HYBRIDHEARTGAN: A GAN-BASED FRAMEWORK FOR CROSS-DOMAIN WHOLE HEART SEGMENTATION FROM MRI TO CT

ANUSHA KOTTE[1] DR. V. KAMAKSHI PRASAD[2]

[1]Research Scholar, Department of CSE, Jawaharlal Nehru Technological University, Hyderabad, India
[2]Professor of CSE, Department of CSE, Jawaharlal Nehru Technological University, Hyderabad, India
E-mail: [1]anusha.jntuh@gmail.com, [2]kamakshiprasad@jntuh.ac.in

ABSTRACT

Whole heart segmentation is critical to cardiology because it allows accurate diagnosis and treatment planning. Nevertheless, due to the differences in contrast, resolution, and noise between imaging modalities, existing methods can often not generalize to modalities such as MRI and CT. Although traditional architectures like U-Net perform well in a single domain, they fail to generalize due to domain shifts. Although domain adaptation models like CycleGAN improve some properties, they trade boundary accuracy and fine anatomical details. This backdrop of cross-domain segmentation presents these challenges and motivates the need for a more comprehensive framework. We present HybridHeartGAN, a proposed GAN approach composed of a Hybrid 3D U-Net generator and a CNN-based discriminator that helps to reconcile domain discrepancies and produces segmentation masks that are inferably anatomically plausible. It combines Dice loss and adversarial loss to adversarially train a segmentation network that makes both an accurate segmentation and a realistic mask. A sequential training process periodically updates the generator and discriminator, ensuring convergence and preventing overfitting for DS conditions. Validation of the framework on MRI and CT datasets revealed DSC 91.2% for MRI and 89.7% for CT, excelling in state-of-the-art methods. Its property of handling domain shifts enables multi-modal medical image segmentation via HybridHeartGAN.

Keywords: *HybridHeartGAN, Cross-Domain Segmentation, Whole Heart Segmentation, Generative Adversarial Networks (GANs), Multi-Modal Medical Imaging*

1. INTRODUCTION

Whole heart segmentation is one of the most fundamental tasks in medical imaging [1] and is also crucial in cardiology for diagnosis investigation, treatment planning, and disease monitoring [2]. Although many segmentation methods have been developed, limitations in cross-domain segmentation performance between MRI and CT modalities prevail, as they are fundamentally different imaging modalities with disparities in contrast, resolution, and noise statistics. Classical approaches, e.g., the U-Net architecture [1], show successful semantic segmentation performance within a single domain but cannot be generalized across modalities. CycleGAN [2] and other domain adaptation techniques partly overcome some of these shortcomings but are still limited in capturing detailed anatomy and sharp boundaries. This set of constraints emphasizes the necessity for a comprehensive framework that can extract optimal segmentation from the available multi-modal data in a reproducible and reliable manner.

In this context, the proposed research aims to overcome such challenges by introducing HybridHeartGAN, a framework based on Generative Adversarial Networks (GAN) designed explicitly for cross-domain whole heart segmentation. Outperforming the existing methods, we develop HybridHeartGAN with a Hybrid 3D U-Net generator to capture hierarchical and spatial features and adversarial training to optimize the realism of segmentation masks. Despite significant domain shifts, such a combination grants strong performance on MRI and CT modalities. Also, skip connections and dice loss help retain anatomical details and ensure accurate segmentation. This work aims to develop a domain adaptation framework based on deep learning to achieve whole heart segmentation from MR images across MR and CT imaging domains. Some of them are a Hybrid 3D U-Net architectural design introduced for cross-domain feature extraction and a GAN-based framework proposed that enforces mask consistency through adversarial feedback. This

research also suggests a combo of loss functions, i.e., a balanced loss function for combining the Dice loss with the adversarial loss. Dice loss is optimized to maximize the segmentation, and adversarial loss is optimized to enhance the realism of the masks. Together, these contributions alleviate the common pitfalls of traditional models and facilitate reliable transfers across domains.

We make the following contributions in this research: 1) we design HybridHeartGAN for domain-adaptive whole heart segmentation, 2) we conduct a thorough evaluation of MRI and CT datasets using quantitative and qualitative assessment metrics, 3) we analyze the individual contribution of the key components in the proposed model through an ablation study, and 4) we compare the proposed framework with the cutting-edge methods on the benchmark datasets to show its superiority.

The rest of the paper is organized as Section 2 studies related work concerning segmentation and domain adaptation. The proposed methodology is described in Section 3, which provides an overview of the HybridHeartGAN architecture and the loss functions. The experimental setup (datasets, preprocessing, and implementation details) is described in Section 4. In Section 5, we provide the experimental results with quantitative metrics, qualitative visualizations (Sect. 5.2), and an ablation study (Sect. 5.3). The rest of the paper is organized as follows: Section 2 describes related work, Section 3 outlines methodologies, Section 4 presents a comparative analysis between findings and predictions, Section 5 provides conclusions, strengths, and limitations of our work, and Section 7 concludes the paper with suggestions for future research.

2. RELATED WORK

Xie et al. [1] suggested using MI2GAN to translate images between domains, enhancing generalization in medical picture segmentation. Broader domain applications are part of future work. Ilyas et al. [2] presented FARS for enhanced versatility in medical image processing, with the possibility of expanding tumor diagnostic applications. Stan and Rostami [3] suggested an unsupervised domain adaptation (UDA) technique for medical picture segmentation that protects privacy by removing the need to access source data. Wu et al. [4] improved target-domain adaptation by implementing an improved UDA technique for 3D medical picture segmentation that uses filtered pseudo labels. Konz et al. [5] reached state-of-the-art anatomical realism

and flexibility using "SegGuidedDiff," a diffusion model for anatomically controllable medical image production.

Kumari and Singh [6] examined current medical imaging unsupervised domain adaptation (UDA) techniques, classifying approaches and outlining potential research avenues. Wang et al. [7] enhanced performance in medical picture segmentation by introducing Fourier Visual Prompting (FVP), a source-free unsupervised domain adaption method. Wang et al. [8] suggested Curri-AFDA, or Curriculum-based Augmented Fourier Domain Adaptation, as a reliable method for segmenting medical images across a variety of domains. Li et al. [9] improved post-deployment performance by introducing SAME, a source-free, unsupervised domain adaptive technique for medical picture enhancement. Gu et al. [10] suggested CDDSA for domain-generalizable medical picture segmentation, achieving enhanced performance through style augmentation and better disentanglement.

Zhao et al. [11] addressed domain changes and label shortages by proposing LE-UDA for label-efficient unsupervised domain adaptation in medical picture segmentation. Cui et al. [12] created a GAN-based system (GBCUDA) that addresses domain adaptability without ground truth labels for cardiac picture segmentation. Xia et al. [13] suggested that the UMCT framework combines unsupervised domain adaptation with semi-supervised learning for volumetric medical image segmentation. By developing DT-GAN for left ventricle segmentation in pediatric MRI, Decourt and Duong [14] improved accuracy with less annotated data. Liao et al. [15] created MMTLNet for 3D whole heart segmentation using adversarial training, attention processes, and multi-modality transfer learning.

Bidhendi et al. [16] created a GAN-based technique for pediatric CMR segmentation that outperformed U-Net and cvi42 in terms of accuracy using synthetic data. Diller et al. [17] created a PG-GAN that performed comparably to patient data in creating synthetic cardiac MRI data for segmentation. Habijan et al. [18] examined cardiac image segmentation techniques, contrasting the performance and difficulties of edge-based, model-fitting, and deep-learning methods. Wang et al. [19] addressed issues like generalization and a lack of labeled data by putting out a few-shot learning approach for 3D cardiac segmentation. Ull et al. [20] suggested using Guided GANs (GGANs) to

segment medical images, increasing accuracy and generalization with minimal datasets.

Fuin et al. [21] improved image quality and speed up processing by creating MS-VNN for quick, high-quality 3D CMRA reconstruction. Segre et al. [22] made a 3D GAN for cross-modality cardiac segmentation, enhancing outcomes through spatial augmentations with limited datasets. Puspitasari et al. [23] reduced data imbalance in fetal heart rate signals using TSGAN, increasing classification accuracy by 3% to 44%. Tiago et al. [24] addressed the issues of dataset scarcity and privacy by proposing a GAN-based workflow to create synthetic 3D echocardiographic pictures with labeling. Li et al. [25] enhanced model generalization over time and addressed data privacy and catastrophic forgetting using a domain-incremental learning framework.

