# Machine Learning Final Project Report

## （Few-Shot Learning 應用於考古學上的簡化模型問題設計）

許晉　314653002

December 3, 2025

Abstract

本期末專案探討一項二十年後人工智慧在考古學領域上可能具備的重要能力，分析其實現所需的資料、工具與學習架構，並提出一個可操作的模型問題作為邁向未來能力的第一步。其中，我們專注在利用原型網路 (Prototypical Network) 執行少樣本學習（Few-Shot Learning）[1],[6] 的簡化實作與討論。

## Contents

# 1 AI 的未來能力

## 1.1 能力描述及應用場景

人工智慧 (AI) 對於目前的考古學領域已發揮一定的影響力 [2]，例如能處理大量的地形或衛星資料、辨識古文物的材質以推測年代、利用自然語言處理破譯古文字等等。但同時也有一些瓶頸急待解決，例如：

- 同一件文物在不同的背景下可能有不同的解釋（像是同時具有宗教、技術、貿易用途），無法簡單的只用一類標籤去標記。
- 承上，文物上可能留下一些符號與文字，AI 可能可以幫助我們辨認文字，但無法給予解釋。且像是古符號通常帶有主觀解釋空間，對於 AI 來說要理解意義似乎還有難度。
- 雖然透過影像辨識技術能認出物體，但因為考古資料通常是多模態的，包含影像、地層、年代測定、出土紀錄、文本記載等等。因此，以現在最多只能處理一到兩種模態的 AI 模型來說，難以整合時間、空間與文化資訊。
- 由於決策過程很常是黑箱，所以 AI 給出的結果很常是無法解釋的。考古學家不一定能接受沒有原因的推論結果。
- 或者，若 AI 給出錯誤的判斷結果，有可能造成文物盜掘的風險。

其中，我們特別討論「**未來 20 年，AI 除了能辨認古器上面的文字或紋理之外，還能幫助學者分類並標註，並從全部的未標註樣本中挑選最值得學者親自標註的樣本**」這個能力。這其中的瓶頸包括：

- 資料大部分破碎且不完整（像是古器或是損壞的碑文）
- 仰賴深厚的背景知識，才能認出是哪個時期的文物
- 因為不同學者有不同詮釋，所以不一定有標準答案

# 2 所需的成分與資源

要完成上節所提到的任務，AI 除了需要能自動辨識古器文字與紋理之外，還要能主動從大量未標註的古器影像中挑選出「最值得讓學者」標註的樣本。其中涉及了**自動分類**、**主動標註選擇**及**知識構建**三項完整的能力，仍然是一件非常大的工程。

## 2.1 資料需求

要使 AI 能達到上述的能力，資料需求除了單純的影像之外，還需要數種其他資料來支援。例如高解析度的影像，其中還必須包括立體資訊（刻痕深度、紋理與材質）。還有該影像的文字標註，包括但不限於出土地點、文物類型、文化時期等來作為背景知識，以推測出哪些未標註資料最值得學者判斷。以及準備讓 AI 挑選的未標註的大量影像池，還有學者對 AI 所挑選出的樣本所做出的接受或拒絕判斷，以校正 AI 的分類模型。

## 2.2 工具與數學方法

首先，我們必須透過表現學習 (Representation Learning) 建立有意義的嵌入 (Embedding)，使得外觀相似、文化時期相近的兩組影像在空間中是相似的向量，同時也為接下來的主動學習 (Active Learning) 建立基礎。

AI 要從大量未標註資料中選出最值得標註的樣本需要策略，例如利用模型分類時計算出的熵 (Entropy)，若熵愈大代表此樣本蘊含資訊量愈大，就值得學者標註。

同時，因為真實情況中已標註的樣本數量可能不多，因此需要利用少樣本學習 (Few-shot Learning) 從少數已有學者標註的樣本中去學習分類。同時，也是本次專案著重的能力。

最後，因為古器研究有大量的結構關係，例如：文物與地點、文物與時期、紋飾與文化、出土位置與地層的關係。因此，可以使用圖神經網路 (Graphical Neural Network, GNN) 來表示以增強 AI 對文化知識的理解，而非單純的影像分類。

