Temporal Image Registration using deep learning for 3D Fetal Echocardiography

Kazi Saeed Alam, Md Kamrul Hasan, Dr Choon Hwai Yap
Department of Bioengineering, Imperial College London, UK

Introduction

- The fetal heart can experience congenital heart malformation and functional abnormalities. Ultrasound imaging plays a vital role in assessing the heart of the developing fetus due to its non-incisive nature.
- Heart chambers, valves, blood flow patterns, etc. can be used as good identifiers to detect and evaluate several cardiac diseases.

Motivation

- However, the detection of heart problems in fetus via mass screening is only around 50%, suggesting a need for improvement.
- The clinical use of echo is still stuck with 2D, likely because doctors can not visualize 3D, but for machine learning it makes more sense to go 3D, for real-time detection with improved accuracy and precision.
- Most of the works are based on adult hearts, there is less research in the field of fetal echocardiography.

Dataset Description

- Images were acquired in 4D echocardiography dicom format, out of which 4 cases were from healthy patients and 10 were from patients with hearts abnormalities.
- The 4D echo images were carried out with GE Voluson 730 ultrasound connected to the RAB 4-8L transducer (GE Healthcare Inc., Chicago, Illinois, USA) which has approximated 154 µm axial resolution and around 219 µm lateral resolution along with a transducer of 5 MHz.
- The fetuses were of mixed gender and different ethnic groups (Chinese, Indian, Malay). Most of the cases had a gestation age between 22 to 32 weeks.

Data Annotation Steps

- **Step 1 (Data Preparation) :** 4D dicom images were converted to 3D video (.avi) format containing each time points from which slices were later extracted in .png format.
- **Step 2 (Registration) :** Each slice image at a particular time point t_n was registered with respect to the initial time point t_0 and the previous time point image t_{n-1}. The deformation field was computed using cardiac motion estimation library[1].
- **Step 3 (Segmentation) :** The left ventricle chamber and myocardium for all slices at selected end-systolic and end-diastolic time points were manually annotated using lazy-snapping.
- **Step 4 (3D Reconstruction) :** Combining the masks from all slices, the next step is to generate the 3D reconstructed masks for each time point. These 3D masks were corrected and smoothed with the help of an expert using Geomagic wrap. Masks for other time points were generated using the deformation fields from registration step.
- **Step 5 (Artifacts Removal) :** Constant white boxes or arrows in the ultrasound intensity image were removed using interpolation method.
- **Step 6 (Image-Mask Pair Generation) :** In the last step, the inner wall of the reconstructed masks was filled and reconstructed masks were binarized where (class 0 represents the background, 1 for the left ventricle chamber, and 2 for the myocardium. Then they were paired with the intensity image to finalize the dataset annotation process.

