

Final Project Report

Member: A102108林承慶、0850740王昱翔

一、介紹 (Introduction)

本作品的發想源自於我們兩人的朋友，平時兼職做手相算命師，拿到顧客的手部照片後，他必須對手上的線紋(生命線、感情線...等等)其骨幹和細微部分進行描繪，依據描繪結果來算命，由於人工進行線紋描繪和細部判斷讓其工作十分繁重，平時一天最多只能夠分析2-3人，但截至目前為止，已經有超過1000人預約請他看手相，依照他的速度，最快也要1年的時間他才能全部做完。於是，我們希望能夠運用課堂上所學習到的知識和工具，建立一個深度學習模型來幫助他進行手相分析，盡可能地將耗費大量時間、精力且重複的工作交由模型來完成。

二、相關工作 (Related Work)

我們探討了過去有關於手相分析結合深度學習的論文，基本上可分為兩種類型：第一種為直接利用手部照片進行分析，分析結果只有好與壞兩種區別，並沒有額外的命理解釋。在[1]中利用Fast-SCNN對照片進行去背，將照片切割後，再利用CNN來預測命運的好壞。但這麼做會有個缺點，切割手部照片可能會一定程度上讓線紋的細節失真，造成某些特徵流失或無法被模型判斷，例如：「生命線末端呈流蘇」這個特徵可能會在照片被切割後無法被模型捕捉，或是線紋與其他線紋交錯的情況也可能無法被模型判斷。

第二種為掌紋標示，利用深度學習模型在手部照片上標示掌紋，也同樣沒有額外的命理解釋，在[2]中利用UNet結合context fusion來進行掌紋繪製，並沒有對於繪製出來的掌紋進行額外解釋或命運預判。[2]所使用的資料全部都是Kaggle上的手部照片，均已經過去背，沒有背景干擾問題，如果使用這種方法，我們也必須對之後輸入模型的照片進行去背，此外[2]只針對線紋較清晰的主幹部分進行描繪，並沒有在線紋交錯或線紋細節的特徵著墨。

本作品在分析上保留了手部的整體，不需要進行照片切割，除了能夠對手部的線紋進行描繪外，還能夠對掌紋的主幹和細節都能進行完整的分析和判斷，給出詳細命理解釋。

三、問題定義 (Problem formulation)

我們希望做出一個深度學習模型來協助他進行手相算命。以下為手相算命的基本流程：

1. 拿到顧客的手部照片
2. 進行掌紋的描繪
3. 對掌紋的特徵進行分析
4. 根據各個掌紋特徵給予其命理解釋

基本上在步驟2的掌紋描繪和步驟3的特徵分析需要花費大量時間，因此我們希望建立出來的模型能夠幫助他完成這兩步驟。模型在輸入手部照片後，就能自行描繪出掌紋，並捕獲掌紋的特徵，再根據掌紋特徵的不同給出其所對應的命理解釋。圖一為手相算命的流程中，本作品能夠協助的部分。本作品使用的分析資料集為：手部照片1442張，已有命理解釋的照片847張，已描繪掌紋的照片182張。

