



HAL
open science

Shape Constraints in Deep Learning for Robust 2D Echocardiography Analysis

Yingyu Yang, Maxime Sermesant

► **To cite this version:**

Yingyu Yang, Maxime Sermesant. Shape Constraints in Deep Learning for Robust 2D Echocardiography Analysis. FIMH 2021 - 11th International Conference on Functional Imaging and Modeling of the Heart, Jun 2021, Stanford, United States. 10.1007/978-3-030-78710-3_3. hal-03371358

HAL Id: hal-03371358

<https://hal.science/hal-03371358>

Submitted on 8 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Shape Constraints in Deep Learning for Robust 2D Echocardiography Analysis

Yingyu Yang¹ and Maxime Sermesant¹

Inria, Université Côte d’Azur, Sophia Antipolis, France

Contact: {firstname}.{lastname}@inria.fr

Abstract. 2D Echocardiography is a popular and cost-efficient tool for cardiac dysfunction diagnosis. Automatic solutions that could effectively and efficiently analyse cardiac functions are highly desired in clinical situations. Segmentation and motion tracking are two important techniques to extract useful cardiac indexes, such as left ventricle ejection fraction (LVEF), global longitudinal strain (GLS), etc. However, these tasks are non-trivial since ultrasound images usually suffer from poor signal-to-noise ratio, boundary ambiguity and out of view problem. In this paper, we explore how to introduce shape constraints from global, regional and pixel level into a baseline U-Net model for better segmentation and landmark tracking. Our experiments show that all the three propositions perform similarly as the baseline model in terms of geometrical scores, while our pixel-level model, which uses a multi-class contour loss, reduces segmentation outliers and improves the tracking accuracy of 3 landmarks used for GLS computation. With appropriate augmentation techniques, our models also show a good generalisation performance when testing on a larger unseen cohort.

Keywords: Segmentation, Deep Learning, Deformation, Echocardiography

1 Introduction

Echocardiography, a non-invasive and cost-efficient imaging technique, is widely used by cardiologists to evaluate the cardiac function. Segmentation and motion tracking are two essential tasks that can help cardiologists in clinical decision-making. Segmentation offers information of shape and volume while motion tracking provides knowledge of deformation and function.

Deep learning based methods have shown very good performance for medical segmentation and registration. As for segmentation, the U-Net architecture has proved its overwhelming power in large cohort echocardiography segmentation [11]. With appropriate adaptation of U-Net model and data augmentation, the U-Net architecture also demonstrated good generalisation ability in segmenting cardiac magnetic resonance images (CMRI) [5]. However, echocardiography segmentation is still difficult since it usually encounters the problem of out of view, poor signal-to-noise ratio etc, especially for myocardium segmentation.

In the field of cardiac motion tracking, unsupervised deep learning are very popular and these schemes reach similar performance or even outperform traditional registration methods. Krebs et al. proposed a conditional variational autoencoder which learned a diffeomorphic transformation from pairwise CMRI in an unsupervised way [10]. Shawn et al [1] designed a U-Net like network for unsupervised pairwise echocardiography motion tracking. However, the displacement field can be unrealistic without relevant regularisation.

Numerous studies have shown that global longitudinal strain (GLS) is more sensitive than left ventricular ejection fraction (LVEF) as a measure of systolic function and has potential in identifying left ventricle dysfunction in clinical[9][7]. This value can be approximated by measuring the left ventricle length change [14]. Therefore, estimating a dense displacement may not be necessary.

U-Net like deep learning models depend largely on pixel-level classification, which can generate artefacts which are irregular with the organ shape. Researchers are seeking to combine shape constraints with deep learning methods [3]. With the same intention to improve the segmentation consistency with anatomical shapes in 2D echocardiography, in our work, we explore to introduce shape constraints from global, regional and pixel level into a baseline U-Net model. From the segmentation results, useful information such as ejection fraction, landmark based GLS can be extracted. The detailed model architecture will be explained in section 2. We then present the implementation and experiment results for segmentation and landmark detection in section 3.

2 Methods

We use a U-Net model as our baseline model. Its encoder consists of 5 down-sampling (MaxPool + Conv) blocks with ReLU activation after the 3x3 convolution. The corresponding decoder has 5 up-sampling (UpSample + Conv) blocks and is skip-connected with the encoder. Based on this model, we consider to incorporate shape constraints from three levels:

- **Global-level:** estimate a triangle like landmark map in parallel with segmentation (SEG-LM)
- **Regional-level:** add a poly-affine myocardium reconstruction network to constrain the shape of myocardium mask (SEG-AFFINE)
- **Pixel-level:** use a multi-class contour-loss to finely classify the boundary pixel (SEG-CONTOUR)

We will explain the three methods in detail in the following subsections.

