Cardiac Landmark Detection using Generative Adversarial Networks from Cardiac MR Images

Aparna Kanakatte[1]
aparna.kg@tcs.com
Divya Bhatia[1]
bhatia.divya@tcs.com
Pavan Reddy[1]
pavank.reddy@tcs.com
Jayavardhana Gubbi[1]
jay.gubbi@tcs.com
Avik Ghose[2]
avik.ghose@tcs.com

[1] Tata Consultancy Services Research and Innovation, Bengaluru, India
[2] Tata Consultancy Services Research and Innovation, Kolkata, India

Abstract

Anatomical landmarks are very important for the structural and functional analysis of the heart. Cardiac magnetic resonance (CMR) images have advanced to become a powerful non-invasive diagnostic tool in clinical practice. The first step in many medical imaging applications is to detect anatomical landmarks accurately. The manual identification of these landmarks is difficult due to their shape and appearance variations across populations and also in the presence of anomalies making it time-consuming and operator dependent. We present a GAN-based landmark detection network that can detect smaller objects including landmarks with greater accuracy across varied sample sizes using a proposed modified loss function. The proposed method outperforms other methods reported in literature when trained and tested on the STACOM LV landmark detection challenge dataset. This improved performance is achieved by leveraging the power of the GAN architecture to learn more complex features of the objects being detected. The robustness of the proposed approach is demonstrated by obtaining reduced mean error when blind tested on ACDC dataset.

1 Introduction

In cardiology, precise information on both the dimensions and functions of the heart chambers is essential in clinical applications for diagnosis, prognosis, and therapeutic decisions. Cardiac MR is considered the gold standard for the non-invasive characterization of cardiac function, primarily due to its high spatial resolution and 3D capabilities. It has proven to be an invaluable tool for the diagnosis of complex cardiomyopathies. Although cardiac MR imaging technologies have rapidly advanced, image analysis and interpretation of cardiac

images are time-consuming and error-prone due to the involvement of human operators. Reliable anatomical landmark detection is an important first step for many medical imaging algorithms. A landmark or local feature is a specific image location that serves as a fixed reference. Local features can be corners, edges, or image regions. Particularly in medical imaging, these landmark points act as individual anchor points that help in interpreting the image and understanding the location of one anatomical structure in relation to another. These landmarks can be used in registration, motion tracking, segmentation, building 3D models, and other applications. These landmarks facilitate robust and precise functional and structural analysis of the heart and also helps in accurate surgical pre-planning. However, accurate automatic detection of landmarks in medical images is challenging due to anatomical variation among patients and also differences in image acquisition. In clinical practice, manual delineation by cardiologists remains the main approach to quantifying cardiac function. A recent study showed a detailed manual analysis and annotation by an expert can take 9 to 19 minutes [6].

Learning-based object detection approaches have been demonstrated successfully in many applications. However, they still encounter challenges in a cluttered environment, such as landmark detection in cardiac MR long-axis slices, due to large anatomy shape and appearance variations across populations along with different acquisition parameters. Several organs in the body in addition to the heart appear in the same slice. For the same patient, time sampling across the entire heartbeat cycle, with end-systole and end-diastole as two ends, also leads to significantly different myocardium contour shape changes. These variations and ambiguities result in challenges for each landmark detector to identify correct landmarks. The need for accurately detecting the landmarks is very crucial for medical applications as a few pixel error maps to very high millimeters which can alter the outcome of surgical procedures. Aparna *et al*. [2] have shown the effectiveness of GAN in accurate small object segmentation from cardiac CMRI images. GAN is a robust algorithm with 2 complex networks working against each other to ensure better convergence. The need for a robust, reliable, and accurate system motivated us to explore GAN for landmark detection on both long and short-axes imaging views with great consistency. The major contributions of our work are summarized as follows.

- Creation of a Generative Adversarial network (GAN) based framework for accurate and reliable cardiac landmark detection.
- Creation of a new loss function (Foreground pixel loss) in the discriminator to strengthen the real and fake detection which in turn creates a better generator predicting reliable landmarks.
- Obtaining good accuracy when blind tested on a new dataset has proven the robustness of our proposed GAN network.

