

Vibe Coding

AI 新秀計畫 - 林泓錕

Outline

- Project Overview(**Intoduction**)
- Multi-Agents **Related Work**
- Multi-Agents Architecture diagram(**Design & Development**)
- Agent Role Instruction & Behavioral Guidelines (R & B)
- Implementation & Results(**Demonstration & Evaluation**)
- **Conclusion** & Future

Project Overview

進行中



AI CUP 2025秋季賽－電腦斷層心臟肌肉影像分割競賽 I－心臟肌肉影像分割 >

9/16/2025

11/30/2025

開始

結束

23.50 萬元(NTD)

TOTAL REWARD

568

TEAMS

距離比賽結束


3 13 49 49

天 時 分 秒

For the full English version of the competition rules, please click here.

(More)

競賽說明

 [For the full English version of the competition rules, please click here.](#)

本競賽旨在進行電腦斷層心臟肌肉影像分割，分割目標有二：全心臟肌肉以及主動脈瓣膜。使用深度學習方法訓練模型，所預測出的分割結果能夠大幅度地減少人工標記所需的人力成本。肥厚性心肌病變透過去除多餘的心肌手術能夠減緩不適症狀。主動脈瓣膜的功能為幫助心臟可以正常的將血液輸往全身的器官。經導管主動脈瓣膜置換手術能夠解決主動脈瓣膜狹窄問題，手術時需利用醫學影像進行瓣膜位置的定位。為此，我們舉辦了電腦斷層心臟肌肉影像分割與主動脈瓣物件偵測競賽。通過高精準度的影像分割技術，在獲得心臟肌肉分割結果後，能夠建立病患的專屬心臟模型，讓醫生進行術前規劃與模擬，避免手術進行中的相關風險。



568
參賽隊伍



總獎金
新臺幣 23.50 萬元

開始 9/16/2025

結束 11/30/2025

Project Overview

評分方式

會依Dice、IoU 評估模型效能：兩種分數的範圍是[0~1]，總分會為2個分數總和的平均。
心臟肌肉分割主要指標

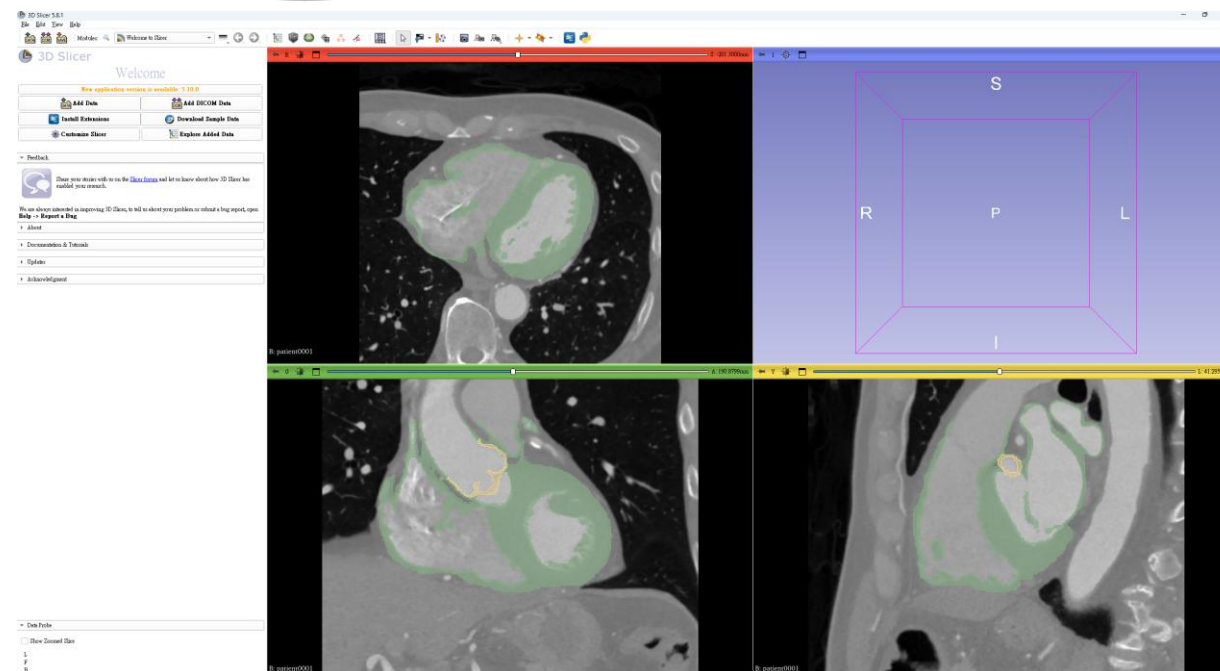
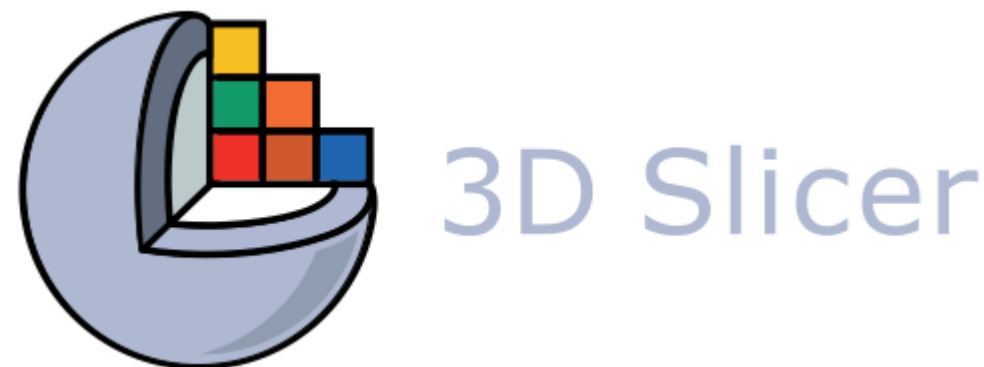
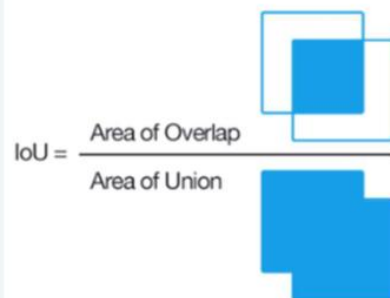
- Dice

