國立中正大學 電機工程學系
Department of Electrical Engineering, National Chung Cheng University

手寫辨識與超解析度模型之CNN架構與量化優化分析

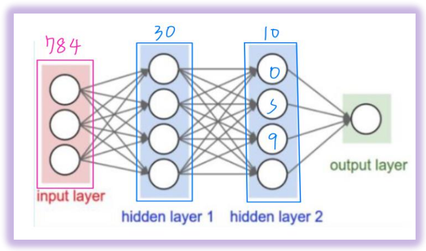
學生:唐妃儀　　指導教授:黃崇勛

摘要

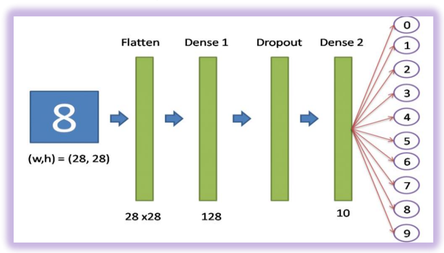
本研究探討卷積神經網路（CNN）於手寫數字辨識與影像超解析度（FSRCNN）的應用。手寫辨識部分，先以多層感知器（MLP）為基準，再與 CNN 比較，驗證其在特徵提取與辨識效能上的優勢。於 FSRCNN 模型中，導入 ReLU、Depthwise Separable Convolution 與 Batch Normalization，以降低參數並提升訓練穩定性。最後實作後訓練量化（PTQ），證明模型在大幅壓縮體積後，仍能維持近乎無損的輸出品質，展現 CNN 於影像任務的應用潛力。

研究方法

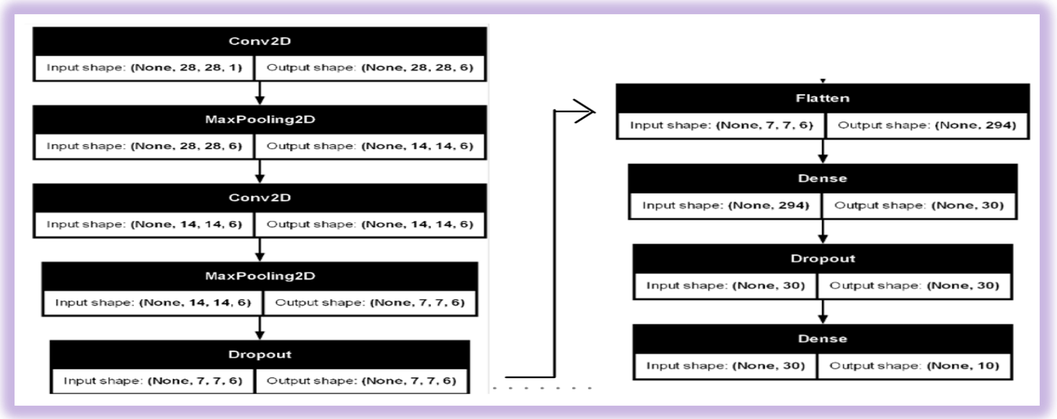
在手寫數字辨識實驗中，本研究先建立多層感知器（MLP）作為基準模型。輸入的 MNIST 影像需展平成 784 維向量，再經全連接層進行分類。此架構雖能達到基本辨識效果，但缺乏空間特徵利用。為提升效能，進一步引入卷積神經網路（CNN），透過卷積與池化層保留二維結構與局部特徵，最後接入全連接層輸出分類結果。整體流程可分為三步驟：MNIST 輸入 → MLP 基準 → CNN 架構比較。



圖一、MLP訓練流程



圖二、MLP概念性架構



圖三、本次專題實作之CNN架構圖

在影像超解析度任務中，本研究採用 FSRCNN 作為主要架構，流程分為三部分：① 特徵提取，以卷積層將低解析度影像轉換為特徵圖；② 中間映射，透過 1x1 與 3x3 卷積進行壓縮與特徵轉換，最後再以 1x1 還原維度；③ 上採樣重建，以反卷積生成高解析度影像。為提升效能，設計上引入 ReLU 取代 PReLU 以降低計算量，使用 Depthwise Separable Convolution 減少參數，並加入 Batch Normalization 提升收斂穩定性。此外，實作 後訓練量化（PTQ），將模型轉換為整數運算格式，以利壓縮與部署的形式。

