**深度學習於醫學影像DLMI – Final Project**

R12725024 資管碩一 蔡鉎驊

這次期末報告我延續之前作業的題目，使用NeurIPS在2023年於Grand Challenge舉辦的競賽當作練習。競賽名稱為Foundation Model Prompting for Medical Image Classification (MedFM)，連結為https://medfm2023.grand-challenge.org。

1. **Introduction**

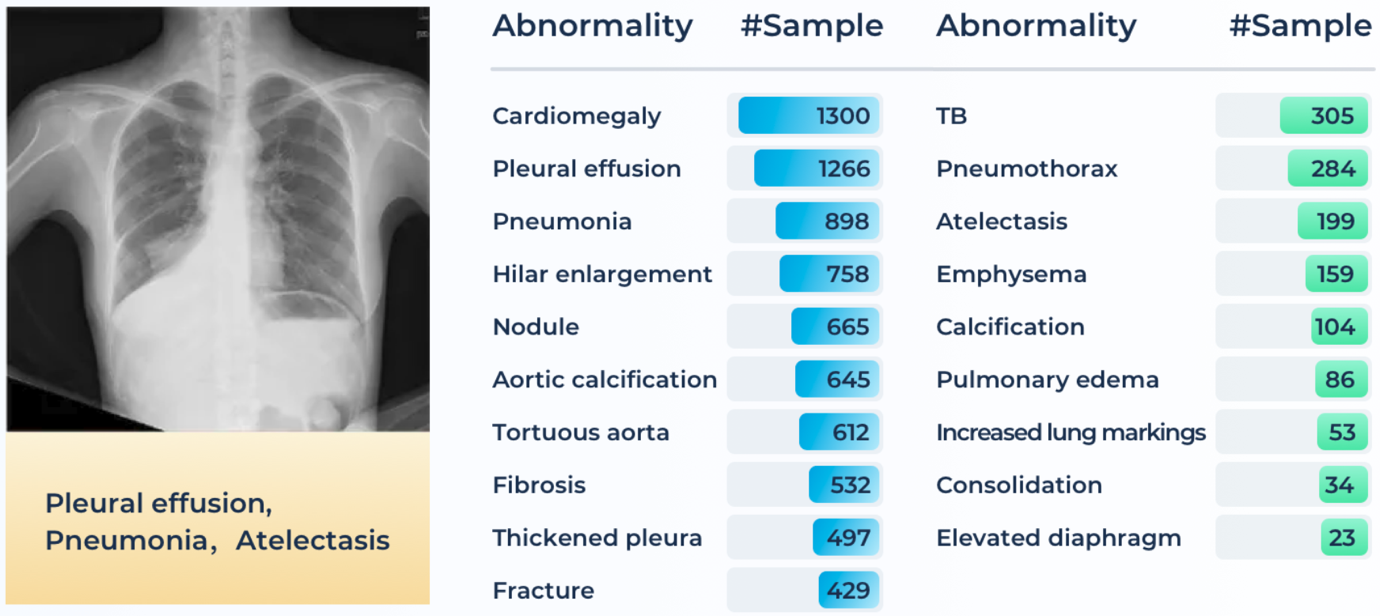
* Data Source

資料來源從他們競賽網站<https://medfm2023.grand-challenge.org/datasets/>即可下載6/2更新公布的資料集，將train與val的影像資料混合再一起，依任務類別放到data/medfmc/MedFMC/chest, data/medfmc/MedFMC/colon, data/medfmc/MedFMC/endo底下。

這個競賽主旨是著重在few-shot learning與prompt engineering，因此不允許參賽者使用全部的data訓練，也不允許使用競賽資料及以外的資料。我在這次的期末報告中，會使用他們示範的baseline裡面切好的train, val, test datasets當作我的訓練以及衡量結果的資料集。挑戰內容也有提到，會有這些限制是希望解決醫學影像中標註不足的問題，希望參與者利用少量私有數據進行初始訓練和驗證，並嘗試提高分類準確性，包括**放射學**、**病理學**和**結腸鏡影像**。