$$Dice = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

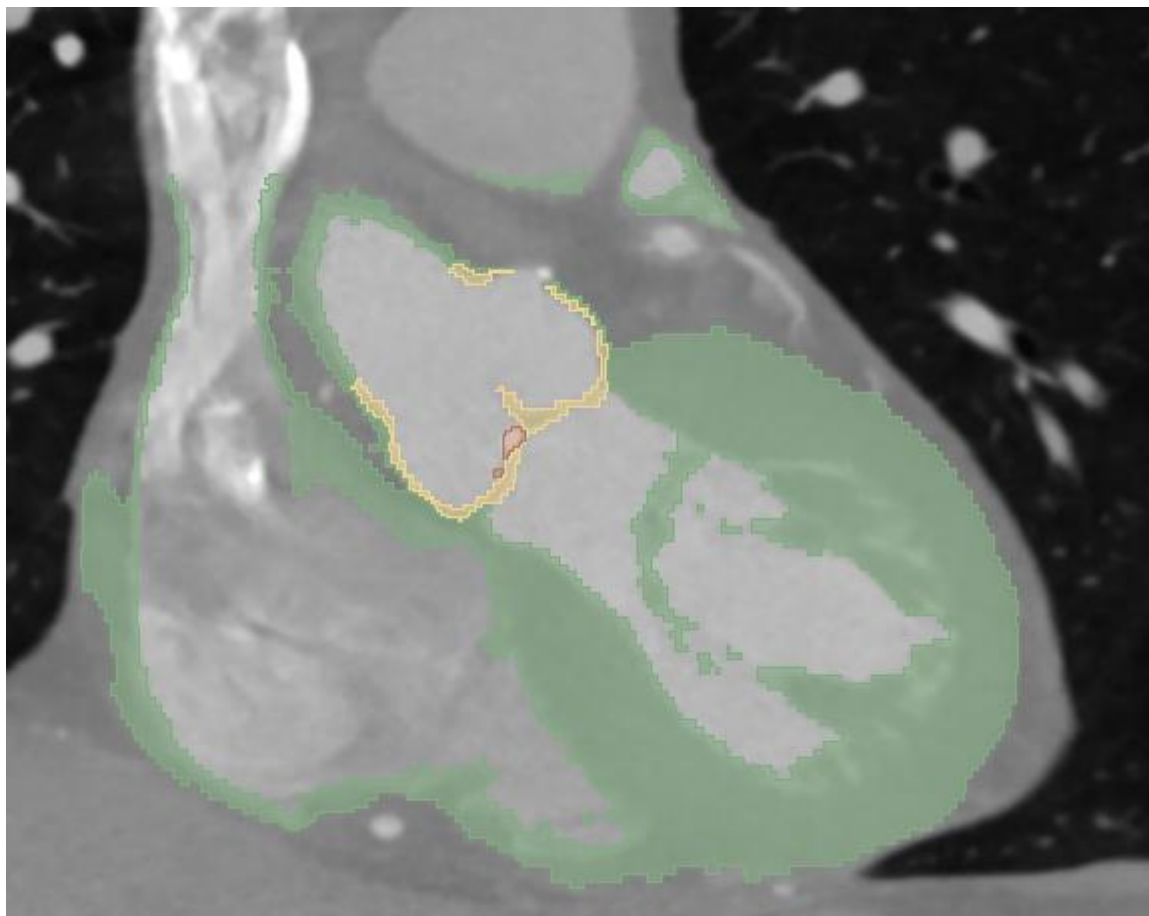


- IoU (Intersection over union)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$



Project Overview



















There are **three** target segments

Green: Myocardium (心肌)

Yellow: Aortic valve (主動脈瓣膜)

Red: Calcification (鈣化)

 patient0001.nii.gz	 patient0001_gt.nii.gz
 patient0002.nii.gz	 patient0002_gt.nii.gz
 patient0003.nii.gz	 patient0003_gt.nii.gz
 patient0004.nii.gz	 patient0004_gt.nii.gz
 patient0005.nii.gz	 patient0005_gt.nii.gz
 patient0006.nii.gz	 patient0006_gt.nii.gz
 patient0007.nii.gz	 patient0007_gt.nii.gz
 patient0008.nii.gz	 patient0008_gt.nii.gz

Training Data & GT * **50** patient

Project Overview

The best result in the current competition.

Public Leaderboard			Private Leaderboard		
#	隊伍名稱	成員	提交次數	分數	上傳時間
1	TEAM_8010	4	129	0.799233	11/26/2025 12:01:11 AM

Multi-Agents Related Work

Multi-Agent Systems: Applications & Deficiencies in AI Competitions

(1) 競賽流程自動化

LLM 作為主控 agent，協調其他子代理：

- 資料預處理 agent
- EDA / 可視化 agent
- 模型訓練 agent
- 超參調整 agent
- Ensemble agent
- 報告與提交檔案生成 agent

這讓整個 AI pipeline 能自動化或半自動化，減少人工調參。

(3) Ensemble / 模型融合

在 AI 競賽中，優勝隊伍幾乎都採用：

- stacking
- weighted blending
- model averaging

這本質上是一種「多代理模型」：每個模型可視為一個 agent，最後由「決策 agent」進行融合。

(2) 平行模型探索

多個 agent 以不同策略探索模型空間，例如：

- 不同架構 (CNN, Transformer, Gradient Boosting...)
- 不同超參 (learning rate, depth, dropout...)
- 不同資料切分策略

