

University of Adelaide

**Deep learning for multi-label scene  
classification**

by

Junjie Zhang

A thesis submitted in fulfillment for the  
degree of Master

Under Supervised by  
Chunhua Shen and Javen Shi  
School of Computer Science

August 2016

# Declaration of Authorship

I, Junjie Zhang, declare that this thesis titled and the work presented in it are my own.  
I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Adelaide.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the University of Adelaide or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---

# *Abstract*

Scene classification is an important topic in computer vision. For similar weather conditions, there are some obstacles for extracting features from outdoor images. In this thesis, I present a novel approach to classify cloudy and sunny weather images. Inspired by recent study of a deep convolutional neural network and the spatial pyramid matching, I generate a model based on the ImageNet dataset. Starting with parameters learned from other classification tasks, I fine-tune the model using outdoor images. Experiments demonstrate that our classifier can achieve state-of-the-art accuracy.

Multi-label learning is a variant of supervised learning where the task is to predict a set of examples, which can belong to multiple classes. This is a variant of popular multi-class classification problems in which each sample has one class label only. It can apply to a wide range of applications, which include text categorisation, semantic image labelling etc.. A lot of research work has been done on multi-label learning with different approaches. In this thesis, I train a neural network from scratch based on the generated artificial images. The model is learned by minimising an error function based on the Hamming distance, through the backpropagation optimisation. The model has high capability of generalisation.

# *Acknowledgements*

I am grateful to my main supervisor, Prof. Chunhua Shen, and co-supervisor, Dr. Qinfeng Shi, whose expertise, understanding, and support made it possible for me to work on the neural network that is of great interest to me. It is a pleasure working with them. Prof. Shen's dedication and keen interest to help his students had been solely and mainly responsible for completing my work. His timely advice, meticulous scrutiny, and scholarly advice have helped me to accomplish the tasks.

I would like to thank for my co-supervisor, Qinfeng Shi, for his time and effort on helping me understand research work and knowledge of machine learning.

I thank PhD candidate Teng Li who cooperated to work on the Weather Classification project. We set up experiment environment, analysed test results and discussed approaches. His prompt inspirations, timely suggestions with kindness, and enthusiasm work attitude have helped to achieve the perfect classification accuracy.

I am extremely thankful to research fellow, Guoshen Lin, for his kind help and discussion about the weather classification and the multi-label classification.

I would also like to thank staffs and visitors in the Australian Center for Visual Technologies (ACVT). Attending the reading group has enriched my knowledge in Computer Vision and learned a lot from them.

Finally, I would like to warmly thank the staffs in the School of Computer Science at the University of Adelaide. Julie Mayo, Sharyn Liersh and Jo Rogers have done excellent jobs on administration which helps students to be focus on their research work.

# Contents

<b>Declaration of Authorship</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Abbreviations</b>	<b>x</b>
<b>I Weather Classification</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Statistical Pattern Recognition . . . . .	2
1.3 Artificial Neural Networks (ANNs) . . . . .	2
1.4 Weather Classification . . . . .	3
<b>2 Background</b>	<b>4</b>
2.1 Related Work . . . . .	4
2.2 Single-Layer ANNs . . . . .	5
2.3 Multi-Layer Networks . . . . .	8
2.4 Stochastic Gradient Descent (SGD) . . . . .	10
2.5 Backpropagation . . . . .	11
2.5.1 Training protocols . . . . .	12
2.6 Overfitting and Regularization . . . . .	13
2.6.1 Weight Decay . . . . .	14
2.6.2 Dropout . . . . .	15
2.7 Softmax Classifier . . . . .	16
2.7.1 Practical issues . . . . .	17
2.7.2 Error function . . . . .	18
2.8 Convolutional Neural Networks (CNN) . . . . .	19
2.8.1 Layers in CNN . . . . .	19
2.9 Spatial Pyramid Matching (SPM) . . . . .	21
2.10 Transfer Learning . . . . .	23