2.1 SEG-LM: Parallel segmentation and landmark detection

We adapt the baseline U-Net model for simultaneous segmentation and landmark prediction by adding a separate branch of decoder for landmark map estimation. The two decoders process the encoder information in parallel. The final layer of segmentation branch and landmark detection have SoftMax activation, Sigmoid

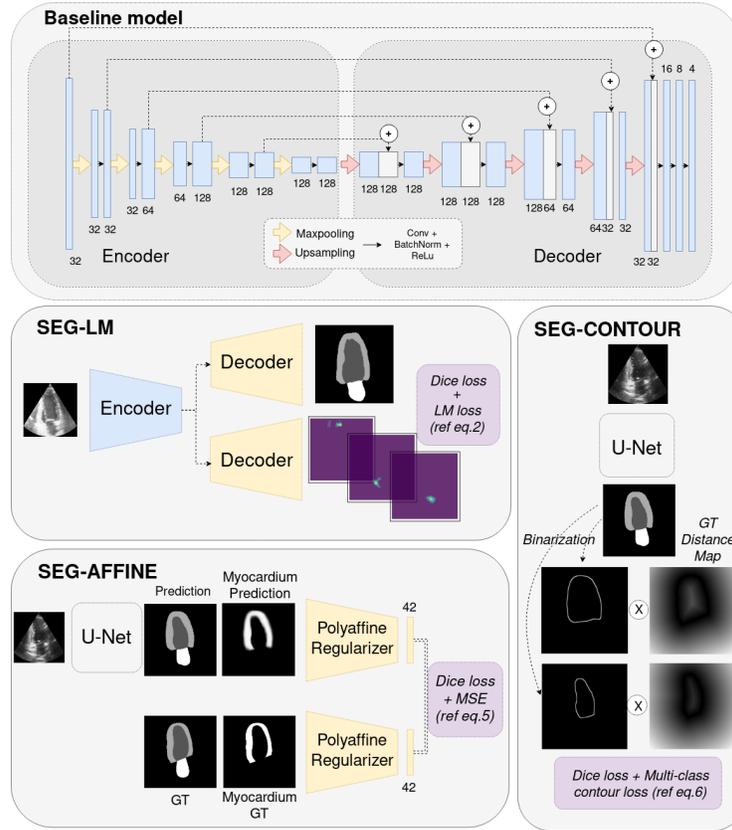


Fig. 1: Detailed information of the 4 explored methods.

activation respectively. In particular, we consider the two end points of mitral valve (basal points) and the apex along the endocardial contour. The basal points are identified as the two end-points of adjacent boundary of both left ventricle and left atrium. The endo-apex point is then calculated as the furthest point to the mid-basal point along the endocardium. The output of landmark detection network is a heatmap of the corresponding target point. From the output heatmap, we extract the landmark position by finding the location of maximum or computing the centroid.

As we have different labels in the ground-truth data (myocardium, blood pool, atrium), a multi-class dice \mathcal{L}_{dice} is used as the segmentation loss. As for landmark detection, we first penalise on the squared error of landmark heat-map (L2 loss: \mathcal{L}_{l_2}). In order to avoid landmark overlapping on different output layers, we regularise the centre distance loss \mathcal{L}_{CD} of different landmark heat-maps as

proposed in [15]:

$$\mathcal{L}_{CD} = \frac{1}{2} \sum_{i=1, j \neq i}^3 1/(C(\mathcal{H}_i) - C(\mathcal{H}_j))^2 \quad (1)$$

with C the operation to obtain the centre position and \mathcal{H}_i the landmark heatmap. Finally, a mean squared distance loss between the predicted heatmap centre and the ground truth point \mathcal{L}_{point} is also applied. Thus we have a total loss \mathcal{L}_{total} for optimisation:

$$\mathcal{L}_{total} = \mathcal{L}_{dice} + \alpha \mathcal{L}_{l2} + \beta \mathcal{L}_{CD} + \gamma \mathcal{L}_{point} \quad (2)$$