2 RELATED WORK

Landmark detection using deep learning has not been extensively tried for CMR images but has been investigated for computer vision applications, such as facial key point detection [8] or human pose estimation [15]. There are some works of literature on GAN being used to synthesize faces using facial landmarks [3, 10]. Payer *et al*. [7] proposed Appearance-Spatial-Combination Network to incorporate local and global information while regressing

landmark coordinates. Pavan *et al*. [13] have combined local and global features in a deep learning framework and demonstrated their approach results in detecting landmarks from skull, spine, and hand X-ray images. A handful of researchers have tried to automate landmark detection from cardiac MR images. Mahapatra [12] has proposed a 2-stage process for landmark detection by first segmenting the left ventricle (LV) or right ventricle (RV) and then examining the regions for landmark points using random forest classifiers. Lu *et al*. [11] used a discriminative joint context for landmark detection. The above two works used the same STACOM dataset [9] as the one we have used. Both of these have reported high pixel errors. Xue [17] and Wang [16] have used CNN for detecting landmarks on their private cardiac MRI dataset.

In spite of the latest developments, the results are not accurate enough for building clinically usable applications. Considering the need for higher accuracy and reliability in biomedical applications and also the increased complexity in identifying pixel-level anatomical landmarks in abnormal or pathological conditions have motivated us to implement a GAN-based framework. We have designed the unique encoder-decoder architecture along with the generative mechanism of image translation by incorporating the newly designed Foreground pixel loss function that can be extended to any small object detection problem. Most reported work trains each landmark as a separate image in case of multiple landmarks detection for better accuracy [5, 13]. However, our proposed GAN-based network takes multiple landmarks in a single image during training thus reducing computation complexity in terms of time and space requirements without compromising on the accuracy.

3 PROPOSED GAN ARCHITECTURE

GAN comprises two competing neural networks called the generator which generates new synthetic data and a discriminator which assess whether the generated data is close to the real data. This network is predominantly used in data creation. However, GANs are being explored for other applications due to their robustness once trained [2]. In this work, we propose a GAN with modified loss functions to predict cardiac landmarks with increased accuracy and precision. The input to the network is the original image and the output is the heatmap of the landmark.

Heatmaps are continuous pixel spreads representing the spatial probability of each landmark point when it is convolved with a Gaussian kernel of some standard deviation [13]. As our network handles multiple heatmaps in a single image during training, we ensured no two landmark heatmap distributions are overlapped, which was one of our challenges when generating the heatmap image. It is prone to localization errors and overfitting when the variance is too small and also there is no global or spatial context. Due to this, depending on the number of landmarks in an image, we choose different values of variance to maintain the trade-off between maintaining the proper spacing between landmarks and also in making the distribution large enough for proper detection. Therefore, by experimenting, we have set the variance of 7 for a single landmark and 5 for multiple landmarks while generating heatmaps. These heatmaps are normalized in the range $0-1$.

The proposed Unet-based GAN generator consists of an encoder and decoder with skip connections along with a feature-filter enhancing block as shown in Fig. 1. The encoder consists of 2D convolutional layers and represents key features from the input image as vectors in latent space. Skip connection concatenates the up-sampled vector in the decoder path with the symmetrically opposite output vector in the encoder path along the channel

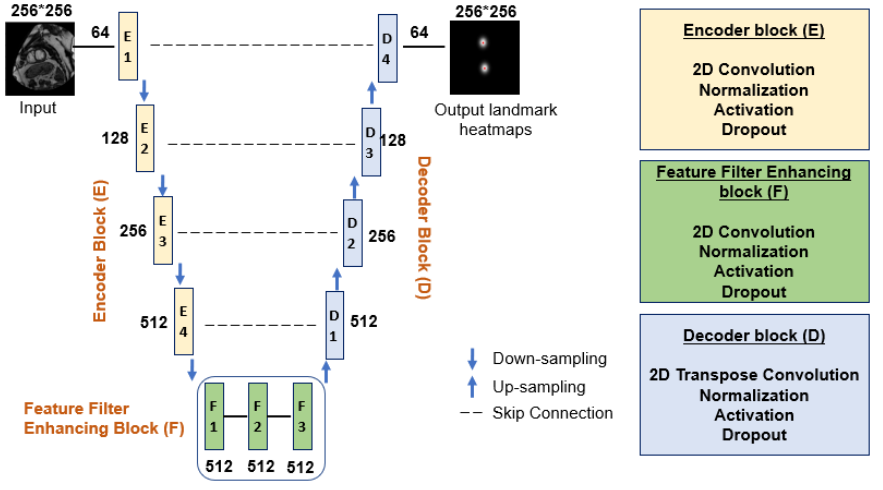


Figure 1: Flow chart of the proposed Unet based GAN Generator

axis. We observed low bias and high variance when modeling the data. This occurs when the data has a variety of features, and the model takes into account all estimated coefficients and attempts to overestimate the actual value. However, only a few features are important and affect the prediction. To select these relevant feature sets, we designed the three-block feature filter enhancing block. This along with skip connections confine the features, decreases parameters, simplifies the model, minimizes the probability of overfitting during training, and regularizes the network. The proposed modified generator loss G_L is given in eq 1.

