UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Deep Learning Based Fine-Grained Species Identification

by Qiyu Liao

Thesis submitted in fulfilment of the requirements for the degree of

C02029 Doctor of Philosophy

under the supervision of Min Xu, Dadong Wang

University of Technology Sydney
Faculty of Engineering and Information Technology

March, 2021

Certificate of Original Authorship Template

Graduate research students are required to make a declaration of original authorship when they submit the thesis for examination and in the final bound copies. Please note, the Research Training Program (RTP) statement is for all students. The Certificate of Original Authorship must be placed within the thesis, immediately after the thesis title page.

Required wording for the certificate of original authorship

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Qiyu Liao declare that this thesis, is submitted in fulfilment of the requirements for the award C02029 Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.
**If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).*

**If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).*

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 18/11/2021

Collaborative doctoral research degree statement

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Indigenous Cultural and Intellectual Property (ICIP) statement

This thesis includes Indigenous Cultural and Intellectual Property (ICIP) belonging to UTS and CSIRO, custodians or traditional owners. Where I have used ICIP, I have followed the relevant protocols and consulted with appropriate Indigenous people/communities about its inclusion in my thesis. ICIP rights are Indigenous heritage and will always remain with these groups. To use, adapt or reference the ICIP contained in this work, you will need to consult with the relevant Indigenous groups and follow cultural protocols.

Abstract

Fine-Grained Visual Categorization (FGVC) is a challenging research topic in computer vision. It deals with the classification of visual data at a subordinate level. This thesis investigates four categories of FGVC methods based on deep learning, including general convolutional neural networks, object part localization methods, approaches using CNN ensemble or higher-order feature encoding, and methods utilizing recurrent visual attention. Overall performance comparison has been conducted to analyse their advantages and disadvantages.

We proposed a new regression-based part detection structure and a novel part-based model, which increased the classification accuracy of PS-CNN from 76.4% to 82.4% on the CUB-200-2011 benchmark dataset.

Inspired by the second-order pooling, we proposed a highly interpretable method with a compressed structure to significantly reduce the computation complexity while improving the fine-grained categorization accuracy. The proposed model provides a supervised selection of the most discriminative second-order channels. With the proposed method, the computation and the feature dimension are linearly reduced to 4% of the original bilinear pooling. By applying matrix normalization and a Fisher-Recurrent-Attention structure, we achieved the best result among the VGG-16 based FGVC models.

Following the conception of attention crop and attention drop in the Fisher-Recurrent-Attention model, we proposed a forcing module to constrain the network to extract more diverse features for FGVC. The forcing module focuses more on confusion regions which are essential for the fine-grained classification. Experimental results show that the proposed forcing module can improve the attention and prediction of the network when an input image is panned or zoomed, and the double prediction performs better than the single prediction.

The existing FGVC methods often come with enormous amounts of computation and require large memory space. This makes these models inadequate for mobile applications. We proposed a Category Attention Transferring Convolutional Neural Network (CAT-CNN) to transfer the attention knowledge from a large-scale FGVC network to a small but efficient network to improve its presentation capability. Using the proposed model, we improved the classification accuracy of the efficient networks by up to 5.7% on the CUB-2011-200 dataset without increasing computation time or memory cost, which makes FGVC feasible on mobile devices.

We also conducted abundant studies to investigate the relationship between attention and classification accuracy of our proposed deep learning models, visualized and analysed the attentional activations of these models. We hope that our findings may inspire further research efforts to advance the FGVC for a wide range of real-world applications.

keywords: Image classification, Fine-grained classification, Deep learning, Species identification.

Contents

List of Figures