2.2 SEG-AFFINE: Poly-affine Regulariser for Myocardium

With the intention to constrain the regularity of predicted myocardium mask, we propose to model the myocardium mask as a combination of 6 AHA regions [4]. We first choose a reference myocardium mask R from the training set. All the N training myocardium masks $(M_i)_{i=1}^N$ are aligned to the reference mask by an affine transform $(T_i)_{i=1}^N$ estimated from the three landmarks (left basal, right basal and endo-apex). Then all aligned masks are averaged to a mean mask \bar{R} . The mean mask \bar{R} is threshold-ed (\bar{R}^f) and split into 6 AHA regions $(\bar{I}_j)_{j=1}^6$. For every myocardium mask M_i , we aim to first find 1 affine matrix A_g that globally transform the reference mask to \hat{M}_i^g . The corresponding reference regions become $(\bar{I}_j^g)_{j=1}^6 = A_g \bar{I}_j$. We then find 6 affine matrices $(A_{ij})_{j=1}^6$, that transform the transformed (globally) mean AHA regions $(\bar{I}_j^g)_{j=1}^6$ into $\hat{M}_i = \sum_{j=1}^6 A_{ij} \bar{I}_j^g$ that best reconstructs the target mask M_i , i.e. $\hat{M}_i \approx M_i$. For better fusion of the transformed 6 regions, we use the spatially weighted regions (multi-variate Gaussian) \tilde{I}_j^w instead of \bar{I}_j , s.t. $\sum_{j=1}^6 \tilde{I}_j^w = \bar{R}^f$.

We use a CNN to estimate the affine parameters and reconstruct the given mask. The proposed network has two sub-networks. The first one seeks to estimate global affine parameters for global alignment which consists of two hidden convolutional layers with down-sampling. The second sub-network is to find the 6 regional affine matrix. It begins with an encoder for high level feature extraction. The extracted features are passed through fully connected layers for affine matrix estimation \hat{A}_{ij} . By affine transform, we could reconstruct \hat{M}^0 from the 6 mean regions $\hat{M}^0 = \sum_{j=1}^6 \hat{A}_{ij} \tilde{I}_j^{wg}$. Two Conv layers are followed to refine the fusion mask \hat{M}^0 and thus we obtain the final output \hat{M}^f (detailed information in fig.2).

In order to regularise the value of the affine parameters, we approach the affine parameter estimation problem using *Maximum A Posteriori* (MAP) with prior probabilities on the parameter values $P(A)$. The MAP aims to optimise: $\arg \max[P(A|M)] \propto P(M|A)P(A)$, i.e. $\arg \min[-\log P(M|A) - \log P(A)]$ We use Gaussian distributions for conditional likelihood and priors. For $P(M|A) \propto \exp(-\frac{1}{2}(M - \hat{M}(A))^T \Sigma^{-1}(M - \hat{M}(A)))$, the variance is identity. For $P(A) \sim$

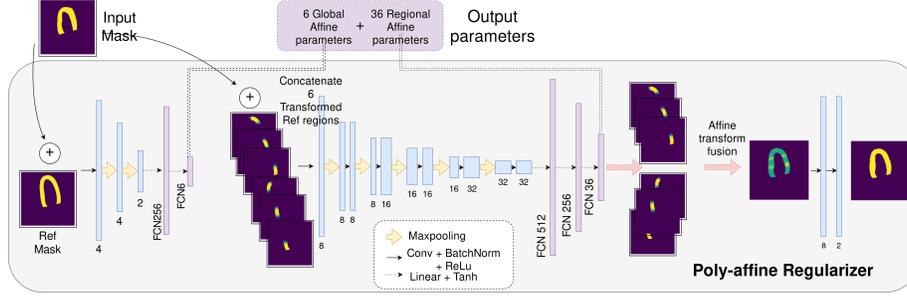


Fig. 2: Architecture of Proposed Poly-affine Regulariser

$\mathcal{N}(\hat{\mu}, \hat{\Sigma})$, $\hat{\mu}$ and $\hat{\Sigma}$ are the maximum likelihood estimate (here we only consider diagonal covariance matrices) from the aligned transformation parameters $(T_i)_{i=1}^N$. The regularisation for affine parameters is $\mathcal{L}_{affine} = \alpha \mathcal{L}_{l2} + \beta \mathcal{L}_{prior}$, where \mathcal{L}_{l2} is the mean square error between the reconstructed image and the input and

$$\mathcal{L}_{prior} = \sum_{k=1}^K \sum_{j=1}^6 \delta_j \frac{(A_j^k - \hat{\mu}_j)^2}{\hat{\Sigma}_{jj}} \quad (3)$$

where $K = 1$ for global affine parameter and $K = 6$ for regional affine parameters.

