RL-U^{2}Net: A Dual-Branch UNet with Reinforcement Learning-Assisted Multimodal Feature Fusion for Accurate 3D Whole-Heart Segmentation

Jierui Qu[1], **Jianchun Zhao**[2*]

[1]College of Design and Engineering, National University of Singapore, 117575, Singapore
[2]School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
zhao_jianchun@stu.xjtu.edu.cn

Abstract

Accurate whole-heart segmentation is a critical component in the precise diagnosis and interventional planning of cardiovascular diseases. Integrating complementary information from modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) can significantly enhance segmentation accuracy and robustness. However, existing multi-modal segmentation methods face several limitations: severe spatial inconsistency between modalities hinders effective feature fusion; fusion strategies are often static and lack adaptability; and the processes of feature alignment and segmentation are decoupled and inefficient. To address these challenges, we propose a dual-branch U-Net architecture enhanced by reinforcement learning for feature alignment, termed RL-U^{2}Net, designed for precise and efficient multi-modal 3D whole-heart segmentation. The model employs a dual-branch U-shaped network to process CT and MRI patches in parallel, and introduces a novel RL-XAlign module between the encoders. The module employs a cross-modal attention mechanism to capture semantic correspondences between modalities and a reinforcement learning agent learns an optimal rotation strategy that consistently aligns anatomical pose and texture features. The aligned features are then reconstructed through their respective decoders. Finally, an ensemble-learning–based decision module integrates the predictions from individual patches to produce the final segmentation result. Experimental results on the publicly available MM-WHS 2017 dataset demonstrate that the proposed RL-U^{2}Net outperforms existing state-of-the-art methods, achieving Dice coefficients of 93.1% on CT and 87.0% on MRI, thereby validating the effectiveness and superiority of the proposed approach.

Code — https://github.com/TantalumKevin/RL-U2NET
Extended version — https://arxiv.org/abs/2508.02557

Introduction

Cardiovascular disease (CVD) represents a leading cause of mortality worldwide. Accurate three-dimensional whole-heart segmentation is essential for quantitative lesion assessment and clinical decision-making. While computed tomography (CT) and magnetic resonance imaging (MRI) serve as primary diagnostic tools, single-modality approaches suffer from inherent limitations in contrast, spatial resolution, and imaging artifacts that compromise comprehensive cardiac characterization. Effective multi-modal feature fusion methods are therefore critical for enhancing segmentation accuracy and robustness in clinical applications (Valsangiacomo Buechel and Mertens 2012; Puyol-Antón et al. 2022).

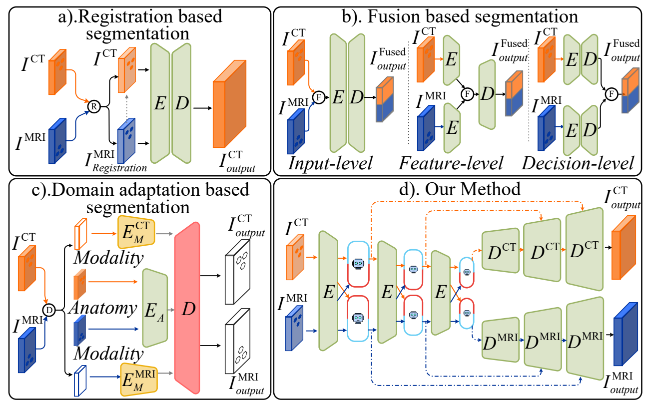

Figure 1: Paradigm comparisons between the existing multimodal medical image segmentation methods and our method

Deep learning has significantly advanced CVD diagnosis through medical imaging analysis. U-shaped architectures have become the dominant paradigm for medical segmentation due to their multi-scale feature aggregation and skip connections (Ronneberger, Fischer, and Brox 2015; Jin et al. 2020). However, traditional 3D CNNs suffer from limited receptive fields, hindering long-range dependency modeling (Çiçek et al. 2016). While Transformers' self-attention mechanisms capture global correlations and achieve notable progress in visual segmentation (Carion et al. 2020; Strudel et al. 2021; Cao et al. 2022), they exhibit weaker fine-grained local feature representation and higher computational costs. To address these limitations, hybrid CNN-Transformer frameworks have emerged (Chen et al. 2021; Wang et al. 2022; Chen et al. 2021). This integration of CNN's local discriminative capabilities with Transformer's global dependency modeling has proven effective for improving cardiac segmentation accuracy.

Cardiac imaging is challenged by the heart's non-rigid dy-

*Corresponding author

namics, often necessitating multi-modal approaches as single modalities provide incomplete information (Freed et al. 2016). Consequently, multi-modal cardiac segmentation has drawn significant interest, typically employing registration, fusion, or domain adaptation to mitigate inter-modal discrepancies (Li et al. 2023a) (Figure 1). However, existing methods suffer from three critical limitations. First, substantial spatial misalignment, induced by cardiac and respiratory motion, renders traditional image-level registration inadequate for precise correspondence. Second, most fusion strategies rely on static concatenation or simple weighting, lacking deep understanding of inter-modal semantic relationships and limiting feature expressiveness. Third, many approaches decouple feature alignment from the segmentation objective, precluding end-to-end optimization and compromising overall performance.

To address these challenges, we propose RL-U^2Net, a reinforcement learning-assisted dual-branch network for multimodal feature alignment. The architecture employs parallel encoders to process CT and MRI patches while preserving modality-specific characteristics. The core RL-XAlign module, inserted between encoders, first establishes semantic correspondences via cross-modal attention, then leverages a reinforcement learning agent to learn optimal spatial alignment strategies for cross-modal features. Aligned features undergo modality-specific reconstruction through dedicated decoders. To ensure training stability, an adaptive gradient weight distributor(AGWD) dynamically balances inter-modal gradient differences, while an ensemble-based decision module integrates patch-level predictions for final segmentation. This network effectively unifies feature alignment, adaptive fusion, and end-to-end optimization. Experiments on MM-WHS 2017 demonstrate state-of-the-art performance, validating our approach's effectiveness. The main contributions are:

- This paper introduces for the first time a cross-modal feature alignment module assisted by reinforcement learning. It extracts semantic correspondences through a cross-modal attention mechanism and utilizes a reinforcement learning agent to dynamically learn the optimal three-dimensional rotation strategy, effectively solving the problem of spatial inconsistency between multimodal images.
- Design an AGWD that dynamically adjusts the gradient weights of the two modalities during the training phase to maintain stability and balance in the optimization process and promote collaborative learning of dual-modality features.
- Construct a dual-branch U-Net structure to process CT and MRI modalities separately, and introduce a decision module based on ensemble learning to fuse patch-level prediction results, thereby improving the accuracy and robustness of multi-modal whole-heart segmentation.

