

# **Evaluating data reduction techniques for supervised training**

*Yuwen Heng*

Master of Science  
Data Science, Technology, and Innovation  
School of Informatics  
University of Edinburgh  
2020

# Abstract

Training deep neural networks can be resources-consuming. The budget required is increasing with the size of the dataset. During the past few decades, many research is dedicated to developing training procedures to accelerate the convergence speed of deep learning. However, we still need the whole dataset to train the network and paying for a large dataset may not pay back well if we can use a smaller subset to achieve an acceptable performance. To solve this issue, we first adapted and evaluated three methods, Patterns by Ordered Projections (POP), Enhanced Global Density-based Instance Selection (EGDIS), and Curriculum Learning (CL), to reduce the size of two image datasets, CIFAR10 and CIFAR100, for the classification task. Based on the analysis, we present our two contributions: the Weighted Curriculum Learning (WCL) and a trade-off framework. The WCL outperforms POP and EGDIS in terms of both classification accuracy and time complexity. It achieves comparable performance compared with CL while keeping a portion of hard examples. The trade-off framework selects a subset of samples according to the acceptable relative accuracy and the dataset. In addition, the framework is also extended to predict the number of samples needed to achieve a particular accuracy with a given subset.

## Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr Yang Cao, for offering me the opportunity to work with him on such an attracting and challenging project. His encouragement and valuable guidance helped me tackle the obstacles in my research path.

Furthermore, special thanks go to Professor Bob Fisher, Dr Pavlos Andreadis at the University of Edinburgh and Dr Jiacheng Ni at IBM for sharing me their knowledge about computer vision and deep learning. The programming skills and coursework experience that I learnt from them helped me to organise the experiments well.

Finally, I would like to send my love to my fiancée Danni Li for her accompany during the past three years . I wouldn't have the chance to study full-time without her full support.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Research Goal . . . . .	1
1.3	Significance . . . . .	1
1.4	Beneficiaries . . . . .	1
<b>2</b>	<b>Background Research</b>	<b>2</b>
2.1	Instance Selection Algorithms . . . . .	2
2.2	Neural Network Training Procedures . . . . .	2
<b>3</b>	<b>Adapted Data Reduction Methods</b>	<b>3</b>
3.1	Patterns by Ordered Projections . . . . .	3
3.2	Enhanced Global Density-based Instance Selection . . . . .	3
3.3	Curriculum Learning . . . . .	3
3.4	Weighted Curriculum Learning . . . . .	3
<b>4</b>	<b>Data Reduction Evaluations</b>	<b>4</b>
4.1	Time Complexity . . . . .	4
4.2	Classification Accuracy . . . . .	4
<b>5</b>	<b>Trade-off Framework</b>	<b>5</b>
5.1	Subset Selection Framework . . . . .	5
<b>6</b>	<b>Trade-off Evaluation</b>	<b>6</b>
6.1	Relative Accuracy Precision . . . . .	6
<b>7</b>	<b>Conclusion and Future Work</b>	<b>7</b>
	<b>Bibliography</b>	<b>8</b>

# **Chapter 1**

## **Introduction**

**1.1 Motivation**

**1.2 Research Goal**

**1.3 Significance**

**1.4 Beneficiaries**

# **Chapter 2**

## **Background Research**

### **2.1 Instance Selection Algorithms**

### **2.2 Neural Network Training Procedures**

# **Chapter 3**

## **Adapted Data Reduction Methods**

**3.1 Patterns by Ordered Projections**

**3.2 Enhanced Global Density-based Instance Selection**

**3.3 Curriculum Learning**

**3.4 Weighted Curriculum Learning**

# **Chapter 4**

## **Data Reduction Evaluations**

### **4.1 Time Complexity**

### **4.2 Classification Accuracy**



# **Chapter 5**

## **Trade-off Framework**

### **5.1 Subset Selection Framework**

# **Chapter 6**

## **Trade-off Evaluation**

### **6.1 Relative Accuracy Precision**

## **Chapter 7**

### **Conclusion and Future Work**

# **Bibliography**