Thus the total loss for the poly-affine reconstruction network is:

$$\mathcal{L}_{total} = \underbrace{\mathcal{L}_{dice}(\hat{M}^g)}_{global\ sub-net} + \beta^g \mathcal{L}_{prior}^g + \underbrace{\mathcal{L}_{dice}(\hat{M}^0) + \mathcal{L}_{dice}(\hat{M}^f)}_{regional\ sub-net} + \alpha \mathcal{L}_{l2} + \beta^r \mathcal{L}_{prior}^r \quad (4)$$

Once the poly-affine regulariser network is trained, the 6+36 affine parameters serve as explicit hidden variables to regularise the shape of myocardium prediction. We train a U-Net model (same as baseline model) which seeks for the best overlapping of mask as well as the minimum distance between the corresponding affine parameters of predicted myocardium and that of ground truth myocardium. The loss function for this method is

$$loss = dice + \alpha MSE(PA(P) - PA(M)) \quad (5)$$

, where MSE is the mean squared error, PA is poly-affine regulariser that outputs the 42 affine parameters.

2.3 SEG-CONTOUR: multi-class contour-loss

In order to increase the classification accuracy on the boundary, we choose to use an adapted multi-class contour loss[8]. Firstly, a distance map $D(M)$ is calculated from ground truth mask and it illustrates the shortest euclidean distance of each pixel to the closest border. Then the contour loss is calculated as

$$loss_{contour} = \sum (D(M) \circ contour(B(P))) \quad (6)$$

where \circ performs element-wise multiplication. P is the prediction output after SoftMax activation of U-Net for a certain class. $B(P)$ represents a differentiable thresholded Sigmoid for binarisation

$$B(P) = \frac{1}{1 + \exp^{-\gamma(P-T)}} \quad (7)$$

where $\gamma = 20$ and $T = 0.5$.

The contour of the binarised mask is obtained by applying a 2D Sobel filter

$$\text{contour}(P) = |G_x * P| + |G_y * P| \quad (8)$$

where $*$ denotes 2D convolution and G_x, G_y are 2D Sobel kernel in x-,y- dimension:

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

3 Experiments and Results

3.1 Datasets

In this work, we work on two public data sets: CAMUS¹ and ECHONET². CAMUS dataset consist of publicly accessible 2D echocardiographies and the corresponding annotations of 450 patients. For each patient, 2D apical 4-chambers (A4C) and 2-chambers (A2C) view sequences are available. Manual annotation of cardiac structures (left endocardium, left epicardium and left atrium) were acquired by expert cardiologists for each patient in each view, at end-diastole (ED) and end-systole (ES) [11]. Along with the image and annotation data, we also have the following information: image quality (good/medium/poor), left ventricle end-diastole volume (LVedv), left ventricle end-systole volume (LVesv) and left ventricle ejection fraction(LVef).

ECHONET dataset contains 10 030 apical 4-chambers echocardiography videos as part of routine clinical care at Stanford University Hospital [13]. Segmentation measurements (left endocardium) at end-diastole and end-systole are available for all videos. The corresponding LVedv, LVesv, LVef are also provided for each video.

3.2 Experiments

We trained the baseline UNet, SEG-LM, SEG-AFFINE, SEG-CONTOUR models on the CAMUS dataset (both 2-chamber and 4 chamber ED/ES frames) using 10-fold cross validation. The 450 patients are randomly split-ed into 10

¹ <https://www.creatis.insa-lyon.fr/Challenge/camus/databases.html>

² <https://echonet.github.io/dynamic/>

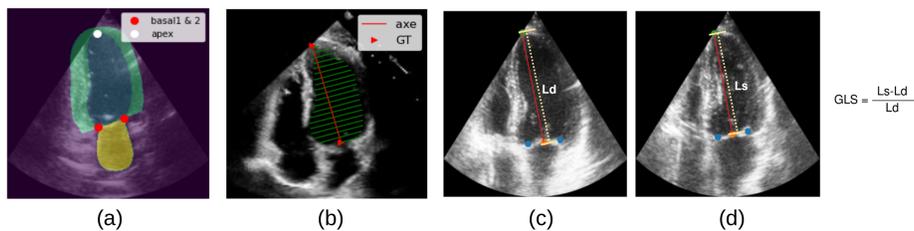


Fig. 3: (a) An example of segmentation ground truth of CAMUS dataset. The two basal and apex landmarks are extracted following the procedures described in section 2.1. (b) An example of annotation provided by ECHONET dataset. We generate the ground truth mask of LV by linearly connecting the border points. The grand axis (in red) is considered as the line connecting the apex and mid-basal point. The length of grand axis is considered as the LV length. (c-d) GLS calculation illustration (background is one echo image from CAMUS). We calculate the GLS from the LV length change by following the approximation method in [14].

