Automatic Myocardial Strain Imaging in Echocardiography Using Deep Learning

Andreas Østvik[1,2,3]($\boxtimes$), Erik Smistad[1,2,3], Torvald Espeland[1,2,4],
Erik Andreas Rye Berg[1,2,4], and Lasse Lovstakken[1,2]

[1] Centre for Innovative Ultrasound Solutions, NTNU, Trondheim, Norway
andreas.ostvik@ntnu.no
[2] Department of Circulation and Medical Imaging, NTNU, Trondheim, Norway
[3] Department of Health, SINTEF Digital, Trondheim, Norway
[4] Clinic of Cardiology, St. Olavs Hospital, Trondheim, Norway

Abstract. Recent studies in the field of deep learning suggest that motion estimation can be treated as a learnable problem. In this paper we propose a pipeline for functional imaging in echocardiography consisting of four central components, (i) classification of cardiac view, (ii) semantic partitioning of the left ventricle (LV) myocardium, (iii) regional motion estimates and (iv) fusion of measurements. A U-Net type of convolutional neural network (CNN) was developed to classify muscle tissue, and partitioned into a semantic measurement kernel based on LV length and ventricular orientation. Dense tissue motion was predicted using stacked U-Net architectures with image warping of intermediate flow, designed to tackle variable displacements. Training was performed on a mixture of real and synthetic data. The resulting segmentation and motion estimates was fused in a Kalman filter and used as basis for measuring global longitudinal strain. For reference, 2D ultrasound images from 21 subjects were acquired using a GE Vivid system. Data was analyzed by two specialists using a semi-automatic tool for longitudinal function estimates in a commercial system, and further compared to output of the proposed method. Qualitative assessment showed comparable deformation trends as the clinical analysis software. The average deviation for the global longitudinal strain was $(-0.6 \pm 1.6)\%$ for apical four-chamber view. The system was implemented with Tensorflow, and working in an end-to-end fashion without any ad-hoc tuning. Using a modern graphics processing unit, the average inference time is estimated to (115 ± 3) ms per frame.

Keywords: Deep learning · Echocardiography · Functional imaging

1 Introduction

Recent years have shown that quantitative assessment of cardiac function has become indispensable in echocardiography. Evaluation of the hearts contractile apparatus has traditionally been limited to geometric measures such as ejection

fraction (EF) and visual estimation (eyeballing) of myocardial morphophysiol-
ogy [3]. Despite being a central part of standard protocol examinations at the
outpatient clinic, the methods tend to have poor inter- and intravariability. With
tissue doppler imaging (TDI) and speckle tracking(ST), the quantification tools
have moved beyond these measures, and enabled new methods for assessing
the myocardial deformation pattern [12]. Myocardial deformation imaging, e.g.
strain and strain rate, derived from TDI and ST, have high sensitivity, and can
allow an earlier detection of cardiac dysfunction. However, these methods also
have several limitations. For instance, TDI is dependent on insonation angle, i.e.
measurements are along the ultrasound beam. A poor parallel alignment with
the myocardium can thus influence the results. Speckle tracking is less angle
dependant (typically dependant on the lateral resolution), but has suffered from
poor temporal resolution and ad-hoc setups. Recent work in the field of deep
learning (DL) suggest that motion estimation can be treated as a learnable prob-
lem [7]. Herein, we investigate this approach in combination with cardiac view
classification and segmentation to achieve fully automatic functional imaging.

1.1   Relevant Work and Perspective

Automatic view classification and segmentation of relevant cardiac structures
in echocardiography has been a topic of great interest and research [2,9]. For
segmentation, work has mainly been conducted on 3D echocardiography, but
2D approaches are also proposed [13]. To the authors' knowledge, no published
study have utilized motion estimation from deep learning in echocardiography.
These methods claim to be more robust in terms of noise and small displace-
ments [7] than traditional optical flow methods, thus appealing for ultrasound
and myocardial motion estimation. Combining the components could potentially
allow fast and fully automated pipelines for calculating clinically relevant param-
eters, with feasibility of on-site analysis. In this study, the goal is to address this,
and measure the global longitudinal strain (GLS) from the four-chamber view
in an end-to-end fashion.