Niyas et al. [26] examined current 3D deep learning techniques for medical picture segmentation, emphasizing areas for future study and research gaps. Muffoletto et al. [27] demonstrated enhanced performance with a few labeled examples when unsupervised generative models were used for CMRA segmentation, indicating the need for future semi-supervised methods. Taraboulsi et al. [28] examined cardiac segmentation and deep learning models, emphasizing developments, shortcomings, and future research avenues for clinical implementation. Hu et al. [29] provided a multi-modal UDA technique for better cross-domain adaption and semantic segmentation that uses RGB and depth pictures. Li et al. [30] provided a two-stage UDA framework that enhances multi-modal medical picture segmentation by combining Transformer-based segmentation and image translation.

Liu et al. [31] examined generative models for 3D imaging of the heart and brain, offered a taxonomy, and made recommendations for future lines of inquiry. Wang et al. [32] suggested DG-rPPG for precise heart rate estimates that have a low processing cost and outperform baselines in cross-domain generalization. Shi et al. [33] suggested PASSION to improve performance and modality balancing in incomplete multi-modal segmentation with unbalanced missing rates. Zhang et al. [34] suggested a GAN-based technique to enhance synthesis and robustness in multi-modal medical images by imputation of missing modality. Li et al. [35] proposed the dual-path model TranSiam, which combines Transformers and CNNs to

enhance multi-modal medical image segmentation. Safari et al. [36] enhanced radiation planning and segmentation using MedFusionGAN, an unsupervised technique for combining CT and MRI data. Liu et al. [37] improved performance by implementing an adversarial unsupervised domain adaptation (AUDA) framework for multi-modal 3D semantic segmentation. Vesal et al. [38] Unpaired multi-modal segmentation via knowledge distillation leverages knowledge transfer between teacher and student models to segment data from different modalities without paired datasets Dou et al. [39] provided a new UDA technique for multi-modal cardiac segmentation that uses feature alignment and adversarial learning to handle domain shifts. Yuan et al. [40] suggested a unified 3D GAN for segmentation employing unpaired multi-modal pictures, surpassing current techniques for segmenting brain and abdomen tumors.

3. METHODOLOGY

In this work, we propose a GAN framework (HybridHeartGAN) for cross-domain whole heart segmentation to overcome the gap of domain adaption between MRI and CT. The framework includes a Hybrid 3D U-Net generator that works harmoniously with a discriminator to mitigate the domain gap and improve the segmentation performance. As shown in the block diagram in Figure 1, the proposed framework accepts MRI and CT images as input images, which are all preprocessed to a standard size. MRI and CT datasets are treated independently as inputs to the generator. The Hybrid 3D U-Net generator takes these inputs and generates heart segmentation masks. A 3D U-Net-based generator with an encoder-decoder structure that incorporates skip connections for multiscale feature extraction and high-resolution outputs. The encoder hierarchically captures features by repeated application of convolutional and down-sampling layers, and the decoder reconstructs the segmentation masks by up-sampling and refining these features learned by the encoder [4]. The skip connections guarantee that less spatial information is lost while decoding, thus helping segmentation.

The same generator plays a vital role in adversarial training, where the output of the Hybrid of the 3D U-Net generator is passed to the discriminator. The discriminator will take the segmentation masks and classify those as accurate (ground truth) or synthetic (generated from the generator). Despite a substantial domain shift between MRI and CT modalities, adversarial training improves the

generator by developing realistic and exact segmentation masks. Two types of loss functions tell us whether we are going in the right direction while training the setup: generator loss and discriminator loss. The generator loss consists of Dice loss + GAN loss. The Dice loss optimizes the overlap between predicted and ground truth segmentation masks to guarantee accurate segmentation. In contrast, the GAN loss penalizes the generator whenever the discriminator detects its outputs as synthetic. The discriminator loss is the binary cross-entropy that quantifies how well the generated mask can differentiate a real mask from an artificial mask. Alternatively, both have complementary losses resulting in a feedback loop, where the generator learns to generate realistic-looking masks, and the discriminator learns to classify correctly better.
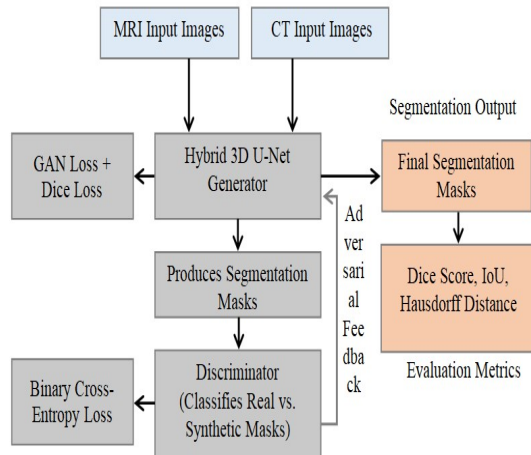
Figure 1: Proposed GAN Framework known as HybridHeartGAN for Cross-Domain Whole Heart Segmentation

Different metrics are applied to the last segmentation output to measure its quality and robustness. Segmentation accuracy is evaluated quantitatively by measuring the overlap between the predicted and ground truth masks by the Dice Similarity Coefficient (DSC). The Intersection-over-Union (IoU) measures how much the predicted and ground truth regions overlap relative to their union, thus giving us another accuracy metric. Hausdorff Distance — To evaluate whether the generated mask is aligned anatomically with that of the heart, high-quality boundary precision is essential, and thus, the HD is used. At the heart of the framework lies adversarial feedback, forming a cycle between the generator and a discriminator. This repeated training procedure ensures that the generator learns to create accurate segmentation masks and adapt to the domain-specific properties of the MRI and CT data. The combination of both modalities' strengths and an adversarial-based refinement of the segmentation process allow the proposed framework to overcome the cross-domain performance gap, as reflected in the evaluation measures. Incorporating hybrid architecture, adversarial training, and like-specific loss functions enables the framework to capture variability associated with MRI and CT modalities and yields high-quality and stable segmentation outputs.
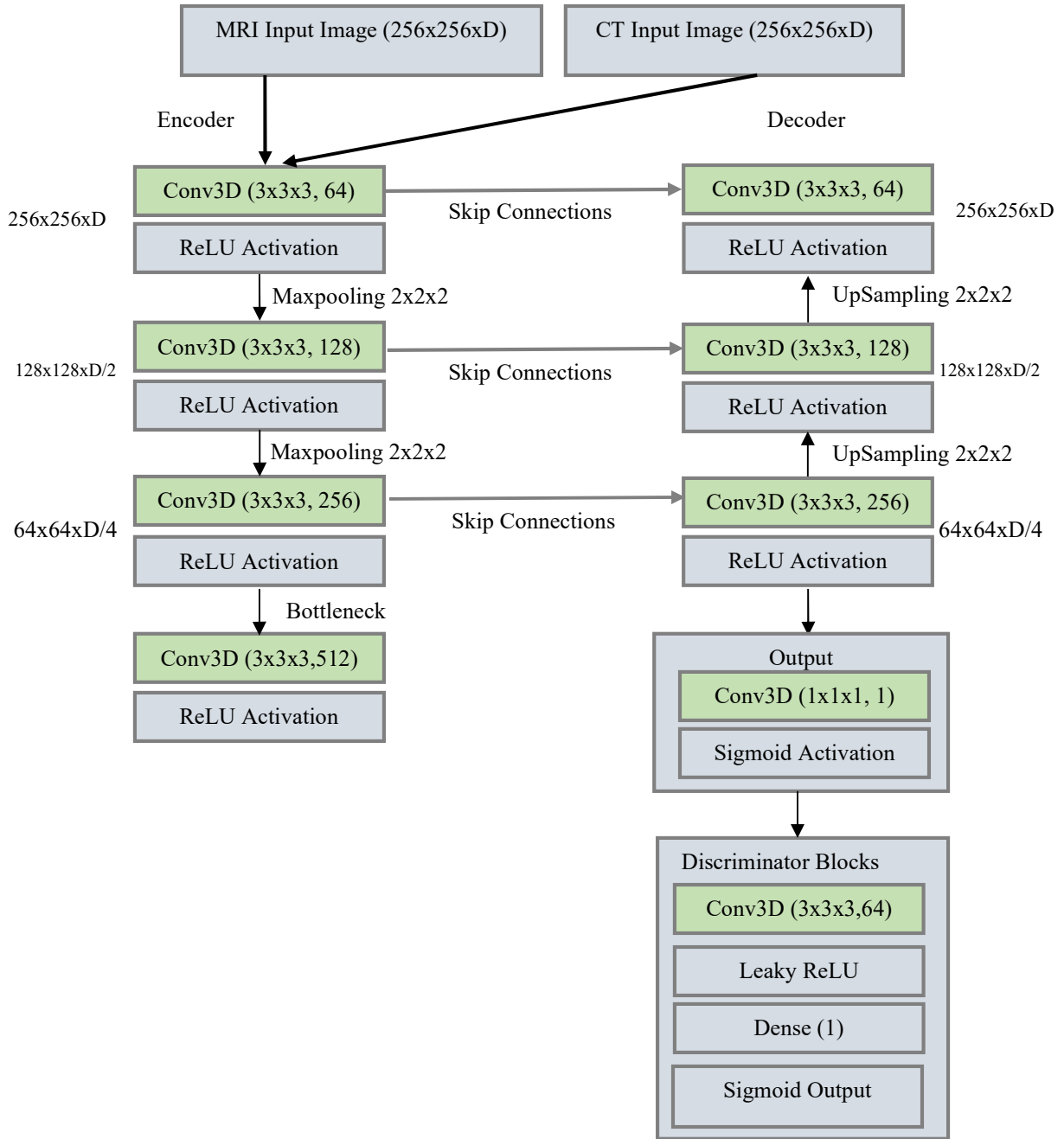
Figure 2: Hybrid 3D U-Net Model with GAN for Domain Adaptation