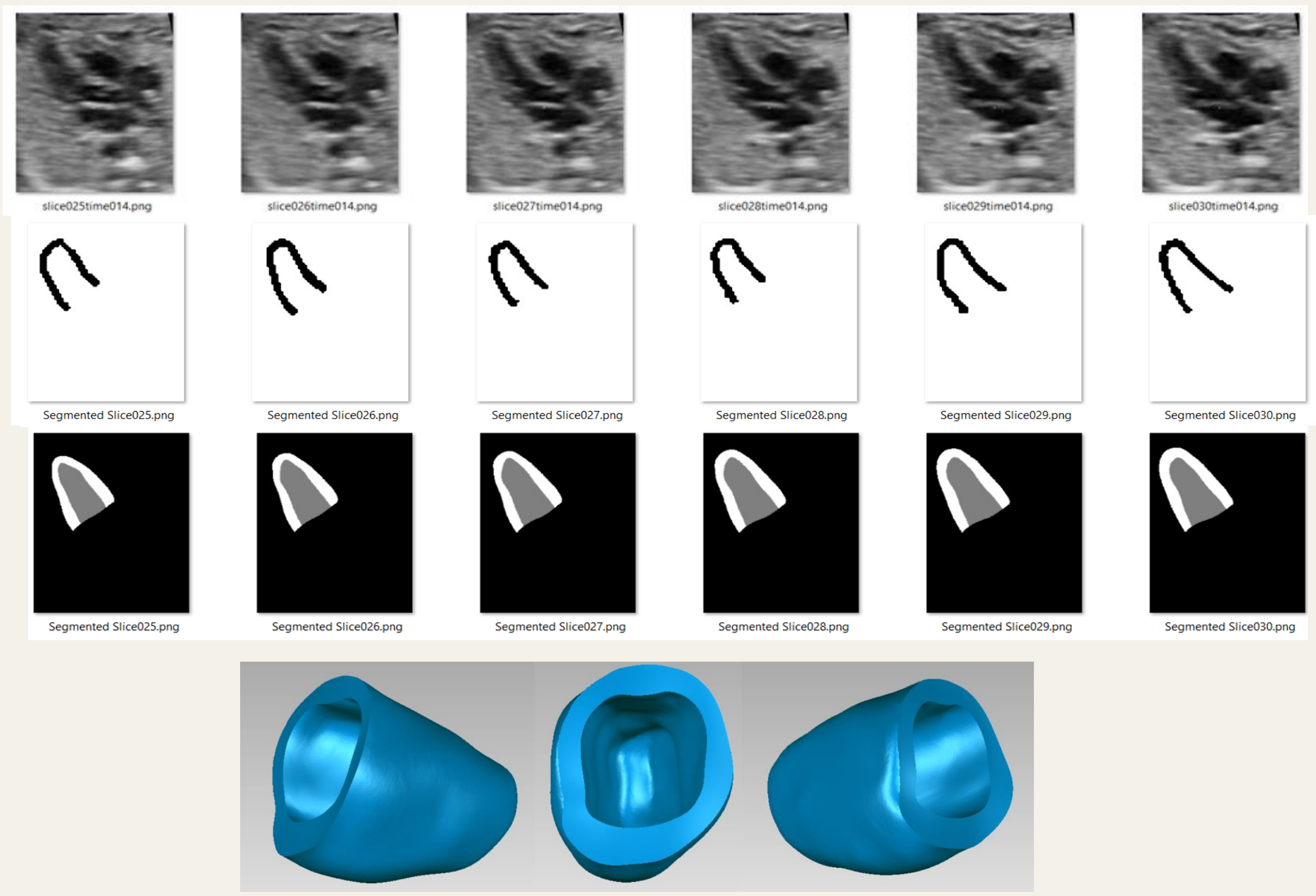

Figure 1: Data annotation steps: i) Preprocessed intensity images, ii) Annotated masks using *Lazysnap*, iii) Masks generated after segmentation step, iv) Final reconstructed 3D left ventricle chamber with myocardium. (In the end, **14** 4D echocardiography images were transformed into a total of **518** 3D images where each of the 3D images holds around **40** 2D slices. As the nifty formatted files are hard to visualize, a sample of slices for image and mask pairs are shown in Figure)

Proposed Approach

The goal is to deform the moving image so that the anatomical location for all the voxels in fixed and moved images will be the same. Deep learning-based image registration (DLIR) neural networks were used to model the displacement field which tried to align the moving image with the fixed image. Several experiments were done and the proposed architecture is shown in Figure 2.