Variability of global longitudinal strain has been discussed in several papers.
Recently, a head-to-head comparison between speckle tracking based GLS mea-
surements of nine commercial vendors [5] was conducted. Results show that the
reproducibility compared to other clinical measurements such as EF is good, but
the intervendor variation is significant. The same expert obtaining GLS in the
apical four-chamber view of 63 patients on nine different systems, gave average
results in the range of -17.9% to -21.4%. The commercial system used as ref-
erence in this study had an offset of -1.7%, i.e. overestimating, from the mean
measurement. The inter- and intraobserver relative mean error was 7.8% and
8.3% respectively.

2   Method

Our proposed pipeline is comprised of four steps, (i) classification of cardiac view,
(ii) segmentation of the left ventricle (LV) myocardium, (iii) regional motion esti-

mates and (iv) fusion of measurements. An illustration of the system after view classification is illustrated in Fig. 1. Step (i)–(iii) utilizes convolutional neural networks (CNNs), while the last step uses a traditional Kalman filter method.
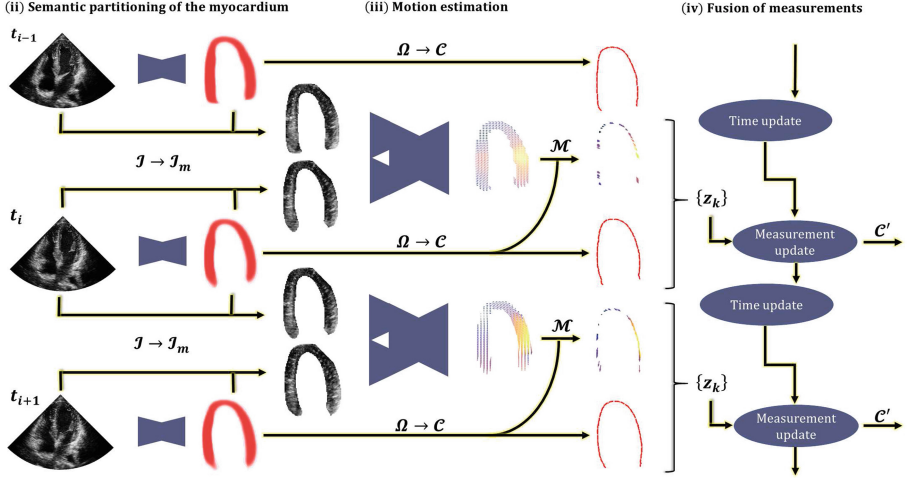
Fig. 1. Visualization of the measurement pipeline. US images are forwarded through a segmentation network, and the resulting masks are used to extract the centerline and relevant parts of the image. The masked US data is further processed through the motion estimation network yielding a map of velocities. The centerline position and velocities of the myocard are used in the measurement update step of a Kalman filter. The updated results are used as a basis for strain measurements.

2.1   Cardiac View Classification

The view classification is the first essential step in the automatic pipeline, and is used to quality assure and sort incoming data. We employ a feed-forward CNN composed of inception blocks and a dense connectivity pattern [6,14]. Initially, input is propagated through two component blocks with (3×3) convolution kernels followed by max pooling. The first and second convolution layer has 16 and 32 filters respectively. We use pooling with size (2×2) and equal strides. After the second pooling layer, data is processed through an inception module with three parallel routes. Each route consist of a bottleneck, two of which were followed by blocks with larger convolution kernels, i.e. (3×3) and (5×5) respectively. The input of the inception module is concatenated with the output and processed into a transition module with bottleneck and max pooling. This step is repeated three times, and we double the amount of filters before every new pooling layer. The dense connectivity pattern alleviates the vanishing gradient problem, and can enhance feature propagation and reusability. After

the third transition, the data is processed through two inception blocks with constant amount of filters and no pooling. The route with (5×5) convolution kernels is omitted in these modules, and dropout regularization was used between them. The final classification block consists of a compressing convolution layer with (1×1) kernels and number of filters equal to the class count. This is activated with another PReLU, before features are spatially averaged and fed into a softmax activation.