<b>3</b>	<b>Methodology</b>	<b>25</b>
3.1	Datasets . . . . .	25
3.2	Data Augmentation . . . . .	26
3.3	Spatial Pyramid Pooling (SPP) . . . . .	26
3.4	Convolutional Neural Networks Architecture . . . . .	27
<b>4</b>	<b>Experiment</b>	<b>30</b>
4.1	Training Neural Networks . . . . .	30
4.2	Fine-tuning Model . . . . .	31
4.3	Companion . . . . .	32
4.4	Experimental Results . . . . .	32
4.5	Architecture Analysis . . . . .	34
4.6	Effects of SPP Layer . . . . .	36
4.7	Error Results . . . . .	37
4.8	Conclusion and Future Work . . . . .	37
<b>II</b>	<b>Multilabel Learning</b>	<b>39</b>
<b>5</b>	<b>Introduction</b>	<b>40</b>
5.1	Overview . . . . .	40
5.2	Multi-Label Learning . . . . .	41
<b>6</b>	<b>Background</b>	<b>44</b>
6.1	Evaluation Metrics . . . . .	44
6.1.1	Example-based Metrics . . . . .	45
6.1.2	Label-based Metrics . . . . .	46
6.2	Learning Algorithms . . . . .	47
6.2.1	Problem Transformation Methods . . . . .	47
6.2.1.1	Binary Relevance (BR) . . . . .	47
6.2.1.2	Classifier Chains (CC) . . . . .	48
6.2.2	Algorithm Adaptation Methods . . . . .	49
6.2.2.1	Multi-label k-Nearest Neighbour (ML-kNN) . . . . .	49
6.2.2.2	Collective Multi-label Classifier (CML) . . . . .	50
<b>7</b>	<b>Methodology</b>	<b>52</b>
7.1	Artificial Dataset . . . . .	52
7.1.1	Generating Images . . . . .	53
7.2	Artificial Neural Networks (ANNs) . . . . .	53
7.2.1	Network Architecture . . . . .	54
7.2.2	Error Function . . . . .	56
7.2.3	Cross Entropy . . . . .	58
7.2.4	Training and Testing . . . . .	60
<b>8</b>	<b>Experiment</b>	<b>63</b>
8.1	Dataset . . . . .	63
8.2	Details of Network . . . . .	63
8.3	Results . . . . .	66

---

8.4 Conclusion and Future Work . . . . .	68
<b>Bibliography</b>	<b>69</b>

# List of Figures

2.1	Diagram of a perceptron [1]. . . . .	5
2.2	Threshold function . . . . .	6
2.3	Linear function . . . . .	6
2.4	Sigmoid function . . . . .	6
2.5	Tanh function . . . . .	6
2.6	The error surface for a single layer neural network [2]. . . . .	8
2.7	Two types of dataset. The left one can be separated by a single layer neural network. The right one cannot be separated by a single neural network. Generated from [3]. . . . .	8
2.8	Diagram of a feedforward neural network [4]. . . . .	9
2.9	The error surface for a multi-layer neural network [2]. . . . .	10
2.10	A multi-layer neural network can separate a complicated dataset. Generated from [3]. . . . .	11
2.11	Overfitting example, the left one has a decent generalisation performance and the right one is overfitting [5]. . . . .	14
2.12	Illustration of dropout [6]. . . . .	15
2.13	The left is a fully connect regular neural network. The right is a CNN in 3 dimensions [7]. . . . .	19
2.14	Diagram for depth in a convolutional layer [7]. . . . .	20
2.15	Diagram of the Spatial Pyramid Matching [8]. . . . .	22
2.16	The left is traditional machine learning method. The right is transfer learning [9]. . . . .	23
3.1	2 Figures from the ImageNet [10]. . . . .	25
3.2	2 Figures from the Weather Dataset [11]. . . . .	26
3.3	A set of cropped patches from original image . . . . .	27
3.4	Diagram of the SPP layer [12] . . . . .	28
3.5	Architecture of the AlexNet [13] . . . . .	28
4.1	Training Process . . . . .	33
4.2	Training Loss . . . . .	33
4.3	ROC Curve . . . . .	34
4.4	A cloudy image and the feature maps from convolutional layers . . . . .	35
4.5	A sunny image and the feature maps from convolutional layers . . . . .	36
4.6	Visualisation of feature maps from the CNN model and the SPP model. The upper images are from the CNN model and the lower images are from the fine-tuned SPP model. . . . .	37
4.7	Histogram distribution of vectors from FC7. The left is from CNN model and the right is from the SPP model. . . . .	37