### 2.3 硬體與環境要求

若要取得高解析度的影像，則需要像是 3D 光掃描儀或者自動旋轉平台等設備，以提供足夠的感測資訊，且自動拍攝出不同角度的文物照片。以及，為了能執行大量的運算，大規模的圖形處理器 (GPU) 或張量處理器 (TPU) 的訓練環境也是必要的。除此之外，也需要有人機互動介面讓學者可以標註影像及對 AI 的挑選做出拒絕或接受的判斷，提供校正模型的回饋。

## 3 涉及的機器學習類型分析

### 3.1 主要的學習方式分類

要達成以上的需求，主要需要以下幾類學習方法：面對如此複雜的任務必須是混合式的學習架構，建立 embedding 需要自監督學習 (Self-supervised Learning)，其中我們將著重在利用少樣本學習來讓模型學習特徵並做出分類、根據學者作出的判斷進行監督式學習 (Supervised Learning) 以及利用主動學習來挑選最值得讓學者標註的影像。最後，藉由學者的回饋讓模型能優化挑選的策略，應用到基於人類回饋的強化學習 (Reinforcement Learning with Human Feedback, RLHF)。

### 3.2 資料來源、目標訊號、回饋與環境互動

資料來源主要來自少量的已標註古器資料、大量未標註的影像以及學者回饋的行為資料。目標訊號則有分類訊號（紋飾圖樣的類別）和主動學習提供的值得學者標註的樣本。透過學者的回饋，更能校正模型挑出值得人工標註的樣本，且能建立文化知識的網路。

## 4 第一步可實作模型問題（Solvable Model Problem）

### 4.1 問題設計與對應到最終能力

回顧最終想要達到的能力「**未來的 AI 能辨認古器紋飾，並從大量未標註影像資料中挑選值得學者標註的樣本**」。光是要達到這項任務就由幾項子任務組成：

- 模型首先要具備分類能力
- 建立好的 embedding 以辨識相似的紋飾圖樣
- 需要在只有少樣本的情況下，快速適應新紋飾的能力
- 必須能從未標註的樣本池中挑選最值得學者標註的樣本

因此，我們設計所謂的「第一步」問題為：「**針對人工合成的多種紋飾上，實作一個 few-shot 分類器，希望模型能在極少的樣本下仍能辨識新的紋飾類型**」。有了分類器，才能做完成後續的任務。

### 4.2 任務定義：輸入、輸出、資料形式

資料集採自己合成的紋飾資料集，都是 $64 \times 64$ 的灰階影像。紋飾總共有十類：水平條紋、斜條紋、點陣、棋盤格、同心圓、波浪紋、螺旋紋、輻射紋、碎片、噪音。每類的資料集切割為訓練集 (Training Set) 150 張、測試集 (Test Set) 30 張以及驗證集 (Validation Set) 20 張。**但這次只使用訓練集內的資料，若後續要調參數才會使用驗證集，本次專案冀在初步瞭解 few-shot learning 的概念**。輸入為一張影像 $x \in \mathbb{R}^{64\times64}$，輸出為紋飾類別 $y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$。