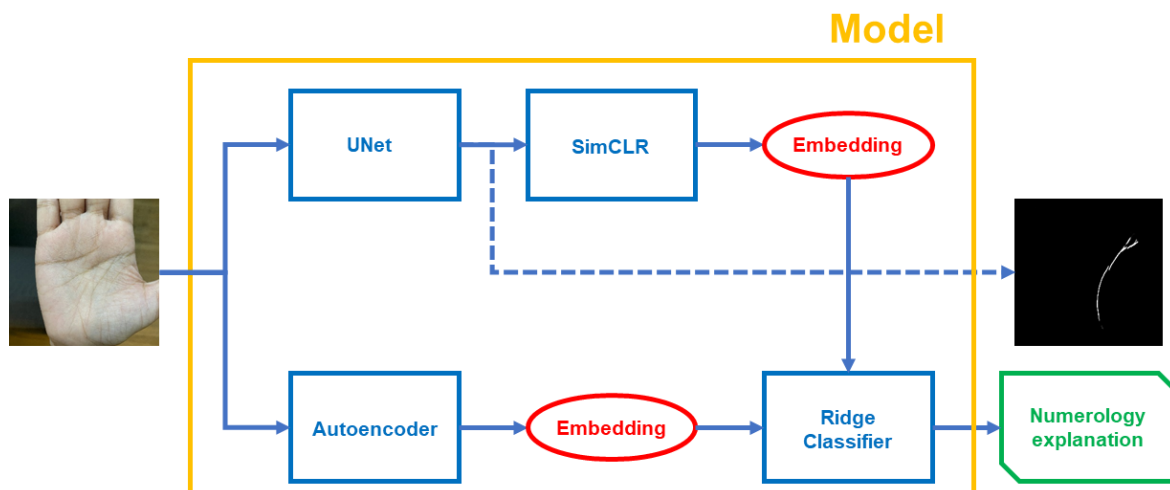


圖一、手相算命流程圖

四、方法論 (Methodology)

本作品所設計出協助手相算命的深度學習模型如圖二所示。由於本作品仍屬於實驗性質，因此所有的影像處理和分析都只針對「生命線」進行。取得顧客手部照片後，將照片輸入模型，本作品利用UNet對影像進行輪廓描繪，此一階段會取得手部線紋的描繪結果，接著將描繪出的輪廓圖經過SimCLR即可得到線紋的Embedding。除上述的分析之外，原始照片在進入模型後，將其放入以手部照片所訓練出

的Autoencoder中，以此獲得手的Embedding，得到線紋和手的Embedding後，將兩者結合放入Ridge Classifier中來對掌部線紋特徵進行判斷，其分類方式為如果線紋擁有該特徵，則標示為1；如果線紋沒有該特徵，則標示為0，並利用Classifier的分析結果給出其對應的命理解釋。下文將針對模型與分析流程進行詳細解釋。



圖二、手相算命深度學習模型

4.1 資料前處理 (Data Preprocessing)

友人所給予已繪製掌紋的資料中，所有的生命線都以綠色進行標註，本作品利用RGB擷取生命線輪廓並做成mask，為下段所敘述之方法UNet進行label。除此之外，因為有label的照片和原始照片有位置上的偏差，因此利用OpenCV抓取兩個照片的特徵點，並且讓兩張照片對齊。本作品利用data augmentation 讓照片旋轉和翻轉，如此一來就能讓1張照片變成8張，train data的總數變為原本的8倍，最後再將所有照片都壓縮到(256,256)後才進行處理。