folds, with each fold has similar distribution of image quality and LVEf distribution. Every turn we use 8 fold data for training, 1 fold for validation (for model selection) and 1 fold for testing.

From the segmentation result, we first find the largest connected component for each class and applied a closing operation to fill the potential hole that could exist inside the predicted mask. We then extract the basal and apex landmark points following protocol described in section 2.1 for the four models except SEG-LM whose landmark is extracted from the landmark branch. The direction of mid-basal to apex will be regarded as grand axis of left ventricle and is used for left ventricle volume estimation by using modified Simpson’s rule [6], thus the calculation of ejection fraction. The distance from mid-basal to apex forms the ventricle length and serves for global longitudinal strain calculation [14]. Because of the dataset limitation, we don’t have the all the 3 apical views for left ventricle, so the GLS estimation will not be very accurate. However, we still calculate the GLS from 2 views (CAMUS) and 1 view (ECHONET) as reference to test the landmark detection accuracy.

All the segmentation models are implemented with Pytorch with a batch size of 8 and input image resized to 256×256 . In order to avoid over-fitting and improve generalisation performance, we apply random data augmentation at training phase for all the networks. A stack of rotation, random cropping, brightness adjustment, contrast change, sharpening, blurring and speckle noise addition is conducted with each a probability of 0.5 for every input image. CAMUS dataset does not contain the ground truth of desired landmarks. We generated the ‘ground truth’ landmark position from ground truth segmentation masks. The Gaussian heatmap of ground-truth landmarks is computed with $\sigma = 4$.

The baseline U-Net is trained with a multi-class dice loss. An Adam optimiser is applied with a learning rate of $1e^{-3}$. The training is early stopped when the dice loss on validation data shows no increase for more than 5 epochs. For SEG-LM model, we set $\alpha = 0.05, \beta = 0.5, \gamma = 0.5$ in eq.2. An Adam optimiser is applied (lr= 10^{-4}). The output heatmap of landmarks is first processed to keep only one point cluster per layer and then the landmark location is extracted as the centroid. The poly-affine regulariser network is first trained with the myocardium ground truth for reconstruction. We set $\alpha = 0.005, \beta^r = \beta^g = 0.01, \sigma_{i=1}^6 = 1$ and learning rate at $1e-4$. We then train a baseline U-Net with the polyaffine regulariser using loss function (eq.5) with $\alpha = 10$. The parameter for global prior is calculated from the training-specific $(T_i)_{i=1}^N$. The parameter for regional prior is set as $\mu_{1\dots 6} = [1, 0, 0, 0, 1, 0]$ and $[\Sigma_{i=1}^6] = 0.1$. As for SEG-CONTOUR model, contour loss is optimised along with the dice loss. The contour loss is easy to fall into a local minimum of 0 so we set a weight of 100, $1e-4$ for dice loss and contour loss respectively.

3.3 Evaluation metrics

For segmentation results, apart from the most used geometrical metrics: Dice coefficient, Hausdorff distance (HD) and Mean Surface Distance (MSD), we also use two anatomical metrics: Convexity(Cx) and Simplicity(Sp) [12].

$$Convexity(Cx) = \frac{Area(P)}{Area(ConvexHull(P))} \quad (9)$$

$$Simplicity(Sp) = \frac{\sqrt{4\pi * Area(P)}}{Perimeter(P)} \quad (10)$$

Based on these metrics, we calculate the number of outliers for algorithm/model robustness evaluation. The outlier of segmentation prediction for CAMUS dataset is established from the inter-variability tests with the upper limit values for HD and MSD, and lower limit values for the simplicity and convexity [12]. A prediction mask is considered as a geometrical outlier if its $HD > 3.5mm$ or $MSD > 8.2mm$ at ED, if $HD > 4mm$ or $MSD > 8.8mm$ at ES. The corresponding limit for anatomical outlier is if $Cx < 0.529$ or $Sp < 0.741$ for endocardium, if $Cx < 0.694$ or $Sp < 0.960$ for epicardium[12].