$$\mathbf{G_L} = A_L + \beta \times L_L \tag{1}$$

where A_L is adversarial loss defined in eq 2, β is the intensity parameter and L_L is learning loss given in eq 3.

$$\mathbf{A_L} = MSE(\mathrm{I}, GP_h) \tag{2}$$

$$\mathbf{L_L} = Huber(GT_h, GP_h, \delta = 0.4) \tag{3}$$

where I is the input image, GP_h is the generator predicted heatmap, GT_h is the ground-truth heatmap and MSE is the mean squared error. In this study, we use the Huber loss function, which combines both MAE and MSE properties. Like the MAE, it is robust to outliers, as it is not heavily influenced by extreme values in the data, and like the MSE, it penalizes large errors heavily. Consequently, the Huber metric optimizes using both the combination of the median (MAE) and the mean (MSE) according to the δ (which is set to 0.4) value and also the size of the errors.

The proposed GAN discriminator has 2D convolution layers with parameters similar to the encoder of the generator as shown in Fig. 2. It has two sets of inputs, the original image and the predicted landmark heatmap from the generator along with the original image and the ground-truth heatmap. This discriminator is built on patch GAN architecture style. It splits the raw input image into small local patches of size 4×4, then runs a general discriminator convolutionally on every patch declaring whether the patch is real or fake. The final prediction is the average of all the patch responses.

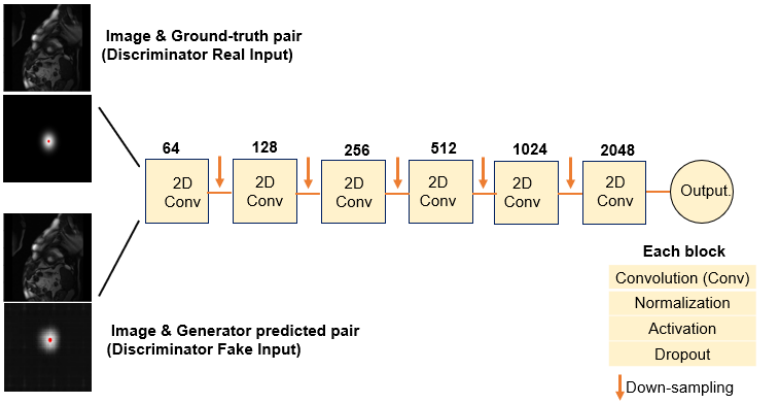


Figure 2: Flow chart of the proposed Discriminator of our GAN

In GAN networks, the generator and discriminator work in tandem to provide better accuracy. Discriminator helps in making the generator efficient by identifying the real and fake data. The standard discriminator loss (SDL) is given by eq 6 which includes a discriminator real loss (DRL) and discriminator fake loss (DFL) as given in Eqs. 4 and 5.

$$\mathbf{DRL} = MSE(\mathrm{I}, GT_h) \tag{4}$$

$$\mathbf{DFL} = MSE(\mathrm{I}, GP_h) \tag{5}$$

$$\mathbf{SDL} = 0.5 \times (DRL + DFL) \tag{6}$$

However, it has to be noted that the majority of pixels will be background when it comes to small object detection. Foreground pixels that contribute to deciding will be very few. This makes the discriminator pass the image or patch as true even if the foreground/key pixels are missing or in the wrong location if we use the standard discriminator loss function as given in Eq 6. This is shown in the second instance in Fig. 3 thereby affecting the landmark detection accuracies.

