pixels. Google Net which is pre trained is used to classify resized butterfly images. Due to good performance of implementation of CNN technique by using Google Net as CNN, Confusion Matrix is used to calculate accuracy after the completion of the process. Lastly butterfly image will be identified with display of the habitat information. This research involves the division of the data into training and testing for which 80% and 20% of the data is dedicated respectively. The Black, Veined Tiger, Chocolate Grain Yellow, Grey Pansy and Plain Lacewing are the 4 divided types of the testing results of the classified butterfly species. Around thirty images are considered for training and testing the count of TRUE and FALSE outcome of identification in confusion matrix form. Around 97.5% is the overall accuracy of the process conducted. Therefore, implementation of the Convolutional Neural Network is concluded as a successful method for the identification of the Butterfly Species as this involves highest accuracy record.

3. Proposed System

In our project we first considered ten categories of 832 butterflies for the study. The images of butterflies that sums up the characteristics of members of the particular category is defined. The butterfly images are stored in the form of dataframes, to differentiate the types of category. Once, optimum characteristics of each species determined the similarity measures that are simpler therefore faster than image texture methods. First we should resize the images to standard pixel i.e. 128x128. Transform the labels to numeric form by using Label Encoding function and One Hot Encoder function to split the columns which contains numerical categorical data. Out of 832 images we should split major amount of images for training and remaining will be directly for testing. Sequential Model containing about 5 layers and CNN network building is done. Batches of the Dataset is created for which epoch is made equivalent to fit the model, from which the outcome is raised which includes loss and accuracy percentage thus classifying images of various species.
4. Result and Discussion

By the analysis of the variety of Butterfly classes we have arrived at the conclusion saying that we have created a model which is trained on about 80% of the dataset, and tested on remaining 20% of the dataset. This Model Identifies the Butterfly class of that particular image which is fed as input. This Model also ensures that it does not identify the objects other than Butterfly.

5. Conclusion

From the above aspects we have observed that there are vast varieties of species of the butterflies found. So it is very difficult to classify all of them. We have taken into the consideration of the dataset which consists of 10 different butterfly classes, Danaus_plexippus, Heliconius_charitonius, Heliconius_erato, Junonia_coenia, Lycaena_phlaeas, Nymphalis_antopia, Papilio_cresphontes Pieris_rapae, Vanessa_atalanta, Vanessa_cardui.

We have used the domain Machine Learning and implemented the Convolutional Neural Network Method (CNN) which is used because it is one of the most efficient methodology under machine learning which has delivered about 97% accuracy. Future work should focus on the following aspects: First, the proposed CNN based method could be employed for real-time butterfly not by considering images of butterfly. Second, the development of a web app or mobile app where the user can use this application anytime anywhere for butterfly species identification. Third, Multi-classification, which focuses on rare species of butterflies and unknown species can also be explored.

Acknowledgment

The authors would like to thank Mr. Manjesh R, Assistant Professor, Department of Computer Science & Engineering, Srinivas Institute of Technology for his valuable suggestions during the planning and development of this project work. And also we thank our parents for their support and encouragement throughout our work.

References