3.4 Results

We show the evaluation results in Tables 1 and 2, computed from 450 patients of CAMUS dataset using 10 fold cross-validation for the four methods detailed in Section 2. Compared with the baseline model, the SEG-LM, SEG-AFFINE models demonstrate only a slight decrease in terms of segmentation metric, ejection fraction (EF) prediction accuracy and landmark detection accuracy. The SEG-CONTOUR model shows a similar performance with baseline model in Dice score (table.1), but reduces greatly the number of geometrical and anatomical outliers (table.1). It’s quite reasonable that with a smaller HD and MSD, the

Method	Endo			Epi			Outlier		
	Dice	HD (mm)	MSD (mm)	Dice	HD (mm)	MSD (mm)	Geo.	Ana.	Both
Baseline	0.931 ± 0.040	5.04 ± 3.00	1.51 ± 0.83	0.951 ± 0.025	5.75 ± 3.61	1.71 ± 0.91	15%	2.8%	2.5%
SEG-LM	0.928 ± 0.042	5.47 ± 3.04	1.59 ± 0.84	0.950 ± 0.025	6.22 ± 3.74	1.79 ± 0.93	20%	7.6%	5.3%
SEG-AFFINE	0.930 ± 0.042	5.14 ± 3.00	1.53 ± 0.88	0.951 ± 0.027	5.84 ± 3.75	1.72 ± 0.95	14.3%	5.1%	3.7%
SEG-CONTOUR	0.931 ± 0.041	4.99 ± 2.95	1.50 ± 0.73	0.952 ± 0.026	5.63 ± 3.32	1.67 ± 0.87	13.8%	1.5%	1.2%

Table 1: **CAMUS Segmentation Metric.** Endo.: endocardium, Epi: epicardium, HD: Hausdorff distance, MSD: mean surface distance, Geo.: Geometrical outlier, Ana.: Anatomical outlier. Values in bold represent the best score.

Method	Basal1	Basal2	Apex	GLS(%)			EF(%)		
	MAE (mm)	MAE (mm)	MAE (mm)	MAE (%)	Corr.	Bias(%) \pm std	MAE (%)	Corr.	Bias(%) \pm std
Baseline	2.36	3.06	4.06	3.97	0.74	-0.73 ± 5.20	4.76	0.86	0.93 ± 7.07
SEG-LM	2.82	3.45	4.48	4.51	0.70	-1.74 ± 5.27	5.26	0.84	2.51 ± 8.04
SEG-AFFINE	2.35	2.97	4.22	4.03	0.74	-0.9 ± 5.15	4.96	0.85	1.16 ± 7.31
SEG-CONTOUR	2.25	2.80	3.97	3.96	0.75	-0.36 ± 5.34	5.06	0.84	0.7 ± 7.50

Table 2: **CAMUS Landmark/GLS and EF Prediction.** Basal1: the left mitral valve end point, Basal2: the right mitral valve end point, EF: ejection fraction, GLS: global longitudinal strain. Values in bold represent the best score.

SEG-CONTOUR reduces the classification error along the boundary area thus less outlier predictions. This is consistent with the observations that a good Dice score does not always guarantee a good HD [2] and in our case, not always leads to anatomically-plausible segmentation. The SEG-CONTOUR model is also capable of tracking more precisely the boundary especially the landmarks (table.2) thus a smaller bias of GLS prediction. In terms of EF calculation, the baseline U-Net model shows smaller mean absolute error but a larger bias than the SEG-CONTOUR loss. We show three CAMUS test examples (fig.4) where each row has a good/medium/bad performance in terms of HD score respectively. Comparing with the the rest 3 models, SEG-CONTOUR has a more fluent border similar as the ground truth annotation.