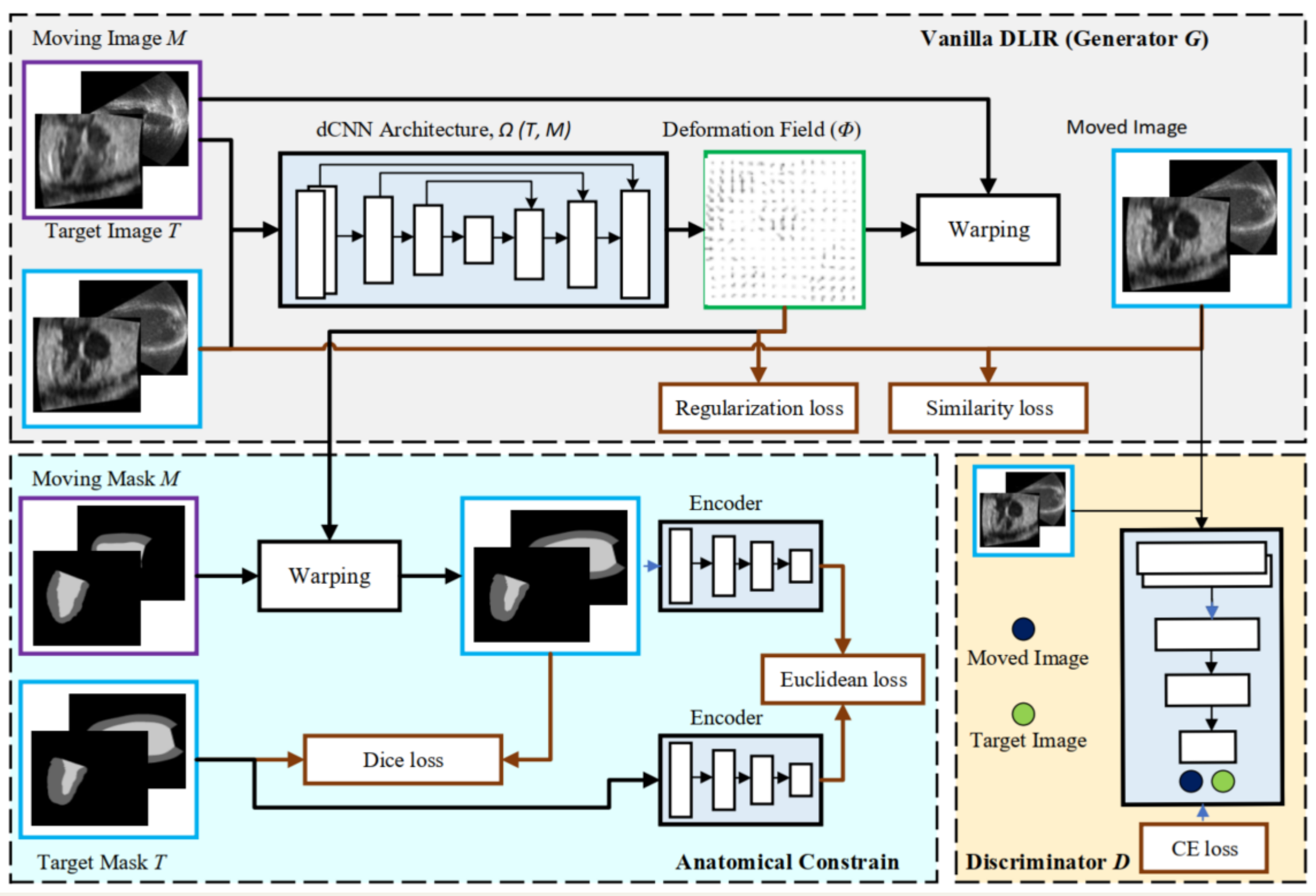

Figure 2: Proposed Architecture of the model to perform temporal image registration using anatomical constraints and adversarial learning.

- **Approach 1 (Vanilla DLIR) :** The underlying architecture of Vanilla DLIR shown in Figure 2 is based on the traditional UNET architecture used for segmentation. The UNET consists of encoding and decoding layers with residual skip connections. The conventional path cannot degrade the features' quality as a non-zero regularizing path will skip over them. On the other hand, the direct skipping of the non-zero regularizing path cannot hamper the performance as it has been added to the conventional path's learned features. The input images and masks are of **256*256*32** size.
- **Approach 2 (AC DLIR) :** To improve the performance of Vanilla-DLIR, global anatomical constraint was added to the model. [4] Also, the latent space was considered as the local segmentation-aware uses pixel-level predictions and may not ensure a satisfactory global match between the warped source and target anatomical masks.[2] Thus, a variational autoencoder was trained to reconstruct the shape of LV and myocardium which was later incorporated in the main model architecture.
- **Approach 3 (Adv DLIR) :** Moreover, the use of the GAN network as a zero-sum game theory could be beneficial for learning deformable fields in image registration. the part of Van-DLIR for generating the deformable images with the produced deformation field was treated as a generator for the adversarial network. In addition to that, a discriminator was also trained which was able to classify the fixed and moved images.
- **Approach 4 (AdvAC DLIR) :** So. Finally both anatomically constrained based variational autoencoder and the adversarial network added to the Van-DLIR to derive the proposed architecture. We've also thought of applying multi-resolution (MACMR) based training where trained parameters on the lower scale will be used to initialize the higher-scale training. We've applied the MACMR for 2D dataset for 3D, it is a part of future plane.