Related Work

U-Net for 3D Medical Image Segmentation

To overcome 3D CNN limitations in long-range dependency modeling, UNETR integrates Transformer encoders with U-shaped decoders (Hatamizadeh et al. 2022), while Swin-UNETR employs hierarchical shifted window attention with self-supervised pre-training (Tang et al. 2022; Hatamizadeh et al. 2021). Recent advances include axial global attention (GASA-UNet) (Sun et al. 2024), Mamba-based state space models (EM-Net) (Chang et al. 2024), and multi-scale convolution-attention fusion (Pan et al. 2025), reflecting efforts to balance computational efficiency with global context modeling. However, these methods primarily target single-modal scenarios and lack explicit handling of spatial inconsistencies and semantic correspondences in multimodal data, making feature alignment and fusion critical performance bottlenecks.

Multimodal Cardiac Segmentation

Multimodal cardiac segmentation has attracted considerable research interest (Zhuang and Li 2020). Existing approaches fall into three categories: registration-based methods that spatially align modalities before segmentation (Luo and Zhuang 2022; Zhuang 2018; Luo and Zhuang 2020); fusion-based methods that exploit complementary CT-MRI characteristics through input-level (Yu et al. 2020; Zhang, Noga, and Punithakumar 2020), feature-level (Zhao, Boutry, and Puybareau 2020; Li et al. 2022), or decision-level fusion (Rokach 2010) with attention mechanisms; and domain adaptation methods using adversarial learning or style transfer for cross-domain generalization (Pei et al. 2021; Koehler et al. 2021; Wang and Zheng 2022). However, these methods typically treat registration and fusion as preprocessing steps (Li et al. 2023b), making unified end-to-end optimization of features alignment, fusion, and segmentation a persistent challenge.

Reinforcement Learning and PPO Algorithms for Vision Tasks

Reinforcement learning (RL) learns optimal policies through agent-environment interactions to maximize long-term rewards (Kaelbling, Littman, and Moore 1996). Deep RL employs neural networks for policy and value function modeling, enabling high-dimensional applications (Arulkumaran et al. 2017). In vision tasks, RL's iterative "perception-correction-feedback" process proves particularly effective for complex organ segmentation with ambiguous boundaries. Recent pixel-level RL methods have improved multi-organ boundary accuracy (Liu et al. 2025), while RL agents excel in image navigation, keypoint localization, and contour refinement (Alansary et al. 2018; Ghesu et al. 2016; Liao et al. 2020). Proximal Policy Optimization (PPO) achieves optimal balance among sample efficiency, stability, and implementation complexity through clipped surrogate objectives that constrain policy updates (Schulman et al. 2017). With the development of multimodal large models, PPO has been widely used for cross-modal policy optimization (Wan et al. 2025; Huang et al. 2025; He et al. 2016). Inspired by this, this paper innovatively introduces the PPO algorithm into medical image segmentation, utilizing reinforcement learning to assist in multimodal feature alignment.

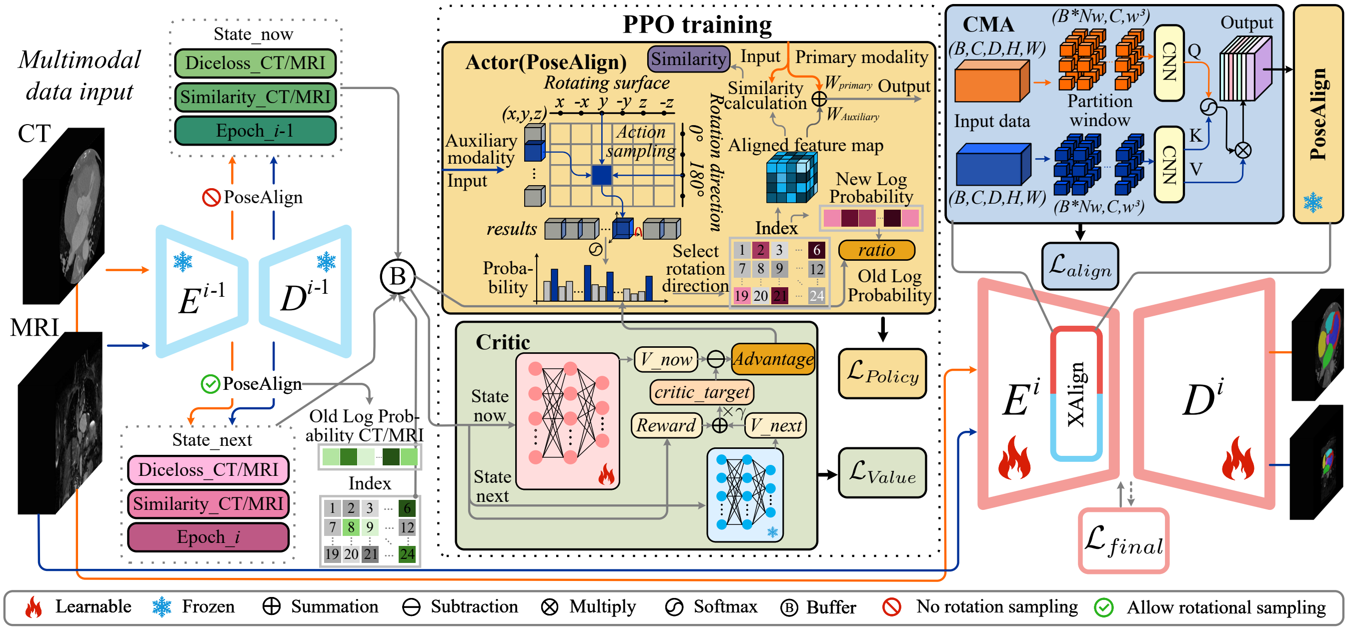

Figure 2: Overall framework of PPO training in RL-XAlign module. The framework demonstrates how CMA establishes semantic correspondences while PPO's Actor-Critic architecture learns optimal spatial alignment strategies through iterative state-action optimization for cross-modal feature fusion.