The evaluation results of applying the trained model on a totally different dataset ECHONET (the same test fold of 1277 patients as in [13]) show the same trend of performance of our 4 methods. The SEG-CONTOUR method

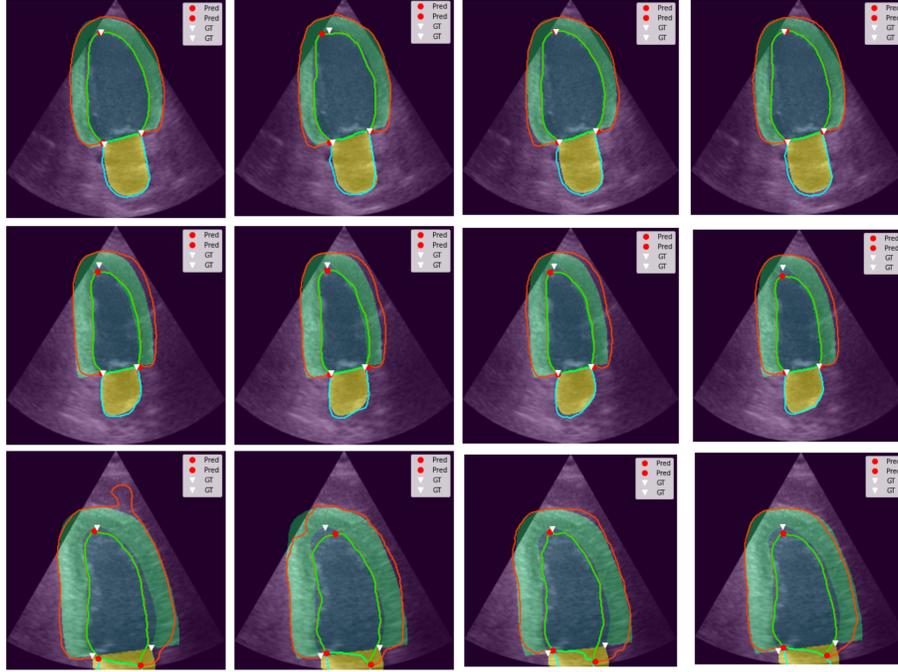


Fig. 4: 3 CAMUS segmentation examples (good/medium/bad in terms of HD). The 4 four columns represent the baseline model, SEG-LM, SEG-AFFINE, SEG-CONTOUR respectively from left to right. The red, green, cyan lines represent the predicted segmentation contours of epicardium, endocardium and left atrium. The transparent green, blue and yellow regions are the ground truth masks of myocardium, left ventricle blood pool and left atrium respectively.

demonstrates a good performance on tasks related to boundary information, for example, lower HD and MSD, better GLS estimation. It's less accurate on area based task, i.e. volume estimation thus ejection fraction prediction. It is noticeable that all of the four models demonstrate a nice Dice coefficient on this different dataset (the Dice coefficient of models trained on ECHONET data is 0.92% [13]), which proves the importance of appropriate image augmentation techniques.

4 Conclusion

In this paper, we explored methods to introduce shape constraints into 2D Echocardiography segmentation models from three levels: global-level (SEG-LM), regional-level (SEG-AFFINE), pixel-level (SEG-CONTOUR). From the evaluation results on CAMUS dataset and its generalisation result on a unseen dataset ECHONET, it is more efficient to introduce pixel-level shape con-

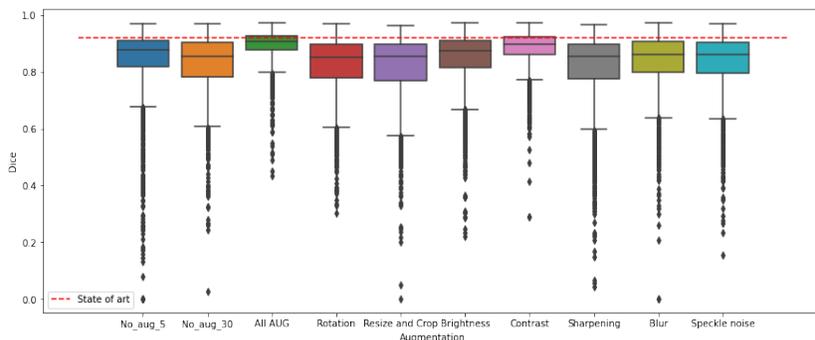


Fig. 5: The baseline U-Net model is trained with only one of the mentioned augmentation methods on CAMUS dataset and then is evaluated on the same test fold of ECHONET segmentation model, whose dice coefficient is 0.92. Contrast adjustment contributes most to the improvement of generalisation result while with all the techniques, we obtain the best Dice score and less variation. No-aug-5: trained model of 5th epoch without augmentation, No-aug-30: trained model of 30th epoch without augmentation. At 30th epoch, the model has already overfitted the CAMUS data.