Variational Autoencoder

To compute the global loss from the observations, the segmented masks needed to be transformed into latent space as in Figure 3. A Variational autoencoders(VAE) provide a probabilistic manner to describe the observations in latent space. Encoders learn effective data encoding from datasets and pass it into bottleneck architectures. The autoencoder's decoder employs latent space in the bottleneck layer to generate dataset-like images. These results backpropagate from the neural network in the form of the loss function.

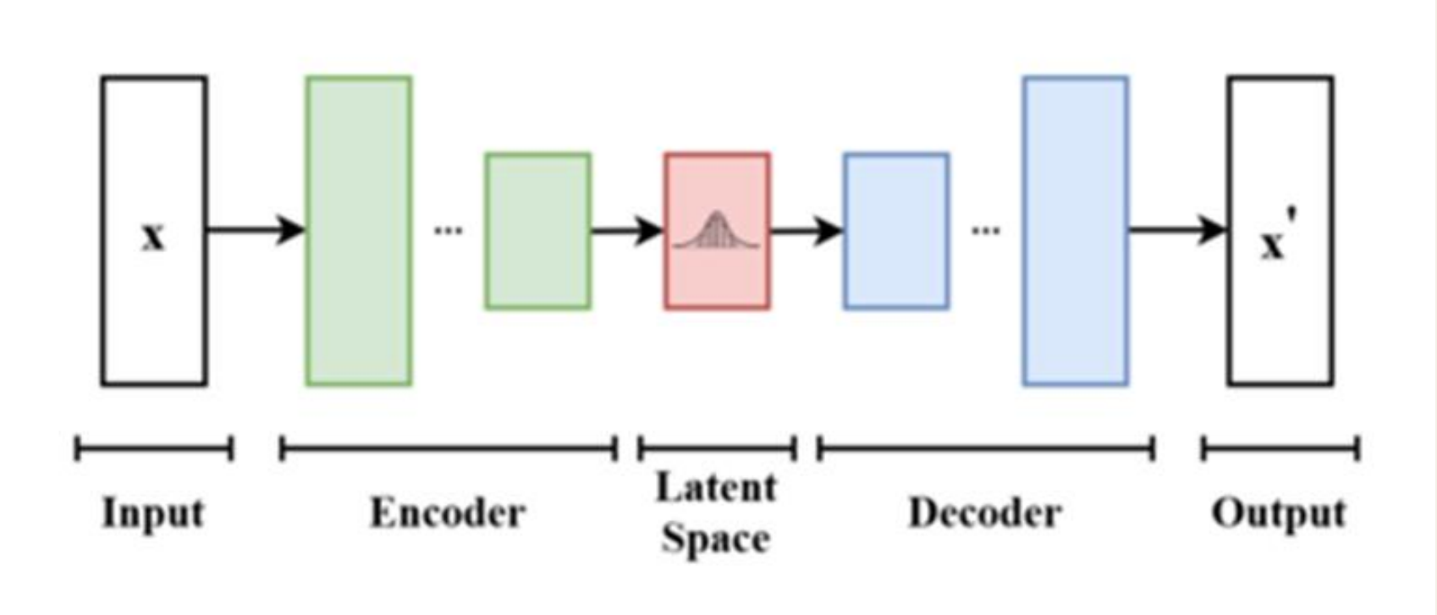

Figure 3: Variational Autoencoder (VAE), latent space consideration.