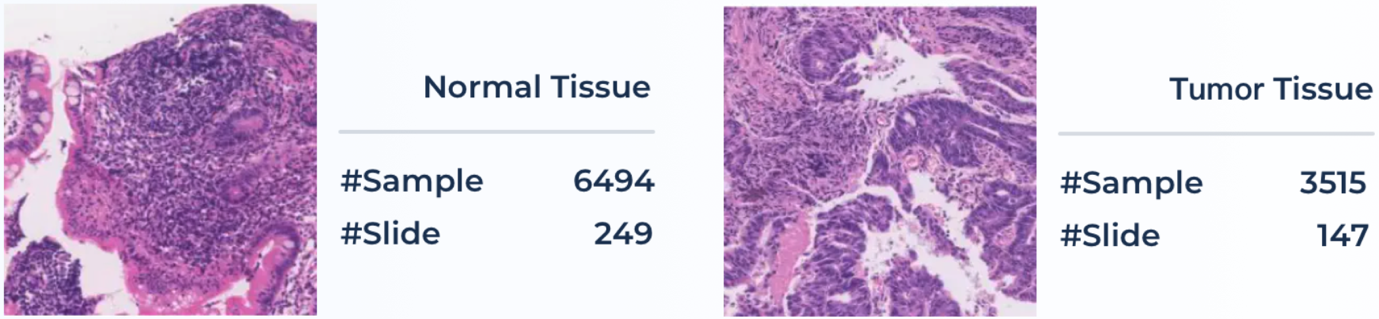
* Dataset

第一項任務：**胸部疾病篩檢 (ChestDR)**，目的在利用基礎模型，通過少量樣本來提供準確的胸部疾病篩查方法。使用胸部X光影像作為日常臨床篩查工具。



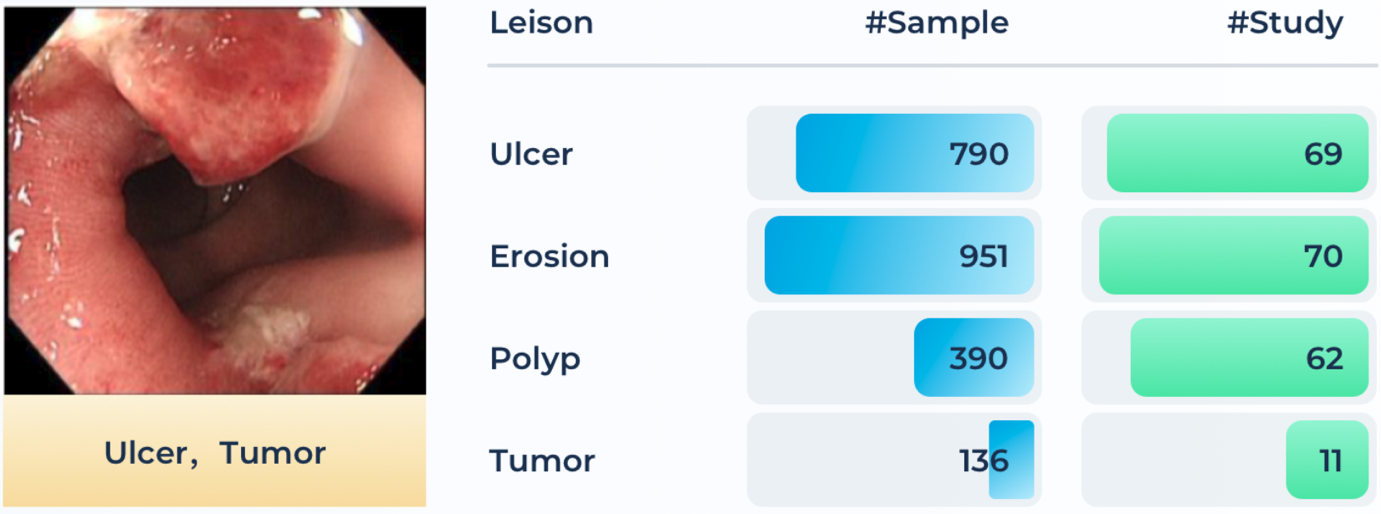
**（圖一，圖片來源：<https://medfm2023.grand-challenge.org/medfm2023/>）**

第二項任務：**病理性腫瘤組織分類 (Colon)**，目的在於利用基礎模型，通過少量樣本提供準確的病理性腫瘤組織分類方法。在病理學家日常的工作中，他們需要檢查許多組織切片，這對他們來說是一項繁瑣的工作。分類病理組織塊的方法有助於簡化這一過程，並幫助篩查整個切片中是否存在惡性細胞區域。



**（圖二，圖片來源：<https://medfm2023.grand-challenge.org/medfm2023/>）**

第三項任務：**病變偵測於結腸鏡影像 (Endo)**，目的是利用基礎模型，在結腸鏡檢查過程中自動檢測病變，以防止漏診並減輕胃腸科醫師的工作量。



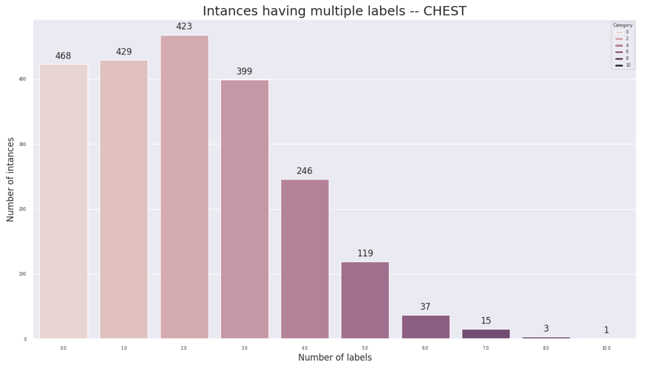
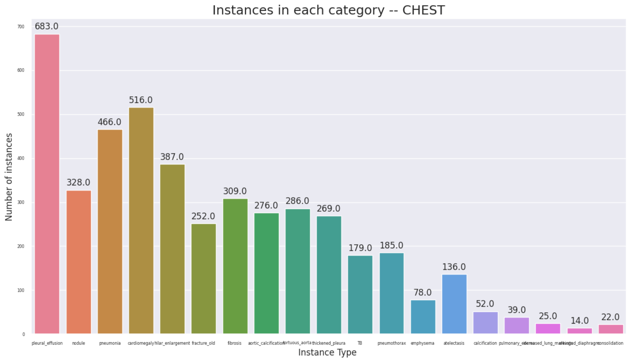
**（圖三，圖片來源：<https://medfm2023.grand-challenge.org/medfm2023/>）**

