Title&Team member

Title：電腦斷層心臟肌肉影像分割競賽I：心臟肌肉影像分割

Team member :

-

Abstract

This work deals with the AI CUP 2025 Fall Competition – Computed Tomography Myocardium Image Segmentation, where aortic valve and cardiac muscle segmentation is targeted over CT images. The motivation is due to the clinical need of accurate myocardial and valve localization for preoperative planning, especially in treatments like hypertrophic cardiomyopathy surgery, and Transcatheter Aortic Valve Replacement (TAVR).
We intend to use deep learning–based 3D U-Net models augmented by advanced pre-processing (e.g. normalization, augmentation) and the fine-tune techniques on the provided dataset (patient001-patient050 for training, patient051-patient100 for testing). We anticipate that the proposed framework can contribute to better segmentation accuracy, decreased workload for manual annotation and patient-specific 3D heart models for assisting clinician in surgical simulation and decision-making. We will compute our performance using the Dice coefficient and Intersection over Union (IoU), consistent with how the competition is evaluated.

Introduction

1. .Motivation

Cardiovascular diseases remain one of the leading causes of death worldwide, and accurate image analysis is critical for diagnosis, treatment planning, and prognosis. Cardiac computed tomography (CT) provides detailed anatomical information, but manual segmentation of myocardial structures is time-consuming, labor-intensive, and subject to inter-observer variability. Although many deep learning–based methods for medical image segmentation, such as U-Net and its subsequent variants, have achieved strong performance in research settings, challenges remain. **Data scarcity, limited generalization ability, and imprecise boundary delineation** continue to hinder direct clinical adoption. These limitations highlight the need for further optimization and innovation in segmentation techniques.

Accordingly, the motivation of this study is threefold:

- **Responding to clinical needs**: To improve the efficiency of cardiac image segmentation through automation, thereby reducing the workload of medical professionals.
- **Addressing technical limitations**: To enhance the baseline 3D U-Net architecture on the competition's real CT dataset, aiming to overcome existing shortcomings.
- **Learning and practical value**: To apply classroom knowledge of artificial intelligence to a real-world challenge through participation in AI CUP 2025, and to cultivate research skills in medical image analysis.

2. Problem Definition

This study focuses on **automatic segmentation of cardiac CT images**, which can be formulated as a semantic segmentation task. The objective is to classify each voxel in three-dimensional CT scans into categories such as myocardium, aortic valve, and calcification.
Input: Three-dimensional cardiac CT scans in compressed NIfTI format (.nii.gz).
Output: The corresponding segmentation mask that accurately labels the myocardium, aortic valve, and calcification regions.
The key challenges include:

- **Anatomical variability**: Differences in cardiac shape and size across patients hinder model generalization.
- **Fuzzy boundaries**: Similar grayscale values between adjacent tissues make boundary delineation difficult.
- **Limited annotated data**:Small datasets and potential inconsistency in ground truth.

Addressing these challenges requires robust deep learning architectures, effective data augmentation, and post-processing strategies.

3. Related Work

Deep learning has become the dominant approach in medical image segmentation. The most influential contributions include:

- **U-Net**: A convolutional encoder–decoder network that introduced skip connections, widely applied in medical imaging tasks (Ronneberger et al., 2015).
- **3D U-Net**: An extension designed for volumetric data, enabling spatial feature extraction across three dimensions (Çiçek et al., 2016). (This architecture also serves as the baseline for the AI CUP 2025 competition dataset.)
- **Attention-based models**: Extensions such as Attention U-Net (Oktay et al., 2018) and transformer-based methods (Chen et al., 2021) improve boundary accuracy by focusing on relevant regions.
- **Post-processing methods**: Conditional Random Fields (CRF) and morphological operations have been used to refine segmentation outputs and correct small isolated errors.