Training is performed from scratch with Adam optimizer and categorical cross entropy loss, with input size of (128×128) greyscale. A total of eight classes were used for training, the apical four chamber, two chamber and long-axis, the parasternal long- and short-axis, subcostal four-chamber and vena cava inferior, as well as a class for unknown data. The final network classifies the different cardiac views, and if applicable, i.e. high confidence of apical four-chamber, the image is processed into the remaining processing chain.

2.2   Semantic Partitioning of the Myocardium

The second step is segmentation of the left ventricle myocardium. A standard U-Net type of CNN [11] is utilized. The architecture consist of a downsampling, and an upsampling part of five levels with concatenating cross-over connection between equally sized feature maps. Each level has two convolution layers with the same amount of filters ranging from 32 to 128 from top to bottom respectively. All filters have a size of (3×3). Max pooling with size (2×2) and equal strides was used for downsampling and nearest neighbour for upsampling. Training was performed with Adam optimizer and Dice loss, and the size of the input image was set to (256×256) greyscale. The output of the network is a segmentation mask Ω.

The segmentation is used a basis for two different tasks, masking the input of the motion estimation network $\mathcal{I}_m$ and centerline extraction. We mask the US image $\mathcal{I}$ to remove redundant input signal. The contour of the segmentation Ω was used to define the endo- and epicardial borders, and further the centerline $\mathcal{C} = \{(x,y)_1, ..., (x,y)_N\}$ was sampled between with $N = 120$ equally spaced points along the myocard. The latter is passed to the Kalman filter.

2.3   Motion Estimation Using Deep Learning

The motion estimation is based on the work done by Ilg *et al.* [7], and the networks referred to as FlowNets. The design involves stacking of multiple U-Net architectures with image warping of intermediate flow and propagation of brightness error. Two parallel routes are created to tackle large and small displacements separately. The prior is solved by stacking three U-Net architectures, the first which includes explicit correlation of feature maps, while the two succeeding are standard U-Net architectures without custom layers. For small displacement, only one U-Net is used, but compared to the networks for large displacements, the kernel size and stride of the first layer is reduced. At the end, the two routes are fused together with a simple CNN. The networks are trained separately, in

a schedule consisting of different synthetic datasets with a wide range of motion vector representations. The small displacement network is fine-tuned on a dataset modified for subpixel motion. Adam optimizer and endpoint error loss is used while training for all the networks. The input size of the network was kept the same as the original implementation, i.e. (512×384).

The output prediction of the network is dense tissue motion vectors in the masked US area. The centerline $\mathcal{C}$ of the current segmentation is used to extract the corresponding set of motion vectors $\mathcal{M} = \{(v_x, v_y)_1, ..., (v_x, v_y)_N\}$.

2.4   Fusion of Measurements

Fusion of measurements was performed employing an ordinary Kalman filter with a constant acceleration model [8] with measurement input $z_k = [x, y, v_x, v_y]_k^T$ for every point-velocity component $k \in \{\mathcal{C}, \mathcal{M}\}$. Essentially, this serves as a simple method for incorporating the temporal domain, which is natural in the context of echocardiography. It adds temporal smoothing, reducing potential pierce noise detectable in image-to-image measurements. The updated centerline $\mathcal{C}' \subseteq \Omega$ is used to calculate the longitudinal ventricular length ι, i.e. the arc length, for each timestep t. Further, this is used to estimate the global longitudinal strain $\epsilon(t) = (\iota(t) - \iota_0)/\iota_0$ along the center of the myocard.