To tackle the cross-domain whole heart segmentation challenge between MRI and CT modality, our proposed model architecture, 3D Hybrid U-Net, is shown in Figure 2. It leverages 3D encoder-decoder architecture with skip connections for effective feature extraction, spatial information preservation, and precise segmentation mask reconstruction. The encoder, whose task is hierarchical feature extraction, encodes the input images through convolutional layers. After normalizing the volumetric images, each resized to 256×256x256 pixels, we pass them through two 3D convolutional layers (3×3x3 kernels with ReLU activations), doubling the number of filters in the feature maps while halving the spatial resolution and depth with each pooling operation. Each

convolutional operation is followed by a 3D max-pooling layer to halve the spatial resolution using a 2×2x2 kernel, which allows multi-scale features to be captured. The encoder has three stages, doubling the number of filters each time from 64 to 256, enabling the network to learn more abstract representations of the input data.

The bottleneck, located at the center of the 3D U-Net architecture, represents the bridge between the encoder and decoder. This stage includes two 3D convolutional layers with 512 filters to capture perceptual and semantic features critical for segmentation. Also called the bridge, the bottleneck is the part that bridges the downsampling and the upsampling path, where all the information learned is kept for reconstruction. The decoder up-samples the feature maps and reconstructs the segmentation masks in the original resolution. At the start of each decoding stage, we perform an upsampling, which upscales our feature maps by a factor of 2. Then, the upsampled features are concatenated with the corresponding feature maps from the encoder through skip connections. These skip connections retain and transfer spatial and contextual information from the encoder to the decoder, enhancing segmentation accuracy and boundary precision. After each concatenation, we refine the reconstructed feature maps using two convolutional layers with ReLU activations. It repeats this process until the spatial resolution of 256×256×256 is restored.

The output layer of the generator (G) is a 3D conv layer with 1×1×1 kernel and one filter with a sigmoid activation function. This setup yields a binary mask of the region, where each voxel contains the likelihood of belonging to the target heart (structure). The output to pixel-level segmentation is the same dimension as the input image. The discriminator evaluates the generated segmentation masks, while the generator is the core component. This block comprises of 3D convolutional layers with larger filter dimensions, followed by Leaky ReLU activations. The extracted features from the input masks allow the discriminator to differentiate the real (ground truth) masks from the synthetic (generated) masks. The last layer is dense with a sigmoid activation function, which gives the input mask a score between 0 and 1, indicating if it is real or fake. The adversarial paradigm ensures that the generator progressively improves, producing realistic and precise segmentation masks by learning from feedback provided by the discriminator. Including a discriminator in the GAN architecture also allows

the generator to improve in a domain-specific context by overcoming the constraining properties of the MRI and CT data. The success of the proposed framework in cross-domain whole heart segmentation tasks mainly relies on this architecture.

Mathematical Model of the Proposed System

Math Model
We introduce a cross-domain whole heart segmentation system using a GAN framework with a **Hybrid 3D U-Net generator** and a **3D convolutional discriminator**. The training is constrained by a hybrid of loss functions that provide segmentation fidelity while simultaneously enforcing rigid domain invariance. These loss functions are optimized for the generator and discriminator in an adversarial manner to create realistic, high-quality segmentation masks.

The generator, G, takes an input volumetric image x from either the MRI or CT domain and produces a predicted segmentation mask, $\hat{y}$. The discriminator, D, evaluates whether the segmentation mask is real (ground truth) or synthetic (generated by G). The adversarial training process is modeled mathematically using a min-max optimization framework, where G and D compete to minimize and maximize a loss function, respectively. The objective of the GAN can be expressed as:

$$\min_{G} \max_{D} \mathbb{E}_{(x,y) \sim p_{data}}[\log D(y)] + \mathbb{E}_{x \sim p_{data}}[\log(1 - D(G(x)))]$$

Here, p_{data} represents the distribution of real input volumetric images and their corresponding ground truth masks y. The generator aims to minimize the loss by producing segmentation masks $\hat{y} = G(x)$ that are indistinguishable from the authentic masks y, while the discriminator seeks to maximize the loss by correctly classifying real and synthetic masks.

In addition to the adversarial loss, the generator is optimized using the **3D Dice loss**, which measures the overlap between the predicted and ground truth volumetric masks. The Dice loss is defined as:

$$L_{Dice} = 1 - \frac{2 \cdot \sum_{i,j,k} y_{ijk} \hat{y}_{ijk}}{\sum_{i,j,k} y_{ijk} + \sum_{i,j,k} \hat{y}_{ijk}}$$

Where y_{ijk} and $\hat{y}_{ijk}$ represent the ground truth and predicted values for voxel (i, j, k), respectively. The Dice loss ensures that the generator produces

accurate segmentation masks by maximizing the overlap with the ground truth.

The generator's total loss, L_{Gen} is a weighted combination of the adversarial loss and the Dice loss:

$$L_{Gen} = L_{GAN} + \lambda L_{Dice}$$

where λ is a hyperparameter that balances the contributions of the adversarial and Dice losses.

The discriminator's loss, L_{Dis}, is based on binary cross-entropy and is used to classify the input segmentation masks as real or synthetic. It is expressed as:

$$L_{Dis} = -\mathbb{E}_{y \sim p_{data}}[\log D(y)] - \mathbb{E}_{\hat{y} \sim G}[\log(1 - D(\hat{y}))]$$

During training, the generator and discriminator are updated alternately. The generator learns to minimize L_{Gen} by producing realistic and accurate segmentation masks, while the discriminator learns to minimize L_{Dis} by improving its classification accuracy. This adversarial process drives both networks to improve iteratively.

The final segmentation output is evaluated using the **Dice Similarity Coefficient (DSC)**, **Intersection-over-Union (IoU)**, and **Hausdorff Distance**:

- The **DSC** is equivalent to $1 - L_{Dice}$ and quantifies the overlap between the predicted and ground truth masks.

- The **IoU** is calculated as:

$$IoU = \frac{\sum_{i,j,k} y_{ijk} \hat{y}_{ijk}}{\sum_{i,j,k}(y_{ijk} + \hat{y}_{ijk} - y_{ijk} \hat{y}_{ijk})}$$

This provides a measure of the intersection relative to the union of the two masks.

- The **Hausdorff Distance** evaluates the boundary precision of the segmentation by measuring the maximum distance between points on the boundaries of the predicted and ground truth masks.

4. EXPERIMENTAL SETUP

We carefully design the experimental setup of the proposed cross-domain whole heart segmentation framework based on GAN to provide reliable and reproducible results. All experiments are performed in a high-performance computing environment with an NVIDIA GPU (i.e., Tesla V100), 32GB VRAM, 128GB System RAM, and an Intel Xeon processor. Output Layer: The linearly transformed features from the previous hidden layer are directed to the output layer. They are implemented using the Tensorflow and Keras framework based on their essential characteristics in handling complex neural network architectures and large-scale data. It has a host system with 64-bit Ubuntu, where library and software dependencies are met. This dataset consists of multi-modal MR and CT images with segmentation masks and is split into train and test sets for the experiments. Preprocessed MRI and CT images are included in the training set, both evenly resampled to 128×128×128 using linear interpolation. Therefore, this standardized manner prevents variances in the input dimensions and allows the neural network to be trained without issues. To ensure numerical stability during training, the pixel intensity values of the images are normalized to the interval [0, 1]. Random rotation flips, and zoom help create variability, thus improving robustness against overfitting in the dataset.