Method

Overview

The main network of RL-U^{2}Net consists of a shared encoder based on Swin Transformer, an RL-XAlign cross-modal alignment module, a Res-Fusion fusion module, and a dual-branch ResU-Net decoder. The following sections will detail the design principles and implementation mechanisms of each core module.

Reinforcement Learning based RL-XAlign module

The RL-XAlign module adopts a reinforcement learning framework, modeling cross-modal feature alignment as a sequential decision-making process, as shown in Figure 2. For encoder layer i with CT and MRI feature maps F_{CT}^{i-1} and F_{MRI}^{i-1}, the module first employs cross-modal attention (CMA) to capture semantic correspondences and construct preliminary cross-modal representations. The PoseAlign component then treats current features as environment states, where an RL agent selects optimal actions from 24 predefined 3D rotations via policy networks for precise spatial alignment. Training utilizes Proximal Policy Optimization (PPO) with Actor-Critic architecture to simultaneously optimize policy and value networks. Through iterative optimization, the module adaptively learns optimal alignment strategies, outputting spatially consistent and semantically aligned features F_{CT}^{i} and F_{MRI}^{i} for subsequent segmentation tasks.

CMA Module The CMA module is the primary component of the RL-XAlign module, responsible for establishing semantic correspondences between different modalities. Considering the high-dimensional characteristics of 3D medical images and computational efficiency requirements, this study designed a CMA module based on segmentation windows. For the input CT and MRI feature maps $F_{\mathrm{CT}}^{(i-1)}, F_{\mathrm{MRI}}^{(i-1)} \in \mathbb{R}^{B\times C\times D\times H\times W}$, with CT as the primary modality and MRI as the auxiliary modality, the CMA module first divides the 3D feature maps into non-overlapping cubic windows of size $w\times w\times w$, converting global attention calculation into local attention calculation within the window, thereby effectively reducing computational complexity.

Within each window, CMA performs cross-modal attention calculations. The query, key, and value matrices are linearly projected through independent 1D convolution layers:

$$Q = \mathrm{Conv1D}_q(W(F_{\mathrm{CT}}^{(i-1)})) \tag{1}$$

$$K = \mathrm{Conv1D}_k(W(F_{\mathrm{MRI}}^{(i-1)})) \tag{2}$$

$$V = \mathrm{Conv1D}_v(W(F_{\mathrm{MRI}}^{(i-1)})) \tag{3}$$

Where $W(\cdot)$ denotes the window segmentation operation. The cross-modal attention calculation formula within the window is:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{Q^T K}{\sqrt{d}}\right) V^T \tag{4}$$

Finally, the original feature map size is restored through output projection layer and window inverse transformation operations:

$$F_{\mathrm{cross}}^{i} = W^{-1}(\mathrm{Conv1D}_o(\mathrm{Attention}(Q,K,V))) \tag{5}$$

This mechanism enables bidirectional information exchange between CT and MRI features, generating feature representations that fuse cross-modal semantic information and provide feature inputs rich in complementary information for the subsequent pose alignment stage.

PoseAlign Module After obtaining cross-modal semantic feature representations, the PoseAlign module is responsible for solving spatial inconsistencies between multimodal modalities. This module abstracts the 3D spatial alignment problem into a discrete rotation transformation selection process, achieving precise feature pose correction through a predefined set of rotation actions.

The core design of the PoseAlign module is based on the theory of cube rotational symmetry, constructing a complete action space comprising 24 rotational transformations (Worrall and Brostow 2018). These 24 rotations encompass all possible orientations of a cube in 3D space. The specific generation process is achieved by combining the six face orientations ($\pm x, \pm y, \pm z$ axis directions) with four rotation angles (0°, 90°, 180°, 270°) for each face. The mathematical representation of the rotation matrix is:

$$\mathcal{R} = \{R_1, R_2, \ldots, R_{24}\} \subset SO(3) \tag{6}$$

Each rotation matrix $R_i \in \mathbb{R}^{3\times3}$ corresponds to a unique 3D rotation transformation. To improve the learning efficiency of reinforcement learning, this module randomly samples K_{sel} from 24 rotation transformations according to the learning weights of each direction during each forward propagation, and uses the mean of the rotated feature map as the output $F_{aligned}$

The RL-XAlign module ultimately fuses the aligned auxiliary modal feature maps into the main mode according to certain weights:

$$F^i_{O_{\mathrm{CT}}} = \lambda \cdot F^i_{\mathrm{aligned}} + (1-\lambda) \cdot F^i_{\mathrm{CT}} \tag{7}$$

Among them, λ is the preset fusion weight, which ensures that the main modal features dominate, while the aligned auxiliary modal features supplement cross-modal information with lower weights.

