

1. 摘要

本報告探討一種結合 **PCA**（線性降維）、**Spectral** 方法（圖拉普拉斯特徵嵌入）與 **CNN**（影像分類）的多階段手寫數字辨識流程，核心想法是：先以 PCA 壓縮高維影像訊號並去除噪聲，再以 kNN 圖構造資料的鄰近關係並透過 **Laplacian** 的特徵向量抽取非線性流形結構（**spectral embedding**），最後將 (i) CNN 從原圖學到的局部空間特徵與 (ii) **spectral embedding** 的全域結構特徵進行融合，期望提升分類的準確率與訓練穩定性

1. 問題設定 (Problem Setting)

1.1 任務

給定 MNIST 手寫數字資料集：

- 輸入影像： $x \in \mathbb{R}^{28 \times 28}$ (攤平後 $x \in \mathbb{R}^{784}$)
- 標籤： $y \in \{0, 1, \dots, 9\}$

目標是學得分類器 $f_\theta(x)$ ，最大化測試集正確率

1.2 監督式學習目標

一般以 cross-entropy 最小化為訓練目標：

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^n \ell(f_\theta(x_i), y_i), \quad \ell = -\log p_\theta(y_i | x_i)$$

2. Method Overview

2.1 Pipeline

1. CNN 分支 (保留影像空間結構)

$$x \rightarrow h_{\text{cnn}}(x)$$

2. PCA + Spectral 分支 (抽取全域結構)

$$x \rightarrow \text{flatten} \rightarrow \text{PCA} \rightarrow z \rightarrow \text{kNN graph} \rightarrow \text{Laplacian eigenspace} \rightarrow e$$

3. 融合分類 (feature fusion)

$$[h_{\text{cnn}}(x), e] \rightarrow \text{FC} \rightarrow \hat{y}$$

3. PCA：線性降維的完整推導 (From Reconstruction Error to Eigenvectors)

3.1 為什麼要 PCA

MNIST 影像攤平後維度 $p = 784$ ，存在：

- 背景冗餘（大量接近 0 的像素）
- 高相關性（筆劃造成的像素相關）
- 雜訊與個別書寫風格差異

PCA 的目的：找低維子空間保留主要變異，等價於最小化重建誤差

3.2 中心化與協方差

令資料矩陣 $X \in \mathbb{R}^{n \times p}$ ，每筆樣本 $x_i \in \mathbb{R}^p$ 。

平均：

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

中心化：

$$\tilde{x}_i = x_i - \mu$$

協方差矩陣：

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i \tilde{x}_i^T$$

3.3 PCA 最佳化形式

要找 k 維子空間的正交基 $U \in \mathbb{R}^{p \times k}$ ，滿足 $U^T U = I$ 。

投影表示：

$$z_i = U^T \tilde{x}_i \in \mathbb{R}^k$$

投影後總變異（能量）為：

$$\sum_{i=1}^n \|U^T \tilde{x}_i\|^2 = \text{trace}(U^T \Sigma U)$$

因此 PCA 等價於：

$$\max_{U^T U = I} \text{trace}(U^T \Sigma U).$$

3.4 取前 k 個特徵向量

對 Σ 做特徵分解：

$$\Sigma v_j = \lambda_j v_j, \quad \lambda_1 \geq \lambda_2 \geq \dots$$

令 $U = [v_1, \dots, v_k]$ ，即可最大化上述 trace

最後得到 PCA 特徵：

$$z_i = U^T(x_i - \mu).$$

3.5 PCA 功用

- 降維與去噪：使得後續建圖的距離更穩定
 - 計算效率：spectral 需用距離/鄰近關係，低維可加速
-

4. Spectral：從圖切割到 Laplacian Eigenvectors

4.1 從資料點到圖 (Graph Construction)