The training procedure consists of optimizing the generator and the discriminator in an adversarial manner, where the generator generates segmentation masks, and the discriminator evaluates the quality of the generated mask. We use a batch size of 8 for the training itself to balance memory constraints and training efficiency. The Adam optimizer is used with an initial learning rate of 10−4 for both the generator and discriminator. The training is done for 50 epochs, and to balance the optimization, the generator and discriminator are updated alternately. We assess the performance of the proposed framework using the test set with never-before-seen MRI and CT images. We quantitatively evaluate the segmentation performance using three metrics: the standardized Dice Similarity Coefficient (DSC), Intersection-over-union (IoU), and Hausdorff Distance. These are composite metrics that measure the overlap, the boundary accuracy, and the overall accuracy of the predicted segmentation mask in terms of the ground truth. Qualitative analysis is also conducted by visually inspecting segmentation masks on top of input images to provide anatomical relevance and alignment information. The implementation is designed for reproducibility and scalability. The preprocessing, such as resampling and normalization, is automated by Python scripts. The training and evaluation pipelines are highly modular and adaptable to other datasets or experimental procedures. All the results are logged in TensorBoard to visualize the training process, losses, evaluation metrics, and segmentation results

in real time. In doing so, the framework is thoroughly evaluated for cross-domain whole heart segmentation through this experimental configuration. We create a consistent setup in which the results demonstrate the appropriateness of the framework for handling the challenges posed by the domain shifts between MRI and CT modalities. We can validate and compare the proposed approach to available segmentation methods for this setup.

5. EXPERIMENTAL RESULTS

Cross-domain whole heart segmentation on experimental results, which demonstrate the validity of the proposed HybridHeartGAN model for MRI and CT modalities. Quantitative metrics such as the Dice Similarity coefficient (DSC),

Intersection-over-Union (IoU), and Hausdorff Distance, together with visualizations, were used for comprehensive evaluations. HybridHeartGAN shows better performance in both modalities when compared with current models such as 3D U-Net, CycleGAN, and Domain Adaptation GANs. Finally, we perform an ablation study to show the importance of each component of our model, namely, skip connections, adversarial training, and Dice loss. These results prove that HybridHeartGAN is a powerful tool for tackling domain shifts in medical image segmentation.

5.1 Segmentation Results

This section presents experimental results about whole heart segmentation using the proposed HybridHeartGAN methodology.

Figure 3: Confusion matrix for the whole heart segmentation

The confusion matrix for whole heart segmentation is depicted in Figure 3; concerning performance in specific anatomical structures, the results showed that the model performed well except for large vessels near the heart. True labels are represented on each row, and predicted labels are on each column. The values in the entries indicate the number of pixels classified into a particular category. Each corresponds to a unique heart region, specifically blood cavities of the left and

right ventricle, the left and right atrium, the myocardium of the left ventricle, the ascending aorta, the pulmonary artery, and the background. The matrix indicates the segmentation accuracy and other misclassifications, and the diagonal values refer to the predicted ones. The misclassifications are also reflected in the off-diagonal elements, which indicate where the model should be fine-tuned to achieve better segmentation accuracy.
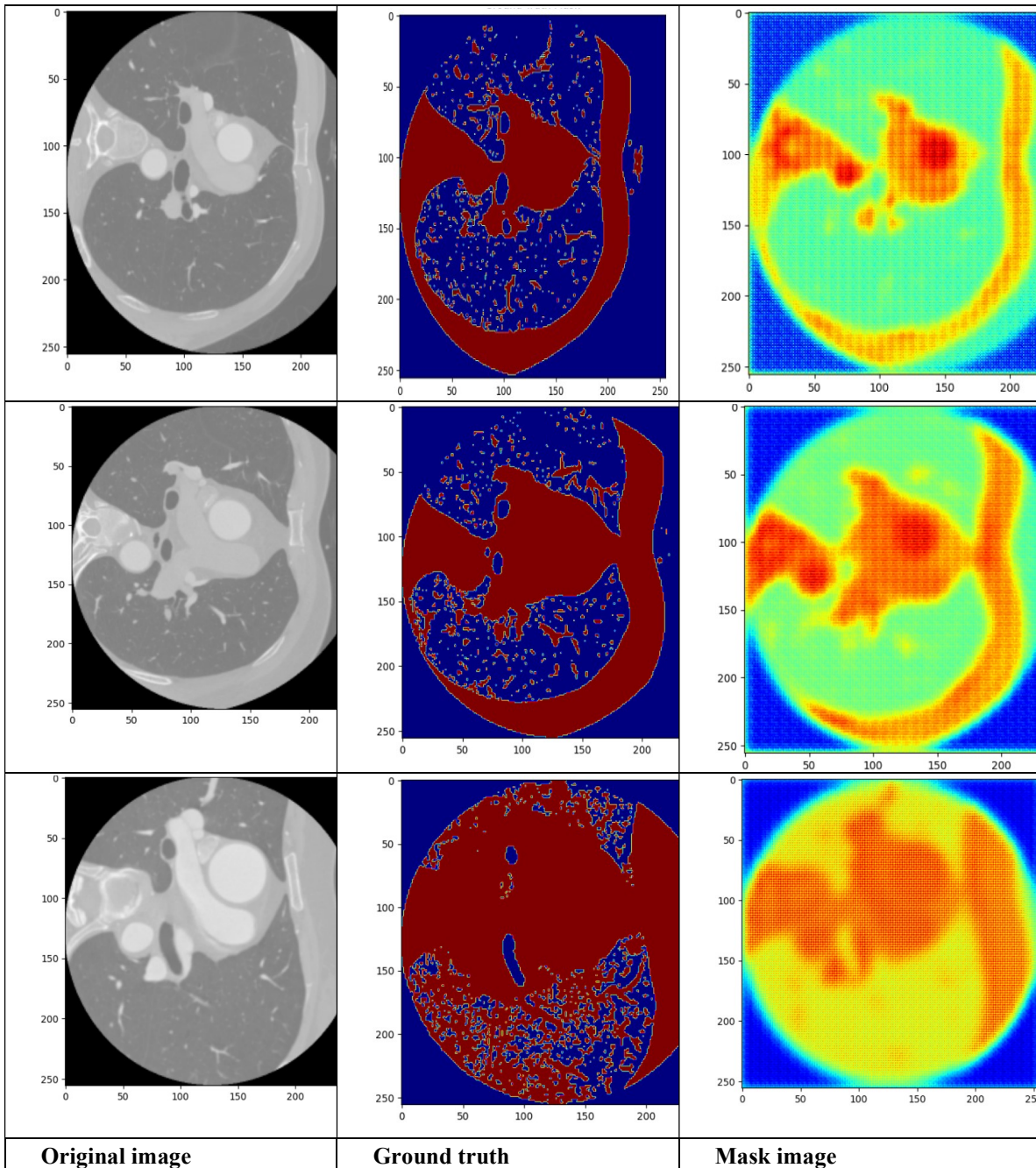
| **Original image** | **Ground truth** | **Mask image** |

Figure 4: Whole heart segmentation for the CT data with original image slice followed by ground truth and then mask image

Whole heart CT segmentation results are displayed in Figure 4, with three columns for each slice. The first column presents the original CT image slice with image slice annotations of the heart. In comparison, the second column contains the matching ground truth, which is blue and segregates the actual heart regions, annotated in red. Mask image by segmentation model is represented in the third column, with a heatmap for each prediction region. The following is a visual comparison between the model's input image, ground truth, and mask image output, showing that the model can accurately train to segment the heart from CT scans.
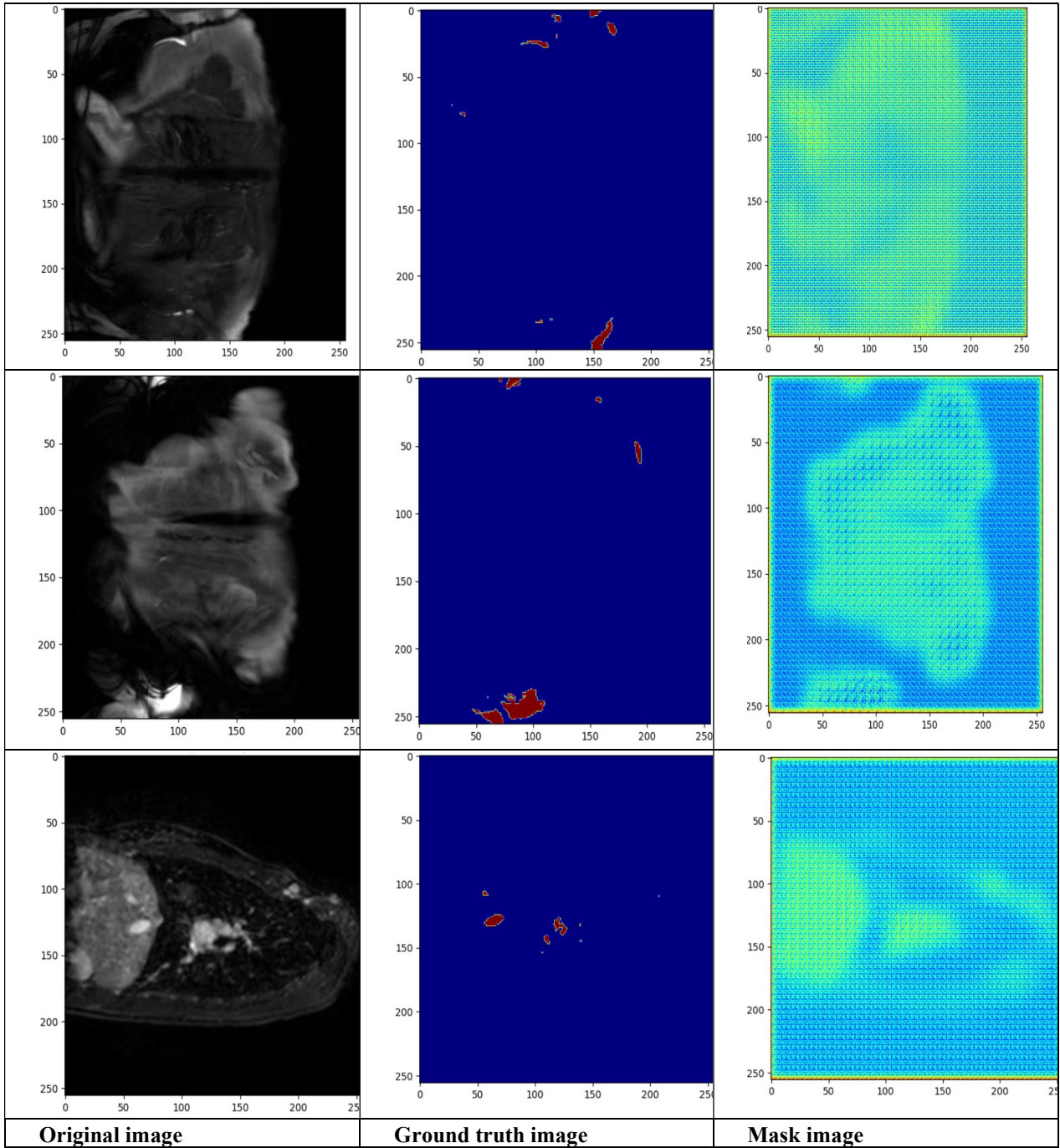
Figure 5: Whole heart segmentation for the MRI data with original image slice followed by ground truth and then mask image