* Exploratory Data Analysis (EDA)

使用eda.ipynb進行資料視覺化，其中參考了此blog：<https://towardsdatascience.com/journey-to-the-center-of-multi-label-classification-384c40229bff>。

從**（圖四）**可以看到，在第一項任務CHEST的訓練資料集當中，以pleural effusion的佔比最高，有683筆，而elevated diaphragm佔比最低，只有14筆。

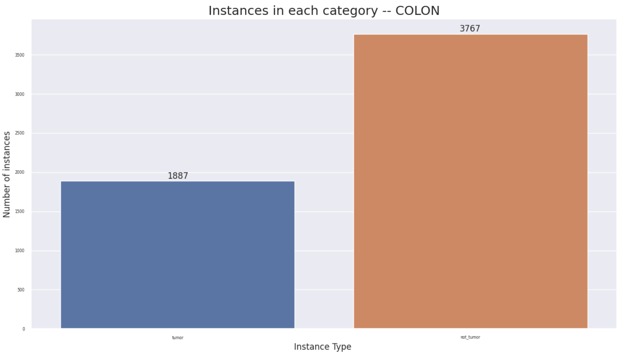
從**（圖五）**可以看到，有1筆資料被標注最多的10種腫瘤偵測，其他大多數的資料都只有被標注3種以下的腫廇。



**（圖四） （圖五）**

從**（圖六）**可以看到，在第二項任務COLON的訓練資料集當中，被標注為無腫瘤組織比有的大約多一倍。

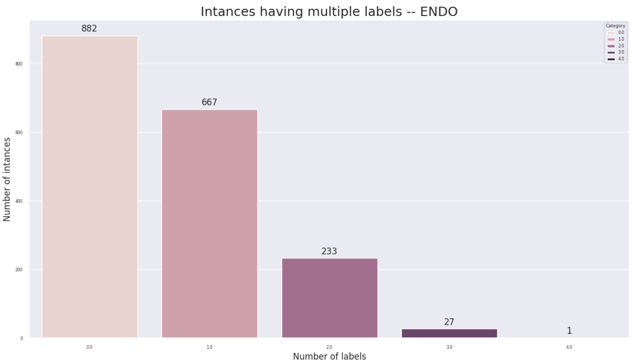
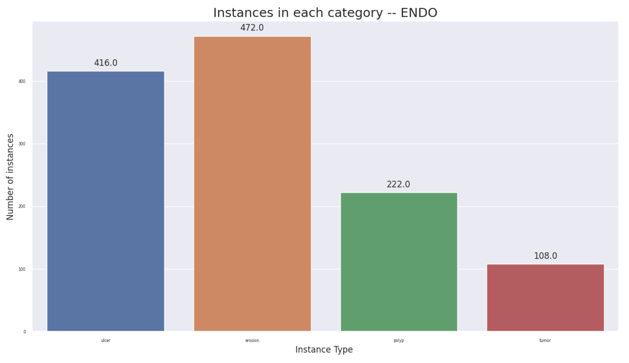
從**（圖七）**可以確認此項任務為binary classification。



**（圖六） （圖七）**

從**（圖八）**可以看到，在第三項任務ENDO的訓練資料集當中，以erosion的佔比最高，有472筆，而tumor佔比最低，不過整體分佈還算不算偏差太嚴重。

從**（圖九）**可以看到，有1筆資料被標注4種病變都有，而有882筆被標注這些病變都沒有。



**（圖八） （圖九）**

1. **Methodology**

* Model and Training Method

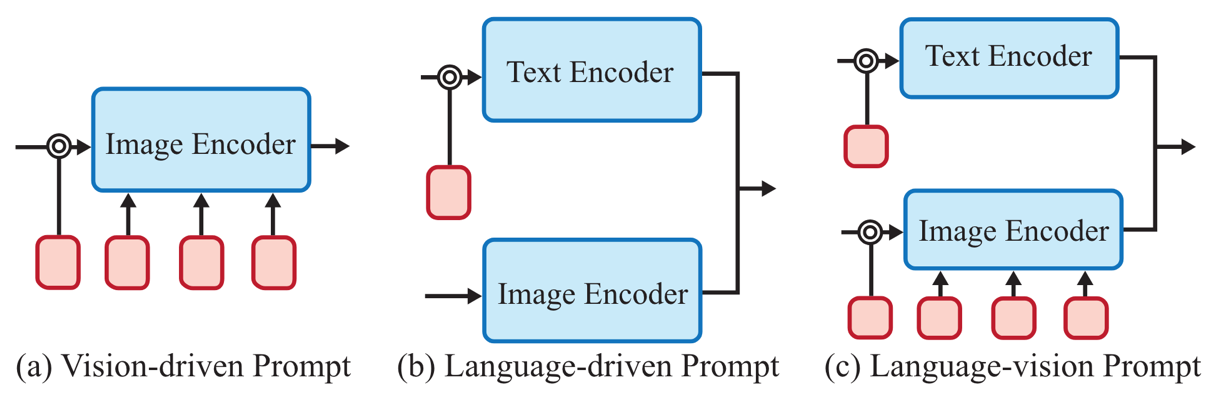
1. *Transfer Learning on 20% of Training Data by Densenet*

