UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

Master in Deep Learning for
Audio and Video Signal Processing

MASTER THESIS

TITLE

Name Surname1 Surname2
Advisor: Name Surname1 Surname2
Lecturer: Name Surname1 Surname2

MONTH 202X

TITLE

Name Surname1 Surname2
Advisor: Name Surname1 Surname2
Lecturer: Name Surname1 Surname2

Dpto. Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
MONTH 202X

Resumen

Este Trabajo Fin de Master… (250-500 palabras).

The spanish summary is needed for the library.

Palabras clave

Palabra 1, Palabra 2,…

Abstract

This Master Thesis... (250-500 words).

Keywords

Word 1, Word 2,...

Acknoledgements

I would like to thank ...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The motivation of this work is...

1.2 Objetivos

The objectives of this Master Thesis are...

1.3 Report structure

This report has the following chapters

- chapter 1 Introducción.
- chapter 2 Related work.
- chapter 3 Design and development.
- chapter 4 Evaluation.
- chapter 5 Conclusions and future work.

Chapter 2

Related work

2.1 Section 1

... Citation example: [?]

2.2 ...

...

Table 2.1: Sample table (long title)

Col1	Col2	Col2	Col3
1	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

Figure 2.1: Sample figure (long title).

Chapter 3

Design and development

3.1 Introduction

Concise introduction outlining approach, and purpose of modifying the activation function in the feedforward network and the expected impact of this modification.

Brief description of each subsection.

3.2 Self-learnable activation functions

Reference fully detailed base networks' architectures on an appendix.

Explain why share the activation subnetwork among all the neurons (same as usual with ReLU, & lower complexity)(comment on potential of per-layer activations).

Explain activation subnetwork architecture. Including explanation of choice of base activation function (ReLU). Explain how in this setup the activation subnetwork is basically a piece-wise linear function. Comment on alternatives.

The ususal ReLU activation from the base models was replaced by a custom activation with learnable parameters. This new activation is shared among all the layers in the base model in the same way as a simple ReLU would. This means that there is only one set of parameters needed to define the activation function, and it is the same no matter where it is called from within the network. In this way, the complexity introduced by the new learnable parameters is kept to a minimum, and compared to the amount of trainable weights, it is usually negligible.

In order to potentially model any arbitrary activation function, a simple fully connected network was used, which we will refer to as (activation) subnetwork hereafter. The only requirement for a neural network to describe a function is that its input and output must be one-dimensional. Apart from that, there is complete freedom to define the architecture as one would please. For our purposes, we chose a network with a single hidden layer, and a ReLU activation. This architecture has the advantage of having a straightforward interpretation: it is a piece-wise linear function with as many pieces as the dimension of the hidden layer.

Mathematical description

3.3 Experimental setup

Actual experiments done. Description of datasets. Architectures tested. Multiple runs, to extract statistics.

Three widely recognized image classification datasets were utilized to study the performance of the custom learnable activation functions: MNIST, FashionMNIST, and CIFAR-10. Each dataset presents its unique challenges and characteristics, making them suitable for evaluating the performance of different neural network models.

- MNIST Dataset: The MNIST dataset is a collection of handwritten digits (0 through 9), comprising 60,000 training images and 10,000 testing images. Each image is a grayscale representation, sized at 28x28 pixels. The dataset is widely used for benchmarking image processing systems and is considered a fundamental dataset for evaluating machine learning algorithms, particularly in the field of image recognition.

- FashionMNIST Dataset: FashionMNIST serves as a more contemporary and challenging alternative to the traditional MNIST dataset. It consists of 60,000 training images and 10,000 test images, each of which is 28x28 pixels. The dataset features 10 classes of clothing items, making it more complex than MNIST but still accessible for benchmarking machine learning algorithms.

- CIFAR-10 Dataset: The CIFAR-10 dataset contains 60,000 32x32 color images in 10 different classes, with 6,000 images per class. There are 50,000 training images and 10,000 test images. The dataset is widely used for machine learning and computer vision research and poses a more challenging problem than MNIST or FashionMNIST due to the complexity of the images and the presence of color.

All three datasets are intended for image classification, with 10 distinct classes. Consequently, the loss function employed in the networks studied was a cross-entropy loss, applied to their output, which is a 10-dimensional logits vector corresponding to the 10 classes in each dataset.

In terms of the models trained:

- Fully Connected Model: This model was trained on both the MNIST and FashionMNIST datasets. Given the relatively simpler nature of these datasets, a fully connected neural network architecture was deemed appropriate.

- Simple CNN: A CNN was employed for the FashionMNIST and CIFAR-10 datasets. The choice of a convolutional network is suitable for these datasets due to the spatial nature of image data.

- More Complex CNN for CIFAR-10: Given the increased complexity of the CIFAR-10 dataset, a more sophisticated CNN architecture was specifically designed and trained for this dataset. This reflects the need for more advanced feature extraction capabilities to effectively handle the more challenging image classification tasks presented by CIFAR-10.

This experimental setup, with its varied models and datasets, allows for a comprehensive evaluation of the neural network architectures under different levels of problem complexity, providing valuable insights into the effectiveness and adaptability of the networks to different image classification tasks.

The fully detailed description of the architectures used can be found in appendix A.

Chapter 4

Evaluation

4.1 Section 1

...

4.2 ...

...

Chapter 5

Conclusions and future work

5.1 Conclusions

...

5.2 Future work

...

Appendix

Appendix A

Appendix-chapter 1