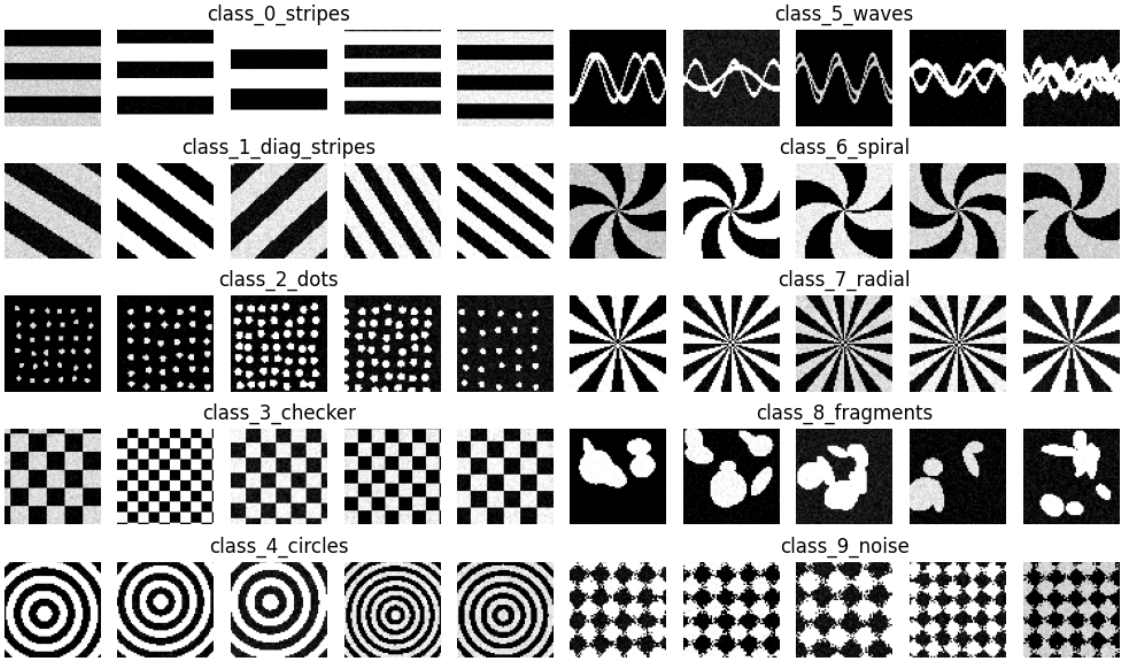


Figure 1: 10 類不同紋飾的示意圖

### 4.3 模型選擇與方法

本專案主要想做的是少樣本學習而非採傳統的監督式學習。因此我們依循此篇論文 [5] 的方法，透過原型網路 (Prototypical Networks) 的架構來做 few-shot learning，接下來簡單介紹 few-shot 的想法：

給定 $N$ 張已被標註紋飾類別的照片，其集合為 $\Omega = \{(x_1, y_1), \ldots, (x_N, y_N)\}$，其中 $x_i \in \mathbb{R}^{64\times 64}$，$y_i \in \{1, \ldots, K\}$ 是對應的類別。我們會將所有的類別分割為 $\mathcal{C}_{\text{train}}, \mathcal{C}_{\text{test}}$。

其中，每一個訓練單位被稱為一個 episode。在一個 episode 中會進行以下的流程：

1. 先從 $\mathcal{C}_{\text{train}}$ 抽出 $n$ 個類別
2. 再從這些類別中各抽 $k$ 張作為支援樣本，形成 support set $S$
3. 再從這些類別中抽 $m$ 張作為查詢樣本，形成 query set $Q$
4. 當模型看到 $S$，會根據 $S$ 當場構造出一個分類規則，然後用這個分類規則分類 $Q$ 中的每一個樣本。以本次使用的原型網路來說：
   (a) 先找一個 encoder 函數 $f_\theta$，將 support set $S$ 中的圖嵌入至特徵空間：
   $$z_i = f_\theta(x_i)$$
   (b) 對每一個類別 $k$，計算該類每樣本在特徵空間中的平均，得到原型中心
   $$c_k = \frac{1}{|S_k|} \sum_{\{x_i : y_i = k\}} f_\theta(x_i)$$
   (c) 對 $Q$ 中的每一個影像 $x^{(q)}$，一樣先做 embedding 為 $f_\theta(x^{(q)})$，再與各原型中心的距離做 softmax：
   $$P_\theta(y = k | x^{(q)}) = \frac{\exp\left(-d(f_\theta(x^{(q)}), c_k)\right)}{\sum_l \exp\left(-d(f_\theta(x^{(q)}), c_l)\right)}$$
   用這個機率決定該圖片要分到哪一類。特別的是這裡計算距離時皆用 L2 norm：$\|f_\theta(x^{(q)}) - c_k\|^2$，而非使用餘弦相似度 $\cos\theta = \dfrac{f_\theta(x^{(q)})^T c_k}{\|f_\theta(x^{(q)})\| \cdot \|c_k\|}$。

(d) 最後計算 Loss，定義為

$$J(\theta) = -\log P_\theta(y = k|x)$$

也就是說，$\mathcal{C}_{\text{test}}$ 中的類別完全不會參與訓練過程，這正是原型網路想做的事：能對訓練過程中沒出現過的類別進行分類。

## 4.4 實作流程

現在將實作細節交代得更清楚。在實作中，我們共有 10 類不同的紋飾，實作 3-way 1-shot 與 3-way 3-shot。$n$-way $k$-shot 的意思是在每個 episode 中訓練模型分辨 $n$ 個類別，且每個類別提供 $k$ 張作為支援樣本。首先，我們從 10 類中隨機選 3 類形成 $\mathcal{C}_{\text{test}}$，**這 3 類不會參與訓練過程**。其餘 7 類則形成 $\mathcal{C}_{\text{train}}$。