模組名稱	包含層數
特徵提取（Feature Extraction）	Conv2d(1→56, 5x5) + PReLU(56)
中間映射層（Mapping/ Mid part）	Shrinking+4×Mapping + Expanding+對應的 PReLU
上採樣重建（Upsampling/Last part）	ConvTranspose2d(56→1, 9x9, stride=4)

圖四、FSRCNN架構模型

原始設計	量化調整後對應	說明
使用 PReLU 激活	替換為 ReLU	PReLU 不支援量化，會導致 PyTorch fuse 出錯
Conv2d + ReLU	融合為 QuantizedConvReLU2d	符合 PTQ fuse 流程，可減少運算與記憶體使用
ReLU 被 fuse，無 Identity	加入 Identity() 層佔位	保持原始層級索引不變，便於還原與 debug

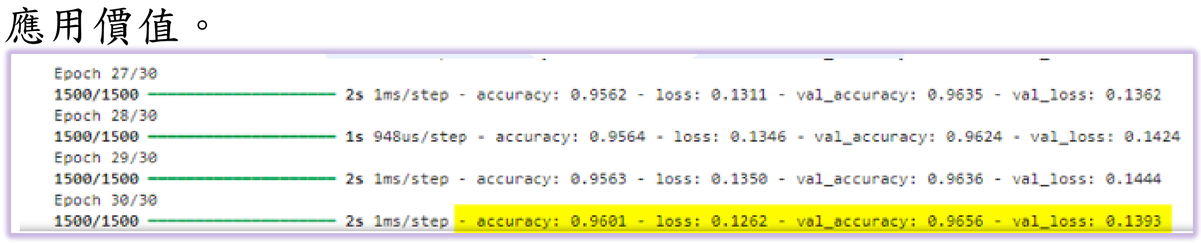
圖五、改良設計表

模組名稱	包含層數(量化前)	包含層數(PTQ 量化後)	功能說明
特徵提取（Feature Extraction）	Conv2d(1→56, 5x5) + PReLU(56)	QuantizedConvReLU2d(1→56)+ Identity()	把低解析度圖像轉成高維特徵圖
中間映射層（Mapping/Mid part）	Shrinking+4×Mapping + Expanding+對應的 PReLU	6×QuantizedConvReLU2d+ 6× Identity()	對特徵圖進行通道壓縮、非線性映射與展開，提取高層抽象特徵
上採樣重建（Upsampling/Last part）	ConvTranspose2d(56→1, 9x9, stride=4)	QuantizedConv Transpose2d(56→1, 9x9)	將中間特徵圖重建成高解析度輸出圖像（Super Resolution）

