Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2016

Automatic segmentation and classification of white blood cells from peripheral blood smear images taken from chickens

Erik H Kjellman

LINKÖPING UNIVERSITY

Master of Science Thesis in Electrical Engineering

Automatic segmentation and classification of white blood cells from peripheral blood smear images taken from chickens

Erik H Kjellman

LiTH-ISY-EX--YY/NNNN--SE

Supervisor: **Gustav Häger**
ISY, Linköpings universitet

Examiner: **Michael Felsberg**
ISY, Linköpings universitet

CVL
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Sammanfattning

Förhållandet mellan två olika sorters vita blodceller, nämligen heterofiler och lymfocyter, är ett användbart mått på stressnivån hos kycklingar. Detta förhållande räknas ut för hand idag, genom att manuellt räkna ett antal blodceller i blodutstryk. Det är ett väldigt tidskrävande arbete, och även ...

Detta examensarbete undersöker möjligheten att automatiskt segmentera och klassificera de vita blodcellerna i blodutstryk från kycklingar för att räkna ut detta förhållande. Detta görs genom maskininlärning, främst genom att använda så kallade Convolutional Neural Networks, faltande neuronnät.

Detta arbete genomförs i samarbete med (((biologiavdelningen))) på Linköpings Universitet, som tillhandahåller blodutstryk från sina kycklingar, och expertis från sina människor.

Resultatet visar att processen att räkna ut förhållandet mellan heterofiler och lymfocyter kan göras semi-automatiskt eller helautomatiskt, beroende på kvaliteten på bilderna och åldern på individerna.

Abstract

The ratio between two different types of blood cells, i.e. heterophils and lymphocytes, is a useful measure to gauge the stress level of domestic chickens. This ratio is calculated by hand today, by manually counting blood cells in peripheral blood smear images. This is a very laborious and time consuming task, and prone to human error.

This master thesis aims to automatically segment and classify the white blood cells in blood smear images taken from chickens in order to calculate the previously mentioned ratio. This is done through machine learning, mainly by using Convolutional Neural Networks.

This thesis was produced at the (((biologiavdelning))) at Linköping University, which provided blood smear image data from their chickens as well as expertise from their humans.

The results show that the process of calculating the ratio can be made semi-automatic or fully automatic, depending on the quality of the images and age of the individual chickens.

Acknowledgments

I want to thank the BIOLOGIAVDELNING for giving the opportunity to make this thesis possible. Also, a big thanks to my supervisor and examiner at CVL, Gustav Häger and Michael Felsberg, respectively.

Finally, I would like to thank my friends and family for supporting me.

Linköping, November 2016
Erik H Kjellman

Contents

Notation

Glossary

Term	Meaning
CNN	Convolutional Neural Network – a ground-breaking type of neural network useful for image classification
CUDA	Nvidia's GPU toolkit, simplifying parallel computing
GPU	Graphical Processing Unit, the highly parallel processor on a graphics card
HL	Heterophil to lymphocyte ratio, i.e. the number of heterophils divided by the number of lymphocytes
PNG	Portable Network Graphics – a lossless image format

1 Introduction

1.1 Background

The ratio between Heterophils and Lymphocytes in chickens is a useful measure of their stress level. As of now, this ratio is calculated by hand, a very laborious and time consuming task. This thesis analyses the possibility of automating this process, in conjunction with an interactive graphical user interface for manual correction.

The aim is that it will be a usable application for researchers in biology and similar fields, without degrees in engineering or other strictly technical fields.

1.2 Problem description

Color images of blood cells from peripheral blood smears are taken with a digital microscope beforehand, and different types of blood cells are to be segmented and classified using a convolutional neural network (referred to as CNN from here on out). The most important blood cells are heterophils and lymphocytes, other white blood cells such as monocytes and OTHERS are not important for the task at hand, but it can be beneficial to detect these as well.

Warning
OTHERSSS

The images are given in the ndpi format, which is basically a proprietary extension of the TIFF file format, with different zoom levels of the blood smears. This application will only use the ones that are taken with the largest zoom available, in which a white blood cell occupies an area of approximately 50x50 pixels.

From the images a ground truth must be established. This is done by manually cropping out the individual cells and saving them as PNG images, with a number

in the name corresponding to its class. Since the cells are seldom isolated in the image, it is expected that almost all individual cell images will contain parts of other cells around it. Since the CNN algorithm is much like a black box in terms of insight into the network from the user's perspective, thorough testing must be done with the trained network so as to verify the accuracy and flexibility. The CNN network will be trained on a graphics card GPU to significantly increase the speed, since graphics cards are highly optimized to parallel tasks, as well as matrix multiplication.

1.3 Limitations

There are several limitations to this thesis, the main ones will be mentioned here.

1.3.1 Data set

There are many large scale images available, but only a few thousand white blood cells are cut out and labeled, since this is a laborious and time consuming activity. The data set also only contain blood smear images from chickens of the age of 9 and 12 weeks, which probably limits the robustness of the application, since the cells vary in size and color during its lifetime.

1.3.2 Loss of depth

When manually analysing a cell in microscopes, one can focus on different depths of it in case of any uncertainties in determining the type of the cell. Since the cells are essentially photographed at a fixed depth, any information about their thickness will be lost.

1.3.3 Image artefacts

Getting a perfect image from a blood smear image is virtually impossible. The concentration of blood cells in the image varies significantly, in some parts the cells may be clumped together in groups of several hundred cells, in other parts the cells are spread very far apart. There are also many damaged cells from the smearing process, and strands of hair and dust is quite common. When analysing the images manually, the most common method is to choose a few regions of interest (here on out referred to as ROIs) where the amount of artefacts is low, and the concentration of cells is at an acceptable level.

2

Theory

2.1 Machine learning

2.1.1 Convolutional Neural Networks

2.2 Problem outline

3

Method

3.1 Data gathering

3.2 Implementation

3.2.1 Neural network structure

3.2.2 Programming language

3.2.3 Preprocessing

3.3 Evaluation

4

Results

4.1 Overview

HEJ

5

Conclusion

5.1 Conclusion

Japp.

5.2 Future Work

5.2.1 Data gathering