4. Proposed Method

In this project, we propose an improved pipeline based on **3D U-Net**, combined with additional enhancements:

- **Data augmentation**: Applying random rotations, flips, translations, and intensity variations to increase diversity and robustness of training samples.
- **Architectural improvements**: Incorporating **attention mechanisms** and **multi-scale feature extraction** to enhance focus on cardiac boundaries and small structures.
- **Post-processing**: Using morphological filtering and CRF-based refinement to improve mask continuity and reduce noise.
- **Evaluation metrics**: Employing Dice coefficient and Intersection-over-Union (IoU) to quantitatively assess segmentation performance.

Through these approaches, we expect to improve segmentation accuracy compared to the baseline model, and contribute a practical framework for clinical cardiac CT analysis.

Method

1. 前處理 (Pre-processing)

我們的輸入為 **CT 影像 (Computed Tomography, CT)**，格式為 **.nii.gz (NIfTI)**。訓練集包含 patient001–patient050 的 CT 與 label ，測試集則為 patient051–patient100 的原始資料。由於不同病人影像分佈差異，我們進行 **強度正規化 (Intensity Normalization)** 以統一 **灰階值 (Gray Intensity Values)**，並搭配 **資料增強 (Data Augmentation)**，包含 **旋轉 (Rotation)**、**翻轉 (Flipping)**、**裁切 (Cropping)** 與 **高斯雜訊 (Gaussian Noise)**，以提升模型的泛化能力。

2. 模型設計 (Model Architecture)

本研究採用 **3D U-Net (Three-Dimensional U-Net)** 架構，由 **編碼器 (Encoder)** 與 **解碼器 (Decoder)** 組成，並透過 **跳躍連結 (Skip Connections)** 保留特徵資訊。輸出為 **多類別分割 (Multi-class Segmentation)**，包含 **心肌 (Myocardium)**、**主動脈瓣 (Aortic Valve)**，以及可能出現的 **鈣化區域 (Calcification)**。

3. 訓練流程 (Training Pipeline)

我們在 **Google Colab + T4 GPU** 環境下使用 **PyTorch** 與 **MONAI (Medical Open Network for AI)** 進行訓練。損失函數為 **Dice Loss** 與 **Cross Entropy Loss** 的組合，以平衡類別不均與分割精度。優化器採用 **Adam Optimizer**，搭配 **Learning Rate Scheduler** 動態調整學習率。由於 CT 體積龐大，我們採用 **Patch-based Training**，並設置 **早停機制 (Early Stopping)**。模型表現以 **Dice Coefficient** 與 **Intersection over Union (IoU)** 評估。

4. 推論流程 (Inference Pipeline)

在推論階段，我們輸入測試集並使用最佳化權重進行分割，輸出 **Segmentation Masks**。隨後進行 **後處理 (Post-processing)**，包含 **連通區域分析 (Connected Component Analysis)** 與 **平滑化 (Smoothing)**，以去除雜訊並改善邊界。最終結果依規範輸出為 **.nii.gz**，合併後壓縮成 **.zip** 上傳。

5. 可視化與應用 (Visualization & Applications)

我們在 **2D 切片 (2D Slice Overlay)** 上將分割結果疊加於 CT 影像，以便直觀檢視；同時利用 **3D Slicer** 生成 **三維模型 (3D Reconstruction)**，呈現心肌、瓣膜與鈣化區域的位置關係。此流程可應用於 **術前規劃 (Preoperative Planning)** 與 **術中模擬 (Intraoperative Simulation)**，並具移植性，可延伸至 **超音波影像分割 (Ultrasound Image Segmentation)**，以降低醫療流程中的人力負擔與手術風險。

Method (NEW)

1. 前處理 (Pre-processing)

為了使模型能應對不同來源的 CT 影像變異，我們使用 `MONAI (Medical Open Network for AI)` 函式庫建立了一套標準化的前處理流程。首先，我們依據 `a_min=-42` 與 `a_max=423` 參數對影像進行**強度裁切 (Intensity Clipping)**，將 CT 的 Hounsfield Unit (HU) 數值鎖定在最能凸顯心臟軟組織的範圍內，過濾掉非必要的極端值。接著，我們執行了最關鍵的**體素重採樣 (Voxel Resampling)**，將所有影像的體素間距 (Voxel spacing) 強制統一為 `(0.7, 0.7, 1.0) mm`，以確保模型學習到的特徵尺度 (Feature scale) 保持一致。最後，由於 3D 影像極為龐大，我們在訓練時採用了 `(128, 128, 120)` 的 3D **區塊取樣 (Patch Sampling)** 策略，並透過 `RandCropByPosNegLabeld` 轉換來確保裁切的區塊中包含足夠的前景標籤，藉此加速模型收斂。