這種「平行搜尋」大幅縮短競賽的模型迭代時間。

(4) Multi-Agent Reinforcement Learning (MARL)

在策略、操控、遊戲型競賽中 (如 NeurIPS MARL Challenge)：

- agents 學習合作或競爭策略
- 支援共享訊息的通信機制
- 使用 centralized training + decentralized execution 架構

Multi-Agents Related Work

Multi-Agent Systems: Applications & Deficiencies in AI Competitions

(1) 可靠性不足 (Hallucination / 不穩定規劃)

LLM 作為主控 agent 時，常出現：

- 任務規劃不一致
- 推理錯誤
- 工具使用順序錯誤
- 生成程式碼無法執行

導致整體 pipeline 崩潰。

(3) 多代理的品質不均

不同 agent：

- 能力差異極大
- 負責任務不均衡
- 有時造成 bottleneck 而非加速

(2) 缺乏強健的 Agent 協作協定

目前 agent 之間的溝通常是：

- 以自然語言互傳訊息
- 缺乏正式的溝通語法
- 缺少一致性驗證

導致 agent 合作容易「跑偏」。

(4) 可重現性 (Reproducibility) 不足

LLM agent 的行為具有隨機性：

- 輸出內容不固定
- 即便同樣 prompt 也可能做不同事情
- 難以保證競賽結果一致

這是學術界目前最大的痛點。

(5) 對競賽任務仍缺乏深層理解

LLM 雖能做 EDA/訓練 code，但：

- 不真正「理解」資料的統計性質
- 不具備 domain knowledge
- 對 overfitting / leakage / metric nuances 的掌握不如人類高手

導致在複雜競賽中通常無法達到前 1% 的成績。

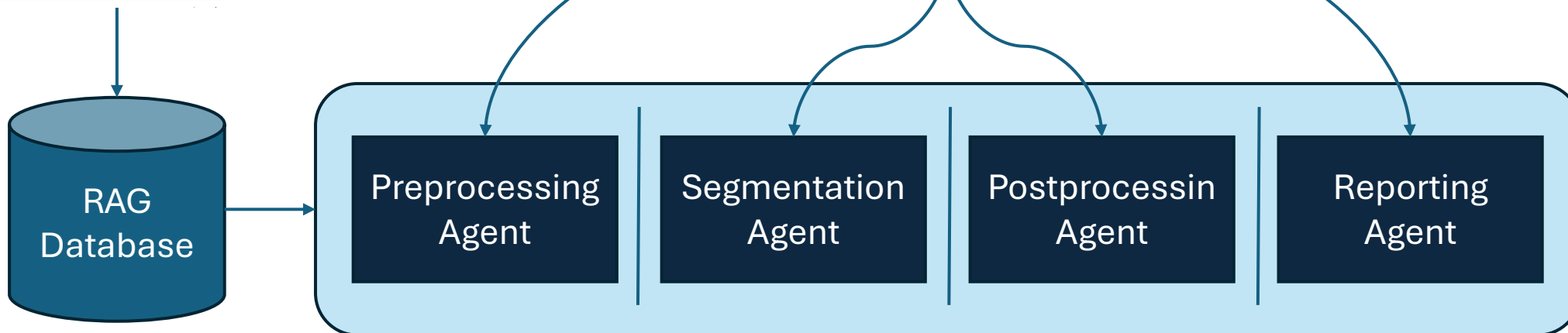
Multi-Agents Architecture diagram

My environment info : windows 、 3080GPU 、 12700K CPU ...

Competition info

Related research info

Like:



提交結果
上傳格式

1. 檔案上傳 *

選擇檔案

Choose a zip file...

2. 檔案說明 *

備註用，方便之後上傳檔案紀錄查詢

Submit

- Train & Test
- Use the Reporting Agent Fit Upload Format
- Fix/ Adjustment/ Integrate code

Agent R & B

Preprocessing Agent

1. Preprocessing Agent (預處理代理)

角色指令: 負責將原始CT影像(.nii.gz格式)轉換為模型可處理的標準化數據格式,確保數據品質和一致性。

行為準則:

- 自動驗證輸入檔案格式是否為.nii.gz,若格式錯誤立即回報
- 執行影像正規化(normalization)至標準灰階範圍
- 進行影像重採樣(resampling)以統一空間解析度
- 檢測並處理異常值或損壞的影像切片
- 應用資料增強技術(如旋轉、翻轉、縮放)以擴充訓練集
- 記錄每個處理步驟的參數和統計資訊
- 若處理失敗,提供具體的錯誤診斷報告
- 將處理後的資料批次化並傳遞給Segmentation Agent

Agent R & B

Segmentation Agent

2. Segmentation Agent (分割代理)

角色指令: 使用3D U-Net模型對心臟CT影像進行精準分割,識別心臟肌肉(Segment_1)、主動脈瓣膜(Segment_2)和鈣化區域(Segment_3)。

行為準則:

- 載入預訓練權重或從頭開始訓練模型
- 根據AICUP_training.json配置進行訓練/驗證/測試集劃分
- 監控訓練過程中的loss和accuracy指標
- 自動保存最佳模型權重checkpoint
- 對patient0051~patient0100執行推論任務
- 確保輸出分割結果符合三類標籤格式(綠色/黃色/紅色)
- 若GPU記憶體不足,自動調整batch size或使用混合精度訓練
- 向RAG Database查詢類似案例以優化分割策略
- 將分割結果傳遞給Postprocessing Agent進行後處理

Agent R & B

3. Postprocessing Agent (後處理代理)

角色指令: 優化分割結果的品質,修正錯誤預測,並將輸出格式化為競賽要求的.nii.gz檔案。

Postprocessin
Agent

行為準則:

- 應用形態學操作(開運算/閉運算)去除雜訊和小區域
- 執行連通域分析,保留最大連通區域
- 平滑邊界以提升視覺品質和解剖合理性
- 驗證三個segment的合理性(如心肌是否連續、瓣膜位置是否正確)
- 檢查是否存在標籤重疊或空洞問題
- 將處理結果儲存為patient0051_predict.nii.gz ~ patient0100_predict.nii.gz
- 若發現異常結果(如極端大小或形狀),標記並建議人工審查
- 通知Reporting Agent生成品質報告

Agent R & B

4. Reporting Agent (報告代理)

角色指令: 生成綜合評估報告,確保提交檔案符合競賽格式,並提供改進建議。

Reporting
Agent

行為準則:

- 驗證所有50個預測檔案(patient0051~0100)是否完整
- 檢查檔案命名規範和.nii.gz格式正確性
- 自動合併所有預測結果並壓縮為單一.zip檔案
- 計算基礎統計資訊(如每個segment的平均體積)
- 與訓練集ground truth進行相似度比對(若可用)
- 生成視覺化報告(可選用3D Slicer格式)
- 標記潛在問題案例並提供修正建議
- 確認提交檔案符合Tbrain平台要求
- 追蹤歷史提交成績,分析改進方向
- 若檢測到格式錯誤,自動回饋給相應Agent進行修正

Implementation & Results

- ◆ code : 存放程式碼
- ◆ data_split : 存放資料切割json檔
- ◆ experiments : 存放訓練權重、log檔案
- ◆ predictions : 存放預測結果
- ◆ test : 存放測試資料
- ◆ train : 存放訓練資料

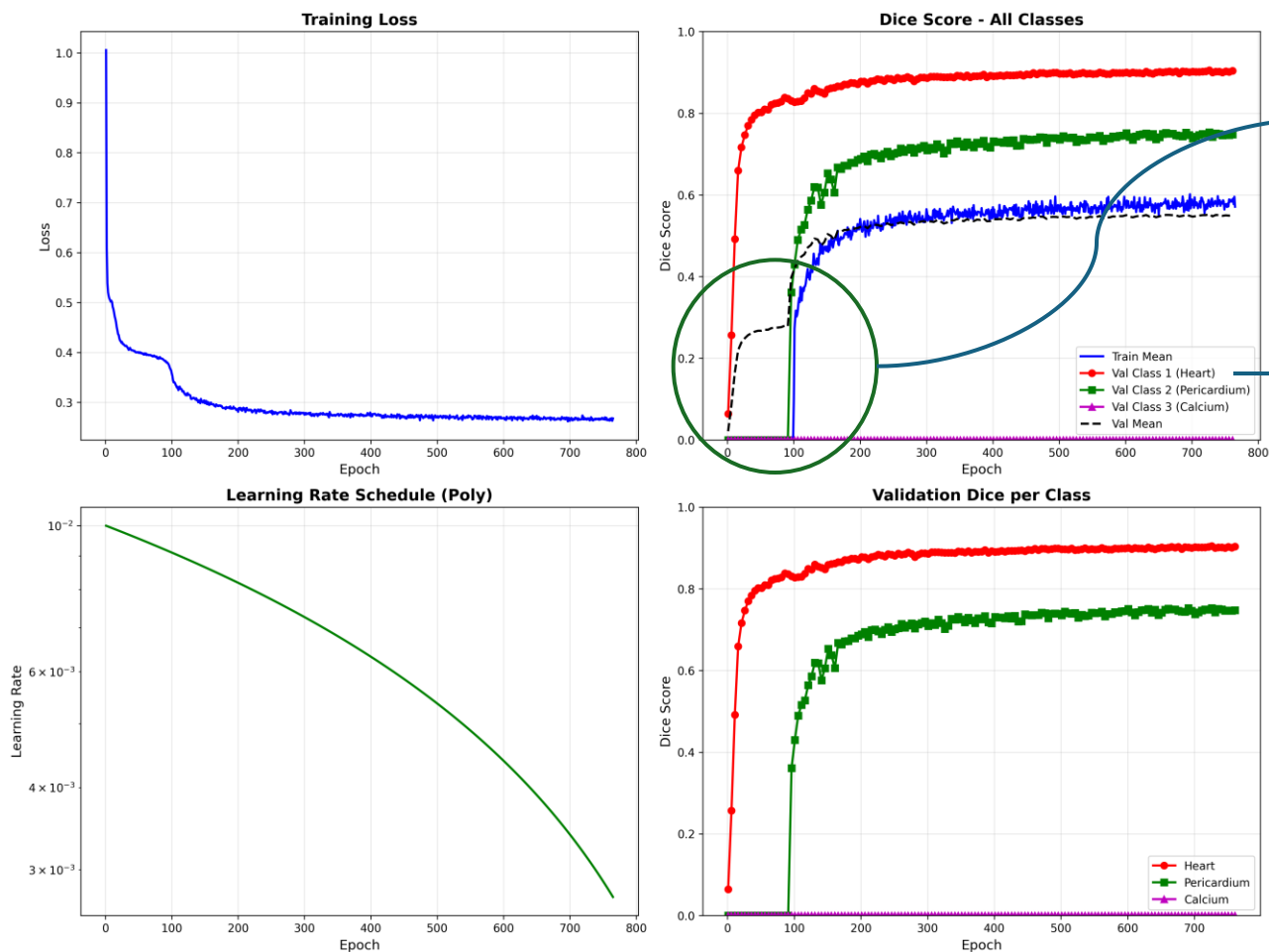
code	2025/11/25 下午 10:58	檔案資料夾
data_split	2025/11/20 下午 07:39	檔案資料夾
experiments	2025/11/24 下午 11:00	檔案資料夾
predictions	2025/11/25 下午 11:20	檔案資料夾
test	2025/11/25 下午 10:00	檔案資料夾
train	2025/11/20 下午 07:12	檔案資料夾

```
AICUP_training.json X
D: > aicup > data_split > {} AICUP_training.json > {} training > {} 4 > abc label

1  {
2    "training": [
3      {
4        "image": "D:/aicup/train/41_training_image/patient0001.nii.gz",
5        "label": "D:/aicup/train/41_training_label/patient0001_gt.nii.gz"
6      },
7      {
8        "image": "D:/aicup/train/41_training_image/patient0002.nii.gz",
9        "label": "D:/aicup/train/41_training_label/patient0002_gt.nii.gz"
10     },
11     {
12       "image": "D:/aicup/train/41_training_image/patient0003.nii.gz",
13       "label": "D:/aicup/train/41_training_label/patient0003_gt.nii.gz"
14     },
15     {
16       "image": "D:/aicup/train/41_training_image/patient0004.nii.gz",
17       "label": "D:/aicup/train/41_training_label/patient0004_gt.nii.gz"
18     },
19     {
20       "image": "D:/aicup/train/41_training_image/patient0005.nii.gz",
21       "label": "D:/aicup/train/41_training_label/patient0005_gt.nii.gz"
22     },
23     {
24       "image": "D:/aicup/train/41_training_image/patient0006.nii.gz",
25       "label": "D:/aicup/train/41_training_label/patient0006_gt.nii.gz"
26     },
27   ]
28 }
```