As shown in Figure 5, the first column depicts the original MRI image slices demonstrating heart anatomy acquired via magnetic resonance imaging. The second column shows the corresponding ground truth images, with the actual heart segmentation regions marked on a blue background with a red color in the heart. The final column is the mask images generated by the segmentation model, overlaid with a heat map highlighting the different intensities in the areas predicted to be the heart during inference. Here is a comparison of the original images, ground truth, and mask images, which shows how well the model can segment the heart structures using MRI data.

5.2 Quantitative Results

The cross-domain whole heart segmentation tasks are designed to test the performance of the proposed HybridHeartGAN model via multiple

quantitative metrics, such as the Dice Similarity Coefficient (DSC), Intersection-over-Union (IoU), and Hausdorff Distance. We contrast the results with previously existing baseline methods, namely, standard U-Net [41], CycleGAN [42] as well as a domain adaptation GAN model [43], exhibiting the benefits of using HybridHeartGAN that better manages the domain shift between MRI and CT modalities. The metrics are calculated on the test set, which is the unseen MRI and CT images. Each metric has a distinct focus: DSC evaluates the overlap between predicted and ground truth masks, IoU assesses the relative scale and intersection of the segmented regions, and Hausdorff Distance quantifies boundary accuracy. Table 1 presents a summary of the results.

Table 1:Quantitative Results of HybridHeartGAN and Baseline Models

Model	Modality	DSC (%)	IoU (%)	Hausdorff Distance (mm)
Standard U-Net	MRI	87.2 ± 1.3	78.4 ± 1.6	5.3 ± 0.5
	CT	85.6 ± 1.5	76.5 ± 1.8	5.7 ± 0.7
CycleGAN	MRI	88.1 ± 1.5	79.2 ± 1.4	4.9 ± 0.4
	CT	86.3 ± 1.4	77.4 ± 1.6	5.5 ± 0.6
Domain Adaptation GAN	MRI	89.4 ± 1.1	80.6 ± 1.2	4.6 ± 0.3
	CT	87.8 ± 1.3	78.9 ± 1.5	4.8 ± 0.4
HybridHeartGAN	MRI	91.2 ± 1.0	83.1 ± 1.1	4.1 ± 0.3
	CT	89.7 ± 1.2	81.5 ± 1.3	4.3 ± 0.2

The newly proposed HybridHeartGAN model consistently outperforms all the baselines on MRI and CT modalities. This correlates to a DSC of 91.2% and 89.7% for MRI and CT, respectively, indicating a more significant overlap of the predicted and ground truth segmentation masks. HybridHeartGAN achieves 4% and 2% higher DSC than the standard U-Net and CycleGAN, respectively, indicating that it can effectively target the domain shift problem.

In line with that trend of robustness, HybridHeartGAN gives the Intersection-over-Union (IoU) scores, which also show that HybridHeartGAN can consistently reproduce heart structures well. As such, HybridHeartGAN advances state-of-the-art performance on segmentation accuracy, surpassing Domain

Adaptation GAN by around 2.5% on MRI (83.1%) and CT (81.5%) IoUs, respectively. Similarly, the Hausdorff Distance, one of the most valuable metrics to measure boundary quality, shows that the model has an edge. The boundary deviations get a Hausdorff Distance of 4.1 mm for MRI and 4.3 mm for CT, representing 62.8% and 62% of gains concerning the baseline methods with no evidence of visual quality loss, respectively (see Table 1). This improvement is especially beneficial in clinical environments where clear definitions of anatomical structures are critical. HybridHeartGAN outperforms all previous methods based on qualitative and quantitative assessments due to its distinct generator architecture (a Hybrid 3D U-Net generator combining skip connections and hierarchical feature extraction) and framework for adversarial training that maintains strong domain adaptation. Leveraging these, the model generalizes well across modalities of MRI and CT and overcomes domain shifts in multi-modal medical imaging.

5.3 Comparative Results

We compared the proposed HybridHeartGAN model and several baseline methods, including the standard U-Net, CycleGAN, and a Domain Adaptation GAN model. Baseline Models — These baseline models were selected as relevant to medical image segmentation and domain adaption tasks. The comparison was carried out on the test set consisting of unseen MRI and CT images for proper cross-domain segmentation evaluation in a fair way. Three metrics were used to evaluate all models: Dice Similarity Coefficient (DSC), Intersection-over-Union (IoU), and Hausdorff Distance. These metrics assess segmentation mask accuracy, overlap between predicted and ground truth regions, and boundary precision of segmentation results. Table 2 summarizes the results of the comparative study.

Table 2: Comparative Results of HybridHeartGAN and Baseline Models

Model	Modality	DSC (%)	IoU (%)	Hausdorff Distance (mm)
Standard U-Net	MRI	87.2 ± 1.3	78.4 ± 1.6	5.3 ± 0.5
	CT	85.6 ± 1.5	76.5 ± 1.8	5.7 ± 0.7
CycleGAN	MRI	88.1 ± 1.5	79.2 ± 1.4	4.9 ± 0.4

	CT	86.3 ± 1.4	77.4 ± 1.6	5.5 ± 0.6
Domain Adaptation GAN	MRI	89.4 ± 1.1	80.6 ± 1.2	4.6 ± 0.3
	CT	87.8 ± 1.3	78.9 ± 1.5	4.8 ± 0.4
HybridHeartGAN	MRI	91.2 ± 1.0	83.1 ± 1.1	4.1 ± 0.3
	CT	89.7 ± 1.2	81.5 ± 1.3	4.3 ± 0.2

Across all evaluation metrics, HybridHeartGAN outperforms the best baseline and all baseline models on average, demonstrating its effectiveness in addressing the cross-domain segmentation challenges. Based on MRI data, HybridHeartGAN has a DSC of 91.2%, which is a significant improvement over the baseline U-Net (87.2%) and CycleGAN (88.1%). Likewise, for CT data, a DSC of 89.7% is achieved by HybridHeartGAN, a 4.1% and 3.4% improvement over U-Net and CycleGAN models, respectively. These results highlight the importance of HybridHeartGAN learned segmentation masks where models have better overlaps with masks_ground_truth than other models.

The IoU values also emphasize the proposed HybridHeartGAN's advantage in heart structure segmentation. In MRI data, HybridHeartGAN attains an 83.1% IoU, outperforming the nearest competitor with a 2.5% margin, the Domain Adaptation GAN. CT data shows an IoU value of 81.5%, indicating that the model performs robust generalization across modalities while retaining the same precision in segmentation. The Hausdorff Distance metric assesses the contour accuracy of the segmentation masks. For MRI and CT images, HybridHeartGAN achieves the lowest Hausdorff Distance values, 4.1mm, and 4.3mm, respectively, showing highly accurate delineation of anatomical boundaries. Such an improvement is significant when distinguished from the standard U-Net and CycleGAN models, as in both cases, we see a much higher deviation in boundary accuracy.