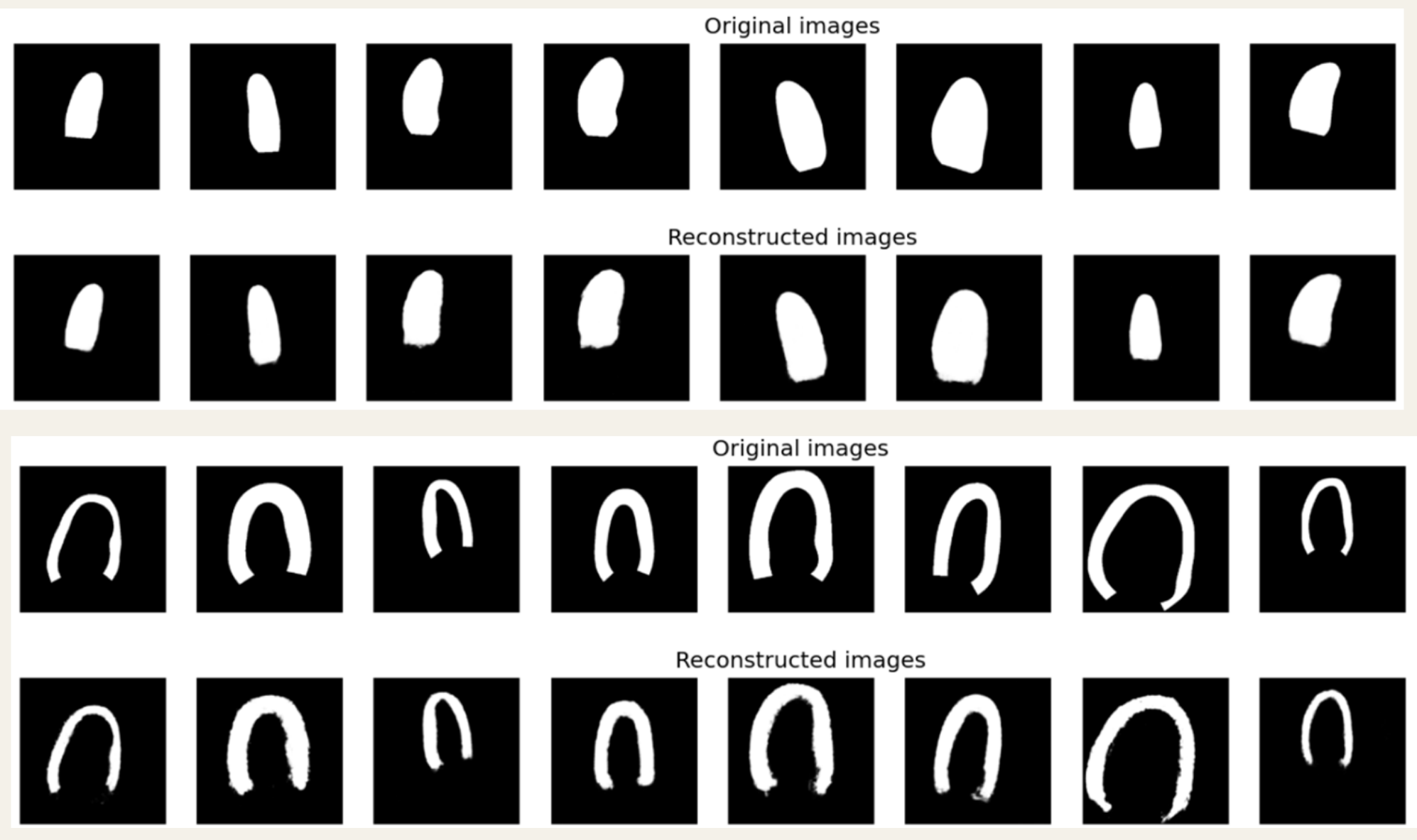

Figure 4 : Results from variational encoder, Reconstructed Myocardium and LV.

Proposed Loss Functions

$$\mathcal{L}_{adac}(f,m,r_f,r_m,d,s_m) = \mathcal{L}_{us}(f,m,d) + \beta\mathcal{L}_{dice}(r_f, r_m \circ d) + \gamma\mathcal{L}_{L2}(r_f,r_m) + \phi\mathcal{L}_g(m,s_m)$$

$$\mathcal{L}_{us}(f,m,d) = \mathcal{L}_{\text{sim}}(f, m \circ d) + \lambda\mathcal{L}_{\text{smooth}}(d)$$

$$\mathcal{L}_{va}(i,r,P(z),Q(z|x)) = \mathcal{L}_{dice}(i,r) + \mathcal{L}_{L2}(i,r) + \mathcal{L}_{SSIM}(i,r) + \mathcal{L}_{KL}(P(z),Q(z|x))$$

Here, β,γ,φ,λ are all regularization parameters.