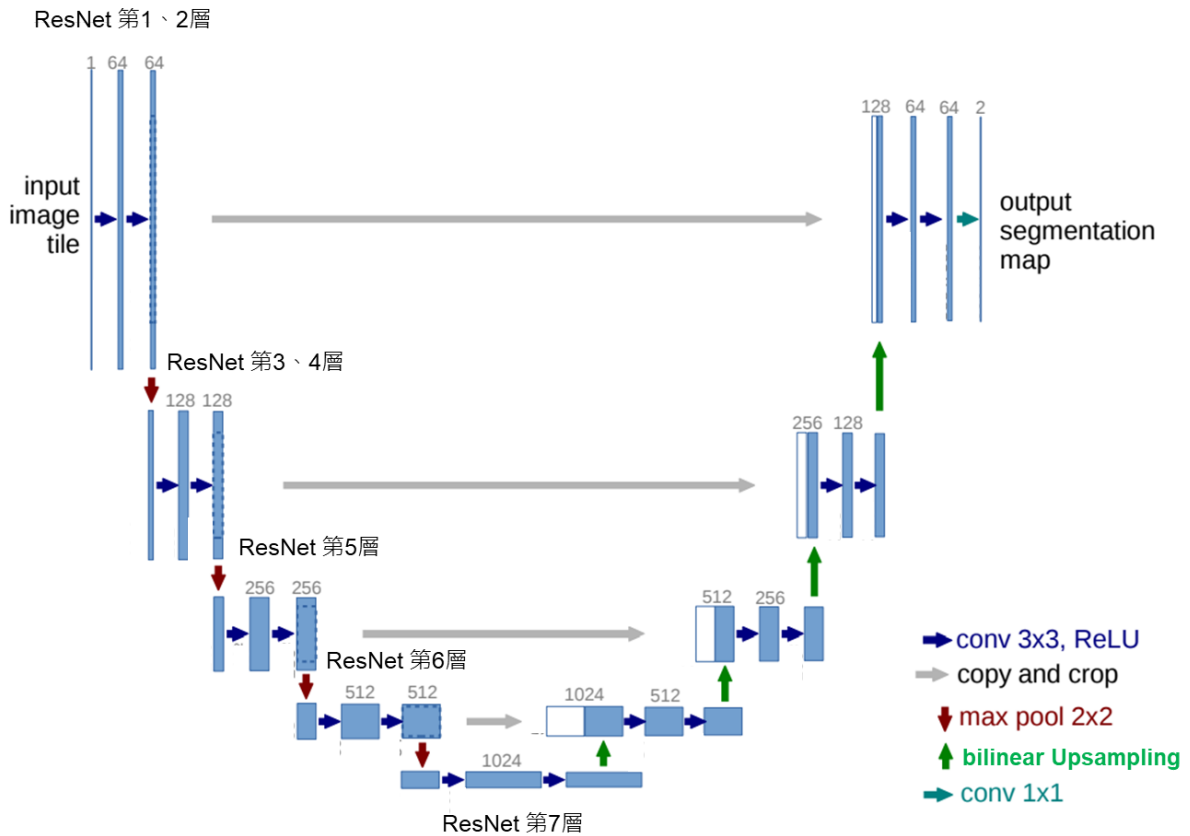
已有命理解釋的照片總數為847張，每張照片都有其對應的ID號碼，與命理解釋的文字說明以CSV的方式儲存，前處理時會將生命線特徵個別拉出獨立成一個欄位，如果ID擁有該特徵則標示為1；否則標示為0。

4.2 UNet

資料在進入模型之前，有先對資料進行標準化 $N \sim ((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))$ 。由於能用來進行訓練的label data數量非常少，而掌紋分析是必須對圖片中細微的物件進行判斷，因此本作品在訓練中使用了semi-supervised learning。其流程大致如下：

1. 利用label data來訓練模型
2. 利用步驟1訓練好的模型來預測unlabeled data，利用sigmoid得到機率值
3. 如果有pixel被模型判定為True的機率大於0.7，則把該照片當作pseudo label，放進training set一起訓練模型

本作品使用的模型為ResNet UNet，模型的一個block由convolution2d + relu + BatchNorm2d所組成，Upsampling的方式則是bilinear，整個UNet模型由block和ResNet18組成的。其結構如圖三所示。