3   Datasets for Training and Validation

Anonymous echocardiography data for training the view classification and segmentation models was acquired from various patient studies with Research Ethics Board (REB) approval. The echocardiography data used are obtained at various outpatient clinics with a GE Vivid E9/95 ultrasound system (GE Vingmed Ultrasound, Horten, Norway), and consist of data from over 250 patients. The health status of subjects is unknown, but representative for a standard outpatient clinic. The data includes manual expert annotation of views, and the epi- and endocard borders of the left ventricle. The view classification and segmentation networks are trained separately on this data, with a significant fraction left out for testing. The motion estimation network was trained on three synthetic datasets, namely FlyingChairs [4], FlyingThings3D [10] and ChairsSDHom [7]. Disregarding the fundamentals of motion, the datasets have no resemblance to echocardiography. However, they can be modified to have representations covering both sub- and superpixel motion, which is necessary to reproduce motion from the whole cardiac cycle.

For validation of GLS, 21 subjects called for evaluation of cardiac disease in two clinical studies were included. Both are REB approved, and informed consent was given. Two specialists in cardiology performed standard strain measurements using a semi-automatic method implemented in GE EchoPAC[1]. The method uses speckle tracking to estimate myocardial deformation, but the methodology is unknown. The results were used as a reference for evaluating the implemented pipeline.

---

[1] http://www3.gehealthcare.com/en/products/categories/ultrasound/vivid/echopac.

4   Results

GLS was obtained successfully in all patients. The results for apical four-chamber views are displayed in Fig. 2, together with the GLS curves of the average and worst case subjects. The average deviation of the systolic GLS between the two methods was $(-0.6\pm1.6)\%$. The average strain on all subjects was $(-17.9\pm2.3)\%$ and $(-17.3 \pm 2.5)\%$, for the reference and proposed method respectively.
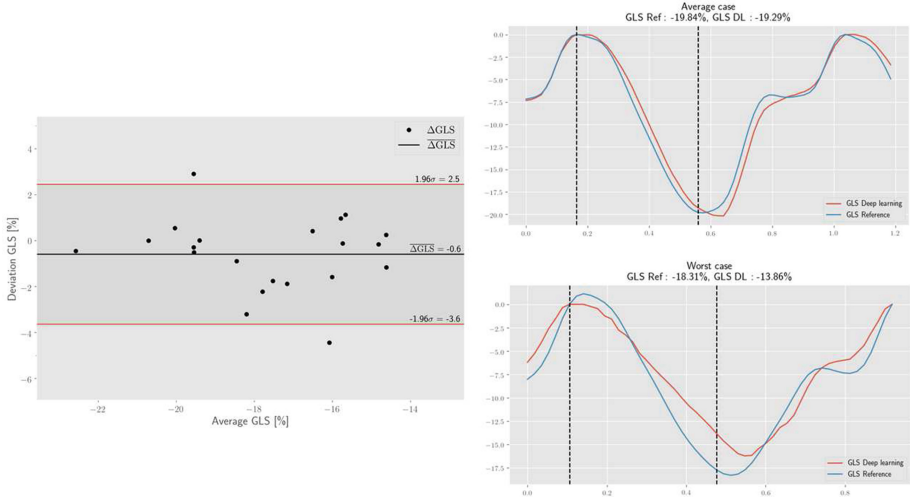
Fig. 2. Bland-Altman plot of global longitudinal strain from all subjects. The estimated GLS traces of the average and worst case, together with the corresponding reference, are displayed to the right.