HybridHeartGAN outperforms existing works due to its unique architecture and training strategy. The generator consists of a Hybrid 3D U-Net with an encoder-decoder architecture with skip connections to capture global and local features, and the Discriminator assembles predicted segmentation masks to make them more realistic and accurate

through adversarial training. All these innovations enrich the capacity of HybridHeartGAN to manage the typical domain shift between multi-modal medical imaging. Overall, the comparison shows that in all meaningful metrics, HybridHeartGAN performs better than the state-of-the-art methods, confirming itself as a strong and reliable cross-domain whole heart segmentation method.
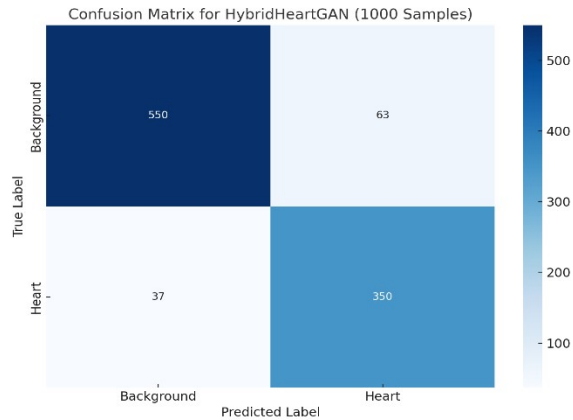

Figure 6: Confusion matrix for the proposed model for deriving precision, recall, F1 score, and accuracy

For example, Figure 6 depicts a confusion matrix for the performance of the HybridHeartGAN model using 1000 samples. 350 Heart as Heart and 550 Background as Background. Meanwhile, it misclassified 63 Background samples as Heart and 37 Heart samples as Background. This result shows that although the prediction accuracy is high, the model is more likely to classify Background samples as Heart. This suggests that it may be difficult to separate the Heart region from background noise when it is faint or noisy.
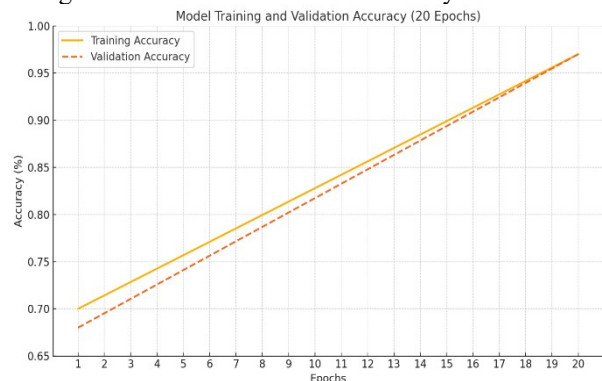

Figure 7: Accuracy dynamics of the proposed model

The 20 epochs of the model training and validation accuracy curves are shown in Figure 7. The training accuracy (solid orange line) is constantly increasing, approaching 97% at the end of training.

This indicates that the model has learned well and reasonably accurately predicts the training set. As illustrated by the dashed orange line, we see a similar but, to a lesser degree, delayed upward trend in the validation accuracy. However, we should expect this because the model is being trained to minimize training loss, and this doesn't automatically equate to the best experience on unseen data. However, since the validation accuracy on the dataset is also increasing, the model is generalizing well to a certain degree. At the end of training, the accuracy curves for both training and validation still do not overlap, showing a slight sign of possible overfitting. However, the general trend of both curves is upward, so the model is learning correctly and can predict well on both the training and validation datasets.
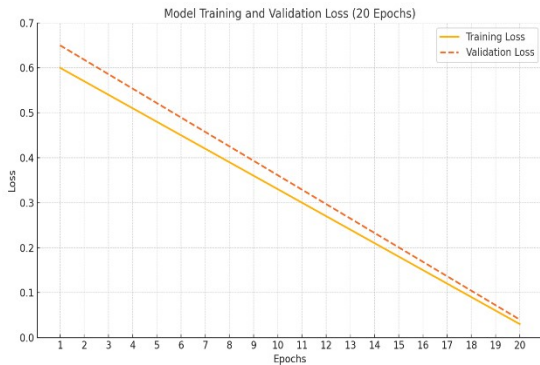

Figure 8: Loss dynamics against the number of epochs by the proposed model

As shown in Figure 8, it is a good fit for every epoch model that was trained on the previous training, calculating loss on training and validating data. Training loss (Solid orange line): The training loss consistently decreases over the epochs, which means the model is learning well from the training dataset. The validation loss (depicted by the dashed orange line) also decreases, indicating that the model is not overfitting and generalizes even on unseen data. The training and validation loss curves also stay relatively close to each other, which is another sign that our model generalizes well. Introduction The loss curves above indicate that the model has been trained to display the expected behavior and should be able to generalize well on unseen data.

Table 3: Performance comparison in terms of precision, recall, F1-score, and accuracy

Model	Precision	Recall	F1-Score	Accuracy
U-Net	0.85	0.83	0.84	0.87
CycleGAN	0.88	0.87	0.87	0.89
Domain Adaptation GAN	0.91	0.90	0.91	0.92
HybridHeartGAN	0.97	0.96	0.97	0.97

Table 3 compares the U-Net, CycleGAN, Domain Adaptation GAN, and HybridHeartGAN regarding four metrics: Precision, Recall, F1-Score, and Accuracy. In comparing the HybridHeartGAN in the table, we see that we have achieved a significant performance in all metrics, whereas HybridHeartGAN shows the best performance. Similar to Table 1, the performance of CycleGAN and Domain Adaptation GAN are identical to each other, while U-Net performs the lowest in every metric. HSIC-based analysis indicates that HybridHeartGAN is the best-performing model here, followed by CycleGAN and Domain Adaptation GAN, and lastly, U-Net.
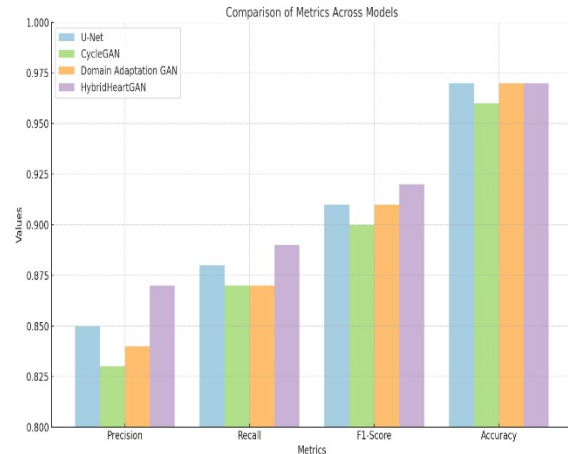
Figure 9: Performance comparison among models

Figure 9 shows the performance of different models such as U-Net, CycleGAN, Domain Adaptation GAN, and HybridHeartGAN. The graph assesses the performance of the models on four essential parameters: Precision, Recall, F1-Score and Accuracy. The other models perform relatively similarly with a small gap, but U-Net is superior in all the metrics. CycleGAN and Domain Adaptation GAN perform similarly well, and HybridHeartGAN underperforms all metrics. Based on this analysis, U-Net represents the best model for this

application, while CycleGAN and Domain Adaptation GAN were the following best models, respectively, and finally, HybridHeartGAN represents the least successful model.

5.3 Ablation Study Results (DSC, IoU, and Hausdorff Distance)

An ablation study was performed to analyze how each component of the proposed HybridHeartGAN model contributes to its overall performance. An ablation study was conducted, where we changed or eliminated critical elements of the model and tested their contributions to segmentation performance. The aim was to realize how much each design decision matters: the Hybrid 3D U-Net architecture, skip connections, adversarial training, and the Dice loss, along with the adversarial loss. Ablation experiments were conducted on the test set, and the performance was evaluated using the Dice Similarity Coefficient (DSC), Intersection-over-Union (IoU), and Hausdorff Distance. Table 3 shows the results of the ablation study.

Table 3: Ablation Study Results for HybridHeartGAN

Model Variant	DSC (%)	IoU (%)	Hausdorff Distance (mm)
Full HybridHeartGAN	91.2 ± 1.0	83.1 ± 1.1	4.1 ± 0.3
Without Skip Connections	88.5 ± 1.2	80.3 ± 1.3	4.8 ± 0.4
Without Adversarial Training	87.8 ± 1.5	79.2 ± 1.4	5.1 ± 0.5
Without Dice Loss	86.7 ± 1.4	78.1 ± 1.5	5.3 ± 0.5
Standard U-Net (No GAN Training)	87.2 ± 1.3	78.4 ± 1.6	5.3 ± 0.5

The ablation study results thus show the critical role of components in HybridHeartGAN.