這次報告使用的競賽中，附有他們自己製作的baseline model ( <https://github.com/openmedlab/MedFM> ) ，使用第三方套件mmpretrain實作。

不過他們是使用較舊的mmpretrain版本，而在這次報告我是使用最新版本的mmpretrain，其文件配置與ground truth annotation file的格式都會與原本的範例baseline不一樣。

1. *Few-shot Learning by CoCoOp*

另外，我尋找了其他visual prompt tuning的方法，找到了CoCoOp (Conditional Prompt Learning for Vision-Language Models, in CVPR, 2022, cited 883 times)，有別於範例程式碼提供的baseline使用的是VPT (Visual Prompt Tuning, in ECCV, 2022, cited 973 times)，根據Yu et al. 發表的Visual Tuning (in ACM Computing Surveys, 2024, cited 16 times)的分類，VPT為Vision-driven Prompt，而CoCoOp為Language-driven Prompt。**（圖十）**為Visual Tuning解釋這兩種prompt method的差異，紅色為tunable block，藍色為frozen block。在這次期末報告我想嘗試使用不同種類的Visual Tuning，並且觀察其在此任務上的差異，因此使用CoCoOp當作我這次實作few-shot learning的方法。

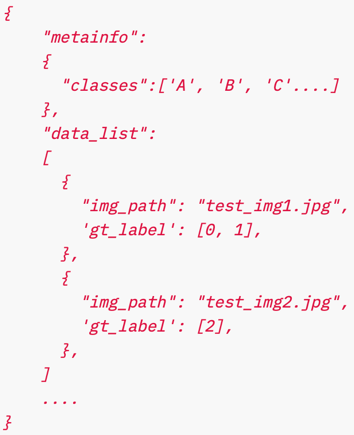


**（圖十）**

* Data Preprocessing

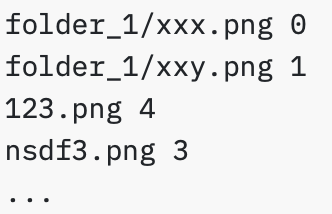
1. *Transfer Learning on 20% of Training Data by Densenet*

在使用mmpretrain方面，由於第一項任務**Chest**與第三項任務**Endo**都是multi-label classification的任務，我因此參考了mmpretrain的文件：<https://mmpretrain.readthedocs.io/en/stable/_modules/mmpretrain/datasets/multi_label.html>，製作出符合mmpretrain需求的annotation files，此文件格式為pickle檔，存取的資料型態是dictionary，以下為範例**（圖十一）**。



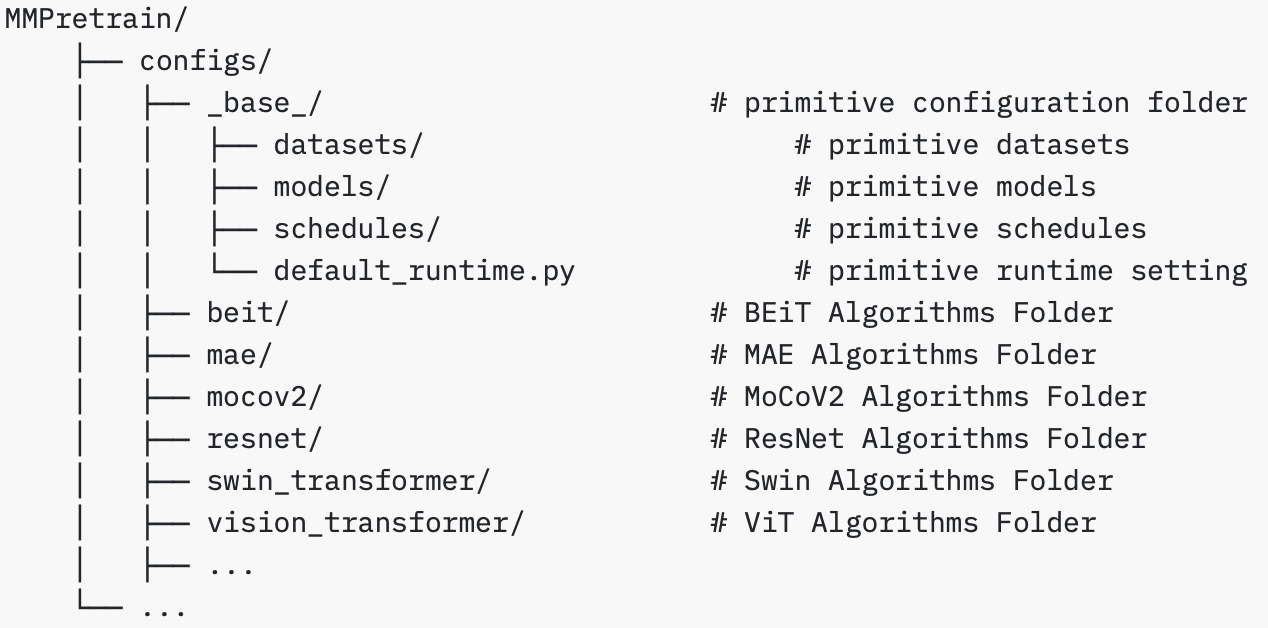
**（圖十一，圖片來源：<https://mmpretrain.readthedocs.io/en/stable/_modules/mmpretrain/datasets/multi_label.html>）**