2. 模型設計 (Model Architecture)

我們採用了 `CardiacSegV2` 專案作為基礎開發框架。雖然此專案在腳本中將模型稱為 `unet3d`，但根據其參數配置（如 `depths`, `norm_name='layer'`, `exp_rate`）及對 `timm` 函式庫的依賴，其核心並非傳統的 U-Net。它實際上是一個更為先進的**混合式 Transformer U-Net 架構 (Hybrid Transformer U-Net)**，其設計理念類似於 Swin UNETR。此架構的**編碼器 (Encoder)** 部分採用了 3D Swin Transformer，它將 3D 影像切割成塊 (Patches) 並利用自注意力機制 (Self-Attention) 捕捉全局的長距離依賴 (Long-range dependency)，這對於理解複雜的器官結構至關重要。其**解碼器 (Decoder)** 則保留了 U-Net 的漸進式上採樣 (Upsampling) 結構，並透過**跨層連接 (Skip Connections)**，將 Transformer 提取的深層語意特徵 (Semantic features) 與淺層的空間細節特徵 (Spatial features) 巧妙融合，實現精確的體素級 (Voxel-level) 定位。模型最終輸出 `out_channels=4`，分別對應背景 (Background) 與競賽指定的三個分割目標（心臟肌肉、主動脈瓣膜、鈣化）。

3. 訓練流程 (Training Pipeline)

在訓練策略上，我們採用了**組合損失函數 (Combined Loss Function)**，即**Dice Loss** 與**交叉熵損失 (Cross-Entropy Loss, CE)** 的加總。Dice Loss 專注於優化分割的重疊率 (Overlap)，能有效應對前景與背景體積極度不平衡的問題；而 CE Loss 則在每個體素上進行懲罰，有助於穩定訓練過程並提升分類的精確性。優化器 (Optimizer) 方面，我們選用了 `AdamW`，並設置學習率 (Learning Rate) 為 `5e-4`、權重衰減 (Weight Decay) 為 `5e-4`。AdamW 是訓練 Transformer 模型時的首選，能更有效地處理權重衰減，防止過擬合。

4. 推論流程 (Inference Pipeline)

推論 (Inference) 階段是本方法的另一重點。為處理完整的 3D 測試影像，我們採用了 **(128, 128, 128)** 的**滑動窗口推論 (Sliding Window Inference)**。此技術會以重疊 (Overlap) 的方式在影像上滑動裁切區塊並分別進行預測，最後再將所有區塊的預測結果平滑地拼接回原始尺寸，以確保分割邊緣的連續性與準確性。此外，我們啟用了 `infer_post_process` 參數，執行了**後處理 (Post-processing)**。這一步驟包含了如「**保留最大連通物件 (Keep Largest Connected Component)**」或「**孔洞填補 (Fill Holes)**」等拓樸學操作，有效地移除了模型預測中可能出現的孤立噪點或內部空洞，使最終的分割結果更平滑且更符合解剖學邏輯，對於提升 Dice 分數至關重要。

Experiment (NEW)

實作程式碼（可以幫忙放上去）：
https://drive.google.com/file/d/1OpPwVLOLTYHku0VgTuBlIC7GNCN5RBF6/view?usp=sharing

本實驗的執行環境基於 Google Colab Pro 平台，並指定使用 NVIDIA A100 GPU 進行加速。核心開發函式庫包括 `MONAI (1.2.0)`、`Ray (2.5.0)` 以及 `PyTorch`。實驗所用的資料集來自 AICUP 競賽官方，然而，根據我們手動修正的 `AICUP_training.json` 設定檔，本次實驗僅使用了一個**極小規模**的資料子集：**9 筆**影像 (`patient0011` 至 `patient0019`) 被劃分為**訓練集 (Training Set)**，**3 筆**影像 (`patient0048` 至 `patient0050`) 被劃分為**驗證集 (Validation Set)**。測試集 (Test Set) 則使用了官方提供的全部 50 筆影像 (`41_testing_all` 資料夾)。