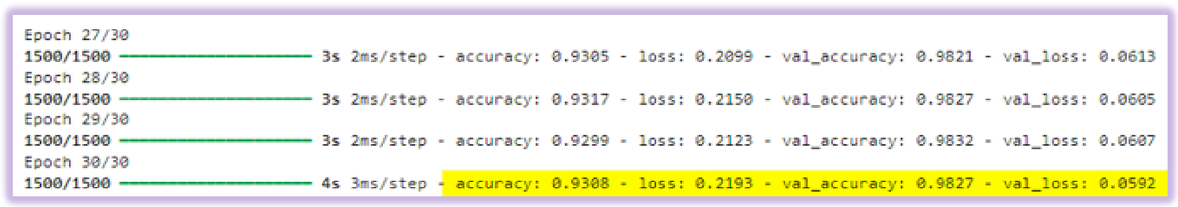
圖六、量化前vs量化後FSRCNN架構模型表

研究結果

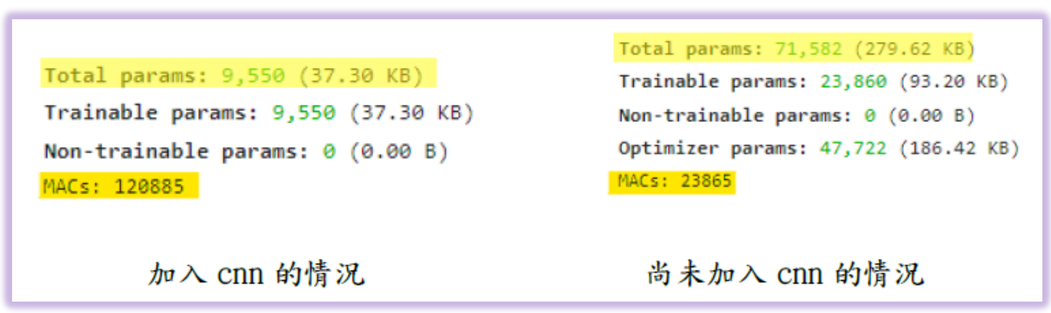
實驗結果顯示，MLP 在 30 個 Epoch 後的準確率約為 96%，能正確辨識大部分手寫數字，但受限於展平輸入的設計，無法有效利用影像的空間特徵，導致表現有瓶頸。相較之下，CNN 藉由卷積層擷取邊緣等局部特徵，並透過池化層降低維度與雜訊，在測試集上的準確率提升至 98%，優於 MLP。在模型複雜度方面，MLP 需要 71,582 個參數，而 CNN 僅需 9,550，參數量大幅減少，記憶體需求也更低，展現出模型壓縮與資源效率的優勢。然而，由於卷積涉及大量乘加運算，CNN 的 MACs 增加至 120,885，遠高於 MLP 的 23,865。這顯示 CNN 雖在效能與準確率上具優勢，但必須承擔更高的計算成本。整體來看，MLP 適合在計算資源有限、追求簡單架構的情境；而 CNN 更適合需要高精度的影像任務，兩者各自展現不同的應用價值。



圖七、MLP 訓練過程結果



圖八、CNN 訓練過程結果



圖九、MLP與CNN模型複雜度比較

實驗結果顯示，量化前後的 FSRCNN 在輸出品質上差異極小。輸出影像與原始模型幾乎一致。模型大小則由 49.36 KB 縮減至 12.95 KB，壓縮率達 73.8%，大幅降低儲存與傳輸需求。結構上，量化模型將卷積與激活融合為整數運算模組（如 QuantizedConvReLU2d），有效提升運算效率。整體而言，量化後 FSRCNN 兼顧效能與效率，證明其在嵌入式裝置與資源受限平台上的應用潛力。

模型版本	大小	壓縮率
原始模型	49.36 KB	
量化後模型	12.95 KB	縮減 73.8%

圖十、FSRCNN模型量化前後比較

結論與心得

本研究以手寫數字辨識與影像超解析度為例，驗證卷積神經網路（CNN）在影像任務中的優勢與潛力。與傳統多層感知器（MLP）相比，CNN 能保留二維結構並自動擷取局部特徵，使模型更能應對複雜的影像模式，展現出明顯的辨識與表現優勢。在超解析度任務中，本研究選用 FSRCNN 作為基礎架構，並嘗試多項優化設計，包括激活函數的調整、卷積結構的輕量化，以及正規化技術的導入。這些改良不僅提升了模型的收斂效率，也展現了在效能與運算需求間取得平衡的可行性。進一步實作的後訓練量化（PTQ）則證明，模型在壓縮後仍能維持近乎一致的輸出品質，展現出高度的實用價值。整體而言，本研究強調 CNN 的核心價值在於其靈活性，透過適當的設計與優化策略，可以針對不同任務調整架構，兼顧效能與效率。未來可將此研究方法延伸至自製資料集或嵌入式應用，進一步驗證其在實務環境中的可行性與應用潛力。