圖三、UNet架構圖

Loss的部分本作品使用Dice Loss(D)和Binary Cross Entropy Loss(BCE)的結合，結合方式為：

$$Loss = 0.4Loss_{BCE} + 0.6Loss_D$$

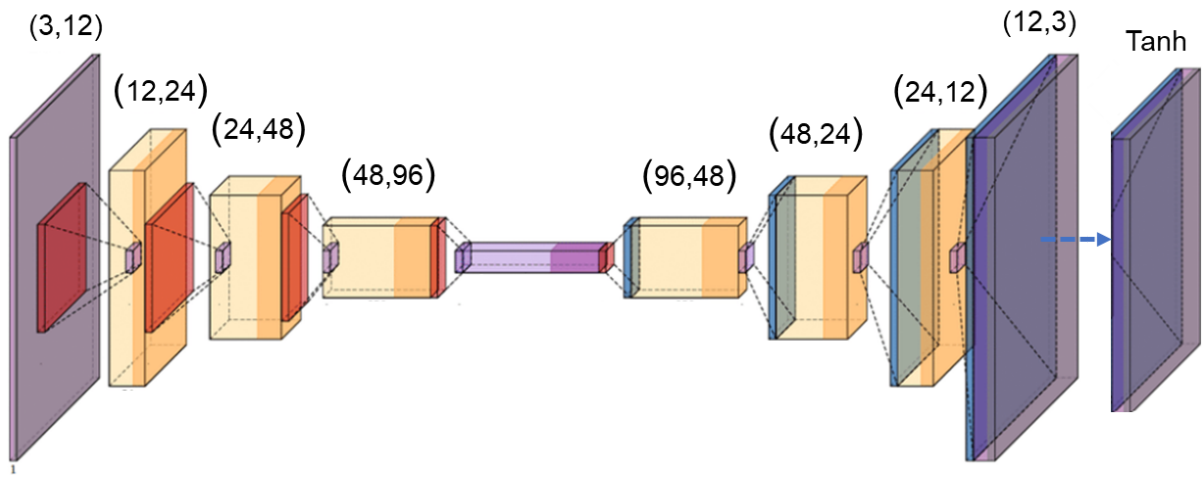
在5.1節會說明決定Loss的詳細過程。在UNet的部分本作品train了100個epochs，optimizer為Adam，learning rate為0.0001，scheduler為StepLR，每20步，learning rate就乘以0.1。

4.3 SimCLR

本作品使用原始SimCLR的架構，改成以ResNet18作為主幹，將UNet產出的生命線mask輸入。在這邊使用self-supervised的原因在於同張通過SimCLR的照片的結果其相似度應該很高，不同照片的結果其相似度應該很低。模型在判斷相似度的時候必須去學習手的細部差異到底在哪，而這就是本作品希望讓模型知道的。我們利用(256,256)的生命線圖，經過SimCLR，得到長度32的embedding。

4.4 Autoencoder

本作品將手掌原始照片輸入，利用Autoencoder取得手掌的照片的embedding。使用Autoencoder的原因在於想要得到手整體的資訊，但如果用SimCLR，那只會得到手的差異資訊。Autoencoder的架構如圖三所示。



圖三、Autoencoder架構圖

Kernel size 都是4, stride 2, padding 1, 每一個(256,256)的照片會得到96*16*16的embedding。

4.5 Ridge Classifier

最後本作品合併了SimCLR的embedding和Autoencoder的embedding, 總長度為24608。將所有生命線的特徵都當作binary classification的問題, 也就是一個multi-label classification的問題, 總共有34個特徵。

五、實驗結果 (Experimental results)

本章節內容在於闡述模型設計時所遇到的問題, 其中間實驗的過程與最後解決的方法。

5.1 UNet

本作品使用的模型為ResNet結合UNet架構, 但也有嘗試過使用EfficientNet結合UNet架構, 發現EfficientNet在10個epochs以內表現很好, 但超過10個epochs之後就會有overfitting的狀況; ResNet則是在10個epochs以內沒有表現得很好, 但超過10個epochs以上的表現不錯。除此之外, 我們也對兩者分別加入semi-supervised learning與否進行比較, 其結果如下:

1. 加入semi-supervised learning
 - (1) ResNet: 模型能夠清楚描繪出生命線主幹的部分, 但無法捕捉細節的部分。
 - (2) EfficientNet: 模型表現的極為不穩定, 在某些照片能夠清晰地描繪出生命線主幹和細節的部分, 但有時候主幹和細節都無法捕捉。
2. 不加入semi-supervised learning
 - (1) ResNet: 模型所描繪出來的生命線會失真, 形成類似橢圓形的區域。
 - (2) EfficientNet: 模型能夠清楚描繪出生命線主幹的部分, 但無法捕捉細節的部分。

