UltraTTT: Self-supervised Test-time Training for Synthetic Aperture Ultrasound Image Denoising

Background, Motivation and Objective

The inherent multiplicative (speckle) and additive (electronic) noises reduce the interpretability of diagnostic ultrasound images. Traditional denoising methods often yield blurry images and are computationally intensive. Existing deep-learning methods rely on massive data and model parameters and lack generalizability beyond their training scenarios. To tackle these, we propose UltraTTT, a pretraining-free ultrasound image denoising model that employs test-time training, with its first application to synthetic aperture (SA) ultrasound.

Statement of Contribution/Methods

In SA with sub-aperture transmissions, distinct sub-apertures share anatomical structures but exhibit different speckle and electronic noises in overlapping areas. Leveraging this, UltraTTT denoises signals by learning similarities and minimizing differences between randomly paired sub-apertures. A lightweight convolutional neural network (Fig. (b)) is initialized for each frame, so no pretraining is needed. Given two distinct sub-apertures, A_1 and A_2, UltraTTT uses in-phase and quadrature signals from A_1 to predict those from A_2, and computes the loss $\mathcal{L}_{1\to2}$ using mean squared error. Then, it swaps the input and output sub-apertures to compute $\mathcal{L}_{2\to1}$. Backpropagation of two losses updates the model and completes an iteration. This design endows the denoising power of UltraTTT by iteratively learning the coherent and incoherent signal components. After 50 iterations, compounding of the last estimates of A_1 and A_2 forms the final denoised image. UltraTTT was evaluated on *in vivo* echocardiography acquired from 46 subjects using a Verasonics Vantage 256 system with a P4-2 probe.

Results and Discussion

Owing to the lack of noise-free *in vivo* images, we regarded full-aperture images compounded from 8 sub-apertures as clean images. Fig. (c-d) shows that UltraTTT effectively denoised image data in all tested conditions and only took 50 iterations (equivalent to 46 ms/frame) to obtain B-mode images with 31.51 PSNR, 0.84 SSIM using two randomly paired sub-apertures. The histogram analysis (Fig. (e)) revealed UltraTTT's capability in reducing electronic noises, while systematic analysis of noises is ongoing. UltraTTT, a self-supervised and pretraining-free method, eliminates the necessity for large datasets and enables zero-shot image denoising.