Standard approaches fail to generate accurate results because they do not incorporate local information about pixels and their positions. So to overcome this we have introduced a new Foreground Pixel Loss (FPL) function in the discriminator which identifies the foreground region or pixels of landmark heatmap in GT data and creates a square bounding box (X_1, Y_1 and X_2, Y_2) patch as shown in Fig 4. To identify this foreground region we create the peak distribution of the entire heatmap image and select the patches that have maximum coverage of foreground pixels. The size of the patch is decided on the run time. The same square coordinates patches will be taken in the predicted heatmap from the generator for comparison towards penalization of the network for better convergence and discriminator gradient updation during training. Three instances of prediction comparison with GT image are shown in Fig 4 wherein the discriminator has passed only the closest one to GT and eliminated the other two as fake. It has to be noted that without the proposed FPL loss second instance in Fig 4 would have passed as true even though the pixel distribution is not similar to the GT image.

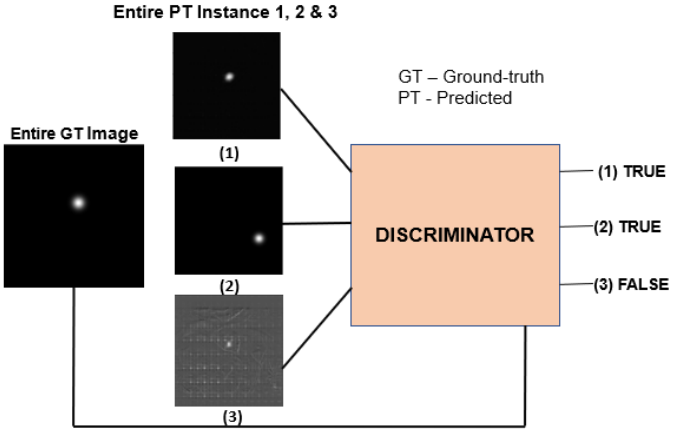


Figure 3: Failure of discriminator in small object detection. Standard discriminator loss has passed instance (2) as true even though the location of the heatmap is in the wrong place.

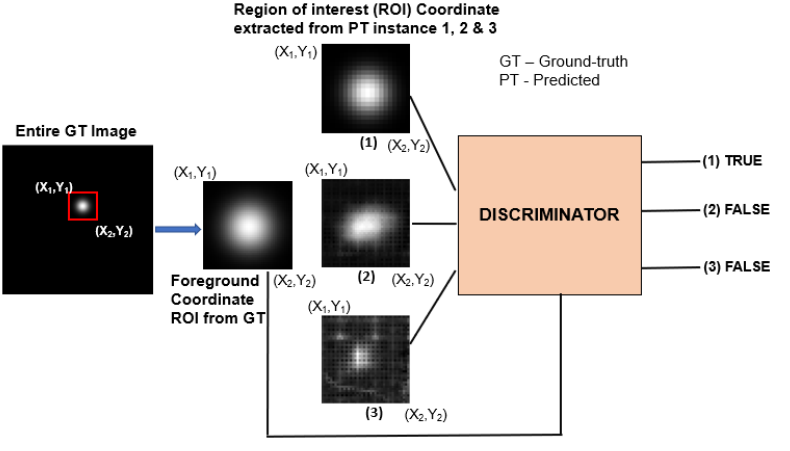


Figure 4: Proposed discriminator FPL loss overcoming the small object detection problem shown in Fig 3. FPL loss has flagged instance (2) as false which would have passed as true if using SDL loss.

The calculation of the proposed FPL is given in Eq. 7

$$\mathbf{FPL} = \alpha \times MSE(P_1, P_2) + MAE(P_1, P_2) \tag{7}$$

where P_1 is the ground-truth patch and P_2 is the predicted patch. α is a crucial scalar coefficient hyperparameter that controls overfitting and works as a regularizer between the two losses. The MAE gives sparse solutions, performs regularization, removes redundant features, and is robust to outliers. MSE learns complex data patterns, has analytical solutions, and penalizes large errors more heavily. The weighted normalized mean loss function of both MAE and MSE allows us to maintain the tradeoff between handling outliers and penal-

izing big error terms. The modified discriminator loss (MDL) is a combination of SDL and FPL as shown in Eq. 8.

$$\text{MDL} = SDL + FPL \tag{8}$$

Thus, by taking the local information using the foreground patch, introducing the FPL loss function, and modifying generator loss using Huber we are collaborating on both the local and global information without any complex network. Dynamic hyperparameters are included by modifying the dropout and learning rate for both generator and discriminator networks based on the loss value during training for better generalization towards increasing precise landmark detection accuracy.