綜合以上結果, 本作品決定使用ResNet結合UNet架構。

在計算Loss的時, 本作品採用混合Dice Loss和Binary Cross Entropy Loss的方式, 會使用此方式的原因在於UNet的BCE loss下降的速度非常快, 但是Dice Loss並不會, 因此權重會偏向BCE, 同時, 因為Dice loss很不穩定, 因此必須拿BCE Loss來穩定Dice loss。有嘗試將BCE Loss的權重調小至0.1, 此時Loss會變成以Dice Loss為主, 生命線會出現失真的狀況, 其位置會被判斷為True的pixel包裹成類似橢圓形的區域。如果將dice loss的權重調小至0.1, 此時Loss會變成以BCE Loss為主, 生命線也會出現失真的狀況, 被判斷為True的pixel會在圖形中間形成類似圓形的區域。

5.2 Autoencoder

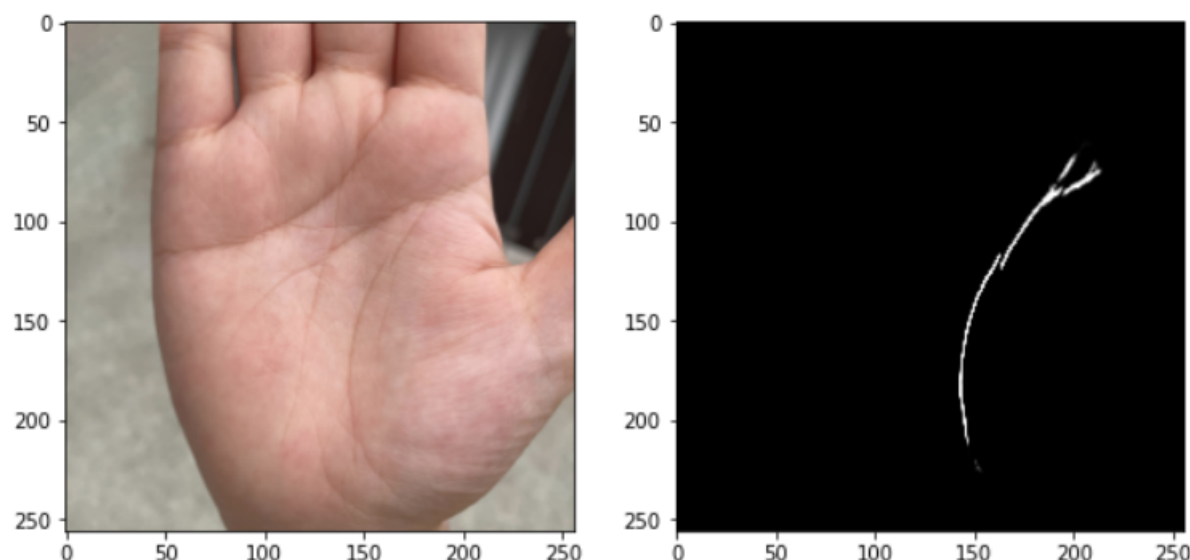
本作品嘗試使用四種方法建立Autoencoder, 分別為CNN、VAE、ResNet和FCN, 其中VAE、ResNet和FCN三種方式, 手部照片經過Encoder壓縮再通過Decoder後, 會變得十分的模糊, 基本上手掌的線紋都以無法辨別, CNN轉換出來的圖片雖然也明顯比原本的照片模糊, 但還可以辨別出手掌上的線紋, 因此本作品最後決定採用CNN架構。此外, 最後一層Tanh的部分有嘗試過替換成Sigmoid, 不過也明顯比Tanh模糊, 因此最後採用Tanh。

5.3 Ridge Classifier

本作品測試了CNN Classifier、Random Forest、KNN和Ridge Classifier。在測試CNN Classifier時，由於有80%的答案都是0，導致模型直接全部猜0的正確率也高達80%，因此這是一個imbalanced的分類問題，於是改採用f1 score來評斷模型好壞。用CNN Classifier、Random Forest、KNN和Ridge Classifier所得到的f1 score分別為 0.01, 0.19, 0.25, 0.37，因此最後決定使用Ridge Classifier。