Reinforcement Learning Training Strategies The training process based on the PPO algorithm is the core driving mechanism of the RL-XAlign module, which models cross-modal feature alignment as a Markov decision process and implements collaborative learning of policy optimization and value assessment through the Actor-Critic architecture (Yao et al. 2024). This training strategy uses experience replay and policy pruning mechanisms to ensure the stability and convergence of the training process. The state space of the reinforcement learning environment is designed as a three-dimensional vector representation:

$$s_t = [(1-\mathcal{L}_{Dice}), S_{similarity}, epoch] \tag{8}$$

The dice coefficient reflects the current segmentation quality, the similarity metric measures the degree of feature alignment between modalities, and the training progress factor provides temporal prior information. The action space corresponds to 24 predefined rotation transformations, and the agent needs to learn to select the optimal rotation strategy in a given state to maximize the cumulative reward. The reward function is designed as a weighted combination of segmentation performance and feature alignment quality:

$$r_t = (1-\mathcal{L}_{Dice}) + S_{similarity} \tag{9}$$

Where $\mathcal{L}_{Dice}$ is the Dice loss and $S_{similarity}$ is the cross-modal feature similarity, the design allows the intelligences to optimize both segmentation accuracy and feature alignment quality.

The PPO training process adopts the Actor-Critic framework, in which the Actor network consists of learnable parameters $\theta \in \mathbb{R}^{24}$ in the PoseAlign module, which generates an action probability distribution $\pi_\theta(a|s)$ through a softmax operation. The Critic network is a simple multi-layer perceptron structure that inputs a state vector and outputs a value estimate $V_\phi(s)$.

The training process is divided into two stages: experience collection and strategy update. In the experience collection stage, the intelligent agent interacts with the environment to generate trajectory samples (s_t, a_t, r_t, s_{t+1}) and stores them in the experience buffer. In the strategy update stage, parameter optimization is performed using the pruning objective function of PPO.

The advantage function estimate is calculated using the time difference method:

$$A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t) \tag{10}$$

Standardized processing is performed to reduce variance. The policy loss uses the pruning objective function of the PPO algorithm:

$$\begin{aligned}\mathcal{L}_{Policy} = -\mathbb{E}\Bigg[A_t \cdot \min\Bigg(&\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, \\ &\mathrm{clip}_\epsilon\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\right)\Bigg)\Bigg]\end{aligned} \tag{11}$$

Where $\mathbb{E}(\cdot)$ denotes the expectation operator (i.e., the mean value), the function $\mathrm{clip}_\epsilon(\cdot)$ ensures that the result remains within the range $[1-\epsilon, 1+\epsilon]$, and ϵ is the clipping parameter, which limits the policy ratio to a reasonable range. The value function loss is expressed as the mean square error:

$$\mathcal{L}_{Value} = \mathbb{E}\left[\left(V_\phi(s_t) - V_t^{target}\right)^2\right] \tag{12}$$

The target value is:

$$V_t^{target} = r_t + \gamma V_\phi(s_{t+1}) \tag{13}$$

The pseudocode of the PPO training is as follows.

The training algorithm employs a multi-round mini-batch update strategy, with each training cycle comprising multiple experience sampling and parameter updates. Specifically, for each RL-XAlign module, the algorithm first collects N trajectory samples, then performs K rounds of mini-batch updates, with each round randomly sampling a batch size of M experiences for gradient descent. This design achieves a good balance between sample efficiency and

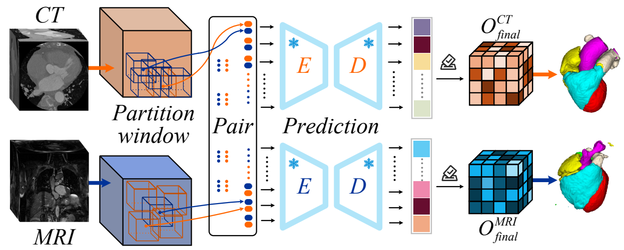

Figure 3: Schematic of the decision-making module based on ensemble learning.

training stability, enabling the agent to quickly converge to the optimal policy in complex multi-modal alignment tasks.

Decision-making Module Based on Ensemble Learning

The final prediction stage of RL-U^{2}Net adopts a decision strategy based on ensemble learning, achieving high-precision whole-heart segmentation through a sliding window inference mechanism and multi-level voting fusion. This module decomposes large-sized 3D medical images into overlapping local windows, with each window acting as an independent weak learner for prediction. The final segmentation result is generated through a weighted voting strategy to achieve a globally consistent segmentation outcome, as shown in the Figure 3.

The ensemble decision module adopts a sliding window inference framework. First, the input image $I \in \mathbb{R}^{H\times W\times D}$ is densely sampled according to the preset region of interest (ROI) size to generate a series of overlapping 3D windows. The scanning interval is adaptively calculated using an overlap ratio parameter to ensure appropriate overlap between adjacent windows and reduce edge artifacts. To further enhance ensemble performance, the module supports cross-slice ensemble mode, which establishes correspondences between different modalities through a spatial mapping mechanism. This mechanism calculates scaling transformation factors based on the spatial resolution differences between CT and MRI images, then constructs spatially transformed window regions, retaining only valid window pairs with spatial overlap for subsequent processing.

To ensure smooth blending of overlapping areas, the module calculates a Gaussian-based importance weight map for each window, giving higher confidence to the center area of the window and gradually decreasing the weight toward the boundary areas:

$$w(x,y,z) = \exp\left(-\frac{(x-x_c)^2+(y-y_c)^2+(z-z_c)^2}{2\sigma^2}\right) \tag{14}$$

where (x_c, y_c, z_c) are the window center coordinates, and σ is the scale parameter. During inference, each window independently generates segmentation predictions $P^{(j)}_{\text{MRI}}$ and $P^{(j)}_{\text{CT}}$ for CT and MRI using the RL-U^{2}Net model, which are then accumulated into the global output buffer based on their spatial positions:

$$O^{\text{CT}}_{final}(x,y,z) = \frac{\sum_j w_j(x,y,z)\cdot P^{(j)}_{\text{CT}}(x,y,z)}{\sum_j w_j(x,y,z)} \tag{15}$$

$$O^{\text{MRI}}_{final}(x,y,z) = \frac{\sum_j w_j(x,y,z)\cdot P^{(j)}_{\text{MRI}}(x,y,z)}{\sum_j w_j(x,y,z)} \tag{16}$$

This weighted averaging mechanism implements a soft voting strategy, which has good numerical stability and boundary continuity. Multiple predictions in overlapping regions are automatically constrained for consistency through weight normalization, effectively eliminating discontinuities at window boundaries and providing stable and reliable prediction results for clinical applications.

