

評估息肉分割模型

Evaluation of Polyp Segmentation Models

指導教授：賀保羅

專題成員：郭昱辰

開發工具：Tensorflow

測試環境：Windows10、WSL

一、簡介：

在醫學影像分析中，特別是結腸鏡息肉分割任務中，自動化方法能夠顯著提升息肉檢測的效率和準確性。結腸直腸癌 (CRC) 是全球主要的癌症死亡原因之一，因此早期檢測和移除息肉對於預防和治療 CRC 至關重要。然而，由於息肉的形狀、大小和質地多樣，準確識別和分割息肉是相當具有挑戰性的任務。

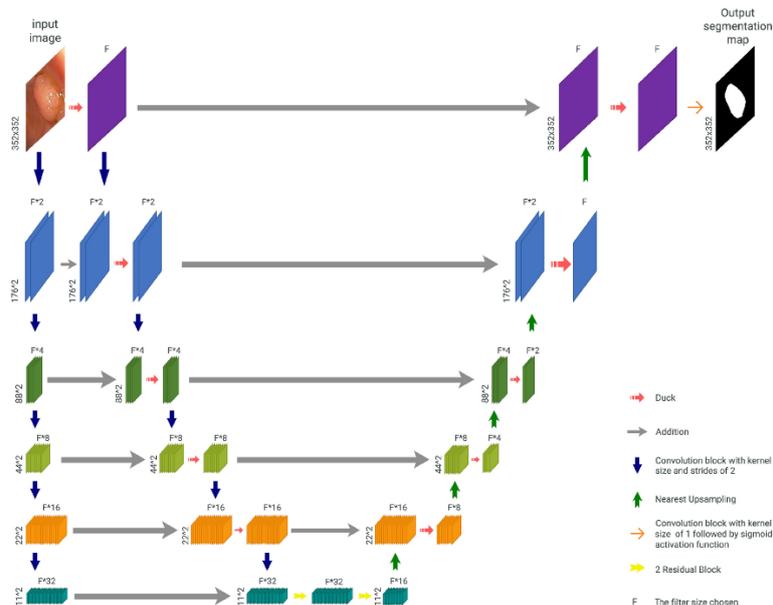
本專題旨在評估現有息肉分割模型在醫學影像資料集上的表現。藉由重現論文中提出的模型結果，我進行了兩個主要測試：

1. 重現論文結果：使用同一個資料集的訓練集進行模型訓練，並測試同一資料集的測試集，分析模型的各項評估指標。

2. 測試泛化能力：使用一個資料集訓練的模型測試另一個資料集，評估模型的泛化能力，並根據生成的評估指標進行分析。

我使用四個常見的息肉分割基準資料集進行評估：Kvasir-SEG、CVC-ClinicDB、CVC-ColonDB 和 ETIS-LARIBPOLYPDB。透過一系列實驗，我們展示了息肉分割模型在這些資料集上的優異表現，包括 Dice Coefficient、Jaccard Index、Precision、Recall 和 Accuracy 等多項指標。

模型架構圖：



二、測試結果：

1. 重現論文結果

根據論文的說明，我各自使用四個息肉分割資料集 (Kvasir-SEG、CVC-ClinicDB、CVC-ColonDB 和 ETIS-LARIBPOLYPDB) 訓練集訓練出的模型來測試同一資料集中的測試集。下面展示使用 Kvasir-SEG 訓練集訓練出的模型測試同資料集中的測試集結果。和其他三個資料集的測試結果一樣，可以展現出此模型優異的能力。

Results for Kvasir-SEG using model saved_model/kvasir-17:

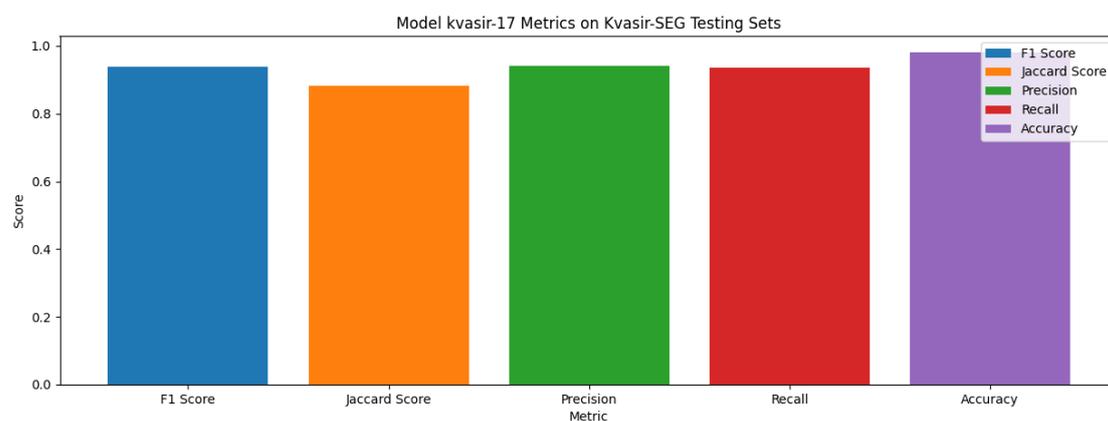
F1 Score: 0.9380582501367194

Jaccard Score: 0.8833424717104206

Precision: 0.9398199834647902

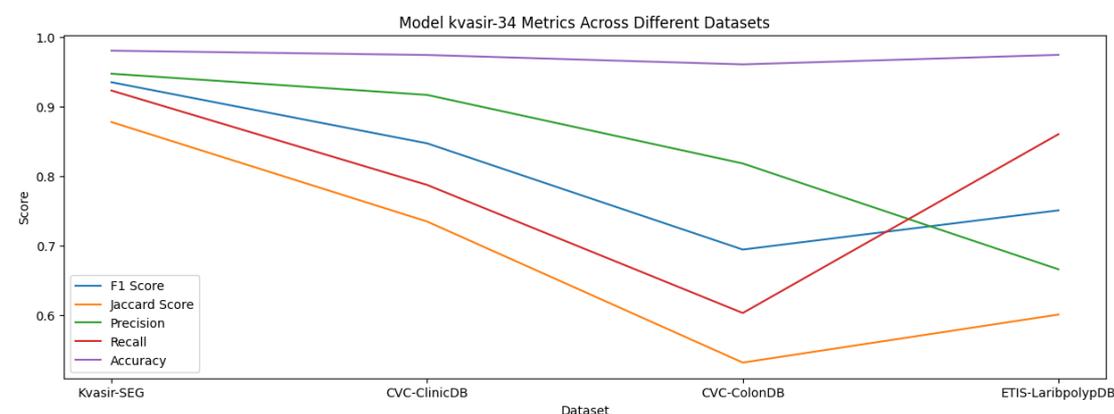
Recall: 0.9363031093417531

Accuracy: 0.9800811111828512



2. 測試泛化能力

我使用一個資料集訓練出來的模型測試其他資料集中的所有圖像，評估模型是否具有泛化能力。下面展示使用 Kvasir-SEG 訓練集訓練出的模型測試其他資料集的結果。這部分模型的能力相較於「1. 重現論文結果」較不強大，但是根據論文的描述，仍然是接近 SOTA 水準。



註: 此頁測試結果由於篇幅原因，只展示其中一部分測試結果。