六、結論 (Conclusion)

在過往的手相分析結合深度學習的研究中，往往需要去背或切割照片，同時也無法捕捉掌紋細節的特徵，本作品改良上述缺陷，照片的輸入不需要經過去背和切割處理，保留手掌完整資訊的同時也能捕捉掌紋的細節特徵。本作品完成後的模型，只要輸入手掌照片，模型就能夠描繪出掌紋，並同時給出命理解釋，圖三為實際輸入照片後的模型所描繪出的生命線圖。圖四為模型給出的命理解釋報告。



圖四、手部照片(左)輸入模型後得到的生命線圖(右)

生命線報告：

發端呈交錯 --> 缺乏耐性，多學少精，創新力不足，個性較為內向。
發端呈尖三角形 --> 個性挑剔感性敏銳、易怒且氣量小，適合做精巧的工作。
發端呈流蘇 --> 兒時多病，心理上比一般人更需要細心照顧。
發端呈鏈狀 --> 兒時多病，壯年時期才會好轉，學習模仿能力較為遲鈍，感情上容易過分獻身。
發端呈斷續小紋 --> 生活獨立性強，易因過於自我而讓人感到冷漠。
發端偏低 --> 缺乏自制力且消極，個性急躁不易冷靜，容易被騙。
間斷在中段 --> 與母親的緣份較薄，中年時期因工作搬家。
間斷在末段 --> 老年後會因為工作而搬家。
上升線在前段外側 --> 年輕時易有成就，或未來會與高收入的對象結婚。
上升線在中段外側 --> 青壯年時期容易有升官或升遷的機會。
上升線在木星丘 --> 只要願意投入便能將自身才能表現於外。
弧度面積較小 --> 包容力不高，常為小事計較，偏執刻板忌妒心強，對情慾較冷漠。
整體較短 --> 元氣不足五臟不堅，易有泌尿系統問題，忍耐力意志力不足。
整體筆直 --> 對疾病的抵抗力較差。
前段斷續連接 --> 年輕時健康不佳，恆心毅力不足，個性悲觀較難有成。
前段雜亂無章 --> 年輕時消化系統欠佳，營養吸收不良，有腸胃疾病。
中段雜亂無章 --> 青壯年時消化系統欠佳，營養吸收不良，有腸胃疾病。
末段深入掌根 --> 有較高機率能長命百歲。

圖五、命理解釋報告

七、任務分配 (Task Assignment)

A102108林承慶: Data Preprocessing data、UNet、SimCLR、Ridge Classifier、Demo code
0850740王昱翔: Autoencoder、Demo presentation、Final Report

八、附錄: 參考文獻

- [1] Van, T. P., Nguyen, S. T., Doan, L. B., Tran, N. N., & Thanh, T. M. (2020, November). Efficient palm-line segmentation with U-net context fusion module. In 2020 International Conference on Advanced Computing and Applications (ACOMP) (pp. 23-28). IEEE.
- [2] Acharjee, S., Nuannimnoi, S., & Huang, C. Y. (2020, December). A Deep Learning Approach for Efficient Palm Reading. In 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI) (pp. 171-174). IEEE.
- [3] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
- [4] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
- [5] Chen, D., Ao, Y., & Liu, S. (2020). Semi-supervised learning method of u-net deep learning network for blood vessel segmentation in retinal images. *Symmetry*, 12(7), 1067.
- [6] Abdollahi, A., Pradhan, B., & Alamri, A. (2020). VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access*, 8, 179424-179436.
- [7] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.

Source Code:

- <https://github.com/usuyama/pytorch-unet>
- <https://speech.ee.ntu.edu.tw/~hylee/ml/2021-spring.php>
- HW8 Anomaly Detection