接著，我的實驗中在每一個 episode 中取 $m = 8$ 張查詢樣本作為 query set，也就是當模型用支援樣本建立完原型中心後，用這 8 張圖片去計算 loss，並以此結果修正 embedding。

至於 encoder，我們挑 Conv-4 網路當作 $f_\theta$，這也是在許多 few-shot 實驗中選用的函數。它包含四個相同的區塊，每一塊中包含四層，依序是 Conv2d（卷積層）、BatchNorm2d（標準化層）、ReLU（非線性層）及 MaxPool2d（池化層），其中池化層將圖像中分區域，取得每個子區域中最明顯的特徵，可以使得原本是 4096 維的資料維度降為一半。因此，經過 Conv-4 運算之後，資料維度會只剩下 64 維。數學上也可以說 $f_\theta : \mathbb{R}^{64\times64} \to \mathbb{R}^{64}$。優化器則普通地挑 Adam[4]。

那麼就可以開始訓練，訓練完之後拿 $\mathcal{C}_{\text{test}}$ 中的類別做 3-way 1-shot 和 3-way 3-shot 測試。由於我們的 $\mathcal{C}_{\text{test}}$ 中就恰好只有 3 類，因此每個 episode 中都是在對這 3 類做分類。

接下來以表格與圖表顯示結果。

## 4.5 結果

首先是 loss curve 與 accuracy curve，測試時以 300 個 episodes 測試，並計算其平均準確率。

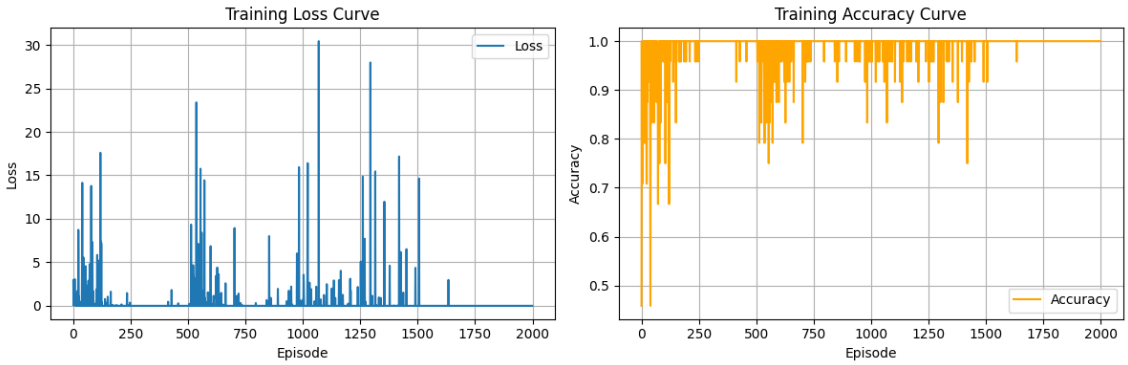


Figure 2: Loss Curve & Accuracy Curve in 2000 training episodes and 300 test episodes