4 Results and Discussions

We have trained and tested the proposed method on the STACOM LV landmark detection challenge dataset [9]. The training data consists of 100 patient images acquired in both the long and short-axis views. The initial split in data is $70:30$ for training and testing, then from the 70% we divide to $80:20$ for training and validation part. We have performed cross-validation to ensure every patient data is part of training and testing. The dataset had 6 distinct landmark annotations; 2 from Mitral valve, 2 from RV insert points and one each from base and apex central axis points. These landmark points are necessary to build a 3D left ventricle model. All the points were annotated by an experienced analyst. The details about these landmark points are provided below.

Mitral valve (MV) points: MV separates the left atrium (LA) and the left ventricle (LV). This is clearly visible in the MRI long-axis view, as this shows both LA and LV. Two endpoints of this valve define the MV points. A line connecting the MV points (base plane) is crucial for LV volume measurement.

RV insert (RVI) points: Two intersections between LV and RV in the short-axis view defining the septum are usually marked as RVI points. The RVI points are important for 3D cardiac modeling, particularly for biventricular models.

Base-to-apex central axis points (BCA and ACA): Base and apex central axis points are essential to define the LV central axis for 3D LV models. For each patient study, one central point at a basal slice and one central point at an apical slice are needed. Both are defined at the middle of the LV cavity on short-axis MRI.

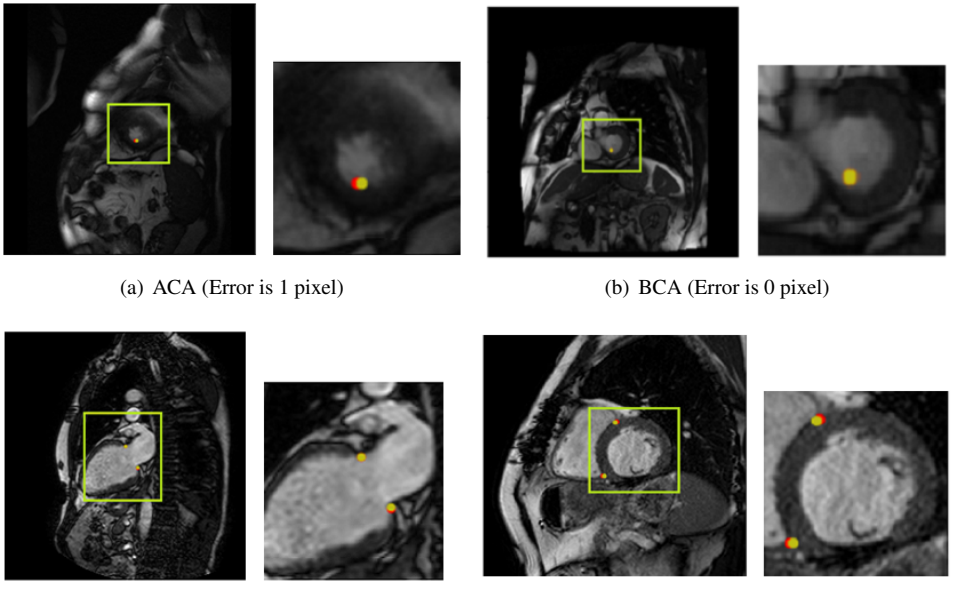
4.1 Preprocessing and Implementation Details

The input image size is normalized to 256×256 as per the network requirement. Images are zero-padded by the boundaries that are less than this size and boundaries cropped if the size is more ensuring that our region of interest is not affected. The pixel values are normalized between $[0,1]$. To increase the training samples and reduce storage dependency, on-the-go elastic, luminance, rotation, and flip data augmentations are applied. The proposed approach is implemented using Tensorflow [1] and OpenCV. In the generator, the encoder has a kernel size of 3×3 with a depth of 4 and stride of 2. We are using a He-Normal kernel initializer with leaky-relu activation and batch normalization. The decoder has a stride of

1. As the proposed architecture generates a single image with N heatmaps for N landmarks, the last layer is modified to have a single filter and stride of 1 with no activation function. The discriminator uses 2D convolution with a depth of 6 and provides an output of patch size 4×4. Both networks use the Adam optimizer with a starting learning rate of $2e-4$ and a starting dropout value of 0.4 for the generator and 0.6 for the discriminator. This gets dynamically adjusted during training [7]. The discriminator is made more dynamic by giving a higher dropout to avoid mode collapse, a common problem while training GAN. Also, a low dropout to the generator helps in convergence and avoids the vanishing gradient problem.