而第二項任務Colon則是binary classification的任務，因此只需遵照一般的annotation，以下為範例。



**（圖十二，圖片來源：<https://mmpretrain.readthedocs.io/zh-cn/latest/user_guides/dataset_prepare.html>）**

另外，configuration也依照mmpretrain配置文件的教學指示，**（圖十三）**為官方範例。

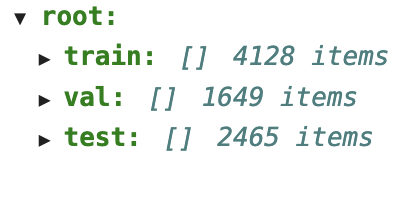


**（圖十三，圖片來源：https://mmpretrain.readthedocs.io/zh-cn/latest/user\_guides/config.html）**

1. *Few-shot learning by CoCoOp*

在CoCoOp方面，我先使用其他範例datasets說明文件：https://github.com/KaiyangZhou/CoOp/blob/main/DATASETS.md，了解我的該如何配置檔案位置，以及annotation files的格式。

在Caltech101這個範例資料中，annotation file: split\_zhou\_Caltech101.json格式為以下範例**（圖十四）**與**（圖十五）**，存取的內容為一個dictionary。



**（圖十四） （圖十五）**

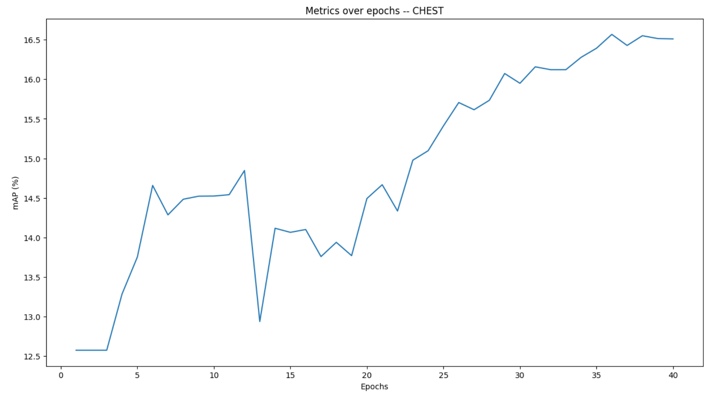
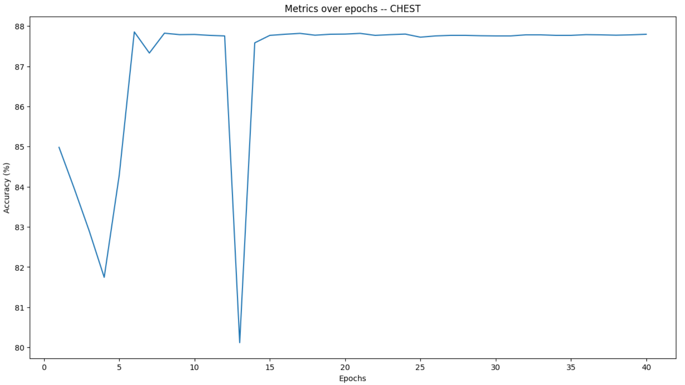
1. **Experiment and Result**

在之前的作業當中，我是自己以80%當作training data，不過在這裡為了配合範例程式碼，我則是使用它已經切好的train, val, test datasets當作訓練以及測試資料。Densenet是以20% data去做transfer learning，而CoCoOp則是以1-shot, 5-shot, 10-shot去做few-shot learning。根據競賽的說明，這裡的few-shot number是以病患的數量決定，而非影像的數量。另外，我使用兩種metrics: mAP & Accuracy去衡量我最後的預測結果。

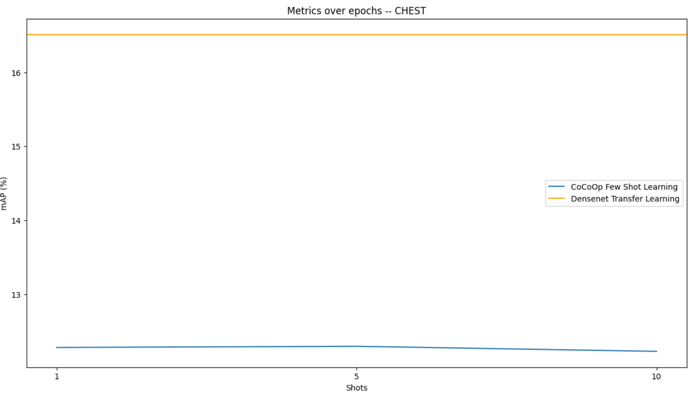
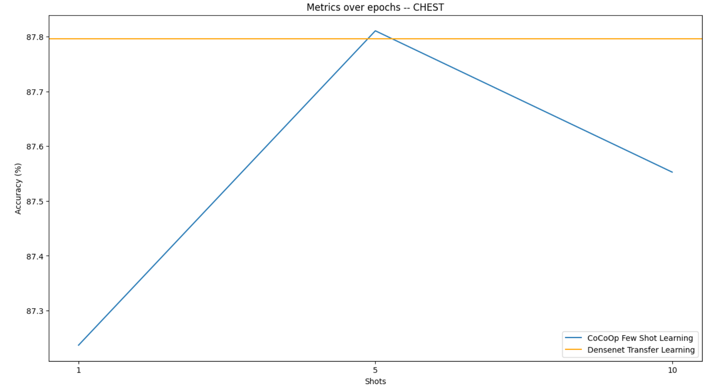
1. 第一項任務：**胸部疾病篩檢 (ChestDR)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number of instances** | **Densenet** | **CoCoOp 1-shot** | **CoCoOp 5-shot** | **CoCoOp 10-shot** |
| Train | 783 | 19 | 95 | 190 |
| Val | 196 | 960 | 884 | 789 |
| Test | 1161 | 1161 | 1161 | 1161 |