用 PCA 特徵 z_i 建立相似圖 G ：

- 權重矩陣 $W = [w_{ij}]$
- 建圖方式： kNN graph (穩定、稀疏)

degree :

$$d_i = \sum_j w_{ij}, \quad D = \text{diag}(d_1, \dots, d_n)$$

Laplacian (unnormalized) :

$$L = D - W$$

4.2 重 $f^T L f$ 幾何意義

對任意向量 $f \in \mathbb{R}^n$ (視為每個點的標量函數值)，有：

$$f^T L f = \frac{1}{2} \sum_{i,j} w_{ij} (f_i - f_j)^2.$$

推導

$$f^T L f = f^T (D - W) f = f^T D f - f^T W f$$

其中

$$f^T D f = \sum_i d_i f_i^2, \quad f^T W f = \sum_{i,j} w_{ij} f_i f_j.$$

得：

$$f^T L f = \frac{1}{2} \sum_{i,j} w_{ij} (f_i^2 + f_j^2 - 2f_i f_j) = \frac{1}{2} \sum_{i,j} w_{ij} (f_i - f_j)^2.$$

4.3 從離散分群到放鬆 (Relaxation) → Eigenvectors 出現

理想二分群可用離散指示向量 $f \in \{-1, +1\}^n$ 表示，但直接最佳化會 NP-hard

Spectral 方法做「連續放鬆」：把 f 放寬為連續向量，並加上約束（避免退化成常數向量）：

$$\min_{f \perp \mathbf{1}, \|f\|=1} f^T L f.$$

此為 Rayleigh quotient 的最小化問題，其解是 L 的第二小特徵向量 (Fiedler vector)。

多群 (k 群) 則取最小的前 k 個特徵向量組成矩陣 $U \in \mathbb{R}^{n \times k}$ ，令：

$$e_i = U_{i,:} \in \mathbb{R}^k$$

作為每個樣本的新表示 (spectral embedding)

4.4 Spectral 在本 pipeline 的角色 (必須定義清楚)

本報告採用最自然、最可落地的定義：

Spectral 的輸出不是「分群標籤」本身，而是「embedding 特徵」
用於分類時與 CNN 特徵融合。

好處：

- 讓「spectral clustering」變成非線性降維/流形特徵抽取

- 對 CNN 來說是額外的「全域結構先驗」
-

5. 實驗設計

5.1 比較的四個模型

Model	CNN(原圖)	PCA feature	Spectral feature	目的
A Baseline	✓	✗	✗	基準
B CNN+PCA	✓	✓	✗	看 PCA 是否提供額外資訊
C CNN+Spectral	✓	✗	✓	看結構特徵是否有效
D CNN+PCA+Spectral	✓	✓	✓	完整方案

5.2 指標

- Test Accuracy
 - Training Time / 收斂速度
 - Confusion Matrix
-

6. Colab 程式碼

- 連結：

<https://colab.research.google.com/drive/1mKg7PTQSJq5INLZHN99mMpLupyUOW0Aj?usp=sharing>

(created by gpt)

7. 結果分析

結合 **PCA**、**Spectral Embedding** 與 **CNN** 的多階段手寫數字辨識方法，其目的為補強其在「全域結構判斷」上的不足，透過理論推導與實驗驗證，本研究可得以下結論

首先，實驗結果顯示 **PCA** 本身對分類準確率的提升極為有限，這與 PCA 作為線性降維方法的本質相符，其主要功能在於去除冗餘與穩定距離結構，而非直接提升判別能力

此結果支持本研究的設計動機：PCA 在本方法中扮演的是「前處理與輔助角色」，為後續的圖結構分析提供較為穩定的低維表示

其次，當在 PCA 表示空間上進一步引入 **Spectral Embedding** 時，模型的分類表現出現明顯且穩定的提升，相較於 baseline CNN，本研究所提出的 **CNN + PCA + Spectral** 架構在 MNIST 測試集上達到約 **0.8% 的準確率增益**

confusion matrix 的分析可觀察到，效能提升主要來自於 **結構相似數字之間的錯誤修正**，例如 3 與 5、4 與 9、7 與 9 等常見混淆類別，這些錯誤通常並非源自局部筆劃資訊的缺失，而是整體形狀或全域結構判斷的不確定性

綜合上述結果，本研究證實 **Spectral Embedding** 並非作為單獨的分類工具，而是作為 CNN 的輔助結構特徵來源 時，才能發揮其實際價值，此方法在 MNIST 上已展現出一致的改善趨勢，顯示該架構具備合理性與可解釋性