我們的實驗設置遵循 `tune.py` 腳本的配置。模型（`unet3d`）的訓練總週期 (Epochs) 設定為 `max_epoch=20`。我們啟用了**早停機制 (Early Stopping)**，設定為 `max_early_stop_count=2`，並配合 `val_every=5` 的驗證頻率，這意味著如果模型在連續兩次驗證（即 10 個 epoch）中 Dice 分數未取得進步，訓練將自動終止。

訓練時的 3D 區塊大小為 `(128, 128, 120)`，根據訓練日誌（每個 epoch 包含 9 個 steps）可推斷**批次大小 (Batch Size)** 為 `1`，這在 3D 醫學影像分割任務中因記憶體限制而十分常見。

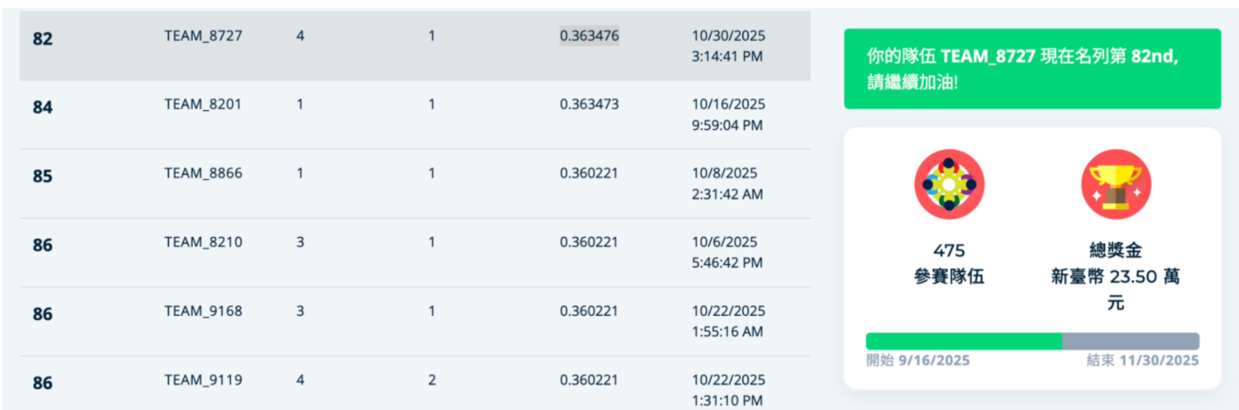
根據訓練日誌，我們的模型在**第 15 個 epoch** 達到了最佳的**內部驗證集 Dice 分數：0.518**，並成功儲存了 `best_model.pth`。然而，將此模型用於 50 筆未見過的測試資料並提交至競賽平台後，獲得的**最終 Public Score (Mean Dice) 為 0.363476**（排名 82）。內部驗證分數 (0.518) 與最終測試分數 (0.363) 之間存在著顯著的差距，這強烈顯示我們的模型出現了嚴重的**過擬合 (Overfitting)**。