從**（圖十六）**與**（圖十七）**可以看到，儘管訓練越多epochs可能無法有效提升Accuracy，不過mAP卻持續在增加，而**（圖十八）**與**（圖十九）**顯示CoCoOp跟Densenet的對比，發現CoCoOp只能在Accuracy方面與Densenet有相當的準確率，不過從mAP來看卻不如Densenet。



**（圖十六） （圖十七）**

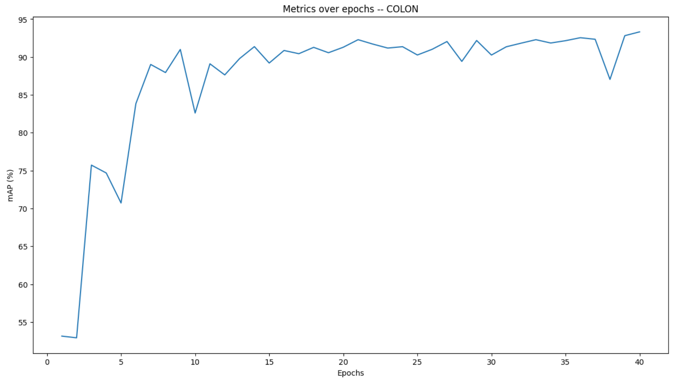
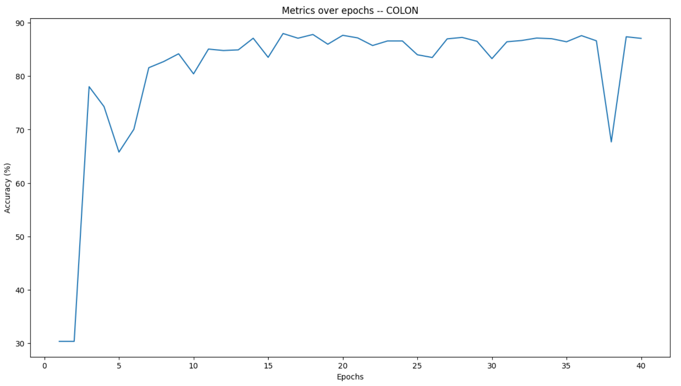


**（圖十八） （圖十九）**

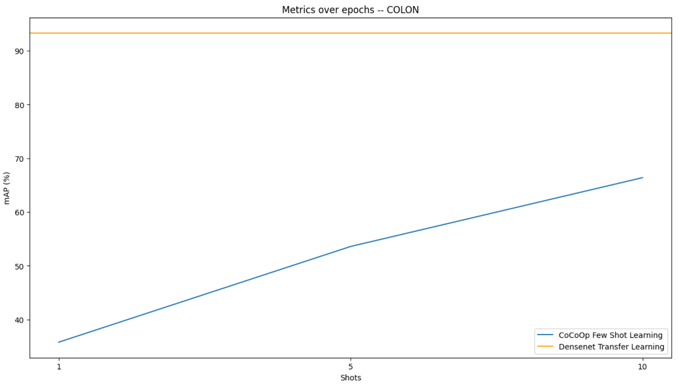
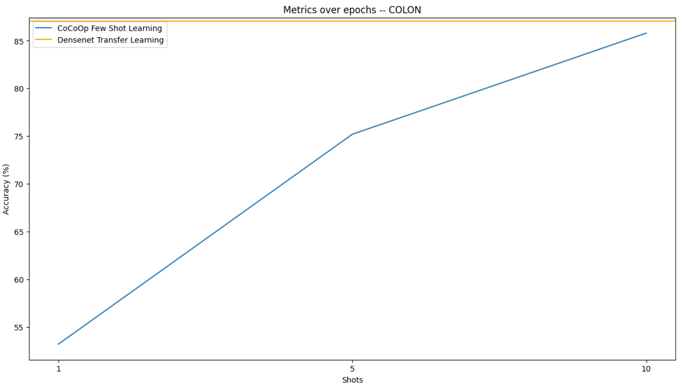
1. 第二項任務：**病理性腫瘤組織分類 (Colon)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number of instances** | **Densenet** | **CoCoOp 1-shot** | **CoCoOp 5-shot** | **CoCoOp 10-shot** |
| Train | 1677 | 30 | 234 | 512 |
| Val | 681 | 2328 | 2124 | 1846 |
| Test | 3296 | 3296 | 3296 | 3296 |

從**（圖二十）**與**（圖二十一）**可以看到，Densenet對於Colon資料集的訓練在早期效果就已經飽和了，而**（圖二十二）**與**（圖二十三）**顯示CoCoOp跟Densenet的對比，發現CoCoOp只能在Accuracy方面與Densenet有相當的準確率，不過從mAP來看卻遠低於Densenet。