The view classification achieved an image-wise F_1 score of 97.9% on four-chamber data of 260 patients, and the segmentation a dice score of (0.87 ± 0.03) on 50 patients, all unknown and independent from the training set. The system was implemented as a Tensorflow dataflow graph [1], enabling easy deployment and optimized inference. Using a modern laptop with a Nvidia GTX 1070 GPU, the average inference time was estimated to (115 ± 1) ms per frame, where flow prediction accounts for approximately 70% of the runtime.

5   Discussion

Compared to reference, the measurements from the proposed pipeline were slightly underestimated. The reference method is not a gold standard for GLS and might not necessarily yield correct results for all cases. Speckle tracking can fail where noise hampers the echogenicity. We could identify poor tracking in the apical area due to noise for some subjects, and this would in turn result in larger

strain. Further, the vendor comparison study [5] shows that the commercial system used in this study on average overestimates the mean of all vendors by 1.7%. This in mind, we note that the results from the current implementations are in the expected range. For individual cases, the deformation have overlapping and synchronized trends, as is prevalent from Fig. 2.

The proposed pipeline involves several sources of error, especially the segmentation and motion networks being the fundamental building blocks of the measurements. Using the segmentation mask to remove redundant signal in the US image seems feasible and useful for removing some noise in the motion network. However, it is not essential when measuring the components of the centerline, as they are far from the borders of the myocard, where the effect is noticable.

Future work will include the addition of multiple views, e.g. apical two- and long-axis, allowing average GLS. This is considered a more robust metric, less prone to regional noise. Also, fusion of models are currently naive, and we expect results to improve inducing models with more relevance to cardiac motion. The same holds for the motion estimation, i.e. the network could benefit from training on more relevant data. Further, we wish to do this for regional strain measurements. For clinical validation, we need to systematically include the subject condition and a larger test material.

6   Conclusion

In this paper we present a novel pipeline for functional assessment of cardiac function using deep learning. We show that motion estimation with convolutional neural networks is generic, and applicable in echocardiography, despite training on synthetic data. Together with cardiac view classification and myocard segmentation, this is incorporated in an automatic pipeline for calculating global longitudinal strain. Results coincide well with relevant work. The methods and validation are still at a preliminary stage in terms of clinical use, and some limitations and future work are briefly mentioned.

References

1. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. OSDI **16**, 265–283 (2016)
2. Bernard, O., et al.: Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography. IEEE Trans. Med. Imaging **35**(4), 967–977 (2016). https://doi.org/10.1109/TMI.2015.2503890
3. Dalen, H., et al.: Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: the hunt study in Norway. Eur. J. Echocardiogr. **11**(2), 176–183 (2009)
4. Dosovitskiy, A., et al.: FlowNet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)

5. Farsalinos, K.E., Daraban, A.M., Ünlü, S., Thomas, J.D., Badano, L.P., Voigt, J.U.: Head-to-head comparison of global longitudinal strain measurements among nine different vendors: the EACVI/ASE inter-vendor comparison study. J. Am. Soc. Echocardiogr. **28**(10), 1171–1181 (2015)
6. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
7. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: FlowNet 2.0: evolution of optical flow estimation with deep networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017. http://lmb.informatik.uni-freiburg.de//Publications/2017/IMKDB17
8. Kalman, R.E.: A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
9. Madani, A., Arnaout, R., Mofrad, M., Arnaout, R.: Fast and accurate view classification of echocardiograms using deep learning. npj Digit. Med. **1**(1), 6 (2018)
10. Mayer, N., et al.: A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4040–4048 (2016)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
12. Smiseth, O.A., Torp, H., Opdahl, A., Haugaa, K.H., Urheim, S.: Myocardial strain imaging: how useful is it in clinical decision making? Eur. Heart J. **37**(15), 1196–1207 (2016). https://doi.org/10.1093/eurheartj/ehv529
13. Smistad, E., Østvik, A., Haugen, B.O., Lovstakken, L.: 2D left ventricle segmentation using deep learning. In: 2017 IEEE International Ultrasonics Symposium (IUS), pp. 1–4. IEEE (2017)
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)