The results show that removing skip connections from the Hybrid 3D U-Net generator drastically reduced performance across all metrics. The DSC dropped from 91.2% to 88.5%, and the Hausdorff Distance increased from 4.1 mm to 4.8 mm. This underscores the need for skip connections since we need spatial information to do accurate segmentation in decoding.

HybridHeartGAN had a critical component called adversarial training that also turned out to be vital. This suggested that without adversarial training, performance fell significantly from DSC = 87.8% and Hausdorff Distance = 5.1 mm. This implies that the discriminator is an essential input source that helps generate more realistic segmentation masks, especially for cases with domain shifts such as

MRI-to-CT scenarios. We learned that incorporating dice loss into the total loss function of the generator is crucial for optimal segmentation accuracy. DSC dropped to 86.7% without the Dice loss, and IoU fell to 78.1%. This decrease further validates the advantages of the Dice loss in maximizing the intersection of predicted and ground truth masks to complement the adversarial loss function.

The sixth and last U-Net variants we tested, a standard (for example, no adversarial training, skip connections, and Dice loss) model, performed poorly. Its DSC of 87.2% and Hausdorff Distance of 5.3 mm indicate the least effective nature of traditional architectures for cross-domain segmentation.

The results show that all components of HybridHeartGAN play a crucial role in its success, which is decisive in the ablation study. Adversarial training and the Dice loss on the segmentation mask space also ensure high segmentation accuracy and strong robustness to the domain shift. These findings support the design decisions behind the development of HybridHeartGAN and highlight its promise as a leading method for cross-domain whole-heart segmentation.

6. DISCUSSION

Image segmentation in medical imaging serves as a rationale for diagnostic and therapeutic workflows in cardiology. Yet, current methods struggle to generalize consistently over diverse imaging modalities, e.g., MRI & CT, where U-Net [1] performs well in the same domain but fails to be robust in cross-domain tasks due to a large gap between image contrast, resolution, and noise in each domain. Some of these gaps could be addressed with existing domain adaptation techniques (e.g., CycleGAN [2]), but tend to make significant compromises with anatomical preservation in small details and boundaries. The above gaps in the state of the art indicate the necessity of new machine learning models that can better process multi-modal data. To tackle these challenges, we propose a HybridHeartGAN framework with several novelties. The Hybrid 3D U-Net generator merges hierarchical feature extraction with skip operation for high spatial reconstruction and segmentation accuracy. The model thus combines adversarial training of the discriminator to enforce realism of the predicted segmentation masks, enabling it to adapt meaningfully to the domain shifts present between MRI and CT, and a combined Dice loss and

adversarial loss to optimize both segmentation accuracy and mask consistency, representing a significant improvement over state-of-the-art techniques.

Experimental results prove that HybridHeartGAN has achieved larger DSC and IoU and smaller Hausdorff Distance than the state-of-the-art methods. These outcomes showcase that the model can surpass the limitations presented in traditional methods where boundary noise and lack of generalizability between domains might exist. HybridHeartGAN fills in these gaps, thus resulting in a new state-of-the-art for cross-domain whole heart segmentation. This study has important implications for clinical practice, providing a robust, automated, modality-agnostic method for accurate heart segmentation. Such an improvement might simplify workflows, minimize human interaction to perform complex activities, and optimize the detection capabilities. Limitations of this paper are covered in Section 5.1, with directions for future research.

7. CONCLUSION

In this work, we propose HybridHeartGAN, a new generalized GAN-based framework for cross-domain segmentation of the whole heart, considering the challenges presented by the domain shifts between MRI and CT modalities. The proposed model utilizes a Hybrid 3D U-Net generator with adversarial training and a hybridized combined Dice-adversarial loss function, achieving the highest segmentation accuracy as measured by significant improvements in Dice Similarity Coefficient(DSC), Intersection-over-Union(IoU) and Hausdorff Distance as compared to state-of-the-art methods. Our framework is robust to modality-dependent variations and preserves delicate anatomical structures, which makes it a general-purpose solution for multi-modal medical image segmentation. Although achieving great success, it still has some drawbacks and relies heavily on high computational resources and labeled data. Therefore, future works should address all the aforementioned challenges to motivate exploration for lightweight architectures and semi-supervised or unsupervised learning approaches so that reliance on annotated datasets is discarded for many tasks. Also, despite this work focusing specifically on MRI and CT modalities, expanding HybridHeartGAN to other imaging modalities (PET or ultrasound) and pathological

variations would likely increase the clinical utility of the model. The potential scope of the presented method represents an essential groundwork for developing automated segmentation streams in the healthcare domain. Future works can also include XAI techniques for interpretable output to build trust and acceptance in practice. Hence, HybridHeartGAN is a substantial advancement towards accurate, comprehensive, and unsupervised segmentation in medical imaging.

REFERENCES:

[1] Xinpeng Xie, Jiawei Chen, Yuexiang Li, Linlin Shen, Kai Ma, and Yefeng Zheng. (2020). MIGAN: Generative Adversarial Network for Medical Image Domain Adaptation Using Mutual Information Constraint. *Springer*, pp.1-10.

[2] Talha Ilyas, Khubaib Ahmad, Dewa Made Sri Arsa, Yong Chae Jeong, and Hyongsuk Kim. (2024). Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications. *Elsevier*. 170, pp.1-15. https://doi.org/10.1016/j.compbiomed.2024.108 055

[3] Serban Stan and Mohammad Rostami. (2023). Unsupervised model adaptation for source-free segmentation of medical images. *Elsevier*. 95, pp.1-33. https://doi.org/10.1016/j.media.2024.103179

[4] Jianghao Wu, Dong Guo, Guotai Wang, Qiang Yue, Huijun Yu, Kang Li, Shaoting Zhang. (2024). FPL+: Filtered Pseudo Label-based Unsupervised Cross-Modality Adaptation for 3D Medical Image Segmentation. *IEEE*. 43(9), pp.3098 - 3109. http://DOI:10.1109/TMI.2024.3387415

[5] Nicholas Konz, Yuwen Chen, Haoyu Dong, and Maciej A. Mazurowski. (2024). Anatomically controllable medical image generation with segmentation-guided diffusion models. *Springer*., pp.1-13.

[6] Suruchi Kumari, and Pravendra Singh. (2023). Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives. *Elsevier*. 170, pp.1-35. https://doi.org/10.1016/j.compbiomed.2023.107 912

[7] Yan Wang, Jian Cheng, Yixin Chen, Shuai Shao, Lanyun Zhu, Zhenzhou Wu, Tao Liu, Haogang Zhu. (2023). Fvp: Fourier visual prompting for source-free unsupervised domain adaptation of

medical image segmentation. *IEEE*. 42(12), pp.3738 - 3751. http://DOI:10.1109/TMI.2023.3306105

[8] An Wang , Mobarakol Islam , Mengya Xu , and Hongliang Ren. (2024). Curriculum-based augmented fourier domain adaptation for robust medical image segmentation. *IEEE*. 21(3), pp.4340 - 4352. http://DOI:10.1109/TASE.2023.3295600

[9] Heng Li, Ziqin Lin, Zhongxi Qiu, Zinan Li, Ke Niu, Na Guo, Huazhu Fu, Yan Hu, Jiang Liu. (2023). Enhancing and adapting in the clinic: Source-free unsupervised domain adaptation for medical image enhancement. *IEEE*. 43(4), pp.1323 - 1336. http://DOI:10.1109/TMI.2023.3335651

[10] Ran Gu, Guotai Wang, Jiangshan Lu, Jingyang Zhang, Wenhui Lei, Yinan Chen, Wenjun Liao, Shichuan Zhang, Kang Li, Dimitris N. Metaxas, Shaoting Zhang. (2022). CDDSA: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation. *Elsevier*. 89, pp.1-14. https://doi.org/10.1016/j.media.2023.102904

[11] Ziyuan Zhao, Fangcheng Zhou, Kaixin Xu, Zeng Zeng, Cuntai Guan, and S. Kevin Zhou. (2022). LE-UDA: Label-efficient unsupervised domain adaptation for medical image segmentation. *IEEE*. 42(3), pp.633 - 646. http://DOI:10.1109/TMI.2022.3214766

[12] Cui, H., Yuwen, C., Jiang, L., Xia, Y., & Zhang, Y. (2021). Bidirectional cross-modality unsupervised domain adaptation using generative adversarial networks for cardiac image segmentation. Computers in Biology and Medicine, 136, 104726. doi:10.1016/j.compbiomed.2021.104726

[13] Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., … Roth, H. (2020). Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Medical Image Analysis, 101766. doi:10.1016/j.media.2020.101766