Implementation & Results

◆ experiments



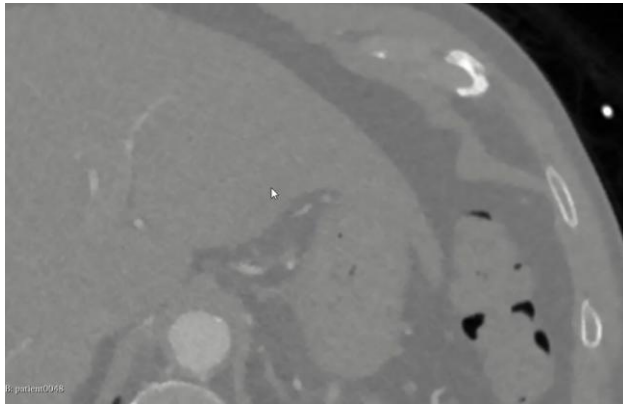
Training Dice 與 Val Dice 計算差異
Dice func 正確但是 Training Dice在整合class123時的機制有誤，但後續還是可以正常呈現

英文錯誤
class1: 心肌
class2: 主動脈瓣膜
class3: 鈣化區域

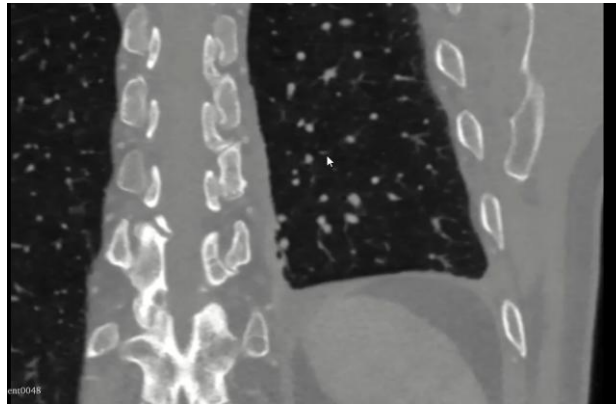
Implementation & Results (video)

- Case: 48- **Myocardium**
- **Blue**: Model Predict
- **Green**: GT

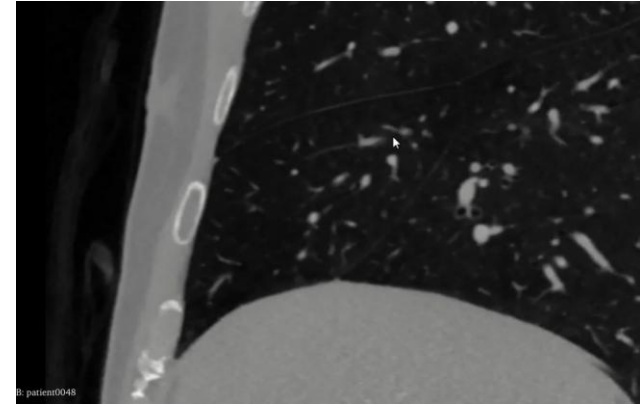
Axis: A



Axis: B



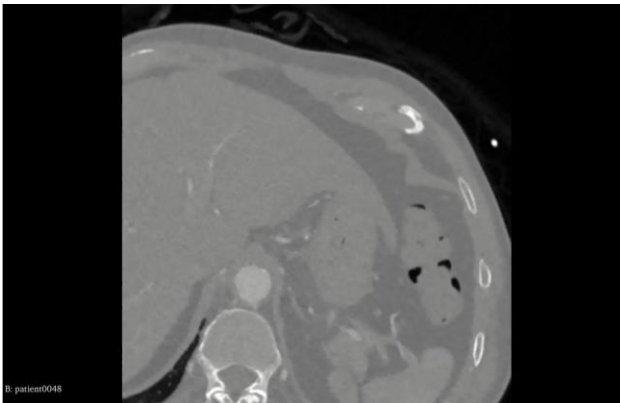
Axis: C



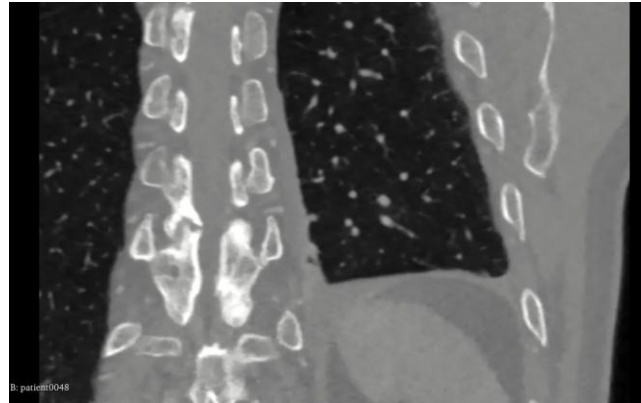
Implementation & Results (video)