Table 3: ECHONET Prediction

Method	Endo-cardium (LV)			EF				GLS			
	Dice	HD (pxls)	MSD (pxls)	MAE (%)	R2	Corr.	Bias(%) ± std	MAE (%)	R2	Corr.	Bias(%) ± std
Baseline	0.892 ± 0.069	12.99 ± 7.96	3.74 ± 3.42	7.97	0.11	0.69	4.46 ± 12.80	4.96	-1.2	0.38	0.36 ± 7.94
SEG-LM	0.886 ± 0.063	13.54 ± 6.86	3.87 ± 2.63	8.79	0.02	0.72	6.21 ± 13.39	5.61	-1.49	0.36	-1.9 ± 7.07
SEG-AFFINE	0.887 ± 0.068	13.12 ± 7.25	3.87 ± 2.97	9.35	-0.14	0.67	7.11 ± 14.41	5.88	-2.44	0.29	-0.75 ± 9.29
SEG-CONTOUR	0.895 ± 0.057	12.56 ± 6.34	3.58 ± 2.16	8.39	0.07	0.69	5.03 ± 13.21	4.39	-0.43	0.51	0.59 ± 6.52

straint than global or regional level constraints for U-Net based models. With a multi-class contour loss, SEG-CONTOUR model achieves better classification on the boundary pixels with a reduced a Hausdorff distance and more accurate landmark detection result. Our experiments showed the good potential of SEG-CONTOUR for a more robust segmentation and deformation analysis.

5 Acknowledgments

This work has been supported by the French government through the National Research Agency (ANR) Investments in the Future with 3IA Côte d’Azur (ANR-19-P3IA-0002) and by Inria PhD funding.

References

1. Ahn, S.S., Ta, K., et al.: Unsupervised motion tracking of left ventricle in echocardiography. In: Byram, B.C., Ruitter, N.V. (eds.) *Medical Imaging 2020: Ultrasonic Imaging and Tomography*. vol. 11319, pp. 196 – 202. International Society for Optics and Photonics, SPIE (2020)
2. Bernard, O., Lalande, A., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging* **37**(11), 2514–2525 (2018)
3. Bohlender, S., Oksuz, I., Mukhopadhyay, A.: A survey on shape-constraint deep learning for medical image segmentation (2021), <http://arxiv.org/abs/2101.07721>
4. Cerqueira, M.D., Weissman, N.J., et al.: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for health-care professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Journal of Nuclear Cardiology* **9**(2), 240–245 (2002)
5. Chen, C., Bai, W., et al.: Improving the generalizability of convolutional neural network-based segmentation on cmr images. *Frontiers in Cardiovascular Medicine* **7**, 105 (2020)
6. Folland, E.D., Parisi, A.F., et al.: Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. A comparison of cineangiographic and radionuclide techniques. *Circulation* **60**(4), 760–766 (1979)
7. Hasselberg, N.E., Haugaa, K.H., et al.: Left ventricular global longitudinal strain is associated with exercise capacity in failing hearts with preserved and reduced ejection fraction. *European Heart Journal - Cardiovascular Imaging* **16**(2), 217–224 (12 2014)
8. Jia, S., Despinasse, A., et al.: Automatically Segmenting the Left Atrium from Cardiac Images Using Successive 3D U-Nets and a Contour Loss. *Lecture Notes in Computer Science* **11395 LNCS**, 221–229 (2019)
9. Kraigher-Krainer, E., Shah, A.M., et al.: Impaired systolic function by strain imaging in heart failure with preserved ejection fraction. *Journal of the American College of Cardiology* **63**(5), 447–456 (2014)
10. Krebs, J., Delingette, H., et al.: Learning a Probabilistic Model for Diffeomorphic Registration. *IEEE transactions on medical imaging* **38**(9), 2165–2176 (2019)
11. Leclerc, S., Smistad, E., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging* **38**(9), 2198–2210 (2019)
12. Leclerc, S., Smistad, E., et al.: Ru-net: A refining segmentation network for 2d echocardiography. In: 2019 IEEE International Ultrasonics Symposium (IUS). pp. 1160–1163 (2019)
13. Ouyang, D., He, B., et al.: Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**(7802), 252–256 (2020)
14. Støylen, A., Mølmen, H.E., Dalen, H.: Left ventricular global strains by linear measurements in three dimensions: Interrelations and relations to age, gender and body size in the HUNT Study. *Open Heart* **6**(2), 1–9 (2019)
15. Wang, X., Yang, X., et al.: Joint segmentation and landmark localization of fetal femur in ultrasound volumes. In: 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI). pp. 1–5 (2019)