- **Equation 1** denotes the total loss function for proposed AdvAC model. Where, L_{us} denotes the loss function of **unsupervised** Van-DLIR, L_{dice} denotes the **dice** score between target and moving mask, L_{l2} denotes the **Euclidian** distance between the target and moving masks in the latent space. And, L_g denotes the loss (Binary Cross Entropy) for the **generator** in Adverserial network.
- **Equation 2** denotes the unsupervised loss which is comprised of the structural similarity loss (L_{sim}) and binding energy loss (L_{smooth}).
- **Equation 3** shows the loss function of variational autoencoder which is the summation of dice-sore, structural similarity loss, Euclidian loss and KL-divergence loss (used for regularization).

Experimental Results

Model	MSE	Dice Score			Mean Dice ±std
		Background	LV	Myo	
Without Registration	0.00377	0.99093	0.78917	0.72605	0.83539± 0.12798
Vanilla-DLIR	0.00296	0.98699	0.70087	0.58543	0.75776± 0.04036
AC-DLIR	0.00251	0.98959	0.73347	0.64435	0.80013± 0.05401
Adv-DLIR	**0.00339**	**0.99031**	**0.73836**	**0.67389**	0.80989± 0.05142
AdvAC-DLIR	**0.00258**	**0.99089**	**0.79884**	**0.73482**	0.84668± 0.04586

Table 1: Comparison of proposed registration models on Fetal 3D Dataset.

Model	MSE	Dice Score			Mean Dice ±std
		Background	Myo	LV	
Without Registration	0.00972	0.96678	0.69391	0.76046	0.80235± 0.05491
Vanilla-DLIR	0.0042	0.97352	0.74977	0.87523	0.88487± 0.03261
AC-DLIR	0.00598	0.97972	0.81437	0.91935	0.90303± 0.03447
Adv-DLIR	0.00533	0.97429	0.79278	0.86842	0.85733± 0.04129
AdvAC-DLIR	**0.00589**	**0.98742**	**0.82751**	**0.93573**	**0.91689± 0.02596**
MACMR	**0.00489**	**0.98779**	**0.84871**	**0.95423**	**0.94245± 0.02474**

Table 2: Comparison of proposed registration models on CAMUS 2D Dataset.[3]