Loss Function

Multi-task Loss Function System RL-U^{2}Net constructs a complete loss function system covering segmentation supervision, cross-modal alignment, and reinforcement learning training. Through the collaborative optimization of multiple subtasks, it achieves joint improvement in feature alignment and segmentation performance.

The segmentation supervision loss employs a combination strategy of Dice loss $\mathcal{L}_{Dice}$ and cross-entropy loss $\mathcal{L}_{CE}$, which ensures overall segmentation performance while enhancing the accuracy of detailed boundaries. The cross-modal alignment loss specifically optimizes the feature alignment quality in the RL-XAlign module, consisting of InfoNCE loss $\mathcal{L}_{InfoNCE}$ and cosine embedding loss$\mathcal{L}_{Cosine}$. InfoNCE loss (van den Oord, Li, and Vinyals 2018) adopts contrastive learning ideas, learning discriminative feature representations by maximizing the similarity of positive samples and minimizing the similarity of negative samples. The cosine embedding loss (Payer et al. 2019) directly constrains the cosine similarity of aligned features.And the total loss for cross-modal alignment is:

$$\mathcal{L}_{align} = \alpha\mathcal{L}_{InfoNCE} + \beta\mathcal{L}_{Cosine} \tag{17}$$

Where α and β are equilibrium weights.

The reinforcement learning loss function is responsible for optimizing the policy network and value network in the PoseAlign module. The policy loss $\mathcal{L}_{Policy}$ and value loss $\mathcal{L}_{Value}$ are shown in equations (11) and (12).

Adaptive Gradient Weight Distributor(AGWD) CT and MRI modalities often exhibit different levels of learning difficulty due to differences in imaging mechanisms, contrast characteristics, and anatomical structure representation. Traditional fixed-weight loss functions cannot adapt to dynamic changes in learning states, often leading to overfitting in one modality and underfitting in another. To address this, this paper proposes an AGWD that dynamically monitors the learning progress of each modality and adjusts loss weights in real-time to achieve adaptive balance in multi-modal learning. Weight calculations use a hyperbolic tangent function for smooth adjustment:

$$w = \tanh\left(\gamma\cdot(\mathcal{L}_{slow} - \mathcal{L}_{fast} - \delta)\right) \tag{18}$$

Where $\mathcal{L}_{slow}$ and $\mathcal{L}_{fast}$ represent the loss values of the slow convergence mode and fast convergence mode, respectively, γ is the temperature parameter controlling the sensitivity of weight adjustment, and δ is the baseline offset providing stability assurance. The selection of the hyperbolic tangent function ensures that the weight values vary within the $(-1, 1)$ interval, avoiding training instability caused by extreme weights. Based on the calculated weight factors, the balanced loss is reconstructed using an adaptive weighting strategy:

$$\mathcal{L}^{balanced} = \mathcal{L}^{fast}(1 - w) + \mathcal{L}^{slow}(1 + w) \tag{19}$$

This weighting strategy is applied to both Dice loss and cross-entropy loss. After integration with the multi-task loss function system, the final loss function is:

$$\mathcal{L}_{final} = (\mathcal{L}^{balanced}_{Dice} + \mathcal{L}^{balanced}_{CE}) + \lambda_1 \sum_i \mathcal{L}^{(i)}_{align} \tag{20}$$

$\mathcal{L}^{(i)}_{align}$ is the alignment loss for layer i, and λ_1 is the weight balancing parameter. This design achieves adaptive balancing of segmentation supervision loss and collaborative optimization of cross-modal alignment loss, ensuring that the entire network maintains a stable and efficient training process in complex multimodal learning tasks.

Results

Datasets and Pre-Processings

The dataset used in this study was obtained from the MM-WHS Challenge 2017 (Zhuang et al. 2019), which includes 60 sets of cardiac CT and MRI image data. Following the data partitioning strategy adopted in previous studies (Cui et al. 2023, 2025), the 40 sets of data were used as the training set, while the 20 sets of labeled data were randomly divided into 15 sets for testing and 5 sets for validation.

Results on MM-WHS 2017 Challenge Dataset

To comprehensively evaluate the segmentation performance of RL-U^{2}Net, we systematically compared it with nine state-of-the-art segmentation models on the MM-WHS 2017 dataset, as shown in the Table 1. To ensure fairness in the comparison, all comparison methods were re-experimented locally using the same dataset and evaluation metrics, with Dice coefficient and Hausdorff distance as the primary evaluation metrics.

In CT image segmentation tasks, RL-U^{2}Net demonstrated outstanding overall performance, with an average Dice coefficient of 93.1%, which is 1 percentage point higher than the second-best method, HRMedSeg (Xu et al. 2025), with 92.1%. In terms of boundary accuracy, RL-U^{2}Net had an average Hausdorff distance of 11.471 mm, which was significantly better than all other methods. Results of MRI image segmentation tasks is included in the supplementary materials.

In Figure 4, we can observe the whole heart segmentation results of our proposed RL-U^{2}Net method and the latest SOTA segmentation model in both CT and MRI modes of MM-WHS 2017. Compared with other models, our segmentation results are closer to the real ones.

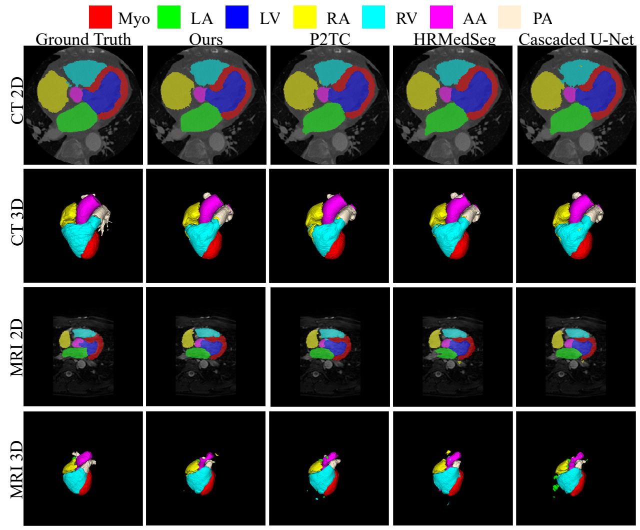

Figure 4: Visualization of methods comparison on MM-WHS 2017 Dataset.