- Case: 48- **Aortic valve**
- **Blue**: Model Predict
- **Green**: GT

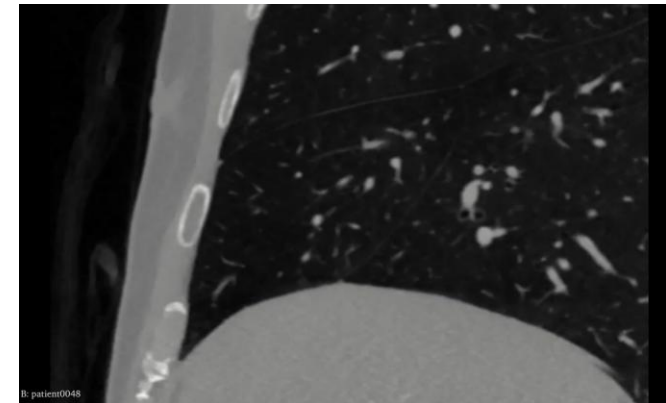
Axis: A



Axis: B

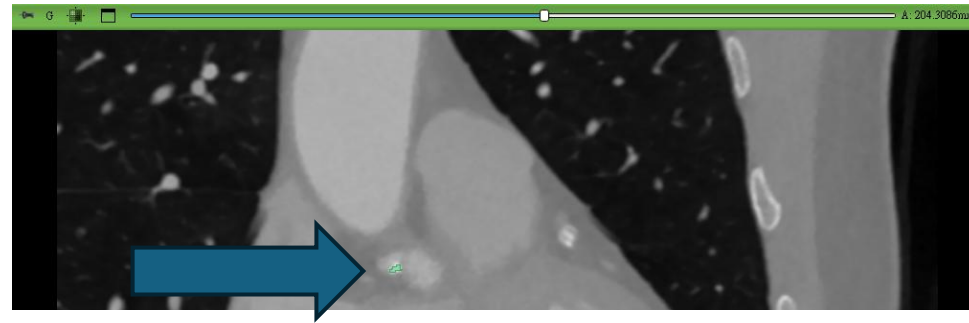


Axis: C

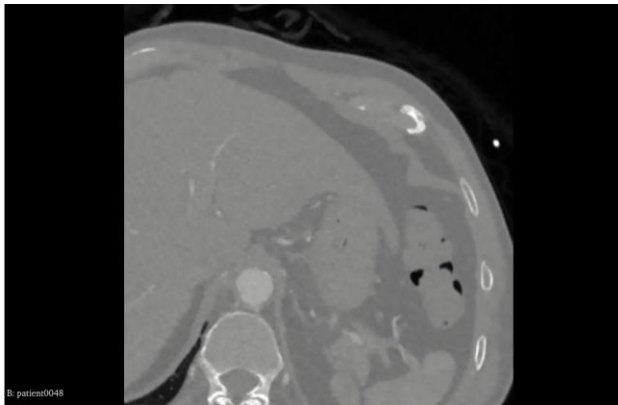


Implementation & Results (video)

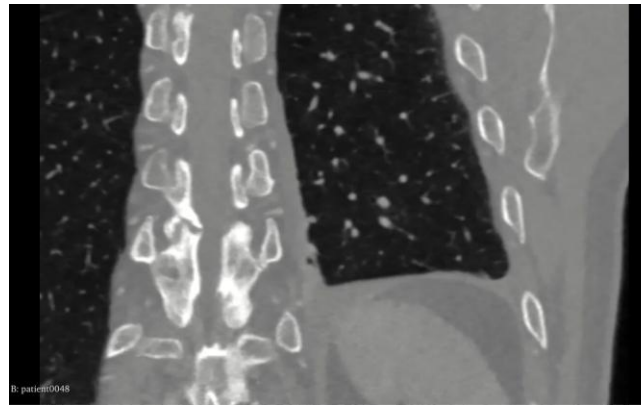
- Case: 48- **Calcification**
- **Blue**: Model Predict
- **Green**: GT



