國　立　中　央　大　學

資訊工程學系

碩士論文

一種以卷積神經網路為基礎的具可解釋性的深度學習模型

A CNN-based Interpretable Deep Learning Model

研 究 生：涂建名

指導教授：蘇木春　博士

中 華 民 國 一百一十三年 六 月

一種以卷積神經網路為基礎的具可解釋性的深度學習模型

摘要

關鍵字：可解釋的人工智慧, 深度學習, 視覺皮質, 自我組織特徵映射, 影像分類

A CNN-based Interpretable Deep Learning Model

Abstract

Keywords: Explainable Artificial Intelligence, Deep Learning, Visual Cortex, Self-Organizing Maps, Image Classification

誌謝

目錄

圖目錄

頁次

表目錄

頁次

一、 緒論

1.1 研究動機

1.2 研究目的

1.3 論文架構

二、 背景知識以及文獻回顧

2.1 背景知識

2.1.1 卷積神經網路

2.1.2 可解釋性人工智慧

Decision Tree: [1] **grinsztajn2022treebased**

2.2 文獻回顧

2.2.1 基於多層自我映射圖之可視覺化深度學習模型

2.2.2 Local Interpretable Model-agnostic Explanations(LIME)

2.2.3 Shapley Additive Explanations(SHAP)

2.2.4 Tabnet: Attentive interpretable tabular learning

2.2.5 Building more explainable artificial intelligence with argumentation

2.2.6 以卷積神經網路為基礎的具可解釋性的深度學習模型

三、 研究方法

3.1 對以卷積神經網路為基礎的具可解釋性的深度學習模型之改進

3.1.1 優化模型流程與新增平行處理

3.1.2 優化空間位置保留機制之設計

3.1.3 優化放射狀基底函數

3.1.4 量化推論成果之方法

3.2 以卷積神經網路為基礎的 RGB 三通道可解釋性模型

3.2.1 模型架構

3.2.2 模型參數

3.2.3 RGB 三通道卷積模組設計與實現

3.2.4 模型流程

四、 實驗設計與結果

4.1 灰階優化模型與以卷積神經網路為基礎的具可解釋性的深度學習模型之比較

4.1.1 資料集介紹

4.1.2 實驗設計

4.1.3 實驗結果

4.2 模型保留空間位置特徵之臉部驗證實驗

4.2.1 實驗背景與目的

4.2.2 資料集介紹

4.2.3 模型架構與參數

4.2.4 實驗結果

4.3 以卷積神經網路為基礎的 RGB 三通道可解釋性模型之效果驗證

4.3.1 資料集介紹

4.3.2 實驗設計

4.3.3 實驗結果

五、 總結

5.1 結論

5.2 未來展望

參考文獻

[1] L. Rokach, “Decision forest: Twenty years of research,” *Information Fusion*, vol. 27, pp. 111–125, 2016.