Qeios

Peer Review

Review of: "Deep Sylvester Posterior Inference for Adaptive Compressed Sensing in Ultrasound Imaging"

Christopher Khan[1]

1. Vanderbilt University, United States

This work presents an approach for performing adaptive subsampling of ultrasound image data in order to reduce the number of scan lines that are required to form an image. It does so by utilizing a Bayesian deep neural network approach to perform reconstruction of an ultrasound image using partial image frame observations. It then uses this framework to estimate the subsampling scheme to use for the next image frame based on maximizing the expected information gain of future observations.

I believe that readers will find the application of the Bayesian framework along with the Sylvester Normalizing Flow to the problem of adaptive subsampling to be quite interesting. Moreover, the results demonstrate the ability of the proposed method to outperform other subsampling methods across several metrics. Therefore, I believe that this work is of both high merit and relevance.

Regarding my rating, I gave a rating of 4 stars instead of 5 to this work in its current state only because I believe that there are some improvements that can be made to further strengthen it. My suggestions are the following:

1. My primary suggestion is to revise Sections II B and II C, along with the algorithm overview, in order to improve clarity regarding the proposed method and to make it more digestible for readers. For example, my impression of the reconstruction aspect of the method is that there are two main components. One component is being able to map the partial frame observations y to the latent variable z, and the other component is to be able to map the latent variable z to the full frame x. However, it was not clear how many encoders and decoders there are exactly. Specifically, it is stated that the generative latent variable model is first trained using a set of full

 doi.org/10.32388/1BHIV2

observations. Does this mean you have an encoder and decoder for the task of going from x to z to x that is trained first? Then, do you train a separate encoder and decoder for the task of going from y to z to y, which seems to be indicated by equation (2) also having a decoder term? Similarly, during actual reconstruction, do you use the encoder that was trained to go from y to z and then the decoder that was trained to go from z to x while not using the other encoder and decoder (this seems to be implied by the description in Section III A regarding there being one encoder and one decoder)? I think a good diagram could clear up this confusion. For example, Fig. 1 could be replaced with a better visualization of all of the different parts of your method (i.e., the architecture).

2. I would be more explicit regarding what exactly the terms generative model and inference model refer to because some readers might not be familiar with the terminology in the context of the Bayesian framework.
3. Consider including DNN training hyperparameters such as the number of epochs, batch size, and so on. These can be included in a table.
4. It is stated that the EchoNet dataset was used for training and testing. Are the imaging parameters used to collect this dataset, such as sequence type, transmit frequency, etc., known? If so, I would consider including them.
5. In Section III A, the parameter beta is mentioned, but it is not stated what exactly it represents. There is a brief description of it later in Section IV, but I would include a description of it where it is first mentioned.
6. For Fig. 2, for the upper bound represented by the rightmost column, is it obtained by passing a full frame x through an encoder that goes from x to z and then passing that through a decoder that goes from z to x? If so, shouldn't the label be Dec[Enc[x]] instead of Enc[Dec[x]]? Please clarify. Also, the absolute difference images are a bit small.
7. For the inference times that were reported for the trace and covariance methods, do those include just the time it takes to reconstruct a full frame from a partial frame and then estimate the sampling mask for the next frame, or does it also include other operations such as beamforming? Please specify.
8. Consider including equations for the L1-Loss, SSIM, and PSNR metrics.
9. For the metric values in Table I, are these average values over several frames? If so, I would include the sample size as well as standard deviations. Also, I would consider using a different font that makes it easier to see which numbers are bolded.

10. One additional analysis to consider that I think would interest readers is regarding the generalizability of the different subsampling scheme reconstructions. For example, in this case, the training data for all of the schemes consisted of echocardiography data. However, if I were to take the trained models and apply them to a different type of target (CIRS phantom, for example) without re-training them, then would the proposed method still outperform the other schemes in terms of reconstruction accuracy?

Declarations

Potential competing interests: No potential competing interests to declare.