Contents

Declaration of Authorship i

Abstract ii

Acknowledgements iii

Contents iv

Abbreviations vii

1 Introduction 1
1.1 How do Artifical Neural Networks compare to the human brain? 1
1.2 Motivation 1

2 Background 2
2.1 Introduction 2
2.2 What are Artifical Neural Networks good at? 2
2.2.1 Classification 3
2.2.2 Generation 3
2.2.3 Reinforcement Learning 3
2.3 What are Artificial Neural Networks bad at? 3
2.3.1 data ineffciency 3
2.4 How to Get Off the Symbol Grounding Merry-Go-Round 3
2.4.1 What is symbol grounding? 3
2.4.2 How do humans do it? 3
2.4.3 How do machines do it? 4
2.5 Why brains are better 4
2.5.1 Embodiment 4
2.5.1.1 Sensory Redundancy 4
2.5.1.2 Biological Filters 4
2.5.2 Development 4
2.5.2.1 Biological Filters, again 4
2.5.3 Pulling yourself up by the bootstraps 4
2.5.4 Machine Equivelancy 4
2.5.4.1 How do we simulate Embodiment for ANNs? 4
2.5.4.2 How do we simulate Development for ANNs? 4

2.6 Where do we go from here? . . . 4
2.6.1 Robot bodies . . . 4
2.6.2 multimodality . . . 4
2.6.3 transfer learning . . . 4

3 A Primer on Artifical Neural Networks 5
3.1 Perceptrons . . . 6
3.1.1 Multi Layer Perceptron . . . 6
3.1.2 Activation Functions . . . 6
3.2 Convolutional Neural Networks . . . 6
3.2.1 What is Convolution? . . . 6
3.2.1.1 Kernels . . . 6
3.2.1.2 Strides . . . 6
3.2.1.3 Dilations . . . 6
3.2.2 Transposed Convolutions . . . 6
3.3 Recurrent Neural Networks . . . 6
3.3.1 Vanilla RNN . . . 6
3.3.2 Gated RNN . . . 6
3.4 Learning Algorithms . . . 6
3.4.1 Gradient Decent . . . 6
3.4.2 Backpropagation . . . 6
3.4.3 Extensions and Improvements . . . 6
3.4.3.1 Learning Rate Schedule . . . 6
3.4.3.2 Momentum . . . 6
3.5 Summary . . . 6

4 Sensory Redundancy: Are two heads better than one? 7
4.1 Why have one sensor when two are better? . . . 7
4.1.1 Aims . . . 7
4.2 Flikr25K . . . 7
4.2.1 dataset description . . . 8
4.2.2 problem description . . . 8
4.2.3 network description . . . 8
4.2.4 results . . . 8
4.2.5 discussion . . . 8
4.3 There and back again: Generating images from natural language . . . 8
4.4 UCU Arabic Spoken Digits and MNIST . . . 8
4.4.1 Dataset Description . . . 8
4.4.1.1 UCU Arabic Spoken Digits . . . 8
4.4.1.2 MNIST Handwritten Digits . . . 9
4.4.2 Problem Description . . . 10
4.4.2.1 Classification . . . 10
4.4.2.2 Bidirectional Symbol Grounding . . . 10
4.5 Experiment Details . . . 11
4.5.1 dataplumbing . . . 11
4.5.1.1 Combining Datasets . . . 11
4.5.2 Merging Modalities . . . 11

4.5.3 Network Description . . . 12
4.5.4 Results . . . 13
4.5.5 Discussion . . . 13
4.6 Conclusion . . . 13

5 Magical Vectors and Where to Find Them 14
5.1 ANN Latent space: the Final Frontier . . . 14
5.1.1 Vector Arithematic Explained . . . 16
5.2 Artifical Shapes Dataset . . . 16
5.2.1 dataset description . . . 16
5.2.2 problem description . . . 16
5.2.3 network description . . . 16
5.2.4 results . . . 16
5.2.4.1 Image Generation . . . 16
5.2.4.2 Multilabel classification . . . 16
5.2.4.3 Vector Arithematic . . . 16
5.2.5 discussion . . . 16
5.3 Real Shapes Dataset . . . 16
5.3.1 dataset description . . . 16
5.3.2 problem description . . . 16
5.3.3 network description . . . 16
5.3.4 results . . . 16
5.3.4.1 Image Generation . . . 16
5.3.4.2 Multilabel classification . . . 16
5.3.4.3 Vector Arithematic . . . 16
5.3.5 discussion . . . 16

6 Transfer Learning: kick-starting the learning process 17
6.1 What is transfer learning . . . 17
6.1.1 A biological analogue . . . 17
6.2 The importance of transfer learning . . . 17
6.3 MultiSense 1 dataset . . . 17
6.3.1 dataset description . . . 17
6.3.2 problem description: image generation from raw speech audio . . . 17
6.3.3 network description . . . 17
6.3.4 results . . . 17
6.3.5 discussion . . . 17

7 What have we learnt and more importantly, what have our neural nets learnt? 18
7.1 Summary of important points . . . 18
7.2 Conclusion . . . 18

8 Future Work 19
8.1 Where do we go now? . . . 19
8.1.1 More Sensors . . . 19
8.1.2 Better Sensors . . . 19
8.1.3 Swarm Robots and the Cloud . . . 19