可以觀察到 loss 與 accuracy 分別確實有隨著 episode 趨於 0 和 1。最後得到的準確率如下表：

| | Accuracy |
|---|---|
| 3-way 1-shot | 94.47% |
| 3-way 3-shot | 97.24% |

Table 1: Accuracy when doing 3-way 1-shot and 3-way 3-shot

或者，我們也可以隨機選一張圖丟入模型，看看其預測結果：

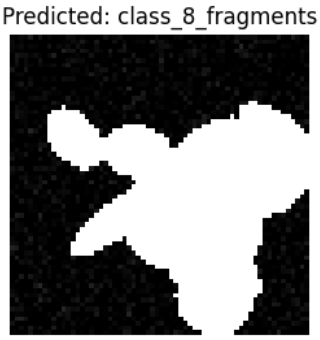


Figure 3: Example Image

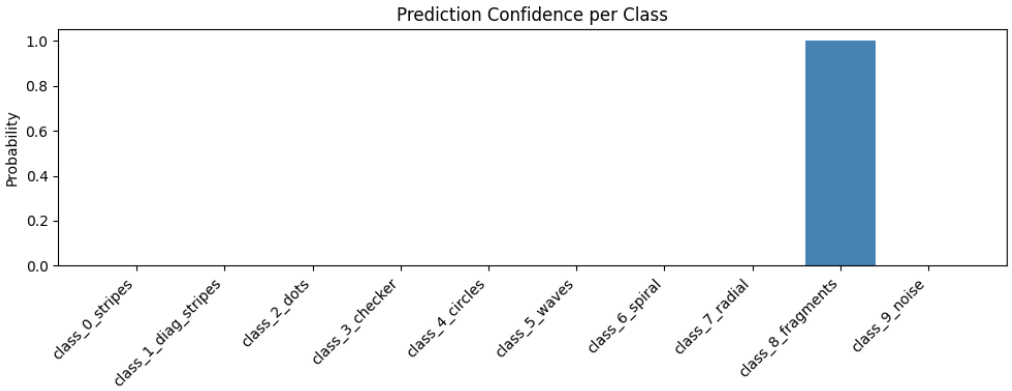


Figure 4: Probability Barchart

### 4.6 討論

從以上結果可以看到預測結果幾乎極好（若隨便從 3 類中亂猜則正確率是 33%），可以推測是因為我們的 $\mathcal{C}_{\text{test}}$ 就恰好只有 3 類，所以在測試時每次都是抽到這 3 類測試，模型可以在這 300 次的 test episodes 內完全地針對這 3 類練習。而且，因為我們人工生成的紋飾圖片即使有手動加上一些噪音，但仍然非常清楚，且每個類別相距甚大，因此會得到很好的準確率也不是很意外的結果。

但同時也可以想像，若應用在真實挖掘到的古器上，觀察到的紋飾可能殘破不堪、有缺損或汙漬，且顏色的分辨度上也沒有人工生成的圖片那麼高，噪音程度也肯定非常高。因此，在真實的例子上除了要有好的設備去得到高解析的影像或更多細膩的資訊之外，也可能需要更複雜的 $f_\theta$ 做 embedding。

並且，在真實的考古情境下，有可能會發現以前從來沒看過的新類別，以這次的 toy model 來說，也就是我們可能在輸入時會餵進第 11 種沒看過的紋飾，希望模型可以告訴我們這是訓練時沒看過的類別，而不是隨便分一類。這一點或許可以靠設定機率閥值來做到 [3]，設定當分到任一類的機率都低於 $\varepsilon$ 時，就不做任何判斷。

## 5 結論

總結來說，在本次的專案中，我們先給定一個 AI 確實對考古學領域會有一定影響力的想法，並發現其中可能分成很多不同的任務可以藉由 AI 獲得更有效率或更優的結果。

其中，對於「未來 20 年，AI 除了給辨認古器上面的文字或紋理之外，還能幫助學者分類並標註，並從全部的未標註樣本中挑選最值得學者親自標註的樣本」這個目標，我們做了一個非常簡化的 toy problem，那就是希望能透過少數樣本做到對人工合成的灰階紋飾圖片分類。技術上則是借用了原型網路做少樣本學習。

# 参考資料


[1] Yinbo Chen et al. “A New Meta-Baseline for Few-Shot Learning”. In: CoRR abs/2003.04390 (2020). arXiv: 2003.04390. URL: https://arxiv.org/abs/2003.04390.

[2] Gabriele Gattiglia. “Managing Artificial Intelligence in Archeology. An overview”. In: Journal of Cultural Heritage 71 (2025), pp. 225–233. ISSN: 1296-2074. DOI: https://doi.org/10.1016/j.culher.2024.11.020. URL: https://www.sciencedirect.com/science/article/pii/S1296207424002516.

[3] Dan Hendrycks and Kevin Gimpel. “A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks”. In: CoRR abs/1610.02136 (2016). arXiv: 1610.02136. URL: http://arxiv.org/abs/1610.02136.

[4] Diederik P Kingma. “Adam: A method for stochastic optimization”. In: arXiv preprint arXiv:1412.6980 (2014).

[5] Jake Snell, Kevin Swersky, and Richard S. Zemel. “Prototypical Networks for Few-shot Learning”. In: CoRR abs/1703.05175 (2017). arXiv: 1703.05175. URL: http://arxiv.org/abs/1703.05175.

[6] Oriol Vinyals et al. “Matching Networks for One Shot Learning”. In: CoRR abs/1606.04080 (2016). arXiv: 1606.04080. URL: http://arxiv.org/abs/1606.04080.