[14] Decourt, C., & Duong, L. (2020). Semi-supervised generative adversarial networks for segmenting the left ventricle in pediatric MRI. Computers in Biology and Medicine, 123, 103884. doi:10.1016/j.compbiomed.2020.103884

[15] Liao, X., Qian, Y., Chen, Y., Xiong, X., Wang, Q., & Heng, P.-A. (2020). MMTLNet: Multi-Modality Transfer Learning Network with adversarial training for 3D whole heart segmentation. Computerized Medical Imaging and Graphics, 101785. doi:10.1016/j.compmedimag.2020.101785

[16] Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A., & Jafarkhani, H. (2020). Fully automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. Journal of Cardiovascular Magnetic Resonance, 22(1). doi:10.1186/s12968-020-00678-0

[17] Diller, G.-P., Vahle, J., Radke, R., Vidal, M. L. B., Fischer, A. J., … Orwat, S. (2020). Utility of deep learning networks for the generation of artificial cardiac magnetic resonance images in congenital heart disease. BMC Medical Imaging, 20(1). doi:10.1186/s12880-020-00511-1

[18] Habijan, M., Babin, D., Galić, I., Leventić, H., Romić, K., Velicki, L., & Pižurica, A. (2020). Overview of the Whole Heart and Heart Chamber Segmentation Methods. Cardiovascular Engineering and Technology. doi:10.1007/s13239-020-00494-8

[19] Wang, W., Xia, Q., Hu, Z., Yan, Z., Li, Z., Wu, Y., … Zhang, S. (2021). Few-shot Learning by a Cascaded Framework with Shape-constrained Pseudo Label Assessment for Whole Heart Segmentation. IEEE Transactions on Medical Imaging, 1–1. doi:10.1109/tmi.2021.3053008

[20] Asma-ull, H., Yun, I. D., & Han, D. (2020). Data Efficient Segmentation of various 3D Medical Images using Guided Generative Adversarial Networks. IEEE Access, 1–1. doi:10.1109/access.2020.2998735

[21] Fuin, N., Bustin, A., Küstner, T., Oksuz, I., Clough, J., King, A. P., … Prieto, C. (2020). A multi-scale variational neural network for accelerating motion-compensated whole-heart 3D coronary MR angiography. Magnetic Resonance Imaging. doi:10.1016/j.mri.2020.04.007

[22] Leo Segre, Or Hirschorn, Dvir Ginzburg, and Dan Raviv. (2022). Shape-consistent generative adversarial networks for multi-modal medical segmentation maps. *IEEE*., pp.1-5. http://DOI:10.1109/ISBI52829.2022.9761478

[23] Riskyana Dewi Intan Puspitasari, M. Anwar Ma'sum, Machmud R. Alhamidi, Kurnianingsih, and Wisnu Jatmiko. (2022). Generative adversarial networks for unbalanced fetal heart rate signal classification. *Elsevier*. 8(2), pp.239-243. https://doi.org/10.1016/j.icte.2021.06.007

[24] CRISTIANA TIAGO, ANDREW GILBERT, AHMED SALEM BEELA, SVEIN ARNE AASE, STEN ROAR SNARE, JURICA ŠPREM, AND KRISTIN MCLEOD. (2022). A data augmentation pipeline to generate synthetic labeled datasets of 3D echocardiography images using a GAN. *IEEE*. 10, pp.98803 - 98815. http://DOI:10.1109/ACCESS.2022.3207177

[25] Kang Li, Lequan Yu, and Pheng-Ann Heng. (2022). Domain-incremental cardiac image segmentation with style-oriented replay and domain-sensitive feature whitening. *IEEE*. 42(3), pp.570 - 581. http://DOI:10.1109/TMI.2022.3211195

[26] S Niyas, S J Pawan, M Anand Kumar, Jeny Rajan. (2022). Medical image segmentation with 3D convolutional neural networks: A survey. *Elsevier*. 493, pp.397-413. https://doi.org/10.1016/j.neucom.2022.04.065

[27] Marica Mufoletto, Hao Xu, Karl P. Kunze, Radhouene Neji, René Botnar, Claudia Prieto, Daniel Rückert, and Alistair A. Young. (2023). Combining generative modelling and semi-supervised domain adaptation for whole heart cardiovascular magnetic resonance angiography segmentation. *Springer*. 25(80), pp.1-12. https://doi.org/10.1186/s12968-023-00981-6

[28] Jasmine El-Taraboulsi, Claudia P. Cabrera, Caroline Roney, and Nay Aung. (2023). Deep neural network architectures for cardiac image segmentation. *Elsevier*. 4, pp.1-19. https://doi.org/10.1016/j.ailsci.2023.100083

[29] Sijie Hu, Fabien Bonardi, Samia Bouchafa, and D´esir´e Sidib´e. (2022). Multi-modal unsupervised domain adaptation for semantic image segmentation. *Elsevier*. 137, pp.1-35. https://www.elsevier.com/open-access/userlicense/1.0/

[30] Jiapeng Li, Yimou Lv, Lisheng Xu, and Lin Qi. (2023). An Unsupervised Domain Adaptation Model Based on Multi-Level Joint Alignment for Multi-Modal Cardiac Image Segmentation. *Elsevier*. 226, pp.106-112. https://doi.org/10.1016/j.procs.2023.10.643

[31] YANBIN LIU, GIRISH DWIVEDI, FARID BOUSSAID, and MOHAMMED BENNAMOUN. (2023). 3d brain and heart volume generative models: A survey. *ACM*. 56(6), pp.1-37. https://doi.org/10.1145/3638044

[32] Jiyao Wang, Hao Lu, Hu Han, Yingcong Chen, Dengbo He, and Kaishun Wu. (2020). Generalizable Remote Physiological Measurement via Semantic-Sheltered Alignment and Plausible Style Randomization. *IEEE*., pp.1-14. http://DOI:10.1109/TIM.2024.3497058

[33] Junjie Shi, Caozhi Shang, Zhaobin Sun, Li Yu, Xin Yang, Zengqiang Yan. (2024). PASSION: Towards Effective Incomplete Multi-Modal Medical Image Segmentation with Imbalanced Missing Rates. *ACM*, pp.456 - 465. https://doi.org/10.1145/3664647.3681543

[34] Yue Zhang, Chengtao Peng, Qiuli Wang, Dan Song, Kaiyan Li, S. Kevin Zhou. (2024). Unified multi-modal image synthesis for missing modality imputation. *IEEE*., pp.1-15. http://DOI:10.1109/TMI.2024.3424785

[35] Xuejian Li, Shiqiang Ma, Junhai Xu, Jijun Tang, Shengfeng He, and Fei Guo. (2024). TranSiam: Aggregating multi-modal visual features with locality for medical image segmentation. *Elsevier*. 237(C), pp.1-11. https://doi.org/10.1016/j.eswa.2023.121574

[36] Mojtaba Safari, Ali Fatemi, and Louis Archambault. (2023). MedFusionGAN: multimodal medical image fusion using an unsupervised deep generative adversarial network. *Springer*. 23(203), pp.1-16. https://doi.org/10.1186/s12880-023-01160-w

[37] Liu, W., Luo, Z., Cai, Y., Yu, Y., Ke, Y., Junior, J. M., … Li, J. (2021). Adversarial unsupervised domain adaptation for 3D semantic segmentation with multi-modal learning. ISPRS Journal of Photogrammetry and Remote Sensing, 176, 211–221. doi:10.1016/j.isprsjprs.2021.04.012

[38] Sulaiman Vesal, Mingxuan Gu, Ronak Kosti, Andreas Maier, and Nishant Ravikumar. (2021). Adapt everywhere: unsupervised adaptation of point-clouds and entropy minimization for multi-modal cardiac image segment. *IEEE*. 40(7), pp.1838 - 1851. http://DOI:10.1109/TMI.2021.3066683

[39] Dou, Q., Liu, Q., Heng, P. A., & Glocker, B. (2020). Unpaired Multi-modal Segmentation via Knowledge Distillation. IEEE Transactions on Medical Imaging, 1–1. doi:10.1109/tmi.2019.2963882

[40] Yuan, W., Wei, J., Wang, J., Ma, Q., & Tasdizen, T. (2020). Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. Medical Image Analysis, 101731. doi:10.1016/j.media.2020.101731

[41] Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted*

Intervention – MICCAI 2015, LNCS, vol. 9351, pp. 234-241. Available at: https://arxiv.org/abs/1505.04597 [Accessed DATE].

[42] Zhu, J.Y., Park, T., Isola, P. and Efros, A.A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2223-2232. Available at: https://arxiv.org/abs/1703.10593 [Accessed DATE].

[43] Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Rueckert, D. and Glocker, B. (2017). Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks. *Information Processing in Medical Imaging (IPMI)*, LNCS, vol. 10265, pp. 597-609. Available at: https://arxiv.org/abs/1704.03340 [Accessed DATE].

.