Ablation Studies

To thoroughly validate the effectiveness of each key component of RL-U^{2}Net, we designed a series of ablation experiments to assess the contribution of each core module to the overall segmentation performance by removing them one by one (see Table 2 for details).

Ablation Experiments for CMA Removing the CMA module yields asymmetric performance impacts: The CT segmentation results fluctuate slightly with Dice coefficient declining to 88%, while MRI performance degrades severely to 60%. Without semantic correspondence guidance, auxiliary modal features integrate chaotically into the primary modal space, severely disrupting network decisions. This feature confusion particularly affects MRI due to its inherently lower imaging contrast, confirming the critical role of cross-modal attention in structured feature fusion.

Ablation Experiments for RL-XAlign Removing the RL-XAlign module reduces CT and MRI Dice coefficients to 90% and 83%, respectively. This degradation stems from the loss of cross-modal interaction, causing the dual-branch network to degenerate into two independent single-modal UNets. Each branch then relies solely on its own modality's limited information, unable to exploit complementary features from the other modality, confirming cross-modal interaction's critical role in segmentation performance. Figure 5 shows PPO reward curves during the first 100 training epochs, where both modalities exhibit steady upward trends, demonstrating successful learning of effective alignment strategies that continuously improve spatial consistency and semantic correspondence.

Ablation Experiments for Auxiliary Loss After removing the auxiliary alignment loss, the performance of CT and

Method	Dice↑								HD95(mm)↓
	Myo	LA	LV	RA	RV	AA	PA	Average	
3D U-Net (Çiçek et al. 2016)	0.894	0.909	0.917	0.869	0.891	0.933	0.883	0.899	22.988
ConResNet (Lee et al. 2022)	0.918	0.929	0.928	0.883	0.914	0.949	0.852	0.910	26.652
nnformer (Zhou et al. 2023)	0.866	0.916	0.923	0.899	0.917	0.935	0.873	0.904	12.174
D-Former (Wu et al. 2023)	0.860	0.892	0.918	0.903	0.920	0.937	0.886	0.902	14.760
SwinUNETR (Hatamizadeh et al. 2021)	0.875	0.926	0.924	0.891	0.922	0.931	0.885	0.908	17.664
UNETR++ (Shaker et al. 2024)	0.883	0.881	0.924	0.899	0.893	0.934	0.860	0.896	14.850
Cascaded U-Net (Salgado-Garcia et al. 2024)	0.899	0.921	0.927	0.905	0.909	0.946	0.889	0.914	14.163
HRMedSeg (Xu et al. 2025)	0.910	0.924	0.937	0.913	0.920	0.951	0.892	0.921	12.255
P2TC (Cui et al. 2025)	0.907	0.930	0.936	0.894	0.918	0.953	0.889	0.918	21.417
RL-U^{2}Net(Ours)	**0.927**	**0.947**	**0.938**	**0.922**	**0.933**	**0.959**	**0.894**	**0.931**	**11.471**

Table 1: Performance comparison of RL-U^{2}Net and the SOTA segmentation methods on the MM-WHS 2017 CT dataset.

Method	Dice↑	
	CT	MRI
Our model	**0.931**	**0.870**
Our model (w/o CMA)	0.884	0.612
Our model (w/o RL-XAlign)	0.909	0.837
Our model (w/o Auxiliary loss)	0.912	0.834
Our model (w/o AGWD)	0.885	0.768

Table 2: Ablation studies of diffrent modules and methods in RL-U^{2}Net.

MRI decreased slightly to approximately 91% and 83%, respectively, with the smallest but still observable decrease. This result indicates that although the auxiliary loss function is not a decisive factor, it plays a significant regularization role in the fine-tuning of feature alignment, effectively constraining the direction of cross-modal feature learning.

Ablation Experiments for AGWD Removing the AGWD reduces CT and MRI performance to 88% and 76%, respectively, with MRI showing greater degradation. Figure 5 illustrates that AGWD maintains balanced Dice and cross-entropy losses across modalities with stable convergence. Without AGWD, significant imbalance emerges: CT loss decreases rapidly while MRI converges slowly with large fluctuations, confirming AGWD's effectiveness in addressing multimodal training imbalance.

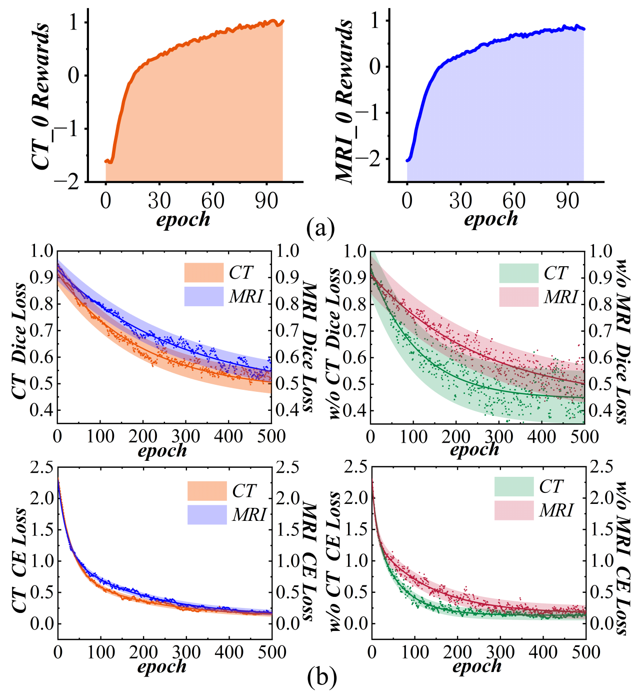

Figure 5: Ablation experiments of RL-XAlign and AGWD. (a).RL-XAlign reward curves(taking the monolayer module as an example). (b). Comparing loss dynamics with/without AGWD.