4.2 Performance Analysis

The proposed network is trained using varied sample numbers. ACA and BCA have 80 samples, RVI has 542 and MV has 5142 samples. The number of epochs used for training also varied with 1500 epochs for ACA and BCA, 1000 for RVI, and 500 epochs for MV. Our proposed network provides a single image with N heatmaps for N landmarks. It can be seen that the proposed predicts landmarks very close to GT with an error of around 1 pixel as shown in Fig. 5. The zoomed image of the squared green region next to the image is shown for better visualization of landmarks.



(a) ACA (Error is 1 pixel)

(b) BCA (Error is 0 pixel)

(c) MV (Error is 1 pixel)

(d) RVI (Error is 1 pixel)

Figure 5: Results of landmark detection. Red is the ground-truth while yellow is the predicted point. The zoomed region are displayed next to the image for better visualization of landmarks.

It was observed that selecting the brightest pixel coordinate points as landmarks did not give appropriate results. To obtain accurate and reliable landmark points we are finding circular contours around the heatmaps. Then, by considering the radius or center of the contour, we are able to localize landmark coordinates and generalize them to any number of

landmarks. Euclidean distance between the predicted landmark and the actual landmark is used to calculate the error measures. All the results obtained were visually validated by a clinical expert cardiologist who is part of our team.

Table 1: Average error measures (in pixels) for landmark detection on training datasets. Figures indicate mean and standard deviation.

	ACA	BCA	MV	RVI
Mahapatra [12]	2.2±1.2	3.0±1.6	9.3±2.5	7.4±2.6
Lu [11]	-	6.2±4.0	3.5±5.6	7.9±11.5
Proposed	**1.8±1.2**	**1.6±1.5**	**3.0±1.4**	**2.8±1.5**

The quantitative performance of our proposed method is given in Table 1 and Table 2. Our method is compared by using average error measures in pixels with the top 2 reported results in the landmark detection challenge [9] in Table 1 . The proposed provides an average mean error of about 1.8 pixels for ACA, 1.6 pixels for BCA, 2.8 pixels for RVI, and 3.0 pixels for MV performing better across all landmarks with significant improvements compared to the other reported results on this challenge dataset. It can be seen that our results are consistent even across varied sample sizes for each trained landmarks. We have also computed successful detection rate (SDR) for less than 2, 3 and 5 pixels in Table 2 which shows that the proposed method provides above 80% for all landmarks within 3 pixel error and above 90% for less than 5 pixel error.

Table 2: Average successful detection rate in % for each landmark.

	ACA	BCA	MV	RVI
$\leq$ 2 pixel error	70	75	62	72
$\leq$ 3 pixel error	85	85	88	86
$\leq$ 5 pixel error	95	90	98	96

As there were only limited work reported on this data for comparison we have also performed blind testing using another opensource dataset to test the robustness of our proposed algorithm.

4.3 Blind-testing on ACDC Dataset

We have blind-tested our model on ACDC data [4], which consists of short-axis CMR images from 100 patients with normal anatomy and pathological cases. The RVI landmarks were manually added as circular regions of 5 pixels by Sven *et al.* [14]. It can be seen in Fig. 6 that our method predicts the landmark points within this circular region consistently for all tested images. To compute the error we are considering the center point of the circular region as the ground-truth (GT) landmark. By comparing the predicted point with this GT landmark we obtained an average mean error of 2.3 pixels with a standard deviation of 1.8 pixels when tested on 1000 images with varied pathologies. It has to be noted that we have not used any of the images from this dataset for training our algorithm thus showing the robustness of our method.

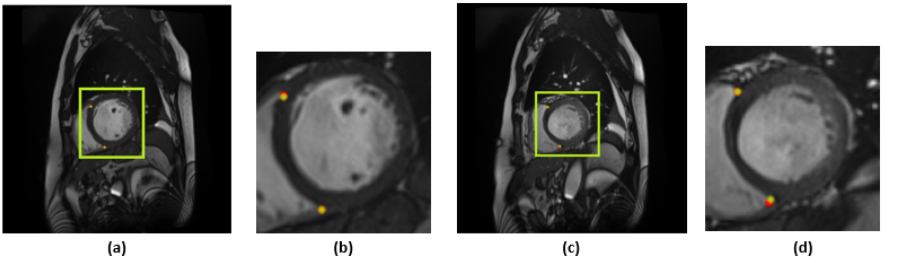


Figure 6: Results of RVI blind-testing with ground-truth (red) and predicted point (yellow). (a) and (c) are input images while (b) and (d) are the zoomed regions for better visualization of landmarks. Both have a 1-pixel error