---

4.8	Misclassified images [11]. . . . .	38
5.1	Example Image . . . . .	40
7.1	Colour Wheel Diagram [14] . . . . .	52
7.2	Multilabel samples and the RGB colour histograms. Three labels mean red, green and blue sequentially. . . . .	54
7.3	Network Topology For Multi-label Classification . . . . .	55
8.1	Network Topology . . . . .	64
8.2	The test results for different number of hidden neurons. Blue circle for Sensitivity. Black plus for Specificity. Cyan star for Harmonic Mean. Red dot for Precision. Green x for F1 Score. . . . .	66
8.3	The learning speed for 200 neurons in the hidden layer. Blue circle for Sensitivity. Black plus for Specificity. Cyan star for Harmonic Mean. Red dot for Precision. Green x for F1 Score. . . . .	67
8.4	The ROC curve for 200 hidden neurons . . . . .	67
8.5	The ROC curve for 200 hidden neurons . . . . .	68

# List of Tables

3.1	Architecture of the model . . . . .	29
4.1	The first test is extracting features from the pre-trained CNN model and training a SVM classifier. The second test is similar with the first except for extracting features from the CNN model with a SPP layer. The third test is fine-tuning the model with AlexNet architecture. The fourth test is fine-tuning the CNN model with a SPP layer . . . . .	32
5.1	Multilabel $Y_1, \dots, Y_L \in 2^L$ . . . . .	41
8.1	Test results for different number of hidden neurons. . . . .	66

# Abbreviations

<b>SVM</b>	<b>S</b> upport <b>V</b> ector <b>M</b> achine
<b>ANNs</b>	<b>A</b> rtificial <b>N</b> eural <b>N</b> etworks
<b>ROI</b>	<b>R</b> egin of <b>I</b> nterest
<b>SIFT</b>	<b>S</b> cale <b>I</b> nvariant <b>F</b> eature <b>T</b> ransform
<b>SGD</b>	<b>S</b> tochastic <b>G</b> radient <b>D</b> escent
<b>BP</b>	<b>B</b> ack <b>P</b> ropagation
<b>MLE</b>	<b>M</b> aximum <b>L</b> ikelihood <b>E</b> stimation
<b>MAP</b>	<b>M</b> aximum <b>A</b> <b>P</b> osteriori
<b>CNN</b>	<b>C</b> onvolutional <b>N</b> eural <b>N</b> etwork
<b>ReLU</b>	<b>R</b> ectified <b>L</b> inear <b>U</b> nit
<b>SPM</b>	<b>S</b> patial <b>P</b> yramid <b>M</b> atching
<b>SPP</b>	<b>S</b> patial <b>P</b> yramid <b>P</b> ooling
<b>GPU</b>	<b>G</b> raphics <b>P</b> rocessing <b>U</b> nit
<b>BR</b>	<b>B</b> inary <b>R</b> elevance <b>C</b> lassifier
<b>CC</b>	<b>C</b> lassifier <b>C</b> hains <b>C</b> lassifier
<b>CML</b>	<b>C</b> ollective <b>M</b> ulti- <b>L</b> abel <b>C</b> lassifier