Conclusion

We propose RL-U^{2}Net, a dual-branch network leveraging reinforcement learning for multimodal feature fusion in 3D whole-heart segmentation. The RL-XAlign module employs cross-modal attention and RL agents to achieve optimal spatial alignment, addressing multimodal spatial inconsistencies. The AGWD ensures training stability through dynamic modality balancing, while ensemble-based decision fusion enhances prediction accuracy. Comprehensive validation on MM-WHS 2017 achieves 93.15% and 86.96% Dice coefficients for CT and MRI, respectively. Experimental results and ablation studies confirm RL-U^{2}Net's superiority over state-of-the-art methods, providing an effective solution for complex medical image analysis.

References

Alansary, A.; Folgoc, L. L.; Vaillant, G.; Oktay, O.; Li, Y.; and et al. 2018. Automatic View Planning with Multi-Scale Deep Reinforcement Learning Agents. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 277–285. Springer.

Arulkumaran, K.; Deisenroth, M. P.; Brundage, M.; and

Bharath, A. A. 2017. A Brief Survey of Deep Reinforcement Learning. *arXiv preprint arXiv:1708.05866*.

Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; and Wang, M. 2022. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. In *European Conference on Computer Vision*, 205–218. Springer.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-End Object Detection with Transformers. In *European Conference on Computer Vision*, 213–229. Springer.

Chang, A.; Zeng, J.; Huang, R.; and Ni, D. 2024. Em-Net: Efficient Channel and Frequency Learning with Mamba for 3d Medical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 266–275. Springer.

Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; and et al. 2021. Transunet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv preprint arXiv:2102.04306*.

Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; and Ronneberger, O. 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 424–432. Springer.

Cui, H.; Wang, Y.; Li, Y.; Xu, D.; Jiang, L.; Xia, Y.; and Zhang, Y. 2023. An Improved Combination of Faster R-CNN and U-Net Network for Accurate Multi-Modality Whole Heart Segmentation. *IEEE journal of biomedical and health informatics*, 27(7): 3408–3419.

Cui, H.; Wang, Y.; Zheng, F.; Li, Y.; Zhang, Y.; and Xia, Y. 2025. P2TC: A Lightweight Pyramid Pooling Transformer-CNN Network for Accurate 3D Whole Heart Segmentation. *IEEE Journal of Biomedical and Health Informatics*.

Freed, B. H.; Collins, J. D.; François, C. J.; Barker, A. J.; Cuttica, M. J.; and et al. 2016. MR and CT Imaging for the Evaluation of Pulmonary Hypertension. *JACC: Cardiovascular Imaging*, 9(6): 715–732.

Ghesu, F. C.; Georgescu, B.; Mansi, T.; Neumann, D.; Hornegger, J.; and Comaniciu, D. 2016. An Artificial Agent for Anatomical Landmark Detection in Medical Images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 229–237. Springer.

Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H. R.; and Xu, D. 2021. Swin Unetr: Swin Transformers for Semantic Segmentation of Brain Tumors in Mri Images. In *International MICCAI Brainlesion Workshop*, 272–284. Springer.

Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; and et al. 2022. Unetr: Transformers for 3d Medical Image Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 574–584.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Huang, X.; Dong, Y.; Tian, W.; Li, B.; Feng, R.; and Liu, Z. 2025. High-Resolution Visual Reasoning via Multi-Turn Grounding-Based Reinforcement Learning. *arXiv preprint arXiv:2507.05920*.

Jin, Q.; Meng, Z.; Sun, C.; Cui, H.; and Su, R. 2020. RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans. *Frontiers in Bioengineering and Biotechnology*, 8: 605132.

Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement Learning: A Survey. *Journal of artificial intelligence research*, 4: 237–285.

Koehler, S.; Hussain, T.; Blair, Z.; Huffaker, T.; Ritzmann, F.; and et al. 2021. Unsupervised Domain Adaptation from Axial to Short-Axis Multi-Slice Cardiac MR Images by Incorporating Pretrained Task Networks. *IEEE transactions on medical imaging*, 40(10): 2939–2953.

Lee, H. H.; Bao, S.; Huo, Y.; and Landman, B. A. 2022. 3d Ux-Net: A Large Kernel Volumetric Convnet Modernizing Hierarchical Transformer for Medical Image Segmentation. *arXiv preprint arXiv:2209.15076*.

Li, L.; Ding, W.; Huang, L.; Zhuang, X.; and Grau, V. 2023a. Multi-Modality Cardiac Image Computing: A Survey. *Medical image analysis*, 88: 102869.

Li, L.; Wu, F.; Wang, S.; Luo, X.; Martín-Isla, C.; and et al. 2023b. MyoPS: A Benchmark of Myocardial Pathology Segmentation Combining Three-Sequence Cardiac Magnetic Resonance Images. *Medical Image Analysis*, 87: 102808.

Li, W.; Wang, L.; Li, F.; Qin, S.; and Xiao, B. 2022. Myocardial Pathology Segmentation of Multi-Modal Cardiac MR Images with a Simple but Efficient Siamese U-Shaped Network. *Biomedical Signal Processing and Control*, 71: 103174.

Liao, X.; Li, W.; Xu, Q.; Wang, X.; Jin, B.; and et al. 2020. Iteratively-Refined Interactive 3D Medical Image Segmentation with Multi-Agent Reinforcement Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9394–9402.

Liu, Y.; Yuan, D.; Xu, Z.; Zhan, Y.; Zhang, H.; Lu, J.; and Lukasiewicz, T. 2025. Pixel Level Deep Reinforcement Learning for Accurate and Robust Medical Image Segmentation. *Scientific Reports*, 15(1): 8213.

Luo, X.; and Zhuang, X. 2020. MvMM-RegNet: A New Image Registration Framework Based on Multivariate Mixture Model and Neural Network Estimation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 149–159. Springer.