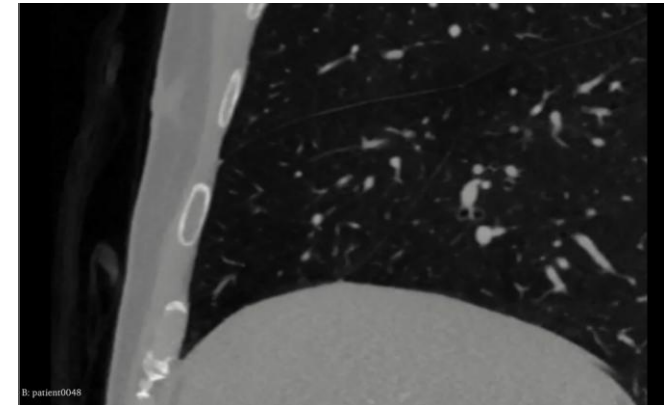
Axis: A



Axis: B



Axis: C



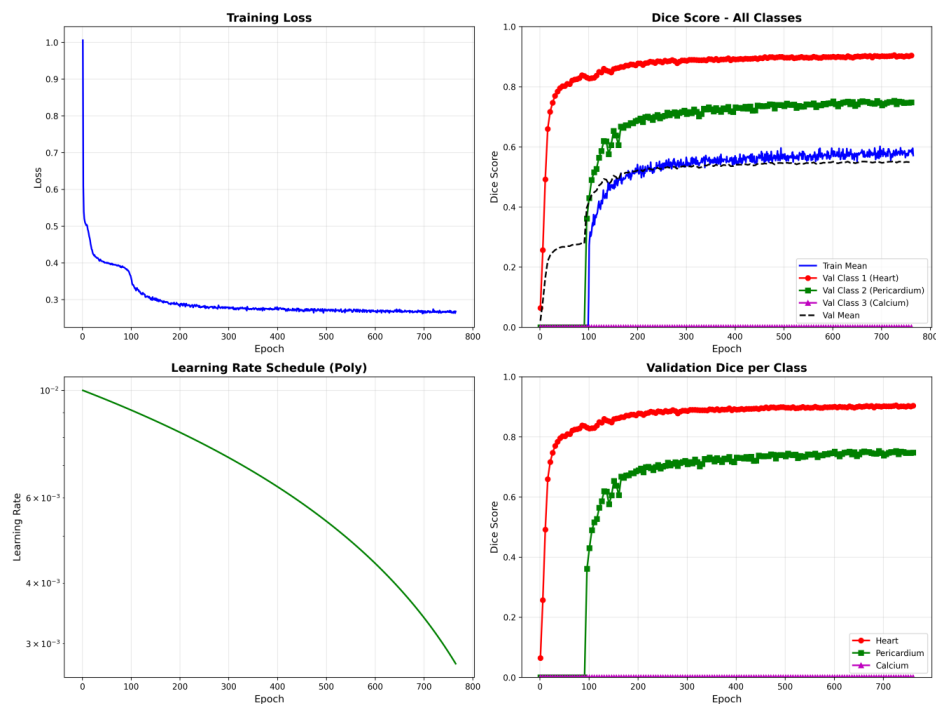
Implementation & Results

Best Val Results:

Class 1 (Heart): 0.9051

Class 2 (Pericardium): 0.7530

Class 3 (Calcium): 0.0000



Best Score in competition:

檔案說明	上傳時間	Public分數	Private分數	訊息
submission_merged.zip post merge 上傳成員 林 泓鋸	2025-11-25 11:14:31	0.70907021		Scoring success.

評分方式

會依Dice、IoU 評估模型效能：兩種分數的範圍是[0~1]，總分會為2個分數總和的平均。
心臟肌肉分割主要指標

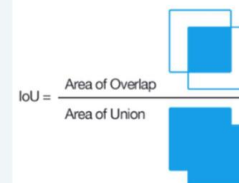
- Dice

$$Dice = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$



- IoU (Intersection over union)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

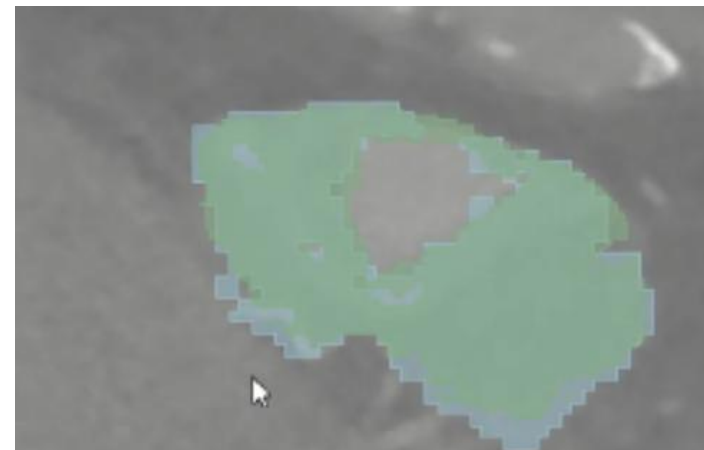


Discussion

Blue: Model Predict
Green: GT

心肌: 藍色 segmentation 看起來邊界粗糙、破碎、局部凹凸明顯。
可能是:

- 邊界辨識困難
- 對 myocardium 外形學習不足
- 對不同 slice 的連續性不穩



後續可行方法:

1. 增加 data augmentation

- random intensity shift
- random contrast / brightness
- elastic deformation
- small rotation/translation