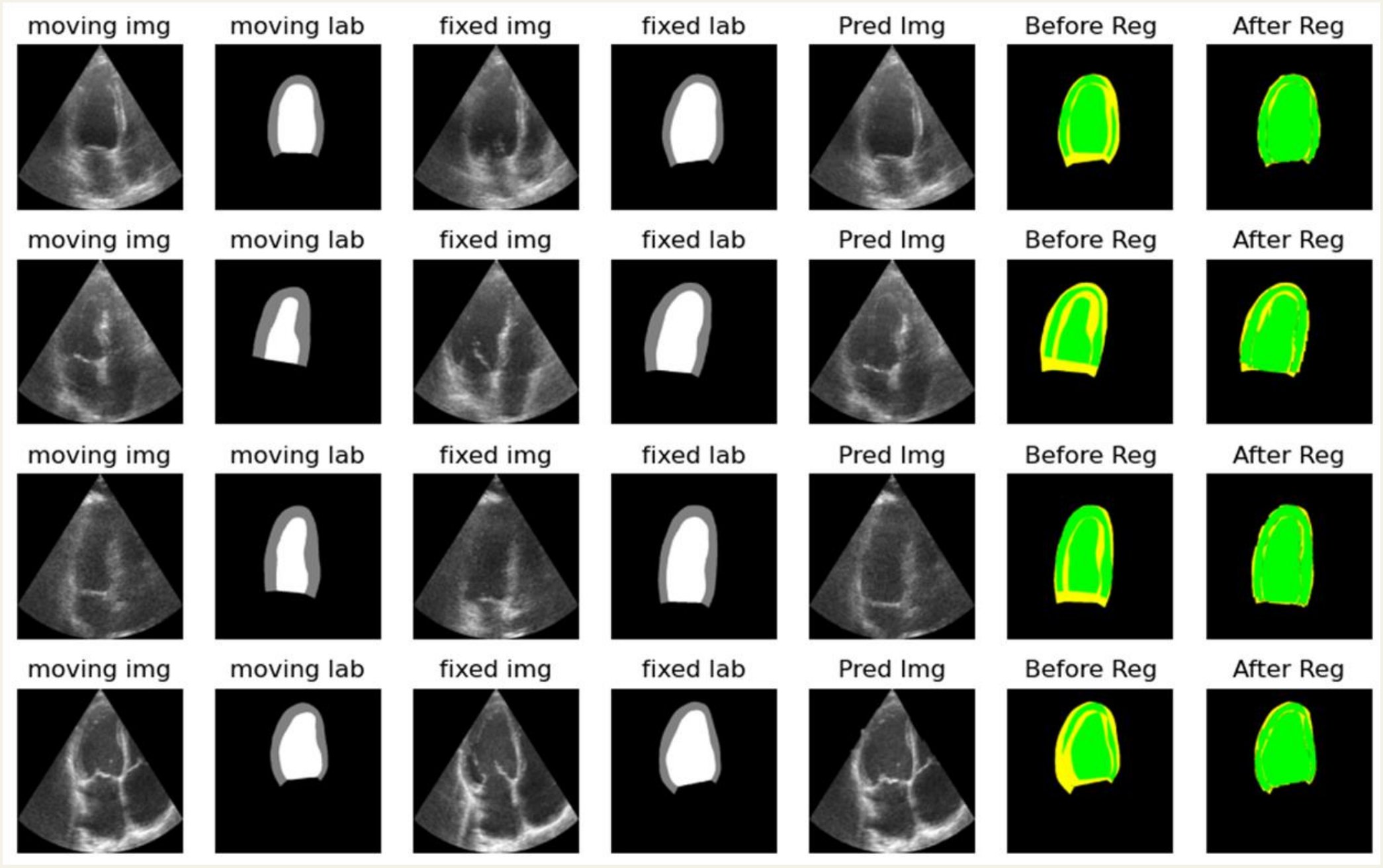

Figure 5: Visualization of the results with overlapping of target and moving mask slices.

Outcomes and Conclusions

- The experiments were done both for 2D Camus and 3D fetal dataset.
- The comparison was shown starting from Vanilla-DLIR to finally the proposed AdvAC model. In case of 2D, the improvement for multi-resolution (MACMR) was also shown in the table.
- We can see, with only the unsupervised Van-DLIR, without considering the anatomy, the similarity between two intensity images increases, but the similarity between fixed and moved masks does not improve satisfactorily or fail in some cases.
- In case of AC-DLIR, Local and global anatomical constraints were added both using latent space consideration which improved the dice and reduced the MSE of simple Van-DLIR. So, adding anatomically constraint is beneficial both for 2D and 3D.
- Finally, after adding the adversarial network's discriminator the dice for LV and Myocardium further improves as generator is trying to generate images more like the target image.
- The results for 2D improves significantly and thus the similar trend of improvement can be observed for 3D as well. Fewer data for 3D could be the reason of slightly low dice as there is still space for further improvement for 3D cases.
- Also, applying multi-resolution based model for 3D is part of future plan.

References

[1]. Wiputra, H., Chan, W.X., Foo, Y.Y., Ho, S., Yap, C.H., 2020. Cardiac motion estimation from medical images: a regularisation frame-work applied on pairwise image registration displacement fields. Scientific Reports 10

[2] Oktay, O., Ferrante, E. et al., 2017. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. IEEE Transactions on Medical Imaging 37, 384–395

[3] Leclerc, S., Smistad, E., et al., 2019. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Transactions on Medical Imaging 38, 2198–2210. doi:10.1109/TMI.2019.2900516

[4] Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J.V., Dalca, A.V., 2018. Voxelmorph: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging 38, 1788–1800.