****

**（圖二十） （圖二十一）**

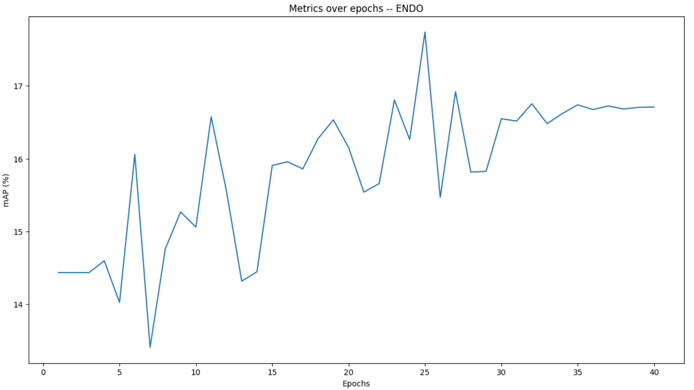
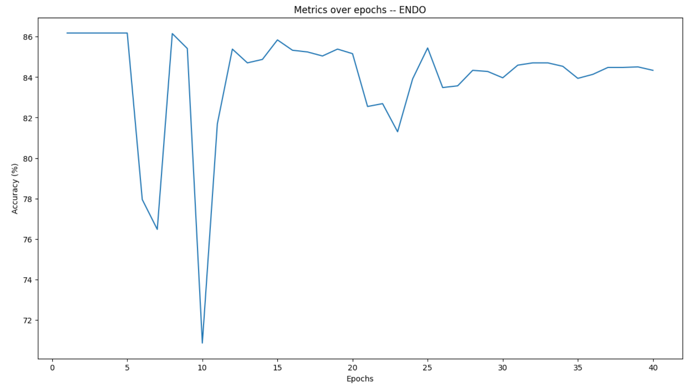
****

**（圖二十二） （圖二十三）**

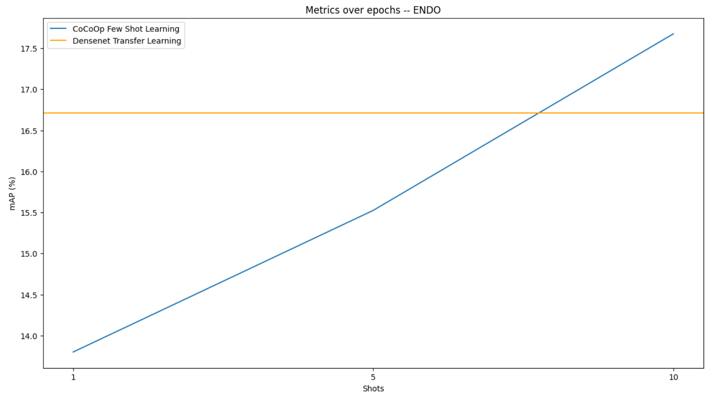
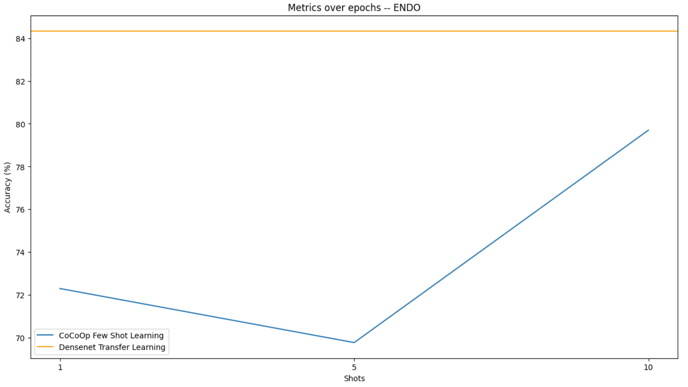
1. 第三項任務：**病變偵測於結腸鏡影像 (Endo)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number of instances** | **Densenet** | **CoCoOp 1-shot** | **CoCoOp 5-shot** | **CoCoOp 10-shot** |
| Train | 743 | 46 | 174 | 422 |
| Val | 186 | 883 | 755 | 507 |
| Test | 881 | 881 | 881 | 881 |

從**（圖二十四）**與**（圖二十五）**可以看到，Densenet對於Endo資料集的訓練在早期的Accuracy就已經飽和了，不過mAP還是有微幅上升，而**（圖二十六）**與**（圖二十七）**顯示CoCoOp跟Densenet的對比，這裡的結果有別於前面Chest與Colon，反而是CoCoOp在mAP方面有機會高於Densenet，不過從Accuracy來看卻不及Densenet。