我們分析其根本原因，在於訓練所用的資料集規模**過於微小**。僅使用 9 筆資料訓練、3 筆資料驗證，對於 Transformer 這樣強大且參數眾多的模型而言是遠遠不足的。模型很快地「背誦」了這 12 筆資料的特徵，但並未學習到足以**泛化 (Generalize)** 到 50 筆全新測試資料的通用解剖學規則。因此，儘管我們選用的模型架構、前處理流程與損失函數均是先進且正確的，但受限於極端的資料稀缺性，導致模型泛化能力不足，是最終得分不如預期的主因。

```
(func pid=8662)
(func pid=8662) eval result:
(func pid=8662) avg tt dice: 0.28358677
(func pid=8662) avg tt iou: 0.20835079
(func pid=8662) avg inf dice: 0.28422248
(func pid=8662) avg inf iou: 0.20904493
(func pid=8662) avg inf sensitivity: 0.36958456
(func pid=8662) avg inf specificity: 0.99020344
(func pid=8662) avg inf time: 0.0258938471476237
(func pid=8662)      patientId  tt_diceC  tt_diceAO  tt_diceCA   tt_iouC  tt_iouAO  tt_iouCA  inf_diceC  inf_diceAO  inf_diceCA  inf_iouC  inf_iouAO  inf_
(func pid=8662) 0  patient0001  0.662962        0.0        0.0  0.495844       0.0       0.0   0.665150         0.0         0.0  0.498296        0.0
(func pid=8662) 1  patient0002  0.644071        0.0        NaN  0.475004       0.0       NaN   0.645647         0.0         NaN  0.476719        0.0
(func pid=8662) 2  patient0003  0.615474        0.0        NaN  0.444538       0.0       NaN   0.616255         0.0         NaN  0.445353        0.0
Trial main_015bf_00000 completed.
== Status ==
Current time: 2025-10-30 06:33:39 (running for 00:02:36.82)
Using FIFO scheduling algorithm.
Logical resource usage: 1.0/12 CPUs, 1.0/1 GPUs (0.0/1.0 accelerator_type:A100)
Result logdir: /content/CardiacSegV2/exps/exps/unet3d/chgh/tune_results/AICUP_training
Number of trials: 1/1 (1 TERMINATED)
+------------------+------------+-----------------+-----------------------+----------+----------+------------+-----------+---------------+------------+
| Trial name       | status     | loc             | exp                   |  tt_dice |   tt_iou |   inf_dice |   inf_iou |   val_bst_acc |   inf_time |
|------------------+------------+-----------------+-----------------------+----------+----------+------------+-----------+---------------+------------|
| main_015bf_00000 | TERMINATED | 172.30.0.2:8662 | {'exp': 'AICUP__ccc0 | 0.283587 | 0.208351 |   0.284222 |  0.209045 |      0.518285 |  0.0258938 |
+------------------+------------+-----------------+-----------------------+----------+----------+------------+-----------+---------------+------------+


2025-10-30 06:33:39,122 INFO tune.py:1111 -- Total run time: 156.84 seconds (156.81 seconds for the tuning loop).
```

（圖：**第 15 個 epoch** 達到了最佳的**內部驗證集 Dice 分數：0.518**）



82	TEAM_8727	4	1	0.363476	10/30/2025 3:14:41 PM
84	TEAM_8201	1	1	0.363473	10/16/2025 9:59:04 PM
85	TEAM_8866	1	1	0.360221	10/8/2025 2:31:42 AM
86	TEAM_8210	3	1	0.360221	10/6/2025 5:46:42 PM
86	TEAM_9168	3	1	0.360221	10/22/2025 1:55:16 AM
86	TEAM_9119	4	2	0.360221	10/22/2025 1:31:10 AM

（圖：平台最終測試分數 (0.363) ）

Expected Results

We're expecting to see dice and IoU scores that are on par with the official baseline, in the range of 0.70–0.80, when segmenting heart images. We're also counting on the segmentation to clearly separate the myocardium and aortic valve, and have the option to pick up calcified areas too, colouring them red.

One of the ways we can visualise this is by generating clean 3D models of the heart muscle and valve from the input, something we could do using 3D Slicer and then examining the quality of the segmentation. Outputs will show a blend of the predicted masks over the slices and 3D renderings for preoperative planning.

We're able to significantly cut down the time and resources required for manual annotation. A task that's crucial for the medical community, when developing the segmentation pipeline for cardiac ultrasound images. This pipeline is also poised to be used in building patient-specific heart models, and will be used in intraoperative simulations to cut down intraoperative risks.

We made a Colab reproducible pipeline with trained model weights and inference scripts, which are provided. The.zip file we've provided contains.nii.gz predictions for patients 051 through 100 in the official submission format.

Reference