5 CONCLUSIONS AND FUTURE WORK

By incorporating the newly designed Foreground pixel loss function, we have developed a unique encoder-decoder architecture and a generative mechanism for image translation that can be applied to any small object detection problem. Some related works use higher model parameters that are complex and time-consuming. In contrast our proposed uses Detection-GAN and FPL loss to integrate relevant information about landmark localization and global context with less data and fewer trainable parameters. Since our design is minimalist and simple, our network generates a single heatmap image containing N heatmaps if multiple landmarks are present in the image instead of generating separate heatmap image for every landmark in that image. We have achieved state-of-the-art landmark localization accuracy, and the results are consistent with a limited amount of training data. The method's robustness is shown by achieving the reduced mean error when blind-tested on the ACDC dataset. Our future work includes extending this to detect 3D landmarks.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z Chen, and et al. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015.

[2] K Aparna, B Divya, and G Avik. 3d cardiac substructures segmentation from cmri using generative adversarial network (gan). In *Proc. IEEE EMBS*, pages 1698–1698, 2022.

[3] S Bazrafkan, H Javidnia, and P Corcoran. Face synthesis with landmark points from generative adversarial networks and inverse latent space mapping, 2018.

[4] O Bernard, A Lalande, F Cervenansky, and et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? In *Proc. IEEE Trans on Medical Imaging*, volume 11, pages 2514–2525, 37.

[5] J.B Bhargav, S.N Vighnesh, V.S. Sathvik, and R.A Vijaya. Fundusposnet: A deep learning driven heatmap regression model for the joint localization of optic disc and fovea centers in color fundus images, 2021.

[6] A.N Bhuva, W Bai, C Lau, and et al. A multicenter, scan-rescan, human and machine learning cmr study to test generalizability and precision in imaging biomarker analysis. In *Proc. Circulation: Cardiovascular Imaging*, volume 12, pages 1–11, 2019.

[7] P Christian, S Darko, B Horst, and U Martin. Regressing heatmaps for multiple landmark localization using cnns. In *Proc. of MICCAI*, pages 230–238, 2016.

[8] S Colaco and D.S Han. Facial keypoint detection with convolutional neural networks. In *Proc. IEEE Int Conf Artif Intell Inf Commun*, pages 671–674, 2020.

[9] C Fonseca, M Backhaus, D Bluemke, and et al. The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. In *Proc. Bioinformatics*, volume 27, pages 2288–2295, 2011.

[10] L Hongzhe, Z Weicheng, X Cheng, L Teng, and Z Min. Facial landmark detection using generative adversarial network combined with autoencoder for occlusion. In *Mathematical Problems in Engineering, Hindawi*, 2020.

[11] X Lu and J Marie-Pierre. Discriminative context modeling using auxiliary markers for lv landmark detection from a single mr image. In *Proc. Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, pages 105–114, 2013.

[12] D Mahapatra. Landmark detection in cardiac mri using learned local image statistics. In *Proc. Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, pages 115–124, 2013.

[13] R Pavan, K Aparna, G Jayavardhana, P Murali, G Avik, and P Balamuralidhar. Anatomical landmark detection using deep appearance context network. In *Proc. IEEE EMBS*, pages 3569–3572, 2021.

[14] K Sven, S Lalith, and et al. Comparison of evaluation metrics for landmark detection in cmr images. In *Proc. Springer Workshop on Medical Image Computing, Heidelberg*, pages 198–203, 2022.

[15] J Tompson, R Goroshin, A Jain, Y LeCun, and C Bregler. Efficient object localization using convolutional networks. In *Proc. IEEE Comput Soc Comput Vis Pattern Recognit*, pages 648–656, 2015.

[16] X Wang, S Zhai, and Y Niu. Left ventricle landmark localization and identification in cardiac mri by deep metric learning-assisted cnn regression. In *Proc. Neurocomputing*, volume 399, pages 153–170, 2020.

[17] H Xue, J Artico, M Fontana, and et al. Landmark detection in cardiac mri using a convolutional neural network. In *Proc. Radiol Artif Intell*, volume 3:e20019, 2021.