**（圖二十四） （圖二十五）**

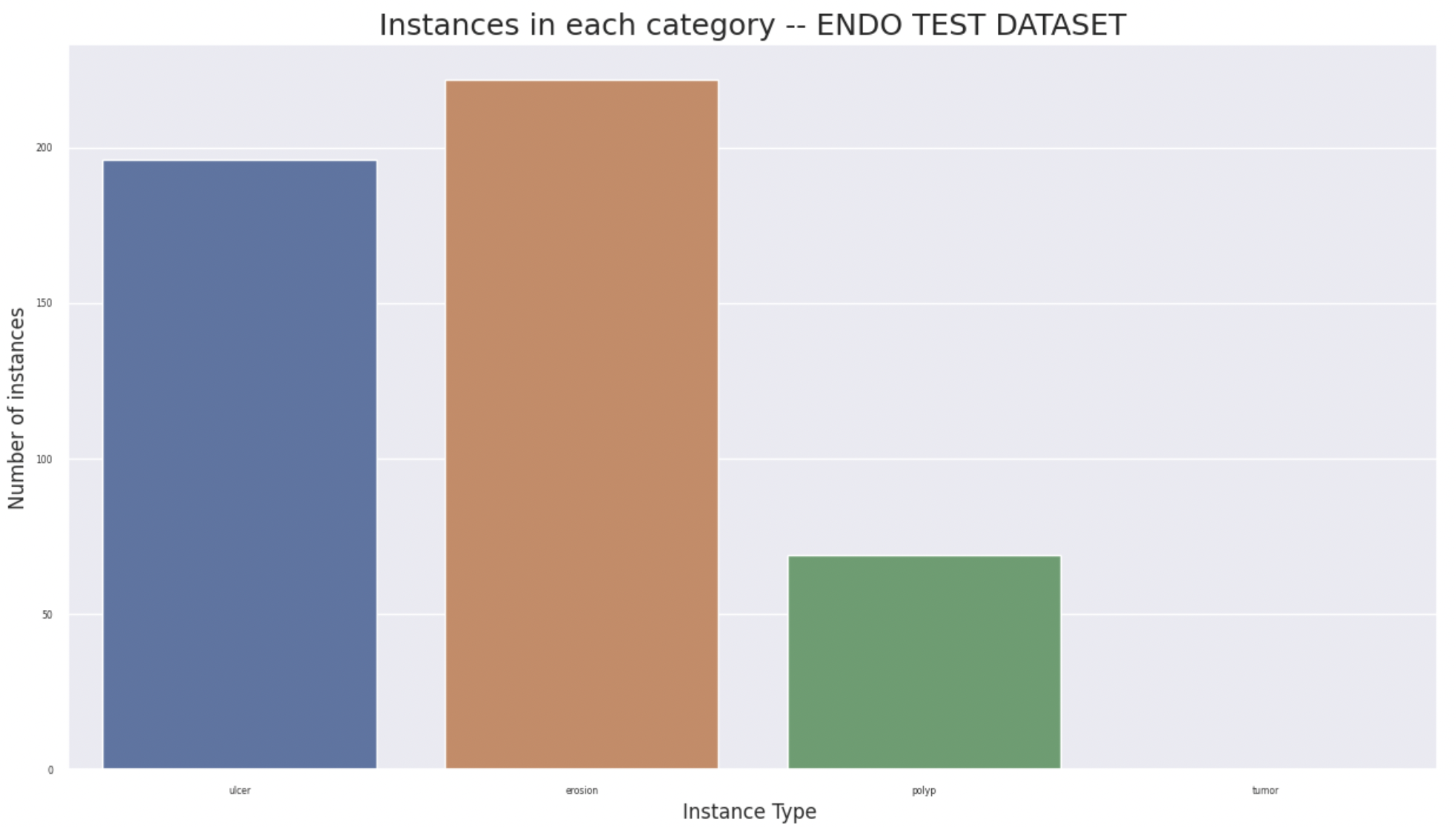


**（圖二十六） （圖二十七）**

1. **Conclusion**

* Discussion: Distribution of labels in Endo test data

我在檢視各類別的mAP時發現，沒有任何資料有被標注tumor，代表不管如何tumor的mAP都會是0，進而導致Endo test data的mAP都很低，也可以從前面的EDA發現tumor的標注數量本來就是最少的，或許範例是希望將少量帶有tumor的圖像放在training以及validation data裡面，去增加模型的效能。



* Future Work
  + 使用DualCoOp以及其他multilabel CoOp

這次挑戰的三個資料集當中就有兩個是屬於multilabel classification，不過其實CoCoOp只能執行binary classification，因此我依照class數量把他們拆分成多個binary classification，這會大大增加運算量以及overfitting的可能行，也會降低它的generalization能力，可能與原挑戰的宗旨背離，因此應該嘗試其他針對multilabel classification的改良，像是DualCoOp以及TAI++。

* + 測試prompt的有效度

在實作CoCoOp時，預設的prompt為「a photo of a {}.」，{}會被替換成圖像所屬的class，不過他們也有針對一些資料集去設計prompt，像是OxfordPets的prompt為「a photo of a {}, a type of pet.」，因此我也針對這三個資料集去設計他們的prompt。

|  |  |
| --- | --- |
| **Dataset** | **Prompt** |
| Chest | A chest X-ray with thoracic disease {}. |
| Colon | A tissue slide of pathological tissue patches with {}. |
| Endo | A colonoscopy image with abnormalities {}. |

不過不能保證這樣的prompt對於預測任務是有幫助的，因此應該先使用預設prompt測試一下結果，再使用新的prompt去驗證這樣設計的prompt是否有比較好。

* + 使用Gradient-weighted Class Activation Mapping等方法視覺化那些在圖像中造成重要影響的部分。
  + 使用experiment tracking tools去追蹤我的實驗過程與結果，例如Neptune AI與Comet ML等等。