讓模型學到心肌外形不會因影像 noise 改變

2. 多階段模型

- Stage 1: 找心臟 ROI
- Stage 2: 在 ROI 裡做精細 segmentation
降低模型 segmentation 的空間錯誤

3. 後處理

- Conditional Random Field (CRF) refinement
能讓 segmentation 邊界貼近真實器官輪廓。

4. 修改模型 or 超參數

- 使用其他模型預測並進行 bagging
- 使用不同的 Loss function

Discussion

Blue: Model Predict

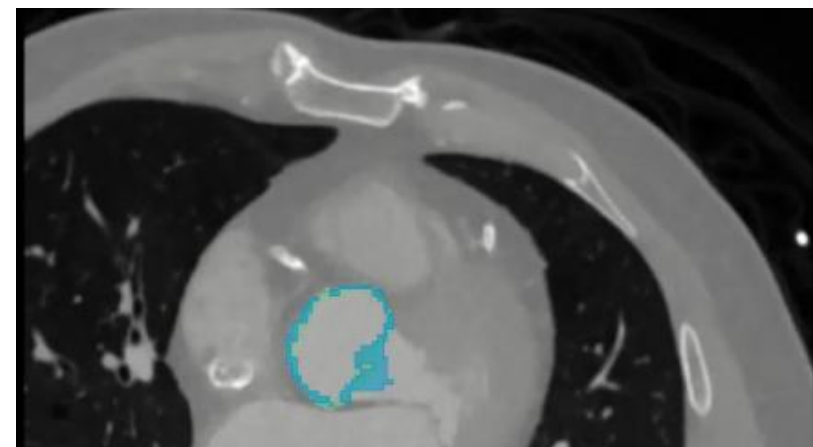
Green: GT

Aortic valve : 藍色 segmentation 看起來邊界粗糙，破碎
可能是：

- 主動脈瓣根部（aortic valve / aortic root）的邊界在 CT 上非常弱，尤其是：

- 軟組織密度變化不明顯
- partial volume effect（部分體積效應）
- 金屬支架、鈣化造成 noise

造成對 segmentation 模型來說，邊界不明確導致預測片段化



後續可行方法：

1. 預處理

- 用心臟/血管適合的 HU 範圍如 $[-300, 700]$ or min-max normalize or z-score

2. 後處理

- 取最大連通
- closing
- remove small objects

HU = Hounsfield Unit（海氏單位）

是 CT（電腦斷層）專用的密度單位，用來表示不同物質對 X 光衰減的程度。

★ 常見物質的 HU

物質	HU
空氣	-1000
肺部	-900 ~ -600
脂肪	-150 ~ -50
水	0
血液 / 軟組織	+20 ~ +60
主動脈血流（對比劑）	+150 ~ +300
肌肉	+20 ~ +80
骨頭（鬆質骨）	+200 ~ +700
骨頭（皮質骨）	+700 ~ +1500
鈣化	+1000 ~ +2000
金屬（支架、牙套等）	+2000 以上

Blue: Model Predict
Green: GT

Discussion

Calcification :從Class3 dice曲線可看出模型沒有預測出鈣化區域可能是:

- 目標過小
- 缺少樣本(僅有少數case都有鈣化)
- GT標註不良

後續可行方法:

1. 調整類別權重比例

- 賦予class3較大權重使模型趨向找出class3

2. 獨立訓練

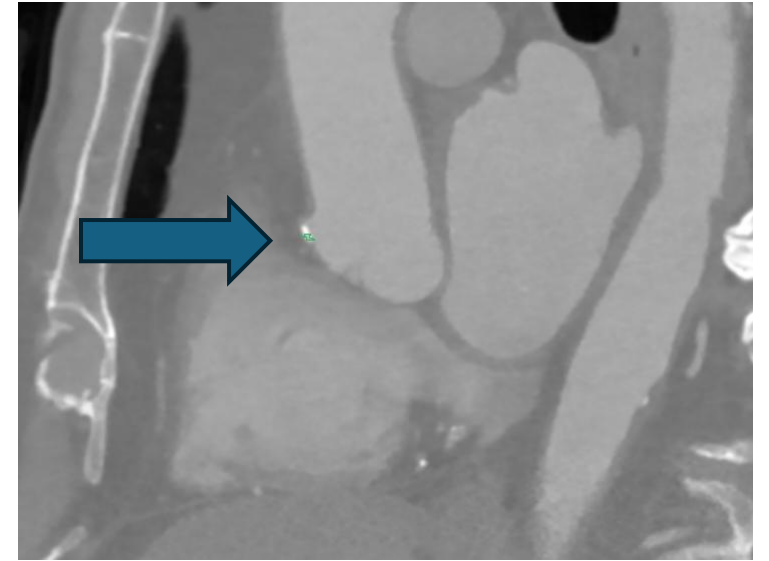
- 由於不是每筆case都有鈣化，嘗試只使用有鈣化的case進行訓練，針對每筆Case再移除沒有鈣化的slice

3. 增加資料

- 由於比賽規定不能自己Gan資料，可找有沒有公開資料，或是做資料增強

4. 使用預訓練權重

- 尋找他人提供的預訓練權重進行訓練



Conclusion & Future

本專案成功設計並實現了一個多代理協作系統,用於自動化處理 AI CUP 2025 心臟 CT 影像分割競賽的完整工作流程。

系統架構包含四個專門化的 AI Agents,各自負責資料預處理、nnU-Net 分割、後處理最佳化,實現了從原始 CT 影像到最終提交檔案的端到端半自動化處理。

主要成果

1. 模組化設計

- 將複雜的醫學影像分割
- 每個 Agent 具有明確的職任務分解為四個獨立但協作的代理模組責邊界和標準化的輸入輸出介面
- 支援單一模組的獨立測試和迭代優化

2. 智能化處理流程

- Preprocessing Agent 自動執行影像正規化、重採樣和資料增強
- Segmentation Agent 利用 nnU-Net 精準識別心肌、瓣膜和鈣化區域
- Postprocessing Agent 透過形態學操作提升分割品質
- Reporting Agent 確保輸出符合競賽格式要求

3. RAG 增強決策

- 整合 RAG Database 儲存競賽規則、研究論文和最佳實踐
- Agents 可查詢歷史案例和領域知識以優化處理策略
- 支援知識庫的持續更新和擴充

4. 人機協作機制

- 保留使用者在關鍵節點的介入能力
- 異常情況自動通知並提供詳細診斷
- 支援手動調整訓練參數和程式碼

Conclusion & Future

打造一個通用、智能、可信賴的多代理醫學影像分析系統,不僅在 AI CUP 競賽中取得優異成績,更能推動 AI 技術在臨床實踐中的落地應用,最終實現提升診斷效率、減輕醫師負擔、改善患者預後的社會價值。

未來規劃

1. 增強 Segmentation Agent 的模型能力

- 整合 SwinUNETR、nnU-Net 等先進架構進行模型集成
- 實現多模型投票機制提升分割準確度
- 引入不確定性估計(Uncertainty Estimation)標記低信心預測

2. 優化 Postprocessing Agent

- 加入基於解剖學知識的約束規則(如心肌厚度合理範圍)
- 實現跨切片的 3D 一致性檢查
- 開發自動品質評分系統(QA Score)

3. 提升 Reporting Agent 的可視化能力

- 自動生成 3D Slicer 格式的視覺化報告
- 提供與 Ground Truth 的逐案例對比分析
- 整合競賽排行榜追蹤和趨勢分析

4. RAG Database 擴充

- 爬取並索引最新的醫學影像分割論文
- 儲存每次提交的結果和反饋形成經驗庫
- 建立常見錯誤案例和解決方案知識圖譜