Luo, X.; and Zhuang, X. 2022. X-Metric: An N-Dimensional Information-Theoretic Framework for Groupwise Registration and Deep Combined Computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 9206–9224.

Pan, P.; Zhang, C.; Sun, J.; and Guo, L. 2025. Multi-Scale Conv-Attention U-Net for Medical Image Segmentation. *Scientific Reports*, 15(1): 12041.

Payer, C.; Štern, D.; Feiner, M.; Bischof, H.; and Urschler, M. 2019. Segmenting and Tracking Cell Instances with Cosine Embeddings and Recurrent Hourglass Networks. *Medical image analysis*, 57: 106–119.

Pei, C.; Wu, F.; Huang, L.; and Zhuang, X. 2021. Disentangle Domain Features for Cross-Modality Cardiac Image Segmentation. *Medical Image Analysis*, 71: 102078.

Puyol-Antón, E.; Sidhu, B. S.; Gould, J.; Porter, B.; Elliott, M. K.; and et al. 2022. A Multimodal Deep Learning Model for Cardiac Resynchronisation Therapy Response Prediction. *Medical Image Analysis*, 79: 102465.

Rokach, L. 2010. Ensemble-Based Classifiers. *Artificial intelligence review*, 33(1): 1–39.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241. Springer.

Salgado-Garcia, R. J.; Vila-Blanco, N.; Carreira, M. J.; and Nuñez-Garcia, M. 2024. Efficient Multi-Modal Whole Heart Segmentation via Cascaded U-Net: A Practical Solution for Clinical Settings. In *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*, 158–167. Springer.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.

Shaker, A.; Maaz, M.; Rasheed, H.; Khan, S.; Yang, M.-H.; and Khan, F. S. 2024. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 43(9): 3377–3390.

Strudel, R.; Garcia, R.; Laptev, I.; and Schmid, C. 2021. Segmenter: Transformer for Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7262–7272.

Sun, C.; Terry, R. S.; Bian, J.; and Xu, J. 2024. GASA-UNet: Global Axial Self-Attention U-Net for 3D Medical Image Segmentation. *arXiv preprint arXiv:2409.13146*.

Tang, Y.; Yang, D.; Li, W.; Roth, H. R.; Landman, B.; and et al. 2022. Self-Supervised Pre-Training of Swin Transformers for 3d Medical Image Analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20730–20740.

Valsangiacomo Buechel, E. R.; and Mertens, L. L. 2012. Imaging the Right Heart: The Use of Integrated Multimodality Imaging. *European heart journal*, 33(8): 949–960.

van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *arXiv preprint arXiv:1807.03748*.

Wan, Z.; Dou, Z.; Liu, C.; Zhang, Y.; Cui, D.; and et al. 2025. Srpo: Enhancing Multimodal Llm Reasoning via Reflection-Aware Reinforcement Learning. *arXiv preprint arXiv:2506.01713*.

Wang, H.; Xie, S.; Lin, L.; Iwamoto, Y.; Han, X.-H.; Chen, Y.-W.; and Tong, R. 2022. Mixed Transformer U-Net for Medical Image Segmentation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2390–2394. IEEE. ISBN 1-6654-0540-6.

Wang, R.; and Zheng, G. 2022. CyCMIS: Cycle-Consistent Cross-Domain Medical Image Segmentation via Diverse Image Augmentation. *Medical Image Analysis*, 76: 102328.

Worrall, D.; and Brostow, G. 2018. Cubenet: Equivariance to 3d Rotation and Translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 567–584.

Wu, Y.; Liao, K.; Chen, J.; Wang, J.; Chen, D. Z.; Gao, H.; and Wu, J. 2023. D-Former: A u-Shaped Dilated Transformer for 3d Medical Image Segmentation. *Neural Computing and Applications*, 35(2): 1931–1944.

Xu, Q.; Lou, Z.; Li, C.; He, X.; Qu, R.; and et al. 2025. HRMedSeg: Unlocking High-Resolution Medical Image Segmentation via Memory-Efficient Attention Modeling. *arXiv preprint arXiv:2504.06205*.

Yao, G.; Xuan, Y.; Li, X.; and Pan, Y. 2024. CMR-Agent: Learning a Cross-Modal Agent for Iterative Image-to-Point Cloud Registration. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 13458–13465. IEEE. ISBN 979-8-3503-7770-5.

Yu, H.; Zha, S.; Huangfu, Y.; Chen, C.; Ding, M.; and Li, J. 2020. Dual Attention U-Net for Multi-Sequence Cardiac MR Images Segmentation. In *Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge*, 118–127. Springer.

Zhang, X.; Noga, M.; and Punithakumar, K. 2020. Fully Automated Deep Learning Based Segmentation of Normal, Infarcted and Edema Regions from Multiple Cardiac MRI Sequences. In *Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge*, 82–91. Springer.

Zhao, Z.; Boutry, N.; and Puybareau, É. 2020. Stacked and Parallel U-Nets with Multi-Output for Myocardial Pathology Segmentation. In *Myocardial Pathology Segmentation Combining Multi-Sequence CMR Challenge*, 138–145. Springer.

Zhou, H.-Y.; Guo, J.; Zhang, Y.; Han, X.; Yu, L.; Wang, L.; and Yu, Y. 2023. Nnformer: Volumetric Medical Image Segmentation via a 3d Transformer. *IEEE transactions on image processing*, 32: 4036–4045.

Zhuang, X. 2018. Multivariate Mixture Model for Myocardial Segmentation Combining Multi-Source Images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12): 2933–2946.

Zhuang, X.; and Li, L. 2020. *Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings*, volume 12554. Springer Nature. ISBN 3-030-65651-9.

Zhuang, X.; Li, L.; Payer, C.; Štern, D.; Urschler, M.; and et al. 2019. Evaluation of Algorithms for Multi-Modality Whole Heart Segmentation: An Open-Access Grand Challenge. *Medical image analysis*, 58: 101537.